NTNU – Trondheim
Norwegian University of
Science and Technology

Calculation of Terminal Currents in Single Photon Excited Avalanche Photodiodes.

Aksel Jan Verne Vestby

Physics
Submission date: November 2012
Supervisor: Jon Andreas Støvneng, IFY

Norwegian University of Science and Technology
Department of Physics

Preface & Acknowledgements

This thesis has been written as final work of the master's degree in Physics at the Norwegian University of Science and Technology (NTNU). The work has been carried out in cooperation with the Norwegian Defence Research Establishment (FFI) under the supervision of Trond Brudevoll (FFI), Asta Storebø (FFI), and Jan Andreas Støvneng (NTNU).

Their support and help has been highly valuable, and greatly appreciated.

I also with to extend thanks to my girlfriend, whose ability to keep me disciplined has been much needed, Malin Forsberg, and to my father, Jan Vestby for moral support.

Table of Contents

1 Introduction

In this thesis, we have used the Monte Carlo (MC) method for semiconductor device simulation to simulate an avalanche photodiode (APD). A diode is a rectifying device which allows current to pass in one direction while current in the opposite direction is blocked. The APD is a highly sensitive photo-detector which relies on the photoelectric effect; a photon excites an electron into the conduction band, an external reverse bias accelerates the electron, and the APD then exploits the physical process of impact ionization to generate an avalanche of carriers. This gives rise to a current pulse which is large enough to be detectable in the electrical circuit surrounding the diode. APDs are increasingly important for technological applications; they are used as detecting components of LIDARs, which are used for optical remote sensing, and for scintillation detectors, which measure ionizing radiation.

The APD and the pn-diode are both relatively simple two-terminal devices in which the pn-junction constitutes the main building block. Simple components like these have traditionally been analysed using drift-diffusion or hydrodynamic models. Particle simulation techniques have not, to any great extent, been employed for this task.

A particular challenge when using MC methods is the calculation of the small terminal currents that occur in APD's. The superparticle picture must be, at least in part, replaced by a real electron-hole picture in a rather large device. To achieve this, photogenerated and impact ionization generated carriers are treated as a perturbation upon a background "frozen field", generated by the superparticles already present in the reverse biased APD. The resulting perturbation in terminal currents was calculated by the Shockley-Ramo (S-R) theorem. To that end, a brief survey of the literature on that subject was carried out. The lack of a consistent notation for the quantities involved makes reading of the various derivations found in the literature difficult and time-consuming. New proposed "generalizations" of the theorem continue to appear, so this situation goes on. In addition, misconceptions about the range of applicability of the theorem have flourished. A couple of authors

have tried to tidy up the literature, which has been very helpful for the present work. We shall introduce a unified notation and derive the theorem in those particular ways we have found useful for displaying its limits of applicability, and then study some analytically solvable examples that further clarify the implications of the theorem. Finally, some tests with this method on a simulated APD have been carried out, using an existing Monte Carlo-program, and the results are discussed.

2 Innledning

I denne masteroppgaven har vi benyttet Monte Carlo (MC) metoden for simulering av halvledere til å simulere en avalanche fotodiode (APD). En diode fungerer som en likeretter som tillater strømledning i en retning, men blokkerer strøm i motsatt retning.

En APD er en høy-sensitiv fotodetektor som bygger på den fotoelektriske effekten; et foton eksiterer et elektron inn i ledningsbåndet, deretter akselereres elektronet av en ekstern revers forspenning, hvorpå APD-komponenten dermed utnytter det fysiske fenomenet støtionisering til å generere et skred av ladningsbærere. Dette gir opphav til en strømpuls som er stor nok til å være målbar i en omkringliggende elektrisk krets. APD-er blir i stor grad viktigere og viktigere for teknologiske bruksområder; de er anvendt som deteksjonskomponenter i LIDAR-teknologier, som benyttes i optisk fjernmåling, og scintilleringsdetektorer, som måler ioniserende stråling.

En APD og en pn-diode er begge relativt enkle to-terminalskomponenter, hvorav pn-overgangen utgjør hoveddelen av komponenten. Simple komponenter som disse har tradisjonelt blitt analysert med hjelp av drift-diffusjon eller hydrodynamiske modeller. Teknikker for partikkelsimulering har ikke i noen stor grad blitt anvendt på dette området tidligere.

En spesiell utfordring ved å benytte MC metoder til dette er beregningen av den svake terminalstrømmen som forekommer i APD-er. Teknikken som benytter superpartikler må, i hvert fall delvis, erstattes av en reell elektron-hull modell i en relativt stor komponent. For å oppnå dette, behandles fotogenererte ladningsbærere som en perturbasjon over et statisk «frosset felt» (“frozen field”) i bakgrunnen som er generert av superpartiklene allerede tilstede i den revers forspente APD-en. Den resulterende perturbasjonen i terminalstrømmene blir beregnet av Shockley-Ramo’s (S-R) teorem.

Til dette formålet har en kort litteraturstudie på feltet blitt gjennomført. Mangelen på en konsistent notasjon for de involverte begreper og størrelser gjør lesing av forskjellige utledninger unødvendig vanskelig og tidkrevende. Nye foreslåtte «generaliseringer» av teoremet fortsetter å dukke opp, så denne situasjonen pågår fremdeles. Videre har det også vært flere misforståelser i dette fagfeltet knyttet til hvilke områder dette teoremet kan anvendes på.

I nyere tid har noen artikkelforfattere prøvd å rydde opp i litteraturen, noe som har vært svært nyttig for nyere arbeid i feltet. Vi introduserer en konsistent notasjon, og utleder teoremet slik at dets bruksområder og begrensninger kommer klarere fram. Deretter vil vi studere noen analytisk løsbare eksempelproblemer som videre klargjør følgene av dette teoremet. Til slutt har vi også utført tester med denne metoden på en simulert APD ved hjelp av et eksisterende Monte Carlo-program som har vært utviklet i tidligere studentarbeider, og resultatene blir diskutert.

3 On The Basics of Semiconductors

This chapter gives a short introduction to the some of the topics of solid state physics which are needed for modelling semiconductor crystals. The aim is to present an introduction to semiconductors and some terms used frequently, and go into how the theory is used in a Monte Carlo particle simulator. First, the properties of semiconductors in general are briefly summed up in 3.1. Later, we will go over a few properties of $\mathbf{Cd_xHg_{1-x}Te}$ in section 3.2.

3.1 Semiconductor Materials

Solid state materials are usually categorised as metals, semiconductors or insulators based on the electrical conductivity of the material. Metals have the highest conductivity, insulators have the lowest, and semiconductors comprise a vast range of materials in between. The key point in semiconductors is that the same material can show properties of varying conductivity. This is a crucial element in all electronics, as neither insulators nor metals can show the same degree of versatility of the conductivity based on externally controllable factors, like temperature, introduced external fields, doping, and so forth. At very low temperature, the electrons are bound in the valence band, leaving the conduction band mostly void of any carriers, both holes and electrons. This state leaves the semiconductor as an insulator, whereas at room temperature, the electrons are thermally excited into the conduction band, making the material conductive.

As mentioned, the key point is that the conductivity is more or less easily controlled and regulated by the user. This can either be done reversibly in real-time via external electrical fields or via temperature changing, or permanently, with mixing in impurities into the material.

Most, if not all, semiconductors are comprised of atoms of one or more types, arranged in a periodic lattice in the form of a crystal. The earlier mentioned doping of

the material, or introduction of impurities, can greatly alter the conductivity, and thus the electronic properties of the material.

To envision this, imagine a lattice of carbon, or silicone atoms. Both have 4 valence electrons, and form a perfect periodic lattice. Now, substitute a select few of these with Arsenic or Phosphorus, both with five electrons in the valence band each. Four of these electrons will be bound to the neighbouring lattice atoms, fitting in nicely with covalent bonds to the adjacent four atoms. This will make the last electron bind loosely to the substituted atom, and thus could be easily excited into the conduction band (Kittel 1976 (2005)). The atom left behind is considered an ionized impurity, but still contribute positively to the conductivity of the lattice.

Atoms with one or more surplus electrons compared to the surrounding lattice are called donors, while the opposite case, of atoms lacking electrons compared to the lattice are called acceptors. Doping with donors makes the semiconductor an n-type semiconductor, and doping with acceptors yields a p-type semiconductor. In this thesis both n- and p-type materials will be considered. The material is assumed to be fully ionised at room temperature, thus the density of conduction band electrons is set equal to the density of donors and likewise the hole density is assumed equal to the acceptor density.

3.2 $Cd_xHg_{1-x}Te$

$Cd_xHg_{1-x}Te$, abbreviated CMT, is a compound ternary alloy semiconductor crystallized in the zinc blende lattice structure. It is composed of the semimetal HgTe and the semiconductor CdTe, x denotes the mole fraction of Hg in the alloy. CMT has a direct band gap which increases approximately linearly with x until reaching the band gap of CdTe which is 1.5 eV at room temperature. The band gap is also temperature dependent, an experimentally obtained formula for the band gap which is valid up to $T = 500$ K has been produced in articles (J. P. Laurenti 1990), and is given below:

$$E_g = E_{g0} + \frac{6.3\cdot(1-x) - 3.25x - 5.92x\cdot(1-x)}{11\cdot(1-x) + 78.7x + T} \times 10^{-4} T^2$$

.

Here, $E_{g0} = -0.303\cdot(1-x) + 1.606x - 0.132x\cdot(1-x)$ and E_g are given in eV, and T in K.

The variable band gap and the high absorption coefficient make CMT a very useful material for various applications. It can be tuned to absorb wavelengths in the interval 0.7 — 25 µm, thus covering most of the infrared region of the electromagnetic spectrum. CMT is an important material for the industry of infrared detectors and photodiodes, and especially useful for fabricating highly sensitive avalanche photodiodes, later in the text referred to as APDs.

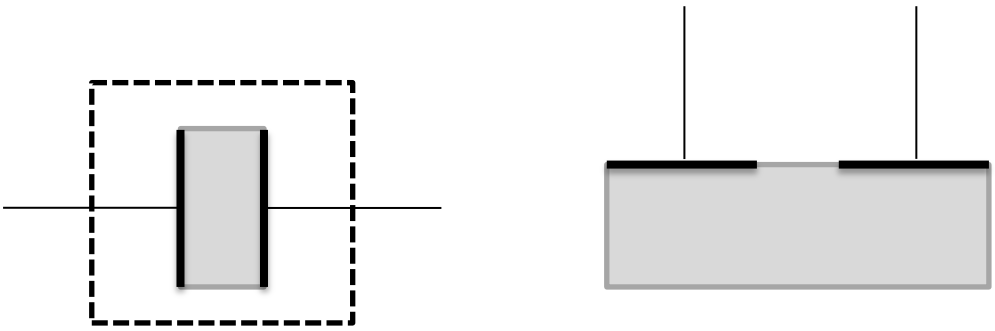
The simulations mentioned and referred to in this thesis are done with an alloy fraction of $x = 0.28$ and, unless otherwise stated, at 77 K. These conditions correspond to a band gap E_g = 0.2 eV.

The different doping densities used in this thesis are:

- $2{,}5\cdot10^{17}\,\mathrm{cm}^{-3}$ in the n+ region, the right corner.
- $5\cdot10^{14}\,\mathrm{cm}^{-3}$ in the i region, the right side of the depletion region.
- $1\cdot10^{16}\,\mathrm{cm}^{-3}$ in the p region, the left side of the depletion region.
- $1{,}95\cdot10^{17}\,\mathrm{cm}^{-3}$ in the p+ region, the left corner.

3.3 On Displacement Currents and Similarity to Capacitors

We will start by going deeper into how our device component behaves, and try to understand it by comparing it to a capacitor. The basic principles apply to both types of devices, the concept of displacement fields – and currents.



Capacitor with dielectric material between electrodes, for illustration purposes enclosed in a «black box».

Setup of electrodes with different geometry.

Figure 3-1. Depicted above is a simplified capacitor, “filled” with a dielectric material between the electrodes. The left one is also presented with a virtual “black box” for understanding the concept of displacement currents

Assume there is a voltage difference $V \neq 0$ over the two ends, and that the dotted line represents a “black box” covering the capacitor. The current coming in at the left end and exiting at the right end will at time $t = 0$ be measured to be the same as it would have been if the capacitor had not been there at all, i.e. short-circuited. The current will rapidly decline, however, proportional to e^{-t}, similar to what can be expected from a regular capacitor.

The reader might ask why this introduction is in place here. The simple answer is, some of the effects observed in the device are similar to what can be observed in a capacitor. This is especially true for the *displacement current*. As mentioned in the introduction, a single photon of appropriate energy arriving onto the APD device will excite an electron and produce a measurable current pulse. This current pulse, however, is measurable *before* the electron and hole avalanches even reach the

contact edges. This is the displacement current mentioned above. In chapter 4, some calculations around this are performed, clarifying this.

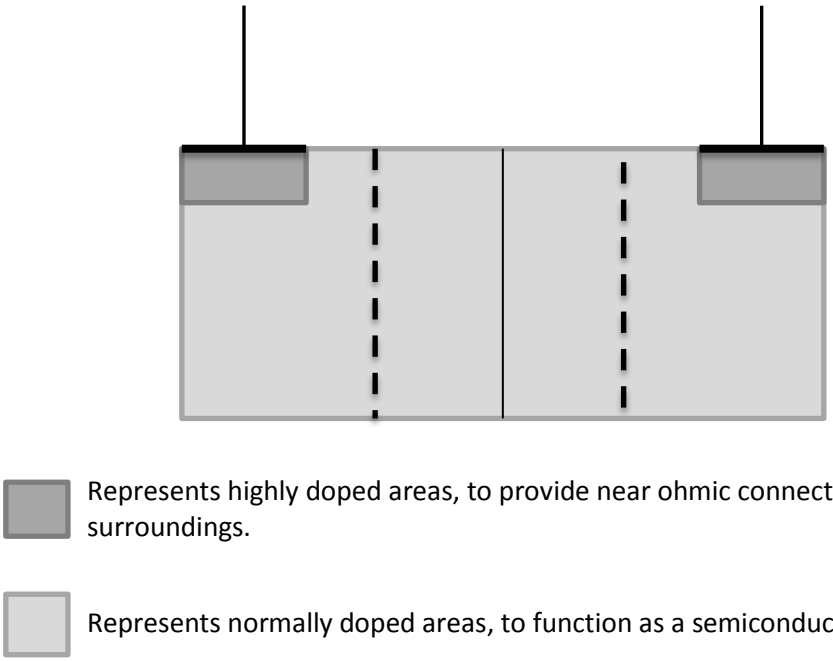


Figure 3-2. An illustration further clarifying the similarities between the device in question and a traditional capacitor. This is the typical design for an APD device. The area between the two dotted lines indicates the *depletion region*.

A photogenerated electron/hole-pair is at time $t = 0$ somewhere in the middle of the device. Due to the external reverse bias field, which is considered static for the duration of the simulations, they start to separate, with the hole moving left, towards the conductor with the lowest potential, while the electron is moving right, towards the higher potential conductor. This process generates in total four different currents, which will be considered independently in later chapters.

- The displacement current induced in the LEFT electrode from the HOLE.
- The displacement current in the LEFT electrode from the ELECTRON.
- The displacement current in the RIGHT electrode from the HOLE.
- The displacement current in the RIGHT electrode from the ELECTRON.

4 Calculating Currents

4.1 Induced Electrode Currents – Basic Observations

The purpose of this chapter is to find a systematic way of calculating the current going into and out of metallic electrode leads due to the movement of charged particles in the vicinity of the electrodes and electrode leads. We shall use the electrostatic equations to form a quasi-static, time dependent approximation. The ambition is not to include currents induced by magnetic fields, nor to calculate the current density or charge distribution on the electrodes themselves. We are only interested in the total current passing through the electrode leads.

We begin this chapter by re-deriving a method described in two essential original papers, Ramo (Ramo 1939) and Shockley (Shockley 1938). The method described in these two papers was at the time motivated by a growing need for calculating transient and high frequency electrode currents in vacuum tubes. During a transient, it was observed that an anode current began to flow long before any electrons could have passed through the vacuum from the cathode to the anode. In other words, the anode current could not be calculated by simply counting the number of electrons travelling inside the device and arriving at the anode per unit time.

A capacitor represents another, more clear-cut case where counting electrons arriving at an electrode from the inside of the device does not properly describe the outside electrode current. If the capacitor has no leakage current, the current inside the device is always zero, even if there can be a substantial transient current to and from the electrodes during charging and discharging of the device. Since an ideal capacitor never has any internal device current, it perfectly illustrates that currents inside a device and currents in electrode leads outside the device can be very different.

During charging of a capacitor, an outside voltage source draws electrons from one electrode plate, becoming positively charged, and forces them to pile up on the other plate, which becomes negatively charged. As electrons begin to pile up on the negative plate, they create an electric field inside the capacitor which gradually forces or displaces the remaining electrons on the positive plate away from that plate, making it more positively charged. The “collaboration” between the outside voltage

source and the internal electric field creates a continuous flow of electrons in the leads. However, the electron flow inside the capacitor is by no means continuous. In the ideal capacitor, the current in the electrode leads stops abruptly on the electrodes, and charge piles up. The density of such charge is described by the volume divergence $\nabla \boldsymbol{D}$. According to Maxwell's equations:

$$\nabla \times \boldsymbol{H} = \boldsymbol{j} + \frac{\partial \boldsymbol{D}}{\partial t}$$

which gives

$$\nabla \cdot \left(\nabla \times \boldsymbol{H} \right) \equiv \mathbf{0} = \nabla \cdot \left(\boldsymbol{j} + \frac{\partial \boldsymbol{D}}{\partial t} \right)$$

This means that the quantity $\boldsymbol{j} + \frac{\partial \boldsymbol{D}}{\partial t}$ has no divergence which again means that a volume integral of this quantity is always zero. If the volume of integration contains one electrode of the capacitor and runs through the area between the electrode plates and through the lead of the contained electrode (where the field ***D*** is zero), we end up with

$$\int_{Volume} \nabla \cdot \left(\boldsymbol{j} + \frac{\partial \boldsymbol{D}}{\partial t} \right) dv \equiv \mathbf{0} = \int_{Surface} \left(\boldsymbol{j} + \frac{\partial \boldsymbol{D}}{\partial t} \right) \cdot \boldsymbol{n} dA$$

where the surface to be integrated over is the surface of the volume of integration, and ***n*** is the outward normal. The main contribution of the surface integral will come from the cross-section of the lead where $\boldsymbol{j} \neq \mathbf{0}$ and $\boldsymbol{D} = \mathbf{0}$, and from the area between the electrode plates where $\boldsymbol{j} = \mathbf{0}$ and $\boldsymbol{D} \neq \mathbf{0}$. The above equation asserts that the divergence of the quantity $\boldsymbol{j} + \frac{\partial \boldsymbol{D}}{\partial t}$ is conserved, which means that if we interpret it as a "current" density, the total current entering the area between the electrode plates is the same as the total current exiting the lead. So, the quantity $\frac{\partial \boldsymbol{D}}{\partial t}$ between the electrode plates eventually becomes a transient, real current density ***j*** elsewhere, namely in the lead, and the surface-integrated sum of ***j*** and $\frac{\partial \boldsymbol{D}}{\partial t}$ is conserved.

In the general cases of a vacuum tube, a leaky capacitor, or a semiconductor device we will have a mixture of a real current ***j*** and a displacement current $\frac{\partial \boldsymbol{D}}{\partial t}$ inside the device. In the leads however, we will only have the real current ***j***. In steady state DC cases, $\frac{\partial \boldsymbol{D}}{\partial t} = 0$, and we will only have the real current ***j***.

If we have a moving lump of electrons travelling the electrodes of a device, they will certainly contribute to $\frac{\partial \boldsymbol{D}}{\partial t}$. Other contributions to $\frac{\partial \boldsymbol{D}}{\partial t}$ many come from a time-varying external voltage source, just as in the case of an ideal capacitor. This contribution can be calculated analytically or using a numerical Poisson solver, and represents the effect of charge carriers piling up at the electrodes. The contribution to the current in the leads from the real current ***j*** inside the device is found by counting the number of electrons in the lump arriving/leaving each electrode inside the device. Contributions from the electrons in the moving lump inside the device can perhaps best be illustrated by the concept of mirror charges or the method of images. We imagine that the alteration of the position of the lump alters the position of the corresponding mirror charges in the electrodes, giving rise to a transient current in the electrode leads. To comply with the voltage boundary conditions prescribed by a given external voltage source, the contribution from the method of images is assumed to obey the external voltage, so no additional voltage change is allowed. That is, the method of images procedure must be performed with all electrodes grounded.

The Shockley-Ramo procedure can be applied in cases where the calculation of the total $\frac{\partial \boldsymbol{D}}{\partial t}$ by a Poisson solver is difficult or impractical. Although we have illustrated the effect of the internal lump of moving electrons by using the method of images, using the Shockley-Ramo theorem historically ended a period of extensive use of the method of images in vacuum tube theory. In fact, if we are only interested in knowing the currents in the electrode leads and not the total ***D***, we shall see that we do not need to consider the method of images at all.

So, in summary, the current due to carriers that never leaves the electrode or lead area and therefore does not enter the vacuum region is NOT calculated by the Shockley-Ramo theorem. Displacement current induced when charging or

discharging an ideal capacitor is a typical example. In steady-state DC conditions this component of the lead current will be zero anyway.

What is calculated, is the specific displacement current in the leads due to carriers moving between the electrodes inside the device.

In a traditional calculation of the current in the leads with a full Poisson Solver, one would calculate the current in the leads by counting the number of carriers arriving/leaving the corresponding electrode and also add the total displacement current just outside the electrode, assuming that the displacement current adjacent to the electrode will turn into a real current inside the leads. Note that these calculations are performed over a surface area adjacent to the electrodes. Assuming that there is no displacement current allowed in the metallic electrodes or leads themselves, this sum will give the total, real electronic current in the leads.

If the Shockley-Ramo theorem is used, however, the real particle current inside the device is the driving force for a corresponding internal displacement current. The calculation is not performed over a surface area adjacent to the electrode in question, but over the whole interior volume of the device. Furthermore, the explicit counting of carriers arriving/leaving an electrode is no longer required, which we shall see simplifies the calculations.

In the Shockley-Ramo scheme, as long as a carrier is moving in the interior of the device, it creates an interior displacement current which turns into a real “evacuation” particle current in the surrounding electrodes. When the carrier arrives at an electrode and is collected by it, the current due to this carrier abruptly stops flowing, with the collected carrier filling in for the evacuated charge. Now this carrier is outside the realm of Shockley-Ramo analysis. The charge evacuated from the electrode is sometimes called the “induced charge”. One way to understand it is that the induced charge represents the continuation of the movement of the collected (stopped) carrier into the leads. For example, if we have a two-terminal device, and another carrier leaves the other electrode at exactly the same moment, the current will continue to flow in both electrodes and can be continued to be described by Shockley-Ramo analysis. In this case the carriers inside the leads behave as if they were placed on a

chain of pearls; when one interior carrier arrives at an electrode, it forces an exterior carrier to pop out from the other electrode and become the new interior carrier sustaining the Shockley-Ramo analysis. The external operating voltages of the Shockley-Ramo (S-R) analysis are assumed to be fixed (i.e. clamped) and superposition is assumed to hold. The S-R theorem will give the contribution from moving internal carriers with the external electrodes short-circuited. Charge build-up on the electrodes is allowed to balance the moving charges in the interior. Additional charge build-up on the electrodes gives the necessary contribution to satisfy the non-zero external operating voltages. This contribution must be treated as an “Add-on” to the S-R analysis, for example by using the capacity coefficients of the device.

If the flow of external electrons is obstructed by a resistance in the lead circuit, or if the external circuit is open, there will certainly be a voltage build-up at the electrodes. As an alternative to using the capacity coefficient approach, there are also later extensions of the S-R theorem that can deal with unclamped voltages on the electrodes. But this subject-matter is beyond the scope of this report. We have seen with pn diodes that internal carriers in equilibrium can create a built-in potential that affects and is only seen by the Poisson solver, but not by an external voltage clamp. Non-equilibrium contributions from photogenerated carriers can be seen both as exterior voltages and by the Poisson solver.

Usually, the field from the extra photogenerated carriers at low-level irradiation will not notably influence total electric field in the device, so in those cases the movement of interior carriers is not affected, and the possible external voltage perturbations are neglected. In cases of high-level irradiation, these effects cannot be neglected, and the lack of self-consistency in our Ramo-Shockley analysis becomes apparent. Counting the internal photocarriers arriving at the electrodes can in those circumstances be helpful for assessing when non self-consistent approaches must be abandoned. In high-level irradiation cases one would also use the superparticle approach with the standard method of calculating the electrode currents by counting arriving carriers and calculating $\frac{\partial \boldsymbol{D}}{\partial t}$. In any case, whatever the approach will be, we see that combining self-consistency and compliance with the actual electrical characteristics of the readout circuit outside the device can become an issue,

especially as these factors are related to what constraints the Poisson solver must obey.

If the external voltage clamp varies with time, an additional internal displacement current contribution $\frac{\partial \boldsymbol{D}}{\partial t}$ with origin in the non-mirror, voltage producing charges on the electrodes is also introduced, just as in an ideal capacitor.

If we study the articles of Shockley and Ramo, we see that only Shockley appears to have any concern for the limits of validity of the theorem and the other aspects that we have discussed in this section. It is therefore appropriate to cite Shockley's viewpoints here:

"Before discussing what effect the moving charge has, we must introduce certain conventions as to what part of the total field is to be attributed to the charge and what part to other causes. It is most convenient to consider that all of the conductors are grounded and to examine the currents to them through the external circuit due to the motion of the charge. If the voltages on the conductors are varying, however, charges will be induced and currents will flow as dictated by the coefficients of capacity. In keeping with the superposition principle, the net current is found by adding the currents induced by the moving charge (or each moving charge if there are several) and the currents due to changing voltages.

As was pointed out above, the current just calculated, although it is expressed in terms of fields produced by potentials on the electrodes, is that induced by the motion of the charge and does not include currents produced by changing potentials upon the conductors.) Since we have used a theorem of electrostatics in our theory, the results will not be valid if retardation effects are important within the volume throughout which the charges move; they will be valid, however, if these effects are small, even for large transit angles of the moving charges." (Shockley 1938)

This will be revisited and further discussed in section 4.3 with examples clearly illustrating the effects mentioned in this section.

4.2 Currents Induced by Electron Motion

We begin this chapter by re-deriving a method for calculating currents due to internal electron motion described in two essential original papers, Shockley (1938) and Ramo (1939).

Ramo's Proof

In 1939 Simon Ramo presented the following simple equation for the contribution of a moving electron in vacuum to the current in the lead of an electrode A,

$$i_A = e\boldsymbol{E} \cdot \boldsymbol{u}$$

Here i_A is the instantaneous current in the lead or wire with positive direction towards the electrode in question, e is the elementary charge, $\boldsymbol{u}$ is the velocity of the electron, and $\boldsymbol{E}$ represents the local electric-field vector at the position of the electron produced by a 1 V "test" or "weighting" potential at the electrode, with all other electrodes grounded.

In this text we shall opt for a slightly different notation to promote clarity and readability, while also making the equation more correct. The reader may observe that the dimensions for this expression clearly do not correlate. This is due to Ramo assuming a unit test/weighting potential, or simply, $V = 1$.

Including the «missing» potential factor, which we shall refer to as V_A^W, the equation would look like this

$$V_A^W i_A = e\boldsymbol{E}^W \cdot \boldsymbol{u}, \text{ or}$$

$$i_A = \frac{e\boldsymbol{E}^W \cdot \boldsymbol{u}}{V_A^W}$$

Our notation differs from Ramos' since we include two indices, whereas his article uses only one, the lower one. The upper index, in this case the W, denotes the "situation", or case, with W being an abbreviation for 'weighting' associated with the

test case. The lower index, A, is the index for the location of each particle or conductor. Thus, V_A^W means the potential in the test case on electrode A. V_e^W similarly represents a potential in the test case, at the position of the electron, wherever that may be. Aiming for a higher level of generality, we shall use a general "source" charge q instead of only electron charge. Therefore,

$$i_A = \frac{e\boldsymbol{E}^W \cdot \boldsymbol{u}}{V_A^W} \Rightarrow i_A = -\frac{q\boldsymbol{E}_q^W \cdot \boldsymbol{u}}{V_A^W}$$

where we have also added the location where the weighting field $\boldsymbol{E}^W$ is to be evaluated. Another important point worth reminding ourselves is that the whole process is viewed as linear, i.e. note the linear relation between $\boldsymbol{E}^W$ and V_A^W. This in turn means that one can find the effect of particles and fields by considering each simple case. Whereas the index W (meaning 'weighting') is interpreted as the case in which other conductors than A are electrically grounded, the calculated current contribution i_A is for the "normal", grounded case, and therefore appears without an upper index.

Technically, the current contribution i_A is one which we can calculate with all electrodes grounded, including electrode *A*. It is actually the mirror current contribution discussed earlier, appearing by virtue of the superposition principle as an independent component adding to the total electrode current, valid also when all voltage sources are re-set back to their actual values. The normal case and the test case for Ramo's analysis are depicted in Figs. 4-1 and 4-2, respectively. In the normal case, all electrodes are grounded and the "source" charge carrier is present, but it is surrounded by an infinitesimal metallic sphere which is just an aid for the proof of the theorem.

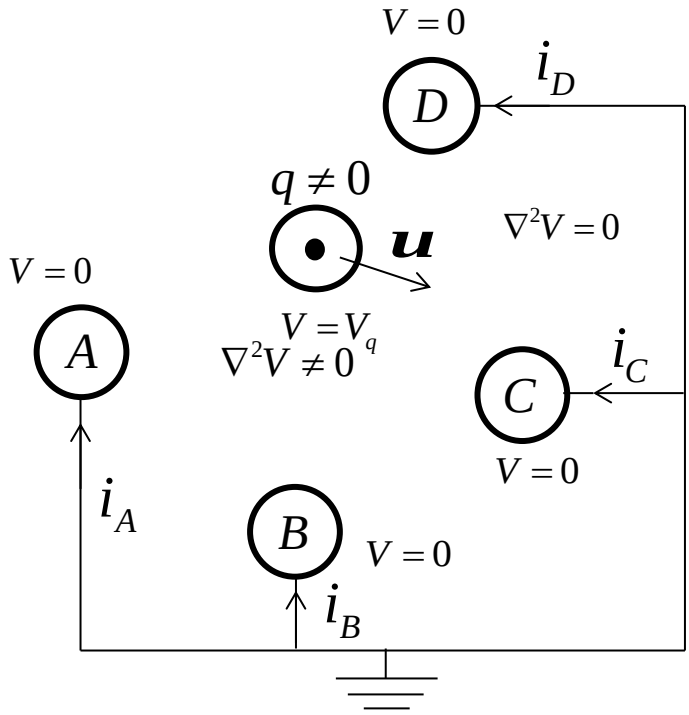


Figure 4-1. Ramo's model in the normal case. Letters *A*-*D* represents different conductors. All conductors are grounded, except for the small metallic sphere surrounding the free charge q. Compare with the "weighting" case, Figure 4-2.

In the weighted case, the electrode A for which we want to calculate the current is raised to the test or weighting potential V_A^W and the source charge carrier is removed, but the tiny metallic sphere is still present.

Before discussing the effect and influence of the sphere, it might be important to point out that the potential V_q actually denotes the potential at the sphere surface, <u>not</u> at the charge q. Therefore, we can also set $V_q \approx V_s$, with "s" denoting "sphere".

As we have seen from the discussion of the capacity coefficients, there is a linear relation between charges at the given conductors and voltages at the conductors. Therefore, in the test case, there is a linear relation between $\boldsymbol{E}_q^W$ and V_A^W so that the actual V_A^W value used becomes irrelevant. One might even be tempted to re-state the

relation and view the test voltages and fields as a small perturbation upon a background of actual electrode voltages and actual fields, i.e.,

$$i_A = -\frac{q\partial \boldsymbol{E}_q^W \cdot \boldsymbol{u}}{\partial V_A^W}$$

Typically, one would do this in situations where the material in the interior of the device is not simply a passive linear medium or vacuum. In principle, all space charge in the interior should be removed when doing the evaluation of the current. But there could be situations with a background current of little interest and a small-signal current caused by extra carriers which was highly interesting. Then a "linearization" upon this background would seem profitable. The response of the medium together with the internal background carriers could be nonlinear, but might well be approximated by a linear relation if the small-signal response indeed was a small perturbation. When the internal background carriers in the medium consist of stationary space charge, we see that the S-R theorem tends to dismiss their direct influence on the current because of their vanishing carrier velocity. Nevertheless, a more indirect influence could persist. With background internal carriers, background currents, background electrode carriers and electrode voltages at their actual values, the task of evaluating the perturbative small-signal response can eventually fail because the interaction between carriers or currents of interest and carriers or currents of "no interest" was not properly accounted for.

From the above discussion we can conclude that the internal particle velocity $\boldsymbol{u}$ is an all-important quantity with quite a direct influence on terminal currents. In the context of the Shockley-Ramo theorem, it is often viewed as "God-given" i.e. the trajectory of the particle is known. This is a convenient approximation for small-signal analysis, where the field contribution from a small group of internal signal electrons can be considered negligible when compared to a dominating "background" contribution from e.g. the electrode voltages. The key to the simplification offered by the S-R theorem clearly lies herein, for this allows a reasonably simple computation of the electrode currents without the need for a fully self-consistent solution of the total electric field.

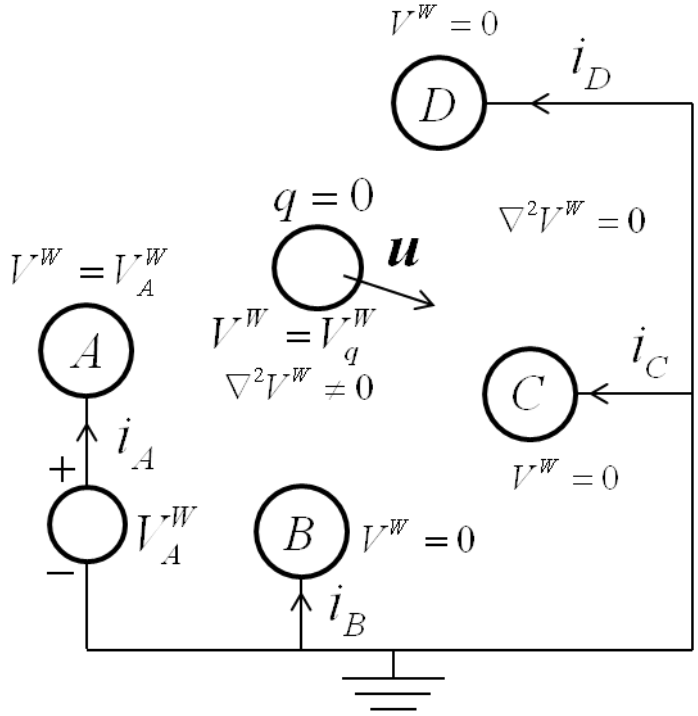


Figure 4-2. Ramo's model in the weighting case. Similar situation as depicted in Fig. 4-1, but with conductor A raised to the weighting voltage V_A^W and the source charge carrier removed.

If there are many signal electrons, a small signal analysis will not be sufficient, as the path of a given signal electron will be affected by the other signal electrons. This does not lead to a breakdown of the S-R theorem per se, it only signalizes that $\boldsymbol{u}$ should be calculated self-consistently according to the total electric field. So, in conclusion, lack of self-consistency in the total electric field governing $\boldsymbol{u}$ manifests itself in $\boldsymbol{u}$ only, since the total electric field does not explicitly appear in the S-R theorem. Indeed, a self-consistent calculation of $\boldsymbol{u}$ can take away the utility of the S-R theorem to such a degree that a full calculation of $\boldsymbol{j} + \frac{\partial \boldsymbol{D}}{\partial t}$ would indeed be preferable.

Before we proceed, it is appropriate to remind the reader that a metallic equipotential sphere of neutral initial charge will not influence the outside electric field stemming from any enclosed internal charge(s). This is due to the fact that an enclosed charge, for example an electron in the following illustration 4-3, will polarize the enclosing sphere, creating a lack of electrons along the inner surface, while creating a surplus

of electrons on the outer surface. This will, in turn, create a field directed inwards. The small sphere can always be centred on the electron so that the symmetry of the outside electric field is still maintained identical to what it would be without the sphere.

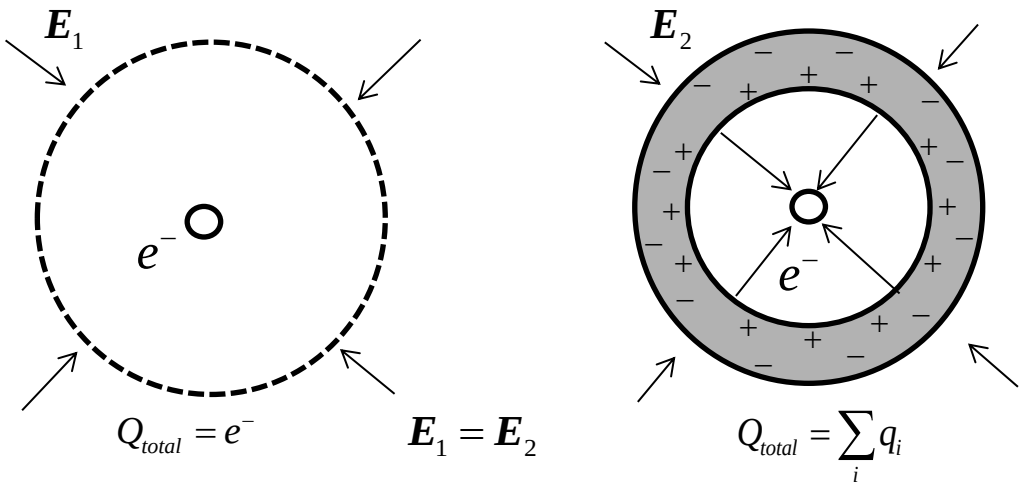


Figure 4-3. Depicting the phenomena mentioned above, regarding the effect of an initially neutral metallic sphere enclosing a charge – the outside electric field will be equivalent to what it would have been without the sphere.

An infinitesimal equipotential metallic sphere will also not influence the contribution from the charges on the electrode conductors when the sphere is empty, since it can be made arbitrarily small. Therefore, in summary, the contribution from the electron and the contribution from the charge on the electrodes are not affected by the presence of the infinitesimal metallic sphere.

Now we shall proceed to prove the S-R theorem following the original approach of Simon Ramo. We first remind ourselves of some mathematical relations,

$$\int \nabla \times \boldsymbol{A} \cdot \boldsymbol{n}\, dA = \oint \boldsymbol{A} \cdot d\boldsymbol{s}$$

$$\int \nabla \boldsymbol{A}\, dv = \int \boldsymbol{A} \cdot \boldsymbol{n}\, dA$$

$$\boldsymbol{E} = -\nabla V$$

and, above all, Green's 2'nd Theorem (for a proof, see the Appendices):

$$\int (V_1 \nabla^2 V_2 - V_2 \nabla^2 V_1)\, dv = -\int (V_1 \nabla V_2 \cdot \boldsymbol{n} - V_2 \nabla V_1 \cdot \boldsymbol{n})\; dA$$

with v representing the integration volume. A lower-case v is used, not to confuse it with the potentials V, which are arbitrary potentials present in the same geometry. The latter may be taken to mean potentials in the normal case and the test case, respectively.

Furthermore, $\boldsymbol{n}$ is the outward normal vector of unit length at the surface of the integration volume. Ramo uses a special integration volume stretching out to infinity, but excluding the conductor surfaces as well as the surface of the tiny metallic sphere. So the surface integral does NOT include the metallic surfaces of the conductors; the integral is actually over the outside borders of the vacuum adjacent to the electrodes. The integration volume is therefore a Swiss cheese-like structure with the outward normal vector $\boldsymbol{n}$ pointing into the electrodes and also directly towards the centre of the small equipotential sphere, see Fig. 4-4.

Now, apply Green's Theorem as stated above. The volume integral becomes zero, since $\nabla^2 V = 0$ and $\nabla^2 V^W = 0$ over the integration volume, which is in fact void of charges. Therefore we must have

$$0 = -\int (V^W \nabla V \cdot \boldsymbol{n} - V \nabla V^W \cdot \boldsymbol{n})\; dA$$

Remember, we are not integrating over the interior of the sphere where the electron actually is located. Any point charge would represent an infinite charge density at that point, $\nabla^2 V = \infty$. It is important to note that the sphere is of zero initial charge, and again, that the surface of the Swiss cheese cavity surrounding the sphere is NOT a part of the sphere surface. Note also that the sphere surface can be charged even if the total sphere charge is zero.

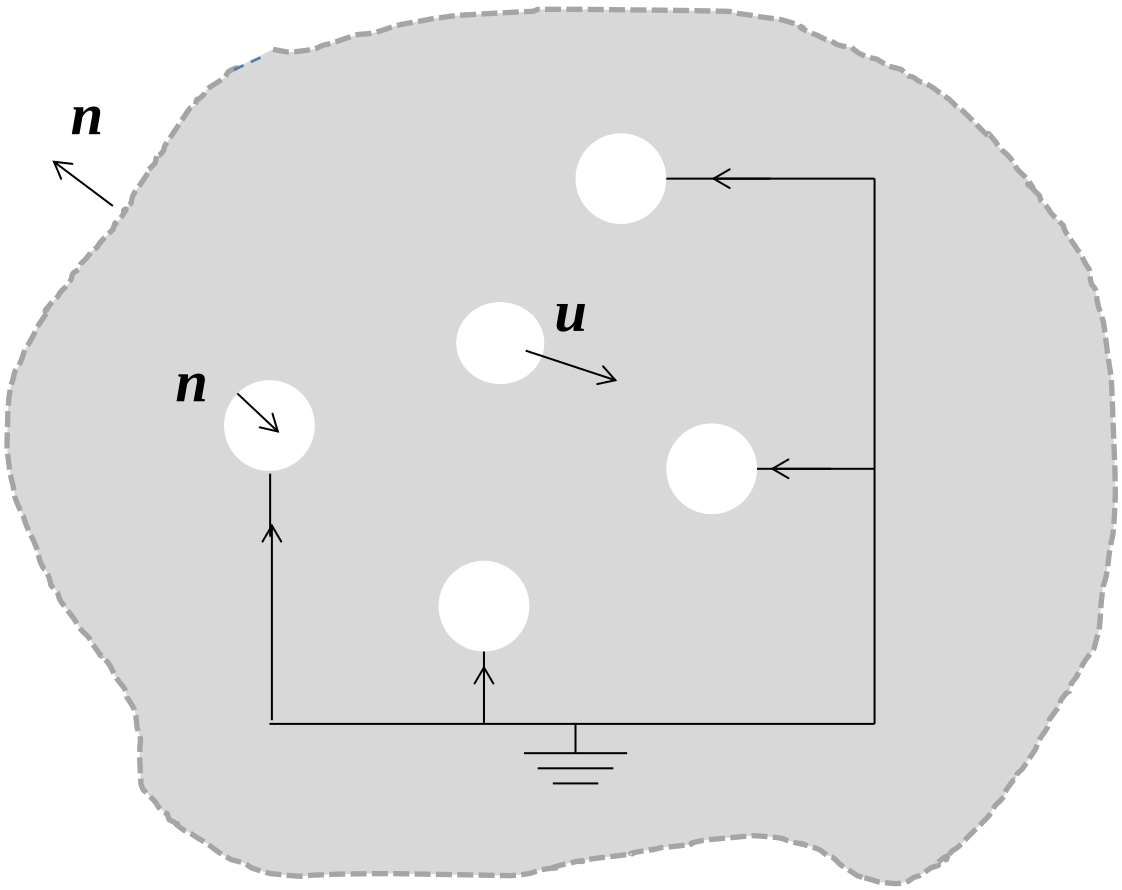


Figure 4-4. Ramo's volume of integration using Greens theorem. The integral is over a "Swiss cheese"-like structure, i.e. a body with holes, where the holes are to be filled with electrodes. Note that it continues to infinity. The grey area indicates the volume integrated over.

If we ignore the integration at infinity, which we shall discuss later, the integral over the surface on the LH side of Greens Theorem can be split into three parts:

1. The integral over surfaces enclosing all conductors except A and the sphere:

This integral is 0 since $V^W = V = 0$ over the grounded conductors, for obvious reasons.

2. Over the surface enclosing A:

The expression $\boldsymbol{n}\cdot(V^W\nabla V - V\nabla V^W)$ reduces to $\boldsymbol{n}\cdot(V_A^W\nabla V - 0\cdot\nabla V^W) = \boldsymbol{n}\cdot V_A^W\nabla V$ since on A we must have $V^W = V_A^W$ in the test case, where the only field present is the field created by the potential on A, and since V on A is zero for the normal case when A and all other electrodes are grounded. This results in a contribution $-V_A^W\int\nabla V\cdot\boldsymbol{n}\,dA$ from A, which, of course, is largest from the side of A that is closest to the other electrodes, where ∇V will be larger.

3. Over the surface enclosing the sphere, noting that ***n*** points <u>into</u> the sphere:

$$-V_q^W\int\nabla V_q\cdot\boldsymbol{n}dA + V_q\int\nabla V_q^W\cdot\boldsymbol{n}\,dA$$

$$= -V_q^W\frac{1}{\epsilon_0}q + V_q\cdot 0,$$

with the latter term becoming zero because integrating ∇V_q^W over this surface gives 0, since we also must have

$$\boldsymbol{E}^W = -\nabla V^W$$

and

$$\frac{1}{\epsilon_0}\int\boldsymbol{D}^W\cdot\boldsymbol{n}\,dA = -q\frac{1}{\epsilon_0}$$

giving

$$-\frac{1}{\epsilon_0}\int\boldsymbol{D}^W\cdot\boldsymbol{n}\,dA = 0$$

for the test case when the charge is removed and there is <u>no</u> charge present within the sphere.

The calculation from 2) is now

$$\int V_A^W(-\nabla V\cdot \boldsymbol{n})dA = -V_A^W\left(\frac{Q_A}{\epsilon_0}\right),$$

taking the special direction of $\boldsymbol{n}$ into account. Here Q_A is the charge on electrode A in the normal case when one electron is present between the electrodes and all electrodes have the same potential, defined to be zero. As we have already elaborated upon earlier, setting the voltages on all electrodes equal (to zero) is the same as removing (short circuiting the leads of) a voltage source, just as when finding the contribution of different sources in linear circuit analysis by the superposition principle. A simple set of such "test cases" - setting various voltage sources in the circuit in turn to 1 V or 0 V defines their individual contributions. During the application of voltage sources the current sources are removed, leaving their leads open, or equivalently in our case, removing the whole interior of the device so that no current can pass. Finding the contribution of a single moving charge inside a device closely corresponds to selectively restoring the interior of the device particle by particle back to a normal conducting state, still with short circuited electrodes eliminating the voltage sources. Here the movement of the interior charges is fixed, just as the current through an ideal current source. Thus, we have considered both voltage source and current source aspects when analysing the device, in the form of external voltage sources and internal particle current sources.

The analysis of the contribution from the test voltages can usually be done analytically or with a numerical Poisson solver. What is then solved is the Laplace equation, since the interior charges are removed. Historically, the analysis of the contribution from the charges inside the device has been done with the method of images. Q_A is nothing more than the mirror charge of the internal charges in the device, mirroring themselves in the different electrodes. As an internal point charge moves, the amount of mirror charge on different electrodes varies, as we have illustrated earlier. When the electrodes are short circuited, the mirror charges can pass freely from electrode to electrode, creating just the transient electrode currents which we want to calculate.

A word of caution is in place here, however. Diodes and transistors are the aim of this study, and they are inherently non-linear circuit elements, only the small signal response is linear. They are also very far from ideal current or voltage sources in

other respects, as they also show responses that can be capacitive, resistive or inductive.

A valid question that often arises in conjunction with the S-R theorem is often put forward: “Is there ***really*** a charge on A when all conductors are grounded? What charge is there on the other conductors then, and what about the integral over them in 1?” After our discussion so far we can truly state that: 1)Yes, there is also charge on A and the others, but not on the sphere. 2) The surface integral nevertheless becomes 0 because $V = 0$ and $V^W = 0$ over the integral in point 1. These questions are obviously also related to the integral at infinity. Ramo makes no comment about this integral, tacitly assuming it is zero. In the original paper by Shockley however, there is a metallic electrode surface enclosing the whole integration region. No comment is made about this, but by taking an integral of the electric field inside the metal we realize that this guarantees the sum of charges inside the integration region to always become zero. In other words, the charge and the mirror charges will sum up to zero. This is also a very common situation for a device. All the charge is gathered from regions close to the device. But as we shall see later, this is not an actual prerequisite for the theorem, because the surface integral at infinity vanishes even if there is some charge at infinity, provided that the charge close to the device is sufficiently well localized.

Since the volume integral in Green's theorem is zero, we are left with only two non-zero contributions from the integral over the surfaces on the right-hand (RH) side of Green's theorem. The equation therefore simplifies to

$$0 = -V_A^W \frac{Q_A}{\epsilon_0} - \frac{q}{\epsilon_0} V_q^W ,$$

The first term is accounting for the contribution over the surface enclosing A, and the second term is over the surface enclosing the sphere.

This equation might seem counterintuitive at first, because the last term with V_q^W is related to the test case with the charge q removed, yet q is still present in the same term. But we must remember that the reason for this is that there are two different situations (and thus two scalar fields V) in the same geometry used in our calculations with Green’s theorem. Each term consists of a product of both situations.

Two situations intertwined like this is what leads to the reciprocity found in these equations.

Our equation can also be rewritten as

$$Q_A = -q\frac{V_q^W}{V_A^W}.$$

In these equations, Q_A is the induced mirror charge on electrode A caused by a charge q at a position $\boldsymbol{r}$ in an electrostatic setting. The only extra assumption in this respect is that the sphere is isolated and charge neutral. This assumption is not a contradiction, since the sphere is not an electrode, as it is not connected to anything. We note that there was no such assumption on charge neutrality of the electrodes. In fact, one may view the finite charge q at $\boldsymbol{r}$ as if it has been taken from the electrodes in the first place, therefore they are not charge neutral.

The charge Q_A, which is a result of the charge q being present, should be independent of the chosen weighting voltage V_A^W on the electrode A. This will be the case if the problem is linear, so that the potential at the sphere's surface (with q removed), V_q^W, is proportional to the weighting voltage V_A^W with a factor of a proportionality constant only dependent on the geometry of the problem. We discussed this in the context of the capacity coefficients earlier.

As can be recalled from electromagnetism, if we now decide to displace the charge q, a corresponding displacement of the mirror charge at the electrode A towards the other electrodes must take place. Thus, the moving electron will contribute to a certain current into electrode A.

This amount is equal to

$$i_A = \frac{dQ_A}{dt} = \frac{-q}{V_A^W}\frac{\partial V_q^W}{\partial \boldsymbol{r}} \cdot \frac{\partial \boldsymbol{r}}{\partial t} = \frac{q}{V_A^W}\boldsymbol{E}_q^W \cdot \boldsymbol{u},$$

where $\boldsymbol{u}$ is the velocity of the electron. This finalizes the proof of the result provided by Ramo in his article. We saw that the contribution from the surfaces of the integration region was the decisive and crucial element in the proof.

Shockley's Proof

In the Shockley proof, on the other hand, the decisive element is the contribution from the volume of the integration region, with no contribution from the surfaces. The reason behind this difference is that Shockley chose a different integration region, and a different approach to the charge outside of the electrodes, see Figs 4-5 and 4-6.

Instead of a point charge enclosed by a metallic, uncharged sphere, Shockley uses a metallic sphere which can either have a charge q in the normal case or be without charge in the test case.

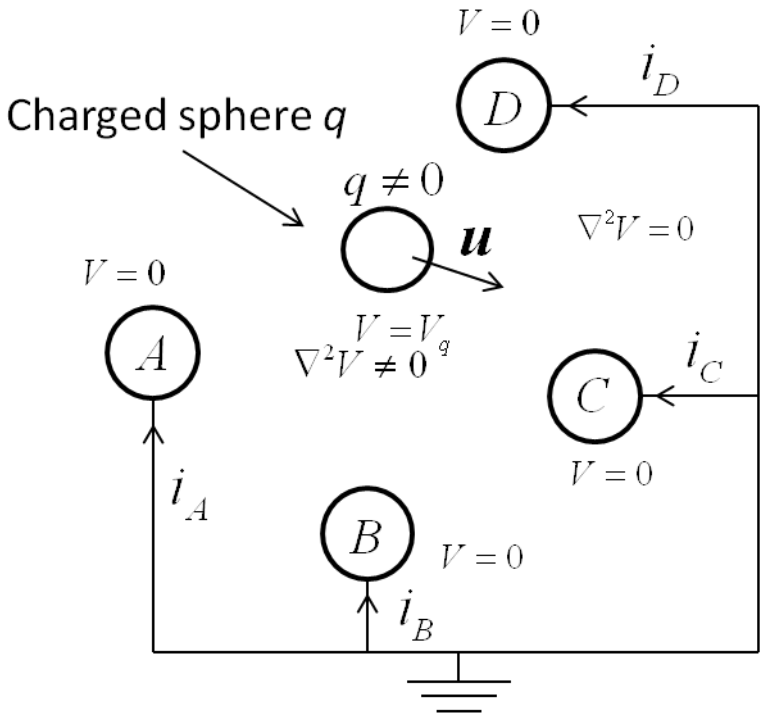


Figure 4-5. This is the layout for Shockley's method, in the normal case.

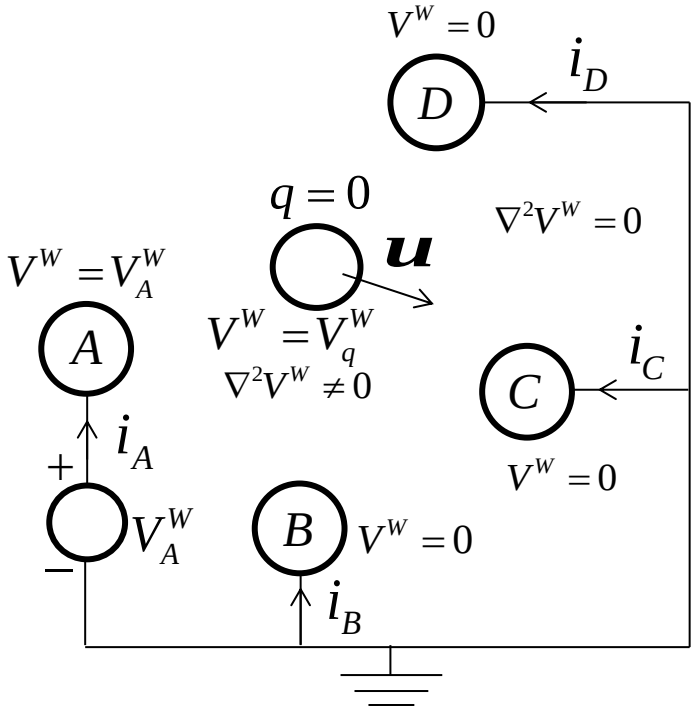


Figure 4-6. This is the layout for Shockley's method, in the weighting case. Note that the weighting cases of Ramo and Shockley are very similar, except for the charged particle.

Shockley's integration region, see 7, includes all the conductors, so that the surface of the integration region is at infinity, where the potential tends to zero, along with the charge density. The surface then consists of $V^W \nabla V$ and $V \nabla V^W$ terms, and since we can consider our device to consist of localized charge, these terms will at most go as

$$\frac{1}{4\pi r} \cdot \frac{1}{4\pi r^2} \sim \frac{1}{r^3}.$$

The first factor stems from the V term, and the second factor is related to the $\nabla V^W = \boldsymbol{E}$ term. The above equation is to be integrated over an infinite size sphere surface $4\pi r^2$, which gives $\sim \frac{1}{r^3} \cdot 4\pi r^2 \simeq \frac{1}{r}$, which obviously tends to zero as r increases. This simultaneously confirms that the outer surface at infinity in Ramo's case also will not contribute.

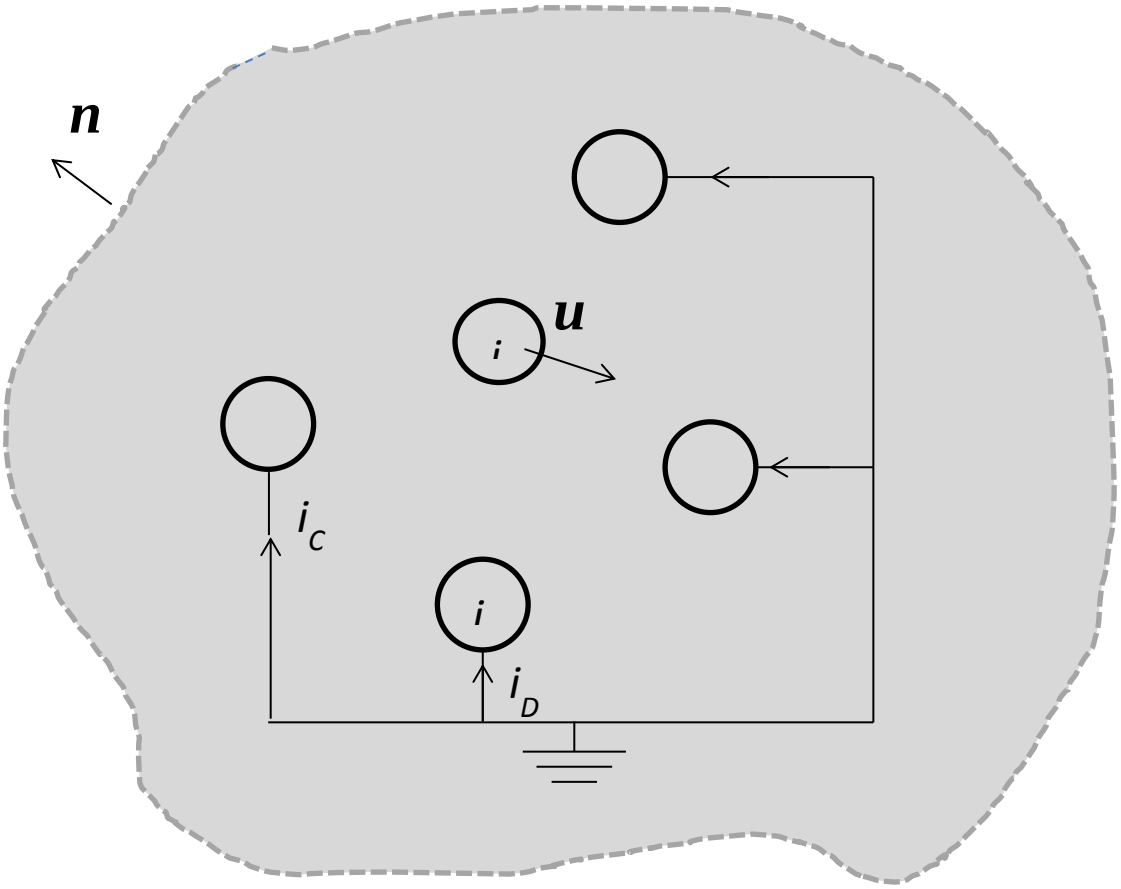


Figure 4-7. Shockley's volume of integration using Greens theorem. The grey area indicates the actual volume integrated over.

So, in this geometry, Green's second theorem reads

$$\int_{\text{all space}} (V^W \nabla^2 V - V \nabla^2 V^W) dV \equiv 0 \quad .$$

One might recall from quantum mechanics that this equation is nothing more than the statement of that the ∇^2 operator (which in QM represents the kinetic energy operator) is Hermitian.

From Poisson's equation, we know that $\nabla^2 V = \frac{-\rho}{\epsilon_0}$, so the integral becomes simplified to

$$\int (V^W \rho - V \rho^W) dv = 0 \,,$$

where ρ^W is the space charge for the test case (with the electron removed), and v the integration volume.

Now, imagine that the charge representing the electron is distributed on the surface of the small metallic sphere, instead of inside the sphere as in Ramo's case. Here, the sphere itself is either charged or neutral. The neutral state represents the test-situation with the electron removed, whereas the charged state mimics the presence of an electron.

Using the fact that the potential on a metal conductor is constant, we thus obtain from volume integration the identity

$$\sum_{i \in \text{all conductors}} (V_i^W Q_i - V_i Q_i^W) = 0$$

This is the starting point of Shockley's proof, and it is often called Green's Reciprocation Theorem.

Following the Ramo-Shockley method with the test potential case and a normal "electron-present" case, it leaves us with

$$V_A^W Q_A + V_q^W q - V_q Q_q^W = 0$$

$$\Rightarrow Q_A = \frac{-V_q^W}{V_A^W} q,$$

since $Q_q^W = 0$, due to the fact that charge on the sphere is removed for the test case. As stated above, it is important to recall that the conducting enclosing sphere is never grounded, let alone electrically coupled to anything at all.

This sums up Shockley's argument, leading to the conclusion

$$i_A = \frac{dQ_A}{dt} = \frac{-q}{V_A^W} \frac{\partial V_q^W}{\partial \boldsymbol{r}} \cdot \frac{\partial \boldsymbol{r}}{\partial t} = \frac{q}{V_A^W} \boldsymbol{E}_q^W \cdot \boldsymbol{u}$$

4.3 Example Situation Using Shockley's Method

In this subchapter, a closer look at an example situation is performed, using Shockley's method to calculate instantaneous current. In the setup below, we have two infinitely large conducting plates, separated by a distance d. Between them is a charged particle moving at an arbitrary speed towards plate number 2.

Using the equation derived above in the most verbose form as starting point:

$$I_i = \boldsymbol{E}_i^W(r) \cdot \boldsymbol{v} \cdot \frac{q}{V_A^W}$$

I_i is the instantaneous current at conductor 1 at any time. $\boldsymbol{E}_i(r)$ is the field experienced at point r from conductor 1. $\boldsymbol{v}$ is the velocity vector of the charge carrier moving, q is its charge, and lastly, V_A^W is the weighting potential the particle experiences.

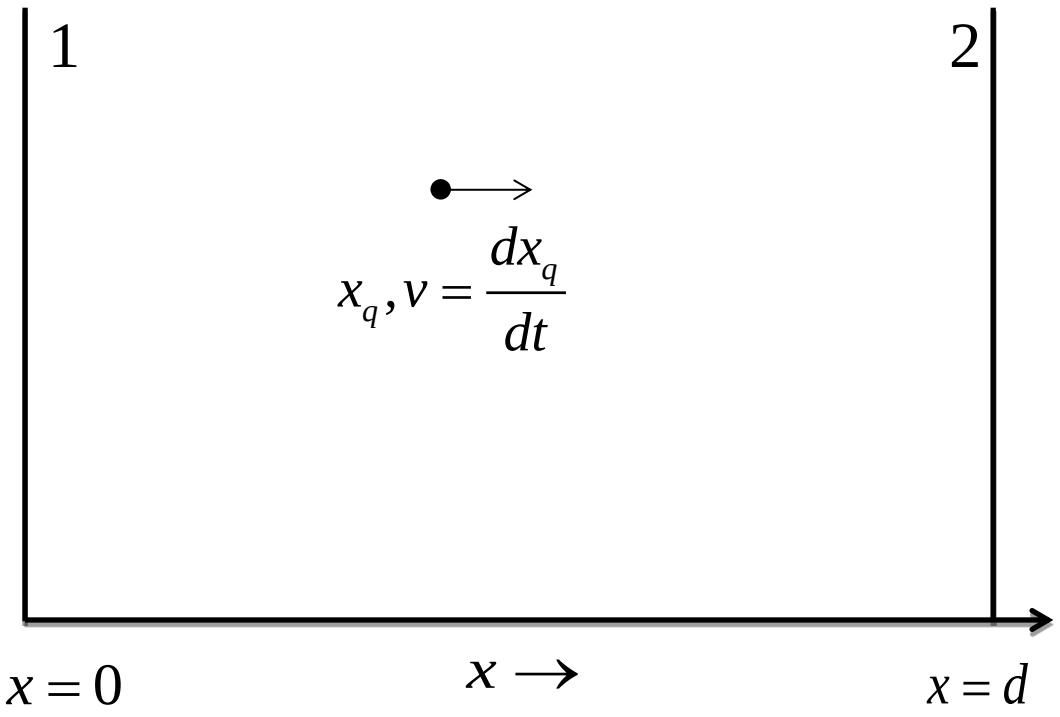


Figure 4-8. A depiction of Shockley's method to calculate current in a relatively trivial, but relevant and instructive nonetheless, setup.

Also, keep in mind that we are working on two separate cases, so that by using the superposition principle we can determine the current on one electrode by ignoring the other completely. Thus, we will start by finding I_1.

Common electromagnetics tells us that the field between two infinite plates separated by a distance of d, is constant and in our situation equal to

$$\boldsymbol{E}_i^W = V_A^W / d.$$

The sign for this equation might seem confusing, but stems from the fact that we have reversed the sign on V, since the particle is moving in the same direction of the $\boldsymbol{E}$-field. The sign will be reversed for I_2.

Substituting for $\boldsymbol{E}_i^W$, we obtain:

$$I_1 = \frac{V_A^W}{d} \cdot v \cdot \frac{q}{V_A^W} = \frac{qv}{d}$$

By inspection, one can similarly obtain

$$I_2 = -\frac{qv}{d}$$

since the only differences are the sign for V.

Note that all of the above derivations are implicitly using the velocity $v = v_x$, meaning we are strictly looking at the part of the total velocity that is parallel to the field, and thus automatically normal to the plates in our geometry.

If a carrier were to scatter in the vacuum between the plates, it would get an angle θ compared to the field, and thus the above derivations would need a slight modification, namely by multiplying a $\cos\theta$ term to the total velocity, or just use vectors from the start, but only look at the component in the x-direction. So if the carrier were scattered by exactly 90 degrees, the current generated would drop to zero, as the velocity of the carrier suddenly would be 90 degrees, and thus normal, to the field.

4.4 Some Theoretical Results on Displacement Current

For a moment, consider the *charge* that passes through this setup for one electron/hole pair generation.

As a simple starting point, we will look at the current induced in a closed circuit between the two (infinitely large) conductor plates, separated by a distance of d as mentioned above. This is the exact same scenario as figure 4-8. First, some simple starting formulas; it is known that

$$-I_{1e} = I_{2e}, -I_{1h} = I_{2h}.$$

This equation states that, although the current induced in plates can be different for electrons and holes, *the current from one charge carrier is the same on each conductor, only of opposite direction.*

$$I_{tot} = I_e + I_h$$
$$I_{tot} = -I_{1e} + I_{2e} + I_{1h} - I_{2h}$$

The signs can be chosen arbitrarily, it depends on which direction the current is set to go. In this current calculations, we are assuming *positive* current towards the conductors, and the electron starting in position x, and then going right. The hole, also starting in the same position, is moving to the left, meaning that those signs are changed. The mere fact that currents are defined positive towards the conductor plates, mean that one of them will have to be negative to make them additive.

Using the formulas for the current I, the equations obtained in the above subchapter, can be further used:

$$I_1 = -\frac{qv}{d}, I_2 = \frac{qv}{d}$$

Note that the signs are switched for the equations compared to the above subchapter. This is due to the geometry of the problem at hand, and since we defined. The current will, as can be seen by inspection, go from right to left, meaning *from* electrode 2 *to* electrode 1, in a closed circuit. As is known from electromagnetism

$$Q_{tot} = \int I_{tot} dt = \sum_{\text{all carriers and conductors}} \int_0^{T_i} I_i dt \,,$$

meaning that the charge is obtained by looking at the current induced from the electron in conductor 1 and 2, and then the same for the hole. This means we have to sum the four integration terms. Luckily, our equation gets simplified to

$$Q_{tot} = -2\int_0^{T_e} \frac{v_e q_e}{d} dt + 2\int_0^{T_h} \frac{v_h q_h}{d} dt$$

Again, d is the distance between the plates, v_e, v_h is the speed of the electron and hole respectively, q_e, q_h represent the different charge on the electron and hole.

$$T_e = \frac{d - x}{v_e}$$

is the time it takes for an electron generated at an arbitrary position x to reach plate 2.

$$T_h = \frac{x}{v_h}$$

is similarly, the time it takes for the hole to reach plate 1. This is assuming constant speed of the electrons, which might not necessarily be the case in reality, but it still illustrates the theory. Performing the integration on the left side yields

$$\begin{aligned} Q_{tot} &= -2\left[\frac{v_e q_e t}{d}\right]_0^{\frac{d-x}{v_e}} + 2\left[\frac{v_h q_h t}{d}\right]_0^{\frac{x}{v_h}} \\ &= -\frac{2v_e q_e (d-x)}{v_e d} + \frac{2v_h q_h x}{v_h d} \\ &= 2\left(\frac{-v_e(-e)d}{v_e d} - \frac{-v_e(-e)x}{v_e d} + \frac{2v_h e\, x}{v_h d}\right) \\ &= 2\left(e - \frac{e\,x}{d} + \frac{e\,x}{d}\right) = 2e \end{aligned}$$

This may seem obvious when given some thought, but it has one interesting consequence. Again, this was done assuming a closed circuit between the two conductor plates. However, due to the symmetry of the electromagnetics involved, one can see that the half of the above determined charge, namely

$$q = e$$

travels through *one* of the conductor plates from $t = 0$ at generation up to the time the electron uses to hit plate 2, or the hole to hit plate 1, whichever takes longest.

This is the core of displacement current mechanics. Imagine that an observer was to sit on plate 2, and that the electron hits plate 2 before the hole reaches plate 1. There would still be an observable current being generated at plate 2, generated by the hole somewhere in the middle moving away from plate 2. The observer would never see the hole, and yet it would affect the current in the conductor.

4.5 Mirror Charges and Capacity Coefficients

The method of mirror charges was developed as an aid in solving problems that were deemed highly nontrivial to calculate purely analytical. For some problems, it is a very suitable approach, especially concerning charges in relation to very large (in comparison) conductor plates.

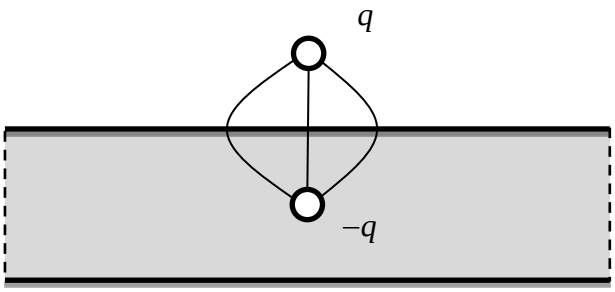


Figure 4-9. A simple setup, with a conducting plate that is infinitely large in the x-direction, with a charge q placed some arbitrary distance d from the plate, thus the distance between the two charges is 2d. The charge given as $-q$ is its *mirror charge*.

In Figure 4-9 given above, a simple example setup is given for how a mirror charge can be used to simplify the problem. The charge q outside the conductor plate generates an electrical field around it, which forces the surface to have a nonzero surface charge density. This charge can be viewed as another carrier, of opposite charge. This obviously does not exist physically, not as a separate charge somewhere inside the material. The charge $-q$ is in reality smeared over the surface out to infinity, but the E-field at any point above the plate is completely identical to how it would be had the conductor plate not been present at all. The reader should note that this assumes a conducting plate, for any non-conducting material it would complicate things.

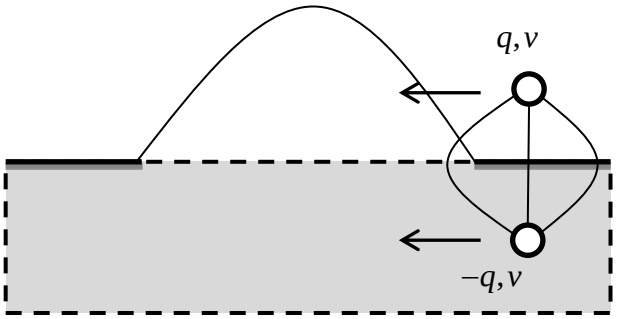


Figure 4-10. This is a slightly alternate setup, similar to the capacitor setup mentioned in some above chapters. The dotted lines are the physical edge of the device, and the thick lines on the top are metallic connection points. The leads are also short-circuited to make the example clearer. Here, the real physical charge is inside the material, and the charge induced on the metallic connection points can be represented by a mirror charge outside.

In the above figure, 4-10, one can again clearly see the resemblance to a capacitor for our device. The physical charge is inside, and is moving towards one of the conductors. This induces a surface charge on the metal conductor edges, but an equivalent virtual mirror charge can be thought to exist outside the material. When the charge inside the device moves, so does the mirror charge. In reality, the charge is spread over the metallic conductor centred around on the position of the charge inside along the x-axis. Although not rigorously correct, it's possible to get an intuitive understanding of the current generated with this method of mirror charge. Imagine the charge in the above illustration is moving left, and the mirror charge following. By the time the real charge has made its way to the other side of the device, so have the mirror charge. And to get there, the mirror charge has necessarily had to go through the circuit connecting the two conductors, thus driving a current.

To move the real charge density on the conductors, represented by the virtual mirror charge, charge, and thus current, *must* have gone through the circuit in the transition from figure 4-10 too.

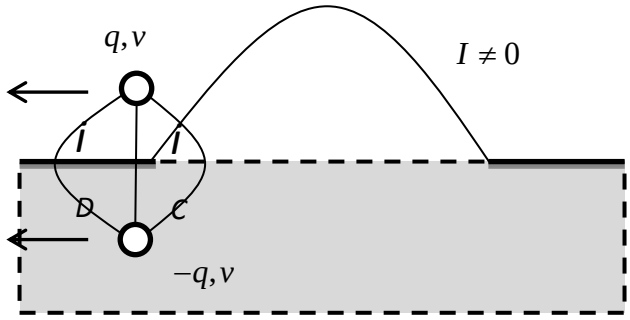


Figure 4-11. Same example as in figure 4-10, after the charge inside the device has moved across the depletion region. Technically, the current I is at any instant zero, but there have been nonzero current at some point in the transition.

The method is usable for more complex geometries, but it will quickly involve advanced series operations, reducing the feasibility of an analytical solution, and due to the fact one often must set up an infinite number of mirror charger for more complex geometries. On a side note, Ramo himself said on using the method of mirror charges for such cases that is “lengthy and requires no little familiarity with methods of handling infinite series.”

Another useful approach is to look at potential coefficients (Svaasand 1981), and capacity coefficients (Svaasand 1981). To introduce these, first let three point charges be placed in any arbitrary points in space.

$$V = \frac{q_1}{4\pi\epsilon_0 r_1} + \frac{q_2}{4\pi\epsilon_0 r_2} + \frac{q_3}{4\pi\epsilon_0 r_3}$$
$$= \alpha_1 q_1 + \alpha_2 q_2 + \alpha_3 q_3$$

Alpha is, as can be easily determined by inspection, a function of the coordinates of the charges. This result is also valid if the charges were not, in fact, point charges, but the charges are given as a distribution over three distinct surfaces. The proof for this is along the lines of this:

$$V = \int \frac{\sigma_1}{4\pi\epsilon_0 r_1} dA_1 + \int \frac{\sigma_2}{4\pi\epsilon_0 r_2} dA_2 + \int \frac{\sigma_3}{4\pi\epsilon_0 r_3} dA_3$$

$$= \alpha_1 \int_{\text{conductor 1}} \sigma_1 dA_1 + \alpha_2 \int_{\text{conductor 2}} \sigma_2 dA_2 + \alpha_3 \int_{\text{conductor 3}} \sigma_3 dA_3$$

$$V = \alpha_1 Q_1 + \alpha_2 Q_2 + \alpha_3 Q_3$$

Q_n is the total charge on each of the conductors. The proof for the above statements goes along the lines of this:

First, let us assume there is only one conductor present. Let the charge on this be Q', with a surface distribution of σ'. The ‘ mark is only present to avoid any ambiguity with the Q used in the previous paragraph. This charge on the conductor sets up a potential V'. We now increase the charge Q' with a constant factor of k, $Q'' = kQ'$.

The result of this will be that $\sigma'' = k\sigma'$. This is necessarily true, since the old potential $V' = \text{constant} = V_1''$ on the surface, meaning the new one have to be too, or $V'' = \text{constant} = V_2''$. Violating this boundary condition is not physically possible, due to the fact that the charge is accumulating on the surface evenly and continuously. If the $\sigma'' \neq k\sigma'$, such an “impossible” potential would arise.

Thus, this shows that any multiplying factors inside the integral can be taken outside of the integral and combined into the α factor, as the proof is valid for the other subscripts as well.

These factors can be used to express the potential V_n on each conductor as a function of all the other conductors. That this can be done is clearly intuitively, but these capacity coefficients allow for much easier manipulation of the equations.

4.6 Connection between the Shockley-Ramo method and the fast multipole moment method

In an earlier project work (Norum, 2009), the fast multipole moment method (FMM) was considered for use as a Poisson solver in a Monte Carlo setting. Hailed as the 10'th most useful algorithm of all time, and as the most useful algorithm invented during the past two decades, it should show great promise for use in our context. The method was reviewed by Norum in his above thesis, and the conclusion was that although it was fast, there was a problem with accuracy. As far as the author is aware, only one MC program uses the FMM method, and that is the program written by Vasileska et al. (Vasileska et al. 2010). In this text the method is only treated very briefly, perhaps since it has not been much tested in the context of MC simulations. Solving for the potential in a 3D case, the speedup with respect to the fastest alternative methods was a factor of 2:20; the average speedup was a factor of 10. Since solving the Poisson equation is the dominant source of CPU consumption in full 3D simulations, this represents considerable savings.

Using a grouping of charge sources and a multipole expansion, the method is well suited for calculating e.g. gravitational attraction problem in astronomy, but the fulfilment of awkward boundary conditions in electrostatics such as a constant voltage at the electrodes can be difficult. The solution here is to divide the problem into two, using superposition. The field contribution from charges inside the device is calculated by a multipole expansion and the method of images, demanding zero voltage and only mirror charge at the electrodes. The non-zero voltage boundary conditions are satisfied separately by assuming no contribution from charges inside the device using another form of 3 D Poisson solver. An obvious candidate for this solver is the Bi-conjugate gradient stabilized method (Bi:CG stab) which is the main alternative to the FMM method. If we have a simulation with a constant external voltage, the time consuming Bi:CG stab method would only have to be used once at the beginning of the simulation to satisfy the boundary conditions, and thereafter only the FMM method would be used.

If the FMM method is chosen, the solver would consist of two parts, a Bi:CG stab solver and an FMM solver. If the FMM method is not chosen, the Bi:CG stab solver would perform the whole task of solving the Poisson equation.

Although the Shockley-Ramo method is not mentioned in Vasileska et al., we see that the lines of thinking leading up to the two methods are quite parallel. The difference is that when using the Shockley-Ramo theorem, we are not interested in finding the actual field contribution from the internal charges and currents. We already have the current if we can estimate the velocities of the charge carriers. This estimate of velocities is of course not fully self-consistent when we use the frozen-field approximation. The action of the frozen internal charges is included in the determination of the velocities, but their field is frozen and the field of photogenerated carriers is neglected, and assumed to be small. If we add the full back action of the internal field on the internal charge carriers, we would have to use the method of images in a self-consistent way. From the viewpoint of the Shockley-Ramo theorem then, we can argue that the FMM method is little more than adding the method of images to the S-R procedure, and we are immediately back into the setting of the 1930's using the method of images for calculation of both fields and currents. The "test" case in the Shockley-Ramo theorem could have been performed with e.g. the Bi:CG stab solver, and the method of images procedure with grounded electrodes corresponds to the "normal" case in the S-R method with grounded electrodes. A difference is of course that in the S-R scheme the respective electrodes are raised to the test potential one by one, so that the normal and test cases are intermixed, whereas in the FMM procedure the electrodes are set at their actual potential all at once and the field contribution is calculated by a Poisson solver, and then the second part with grounding of all the electrodes is performed and an extra use of the method of images is performed to obtain the total electric field.

We see that the S-R method is excellent for calculation of small-signal response, where linearization, superposition and neglect of back-action of the mirror charge (or as in our case, the neglect of the photogenerated part of the mirror charge) is a good approximation. The FMM method shows how to make the calculation fully self-consistent by adding the full back action of the mirror charge, but at the same time this takes us back to a method and an amount of labor similar to the pre S-R period in the 1930's. With full self-consistency included, the S-R method can still be used as a part of the procedure and a way of thinking, but it is not actually needed in the calculations, since in this case every detail of the problem is already fully calculated anyway.

5 Subsequent Developments in Terminal Current Calculations

5.1 Introduction

During the decades that have passed since Shockley and Ramo presented their original results, some progress has been made and new applications have appeared. A short resume over these developments and alternative viewpoints on terminal current calculations are therefore in order. Sometimes used in computational electronics as a generalization of Kirchhoff's current continuity law including an internal displacement current, original derivations of the S-R theorem assumed negligible magnetic and radiation effects (that is, they assumed quasi-electrostatics) and the application was to describe electron transport in vacuum tubes.

Based on the proofs given earlier in this report it is perhaps not very difficult to accept that the theorem also will be valid in homogeneous linear dielectrics. It has subsequently been generalized to systems containing inhomogeneous linear dielectrics (Pellegrini, 1986; Kim et al., 1991).

Perhaps the most common area of use for the S-R theorem today is in various particle accelerator detectors (Fabjan et al., 2004), in radiation detectors, and to a smaller degree in microwave detectors. During the last decade it has also found its way into biophysics, as some MC workers extended their computer programs and fields of interest towards atom transport in biological ion channels. Likewise, the theorem was recently a key tool in a MC analysis of an ion beam induced charge experiment on a Si diode structure (Olivero et al., 2011) which was based upon an earlier work that also discussed a suitable readout circuit (Vittone et al., 2000).

Fairly sophisticated MC simulations of Si photodetectors were carried out over a decade ago (Dubaric et al., 2002, Nilsson et al., 2002).

A useful update on analytical models of modern CMOS photodiodes can be found in the recent literature (Blanco-Filgueira et al., 2012).

We shall not explore the use of the S-R theorem in calculations of noise (Pellegrini, 1986) nor for studying the various effects of charge carrier trapping and recombination (Owens et al., 2006, Kim et al., 2011), as the former are beyond what

we had envisioned as the scope of this report and because the latter has not yet been fully integrated into our MC program.

For an analytical example of treating a depletion layer with the S-R theorem, see the article of Djuric et al. (Djuric et al., 1984.). Analytic examples of induced currents due to sudden changes in permanent polarization or changes in dielectric constant were given by Neyts et al. (Neyts et al., 2007). A very important review of “mistakes in the past” with the S-R theorem was given by de Visschere (de Visschere, 1990), where special emphasis was placed on pn/pin detectors and on *g-r* noise in pn junctions, commenting also upon former reviews by Dabrowski (Dabrowski, 1987, 1989).

Much of the literature deals with (mis)interpretations and generalizations of the S-R theorem. As with other reciprocity theorems of electromagnetism, it fails for non-linear materials. In such situations a linearization can be advocated, but this will usually not be without consequences for the accuracy of the result. Some of the “generalizations” are not even claimed to be generally valid, and only special cases are shown.

We shall confine ourselves to a quasi-electrostatic version of the S-R theorem with clamped electrodes. Although often used, it has never become commonplace in MC simulations, and in a relatively recent book on general semiconductor device simulation (Vasileska et al., 2010) it is not mentioned at all.

5.2 Two Charge Carrier Ensembles - The “Frozen Field” Approach

Initial mistakes in using the S-R theorem apparently occurred both for the basic, constant electrode voltage case and for the more complex constant charge case (i.e. induced voltage on an “isolated” electrode) (Cavalleri et al., 1963). For a long time there was a dispute on whether the theorem was valid in the presence of fixed space charge in the device, such as in a semiconductor diode with depletion regions. The space charge in a pn junction consists of the ionized donors and acceptors responsible for the doping. We discussed this matter earlier, and noted that according to the S-R theorem carriers at rest do not directly influence the terminal currents. As long as the total system is linear, stationary charge can also be treated by the theorem on the same footing as moving charge.

Given the fact that the space charge is actually not totally fixed, but rather slightly polarizable, it would be tempting as a further refinement to consider the background space charge ensemble as a part of the material properties, i.e. by just introducing a modified dielectric constant. Although this issue of validity in presence of space charge was first discussed several decades ago (Cavalleri et al., 1971, Gunn, 1964)), it tends to resurface from time to time. An important point was re-emphasized a few years ago (Gatti et al., 2004), commenting upon another work on this subject that was going to be published at the time (Kotov, 2005). Gatti et al. treated stationary charge (polarization charge) that could react to the presence of free moving charge as described by a variable dielectric constant of the material. They went on to state that considering the space charge as part of the background is acceptable as long as the space charge is restricted from macroscopic movement, i.e. polarizable, but not free. In a fully depleted semiconductor this criterion would be automatically satisfied.

The issue becomes more complicated in cases where there are other free moving charges in addition to the charges of interest. Regions of normal, doped semiconductor outside the depletion zones indeed contain carriers that are totally free to move. Hence there are two moving groups of particles, one group of little interest (the free background charges already present in reverse bias without photogeneration) and a group of interest (for example a photogenerated and impact ionization generated ensemble). The background bias ensemble could screen the

effect of the photogenerated ensemble in terms of induction of current on an electrode. It was concluded that: “In other words, the real velocity field of the particles is not known and should be calculated by means of a complete analysis of the device.”

This is indeed the situation we would have in an MC APD simulation (Bertazzi et al. 2010, Bellotti et al., 2011), where the carrier ensemble is divided into a background bias ensemble which provides a stationary, “frozen field” configuration in which the photo and impact ionization generated carriers move. At low irradiance levels we can describe the latter on a real electron and hole basis, and not as superparticles. In most cases the background ensemble electric field is directly fed into the MC simulation as a stationary electric field obtained from a Fermi-Poisson equation solver (Bertazzi et al. 2010).

Without photogeneration, the free background charge usually does not carry any appreciable amount of current, and especially so if we include the dark current, which is due to thermal carrier generation in and near the depletion zone, into the ensemble of interest. Under photogeneration however, it could happen that some of the photocurrent was actually transferred from the photocarriers and onto the free background carriers. As far as the problem is the changed velocities of both photogenerated and background free carriers as a result of short range interactions, we know from experience that it is only in special rare cases that carrier-carrier interactions change the result of a MC simulation. Since in unipolar transport the crystal momentum generated within a particular, single type of carrier (either electron or hole) is conserved, it is just transferred to another carrier of the same type. And in fact, on both sides of the depletion layer the carrier transport is unipolar, either holes near the p contact or electrons near the n contact.

In low irradiance cases where the photogenerated ensemble is a tiny perturbation upon a large free carrier background, it would be difficult to calculate the photocurrent with a full analysis involving MC methods, as Gatti et al. (Gatti et al., 2004) ideally suggest. Since MC simulations involve a lot of details about many individual particles, and the tiny photocurrent is produced by only a few photocarriers well below the resolution limit of a single superparticle, there are both superparticle resolution issues

and limitations due to the long simulation times needed for the photocarriers to be extracted from these usually large devices.

Admittedly, the rudimentary “Frozen-Field” interaction between the two moving ensembles which we shall use in this report could have been improved. The effect of the photogenerated ensemble upon the background free charge should not have been restricted to a crude form of generation/recombination, but expanded to include more general carrier-carrier interactions and long-range electric field interactions, showing the effect of the photogenerated carriers upon the background free ensemble. But a method for this would have to be developed. It does not seem that the previously mentioned paper (Bellotti et al., 2011) included any such refinements. At high irradiation levels, both photogenerated and background carriers could have been treated self-consistently within the same superparticle ensemble, although the computations would be very time-consuming.

Some possible keys to the ambitious goal of also including effects of the background carriers in the undepleted volume near the electrodes can be found in more recent papers (Heubrandtner et al., 1998, 2002a, 2002b), (Riegler 2002, 2004), building upon a special theory for a weakly conducting medium (Haus et al., 1989).

Potentially useful extensions of the S-R theorem for unclamped or isolated electrodes can be found in an older work (Gatti et al., 1982). In this paper time dependent weighting potentials are used. By “time dependent” it is understood that the voltages at the electrodes are unclamped and free to change according to the physical laws governing the problem.

All these papers mainly refer to particle physics detectors of various kinds, especially ionization chambers consisting of gases or liquids. These detectors have a wide application in X-ray and gamma-ray astronomy, nuclear spectroscopy and nuclear medicine. Remembering that all semiconductor detectors operate as solid-state ionization chambers, the physics of these devices is very similar, with carrier avalanches created by incident particles or photons, which are sensed by electrodes. We do not know if any of the new methods have yet been used to any significant extent for semiconductor detectors. The methods rely upon the assumption that as

we get closer to the contacts the material behaves more like a simple conductor with a finite resistivity in near equilibrium, so that the violent physics is restricted to other regions where MC simulation etc. is needed. Within our present use of the S-R theorem electrodes and leads actually constitute such a near equilibrium region, albeit with zero resistivity.

A recent article (Hamel et al., 2008) takes on the issue of a non-linear S-R theorem for a material with non-linear dielectric constant, and the authors go about to justify the derivative form of the theorem which we mentioned earlier in this report. They introduce the concept of “local linearity” for small voltage increments around the actual operating voltages and claim to have shown that it is valid in a particular example for an n^+p junction.

An extended version for the full electrodynamic regime including radiation has also been derived (Yoder et al., 1996, 1997). Most of the latter article actually dealt with how to reduce uncertainties in the calculation of terminal currents in MC simulations, but the former article presented a collective, particle density S-R like approach, in contrast to the original corpuscular, or particle based approach of Shockley and Ramo. Some would say that a particle distribution averaged theorem would not really be a S-R theorem at all, and that the term “S-R” should be reserved for the corpuscular viewpoint only. Another question is whether, in a complicated case with radiation and retarded potentials, it is useful to even consider an S-R like theorem instead of doing an analysis on a more free basis.

5.3 Practical Consequences of Our Frozen-Field Model

The use of a fixed charge carrier velocity based upon the electric field from the electrodes and the stationary background charges (i.e. a “frozen field”) ignores the contribution to the particle velocity from forces between the electron and hole clouds, from internal forces within the electron and hole clouds, and from forces due to the induced image charges in the electrodes upon the charge carriers. A recent study (Ettenauer, 2008), considers the consequences of neglecting these effects in a Ge

gamma ray tracking detector and concludes that this approximation was well justified. We would however like to add some comments on this problem for two irradiated textbook-type device structures which have been studied with the present MC program.

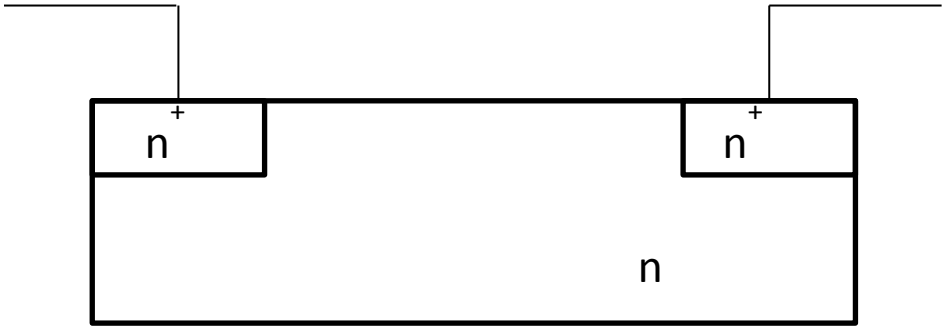


Figure 5-1. n-type HgCdTe slab

The first structure, Fig. 5-1, is an n-type HgCdTe slab with n^+ contacts on both the RH and the LH side, with a positive potential on the RH side and a negative potential on the LH side, irradiated on a spot between the contacts. As the faster electrons are pushed towards the LH electrode and leave the device, a surplus of photogenerated holes are left inside the device. This cloud will act to slow down the escape of electrons through the RH contact, but also to inject more new electrons into the device from the LH contact, since background electrons already present near the LH contact will be drawn further into the device towards the cloud of remaining excess photogenerated holes. All these activities will be governed by the Poisson solver.

When the radiation stops, the device must revert back to the equilibrium situation which existed before the radiation was applied and therefore to a field configuration corresponding to the frozen field. The details of how this happens depend on how strong the recombination process is. If there is no recombination, photogenerated electrons and holes must finally be pushed out of the device in an equal, total amount. In the limit of very strong recombination, photogenerated electron-hole pairs

recombine before any carrier can reach an electrode, and no carriers need to be injected or ejected.

In the frozen field approximation as we shall apply it here, there is no Poisson solver that can correct and supervise this process and the events at the contacts. Carriers are ejected as soon as they reach their respective positive and negative electrodes because they represent an excess charge with respect to a former equilibrium situation. Background electrons are prevented from reacting to a remaining temporary positive excess charge in the device.

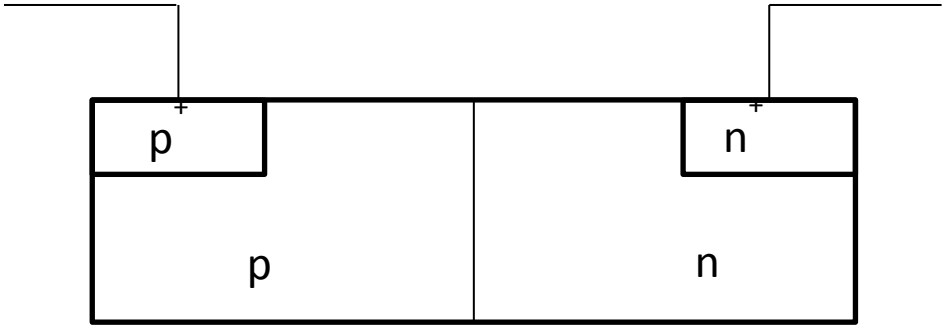


Figure 5-2. pn diode.

The second structure we shall consider is a pn diode in reverse bias, Fig. 5-2, with the hole electrode on the LH side and the electron electrode on the RH side. As the cloud of excess photogenerated holes are driven towards the hole electrode, they will push the background holes near the hole contact onto the electrode where they will exit the device and be neutralized by a corresponding current of incoming electrons from the hole electrode lead. The lead itself has zero capacitance and can therefore not store excess charge. Recombination within the device volume will of course diminish the pushing effect, but once the hole cloud is separated from the electron cloud and the holes find themselves on the p side there are very few electrons to recombine with.

The pushing effect is of course associated with a displacement current and the accumulation of background holes near the p contact. A background hole is expelled from the device and a hole from the advancing photogenerated ensemble can fill in

the empty position. After some time, many photogenerated holes eventually also reach the p electrode themselves and are expelled from the device. To what extent the background free holes are pushed into the electrodes or to what extent photogenerated holes travel to the electrodes themselves cannot be determined unless a self-consistent approach is used, with both carrier types in one ensemble. Holes are in principle indistinguishable, so the question of which hole ensemble arrives first is not the key issue. The point which we would like to make is that with more carriers available, a self-consistent solution would tend to short-circuit the electric field on the path towards the hole electrode, and thereby reduce the internal displacement current. This reduction could be offset by an increased particle current transient, so that the resulting current pulse in the lead would be less influenced. There are also restrictions due to the finite amount of charge generated by the photon, and the free background charge must revert back to its original internal configuration if the irradiation is a short pulse. Thus the photogenerated charge is very much in control in this kind of device.

Arguing that the current response in the lead would be faster if existing background charge near the hole contact was pushed into the contact, there is the immediate counterargument that the displacement current represented by the S-R theorem shows an instant relation already. Surely, in our frozen field approach, the movement of background holes towards the electrode clearly does not take place. As far as the interior hole particle current is concerned, the p electrode must “wait” for the photogenerated holes to physically arrive at the electrode. Therefore, the internal particle current pulse could be slightly delayed and diminished in our model, but this would not necessarily affect the total current response in the leads to the same degree. Neither could the time integral of the two currents be affected, as long as photocarriers avoid recombination or becoming trapped, this is determined by the total photogenerated charge.

In an earlier example we saw that a charge q has not passed through all cross-sections of the device and wires until the slow photogenerated hole had fully reached its hole electrode. Admittedly, the S-R theorem combined with the frozen field approximation successfully describes how the photogenerated hole cloud forces charge around in the exterior leads before these holes physically arrive at their electrode, forcing by means of an internal displacement current. Regarding the

physics at the electrodes, we have assumed that holes are immediately neutralized once they arrive, so there will not be additional accumulation of photocarriers due to poor contact. Should holes accumulate near the hole contact, a corresponding amount of neutralizing electrons accumulate on the hole electrode, by the principle of image charge.

There will also be a small, real transient free background particle hole current in the device which also has an associated displacement current. Both of those are ignored, as well as a correction of the photogenerated hole velocities due to these changes in the background charge carrier configuration.

We remember our result that the current into an electrode *A* is given by

$$i_A = \frac{dQ_A}{dt} = \frac{q}{V_A^W} \boldsymbol{E}_q^W \cdot \boldsymbol{u}$$

An internal particle that moves a macroscopic distance from position $\boldsymbol{r}_0$ to position $\boldsymbol{r}$ simultaneously pushes a charge $-\Delta Q_A$ (called the "induced charge") away from electrode A and onto the outer circuit;

$$-\Delta Q_A = -\int_0^t i_A dt' = -\int_0^t \frac{q}{V_A^W} \boldsymbol{E}_q^W \cdot \boldsymbol{u} dt' = \frac{q}{V_A^W}\left[V_{\boldsymbol{r}}^W - V_{\boldsymbol{r}_0}^W\right]$$

The particle current due to charge pushed away from electrode A is a continuous function in time even though the interior particle current contribution from the charge q is a moving delta function spike. The electrode charge Q_A is a distributed surface charge. Since the sum of the particle current and the displacement current densities is divergence free, the spike in the interior particle current must be offset by a corresponding spike in the interior displacement current, which is provided by the delta function nature of the electric field from the point particle.

Charge buildup on the electrodes is allowed in the S-R analysis because there is interior charge in the device. We see that the point charge q can be fully collected without electrode charge buildup if it has travelled from a position where the weighting potential is 0 and onto the electrode A. In other words, it will be collected

without electrode charge buildup if it has travelled from one electrode to another through the interior of the device. This is just another manifestation of Kirchhoff's current law for the device as a whole.

A charge that has been collected is to be considered as present but no longer moving. It just fills in the void in a stationary way, replacing the charge it has already evacuated from the electrode during the time it was moving towards that electrode. In the metallic leads, there are an abundance of electrons to convey in a continuous way this "evacuation current" created by one single internal charge carrier. The collected charge distributes itself on the surface of the collection electrode, becoming more and more focused as it comes closer. Mathematically, the S-R theorem only describes currents induced by moving internal carriers, but says nothing about the process of removing a carrier from an electrode surface or the collection of a carrier by an electrode. A view of an electron arriving at an electrode and travelling further into the leads without delay must be modified into a view where the arriving electron stops on the surface but pushes another electron into the lead. It is still the "string of pearls" analogy, but with one pearl tallied at a time. In other words, the internal electron can travel through all the internal cross-sections of the device, but it never actually passes through the electrode surface. The charge pushed through all cross-sections of the lead which we spoke of earlier is actually the evacuated charge corresponding to the charge of the stopped electron. All in all, it therefore seems that the corpuscular version of the S-R theorem blends very well in with our current MC simulation approach.

All photogenerated charge carriers are supposed to recombine in order to restore the state of the device before the photon was absorbed. One may say that electrons and holes pushed out of the device "recombine through the contacts". Remaining photocarriers in the device must effectively recombine with each other in order to restore the device to its original state. In practice, this net result is achieved by means of the background carriers. An isolated photocarrier already represents a local non-equilibrium situation. But there might not be another photocarrier nearby to recombine with. Therefore, the isolated photocarrier tends to recombine with a local background charge carrier, and "left over" photocarriers must remain within the

device as a substitute for the background photocarrier which was lost to recombination. As a consequence, in a frozen-field model where the interaction with the background carriers is not explicitly modelled, an electron photocarrier that recombines and is removed must be accompanied by the removal of a hole photocarrier.

5.4 Lessons Learned From Particle Physics Detectors

In detectors of the ionization type, an incident particle generates several pairs of charge carriers, such as electron-ion pairs in ionization chambers or electron-hole pairs in semiconductors. The number of pairs of charge carriers is proportional to the energy deposited in the detector volume. Electric fields in the detector volume move charge carriers around and cause a variation of the induced charge on the detector electrodes. The total induced charge is proportional to the number of charge carrier pairs, and therefore to the energy deposited in the detector volume. If so desired, pulse shaping circuitry can convert the output pulse so that its amplitude corresponds to the energy of the incident particle, as one of several possible readout modes.

Resistive Plate Cambers (RPC's) (Petrovici et al., 2002, Schuttauf et al., 2009) are basically constructed as a two-plate capacitor, with conducting or weakly conducting electrodes. Apart from being a central detector in particle physics experiments, they are also now considered for use in PET tomography, where electron-positron annihilation in the body creates signal photons of several hundreds of keV (Lippmann et al., 2009). Pertaining to the simple geometry of these devices, analytic models of the static electric field of a point charge in an infinite plane condenser (Jackson, 1975) with one or 3 different layers have been presented (Heubrandtner et al., 2002b) as well as considerations on the distribution of the charge on the electrodes (Samedov, 2009). The latter study was based on the method of images and conformal mapping, as the S-R theorem would only give a total charge and not the charge distribution. The need for knowing the actual field caused by the moving charges therefore still seems to reappear, especially as this need has been

suppressed in the past by extensive use of the S-R theorem which readily gives plausible results even with grossly approximated particle velocities. If many charge carriers are generated simultaneously, their actual velocities will be influenced by mutual interactions, represented by the fields of the individual charges.

Based on experience with MC avalanche simulations, even analytic expressions for the time response function of RPC's have been presented (Riegler 2009). Another particle physics detector deserving to be mentioned is the Wire Proportional Chamber (WPC) which has wire electrodes (Gruber et al. 2011).

Many suggestions for making detectors sensitive or insensitive to the position where the incident photon hit have been put forward, and also suggestions for making detectors sensitive to just one photocarrier type.

We realize that trapping of e.g. a hole will lead to a loss of electrode signal, and the remaining signal induced by the electron (the electron induced charge) will then come to depend on the *position* where the electron-hole pair was generated. In many cases this is an unwanted effect, because the induced charge will vary haphazardly with the point of photon impact and no spectroscopic information on the original photon energy can be retrieved.

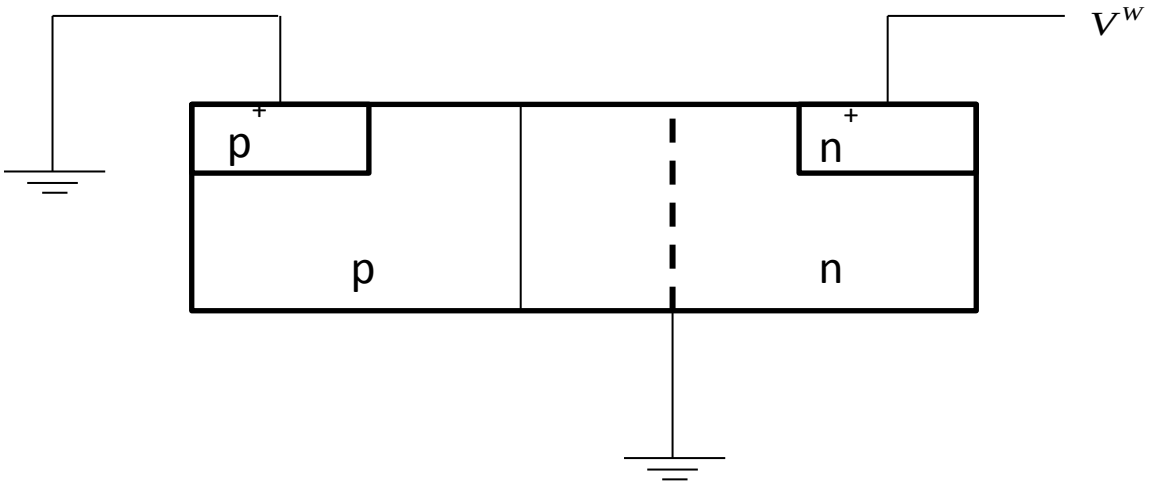


Figure 5-3. Frisch grid detector with weighting field for the electron electrode

By placing an additional, carrier penetrable and grounded internal grid electrode (Fig. 5-3) vertically between the left and right hand electrodes of a two-terminal device, so that the internal grid electrode divides the detector volume in two sections, the signals on the right hand (RH) electrode can be manipulated. Suppose we restrict photogeneration to the area on the LH side of the grid. Lying close to the right hand electrode, the grid causes the weighting field of the RH electrode to be screened and compressed into a small area. Charge moving on the LH side of the grid will not be sensed at all on the RH electrode, because the RH electrode weighting field is zero in that region. Then the signal on the RH electrode would sense only the electrons that have crossed the grid, and this signal will be fairly uniform for all electrons, and independent of whether the hole was trapped or if it completed its journey to the LH electrode. Also, the location of the initial electron-hole pair generation will not influence the signal. In other words, the detector becomes position insensitive and single carrier type sensitive. Since the RH electrode signal from the slow holes wandering towards the left electrode is obliterated, the otherwise slow response from the holes could be suppressed, whereas the rapid response from the electrons is kept. It is worth noting that the complete loss of signal from holes does not reduce the induced charge signal read out from the RH electrode for this three-terminal device, when compared to a conventional two-terminal device. As we have seen, it is the voltage difference in test potential over the electron path from grid to RH electrode (0 to V^w) that matters, not the total distance travelled from the generation site.

In semiconductors there are good reasons to suppress the hole signal, since the slow moving holes are prone to being trapped at defect sites, making the signal linger on for long periods (spectral tailing, (Owens et al., 2006)). This configuration of electrodes is called a Frisch grid detector after its inventor first had introduced it in the 1940's to overcome slow drift and loss of ions in gas-filled ionization chambers (He, 2000, Gook et al., 2012).

Instead of placing a grounded grid as an internal electrode, an invention in the mid 1990's introduced two banks of coplanar, interlaced external strip electrodes to be placed vertically on the RH side to replace both the grid and the single RH electrode (He 1997, 2000). The strips are connected in an alternate, interlacing manner so that one of the banks is the readout electrode (say electrode 2) and the other bank is grounded and has a function that resembles the Frisch grid (electrode 3). Since the

readout electrode bank (electrode 2) always has a positive operating voltage and the other bank is grounded, electrons will always be headed towards one of the strips of the readout bank. Putting weighting potentials on electrodes 2 and 3 in order to calculate the difference in electrode currents $i_2 - i_3$, we must actually calculate the difference between the weighting potentials on paths going from the LH electrode to the readout electrode strips. Far away from the RH side, the weighting potentials from the two banks will be similar, because there is a gradual "fanning out" of an individual electrode weighting potential with distance from the electrode. Closer to the electrodes, however, the individual weighting fields become very localized. Finding themselves on their preferred path close to the readout strip, electrons therefore only sense the weighting field of electrode bank 2.

The difference in weighting potentials between the readout electrode bank 2 and electrode bank 3 will now be zero along the path of the electrons, until they approach the two banks of electrodes, where readout weighting potential suddenly rises to V_2^W whereas the weighting potential from electrode 3 drops to zero. The function of this detector is therefore exactly like that of the Frisch grid detector, with the modification that a difference of weighting potentials replaces the former single RH electrode weighting potential, and a difference between two RH electrode bank signals replaces the former single RH electrode readout of the Frisch grid detector. CdZnTe is a very common semiconductor material used in these types of detectors. For more details, see the reviews of He (He, 2000) and Owens (Owens et al., 2006).

Another intriguing feature in the review of He (He, 2000) is the derivation of the S-R theorem using an energy approach. We do not especially advocate the energy approach here, because it does not represent a simpler path for proving the theorem. But admittedly, it is always useful to study the energy exchange under the conditions set by the theorem. Also, the article presents a clear view on the electrode configuration assumed in the proof. Some would presumably also appreciate a very insightful older presentation using energy methods (Gunn, 1964).

6 Results

The program was used to run several simulations with lattice temperature set at T=77K, while varying reverse bias external fields. We focused on -8V and -18V reverse bias voltage. Based on the Shockley-Ramo method discussed in earlier chapters, the currents generated on the electrodes were calculated by the program, and stored in four different files measuring the total current generated by the 1) electrons on the left electrode, 2) electrons on the right electrode, 3) holes on the left electrode, and 4) holes on the right electrode. This data was fed into MatLab and used to generate the plots.

This plot was generated using the quiver()-function in MatLab (with the MatLab arrow size parameter set to 2), based on the final E-field generated by the simulations. The final E-field files should be identical to the initial E-field and remain constant during the run, as when running an APD simulation the program shouldn’t run any bias field calculations.

It is important to note that the quiver() function in MatLab performs an arbitrary initial scaling, so that these two plots for the different external field strengths is incommensurable as they are presented here. They must thusly be used merely to get a better picture of how the fields appear. The data behind the plots is, of course, unscaled, and can be used in further work.

6.1 Static “Frozen” Field

Let us start by looking at the “frozen field” meaning the background field at $t = 0$.

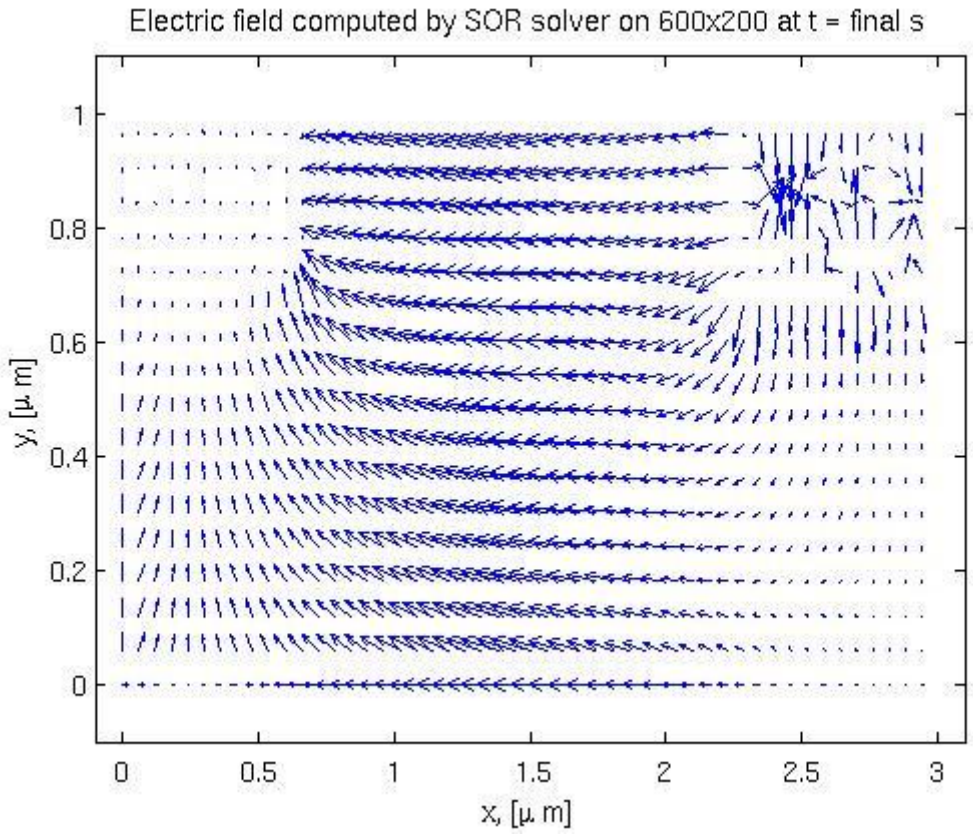


Figure 6-1. This the static bias field for 08V, often referred to as “frozen field”. Note the clearly recognizable areas of higher doping, and the smooth field in the region between the conductors. Also note the small area in the upper right corner where the field is pointing away in all directions.

When running a simulation to generate a frozen field, in this thesis and earlier work, a simulation time of up to 200 ps has been used. The most transient period, however, is only up to around 50 ps. After that, there is very little effective current going, meaning the field is closing in on an equilibrium state. Running a simulation up to 100 ps should be sufficient time to be confident the field is stable, based on work by C. Kirkemo (Kirkemo 2011).

We have experimented with different sampling rates for the Shockley-Ramo calculated current, since it is of paramount importance in an actual application to know what sample frequency or bandwidth the output measuring device needs to have.

The purpose is also to determine which bias voltage is optimal to get a measurable current pulse output and minimize the number of “stuck” electrons.

As can be seen the fields is very linear and “well-behaving” in middle area. The field for both voltage biases appear smooth outside of the highly doped contact zones, while the field there may seem discontinuous, or at least highly non-uniform. Notice again that this is a static field and a stable configuration, and not a transitive state.

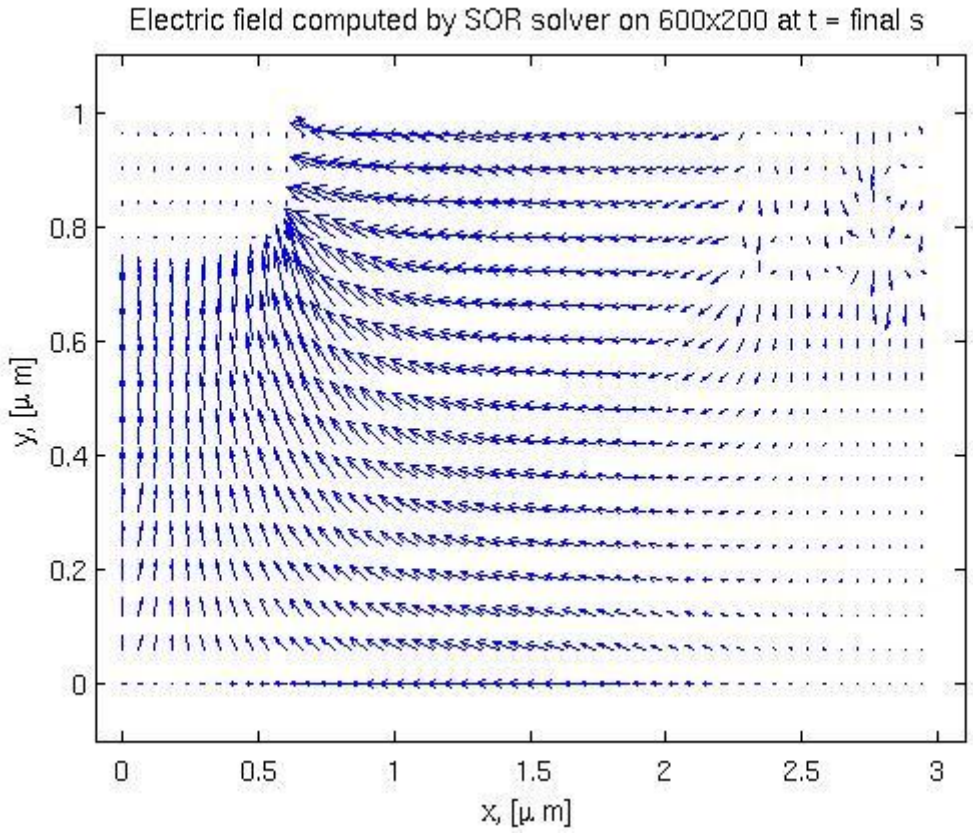


Figure 6-2. This the static bias field for 18V, often referred to as “frozen field”. Note the clearly recognizable areas of higher doping, and the smooth field in the region between the conductors. Also note the small area in the upper right corner where the field is pointing away in all directions.

The graphs can indicate that it might be needed to run a simulation with more mesh points for the Poisson equation solver for a more precise image of the field. As of the moment of writing, the number of mesh points is 600 x 200. The downside of this is that a static higher number would severely affect performance of the program. Still,

some optimization might prove more efficient, either in the form of separating the Poisson solver for the frozen field calculation from the regular part of the program, since that is only needed to be performed once, or in the form of variable mesh, with higher point density in the highly doped corners.

Furthermore, it appears that there are several “dead zones”, i.e. small spots that appear to experience a very low or close to zero E-field. This has, however, not proved to be a problem in practise, as we have very rarely observed any carriers ending up in these remote areas.

6.2 Carrier Positioning

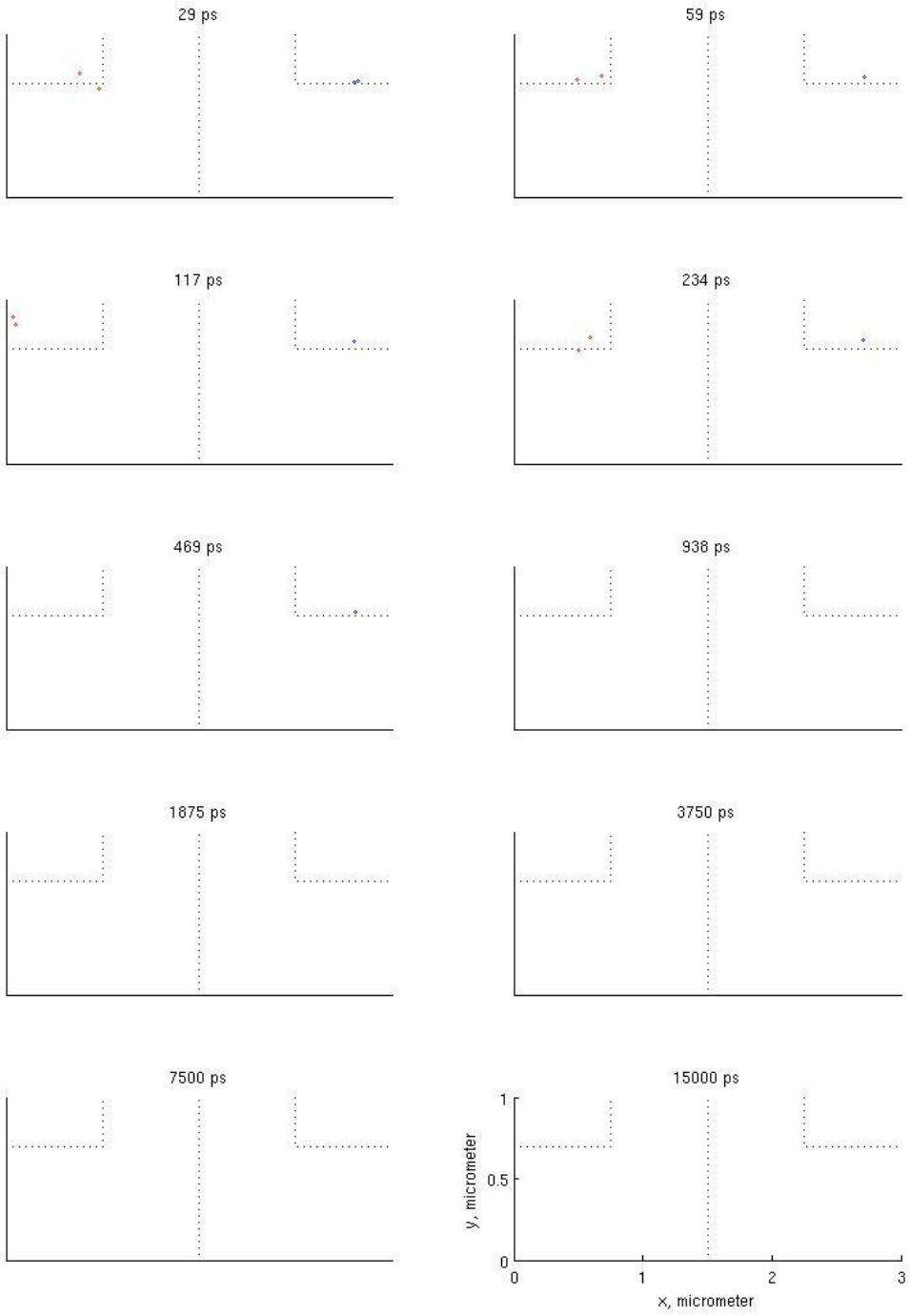


Figure 6-3. Position plot revealing the physical positions for the electrons (blue), and holes (red) at different time steps for the 08V field.

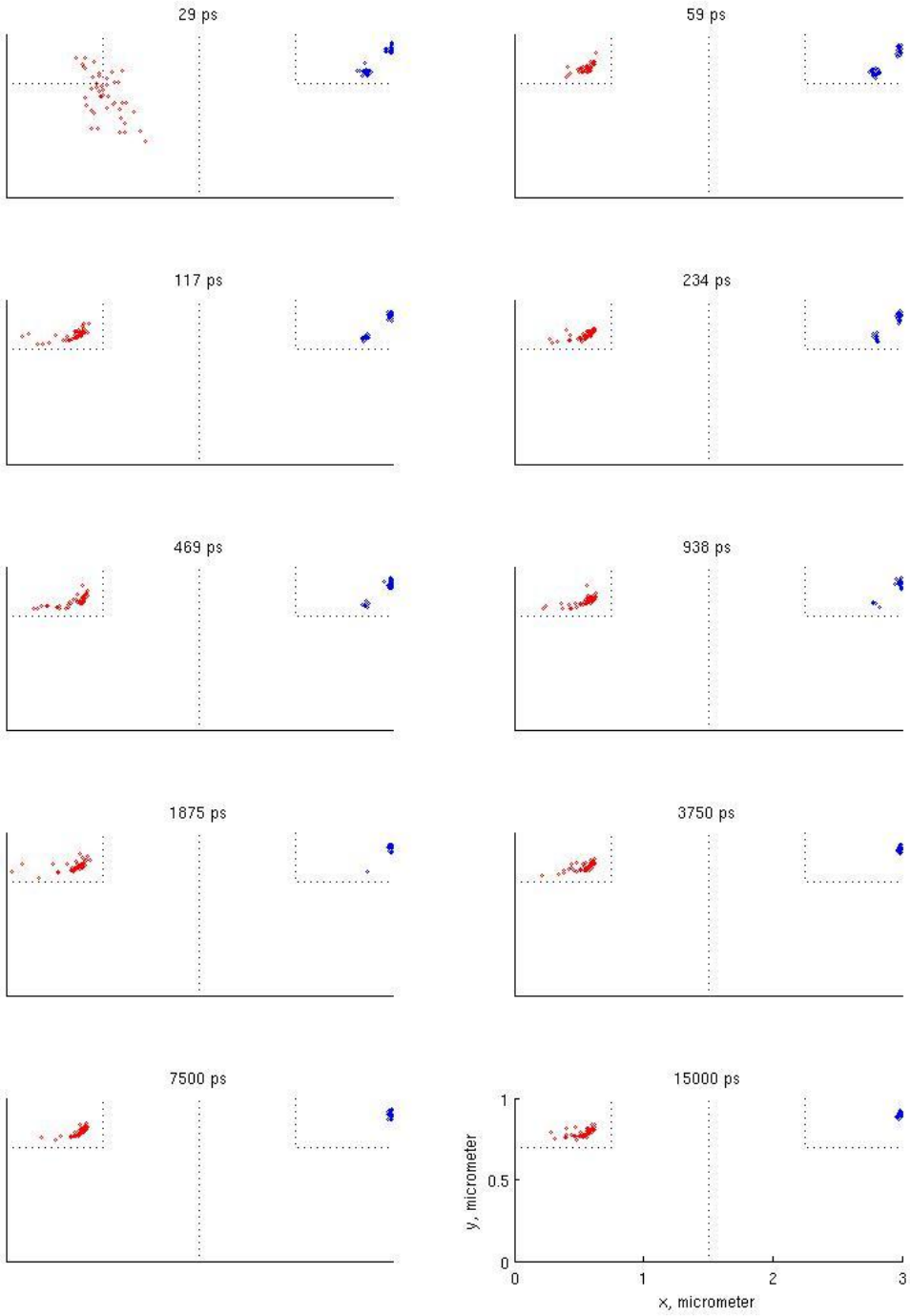


Figure 6-4. Position plot revealing the physical positions for the electrons (blue), and holes (red) at different time steps for the 18V field.

As later plots also will confirm, the particle multiplication can be seen from these position plots to be much higher when the bias voltage is increased. In both cases, the initial scenario is one electron/hole pair generated at time $t = 0$, but the higher field in from the 18V reverse bias in fig 6-4 compared to 8V in fig 6-3 accelerates the electrons further, giving them much more energy, thus allowing them to ionize other atoms they hit, releasing a “new” electron/hole pair, which begin to accelerate due to the bias field. This process is called impact ionization. Keep in mind that in the program, this impact ionization is modelled to only happen for electrons, not holes. It can happen that a hole initiates impact ionization, but the frequency of this is so low that the effect is negligible. We will get back to discussing more on this a bit later in this chapter. Thus, new holes and electrons can appear almost anywhere to the right of the initial electron/hole pair, while there never appear new electron/hole pairs to the left.

Another very interesting effect observed, is the fact that in the higher voltage, the carriers tend to be unable to escape the “pocket” in the connector corners. This correlates to the small location found most prominently in the lower right area of the heavily doped corner for both voltages, and another slightly above and to the right where the field arrows all point away from that small spot. According to the advisor of this thesis work, Trond Brudevoll, this is within reason close enough to what might have been expected. Thus, it is not necessarily an obvious flaw of the program, that it generates such a volatile and non-uniform field within the conductor edges.

There is still the problem of dealing with these “stuck” carriers. As can be seen further on in the results section, carriers that never exit the device contribute a significant amount of noise to the output signal. This is because the output current is generated by the movement of the carriers relative to the connector at the top edge. Even if the carriers are “stuck”, they are still randomly moving back and forth in the trapped region, meaning a significant seemingly random current is generated by this. This current could potentially be filtered away, as we are mainly interested in the initial pulse, but problems could arise due to the build-up of particles in the trapped regions, adversely affecting the conductivity and performance of the device.

6.3 Particle Numbers

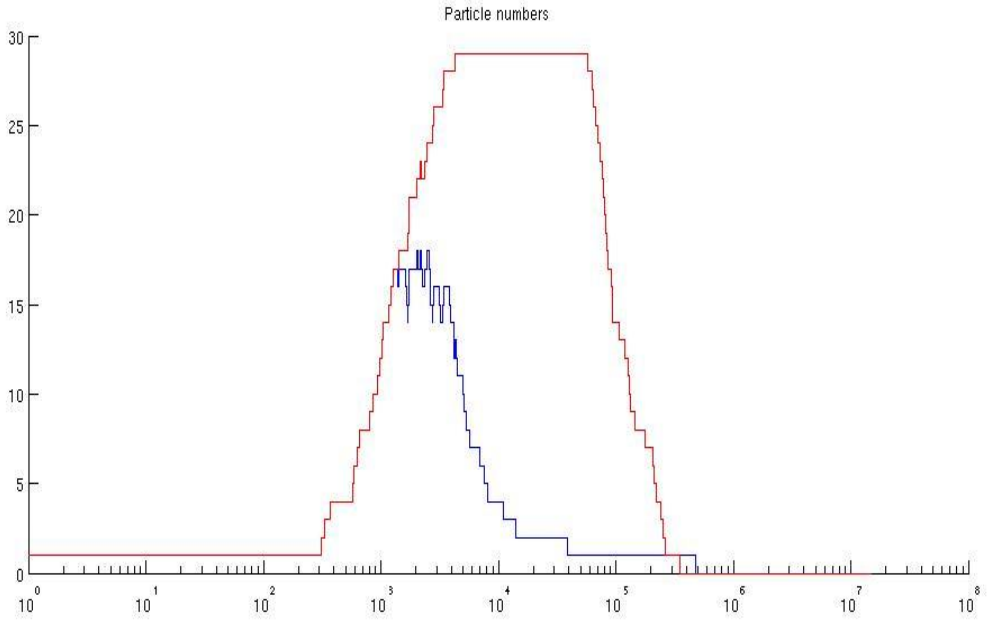


Figure 6-5. Total number of each type of carrier present inside the device with 08V reverse bias.

Above and below are plots for the graphs of the number of each type of charge carrier present at any time in the device for respectively 8V and 18V. The graphs can both look misleading, in their respective ways. The integral over the two graphs is by definition zero, as the number of charges is symmetric. For every impact ionization event, both a hole and an electron are generated, so there is no way to generate a surplus of one type of carrier. The reason the integral over all the electrons and holes doesn't match up here, is due to the sampling rate. The electrons rush out so fast they are not even visible as a spike here. This also introduces another problem in a physical implementation of the APD; what is the required measuring frequency needed for any component or device that is supposed to measure this current pulse and the conducting edge of the APD?

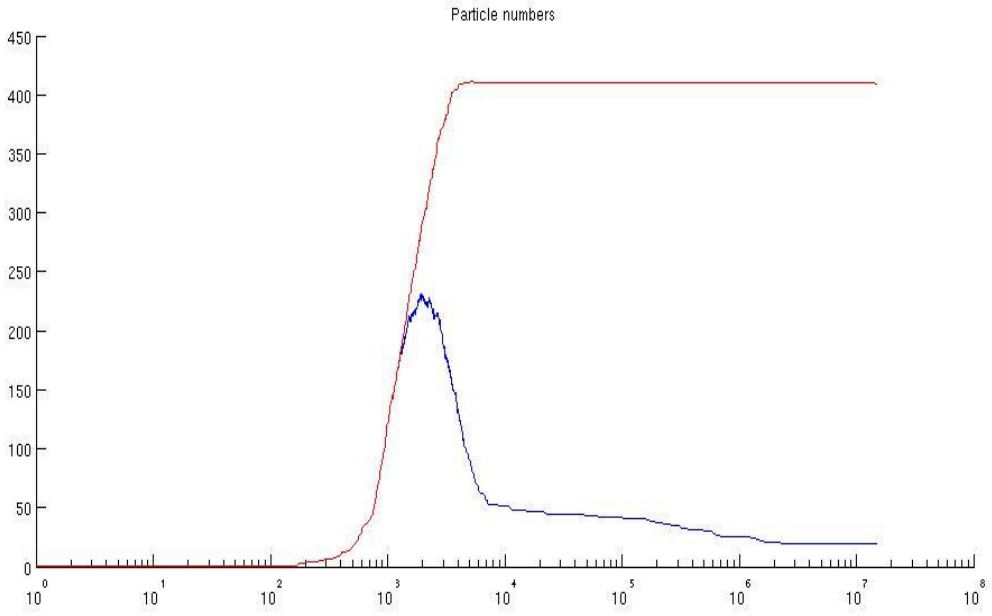


Figure 6-6. Total number of each type of carrier present inside the device with 18V reverse bias. As above, both starts on 1, but due to the higher energies involved, the numbers quickly multiply.

A disturbing observation for the 18V case is that it appears that close to no holes ever escape the APD, and the number of electrons (after the initial spike discussed above) only halves. Keep in mind that the simulation is of length 15 000 ps, or 15 nanoseconds, which is *very* long in this environment and for a simulation of this type. As can be seen, the multiplication degree is much higher, but this side effect of the "stuck" carriers might make such a high voltage inefficient in practise.

6.4 Shockley-Ramo Calculated Currents for 8V

Based on the numerical Shockley-Ramo solver these plots have been generated by the program. We have tried varying the sampling rate for the current calculation. The calculations are of course at every iteration pass, but only sampled, or written to file, with a certain rate. The most notable test sample rates was to sample every iteration (which is to say sample every femtosecond), and every 1 000 iterations, meaning every 1 000 fs, or every 1 ps.

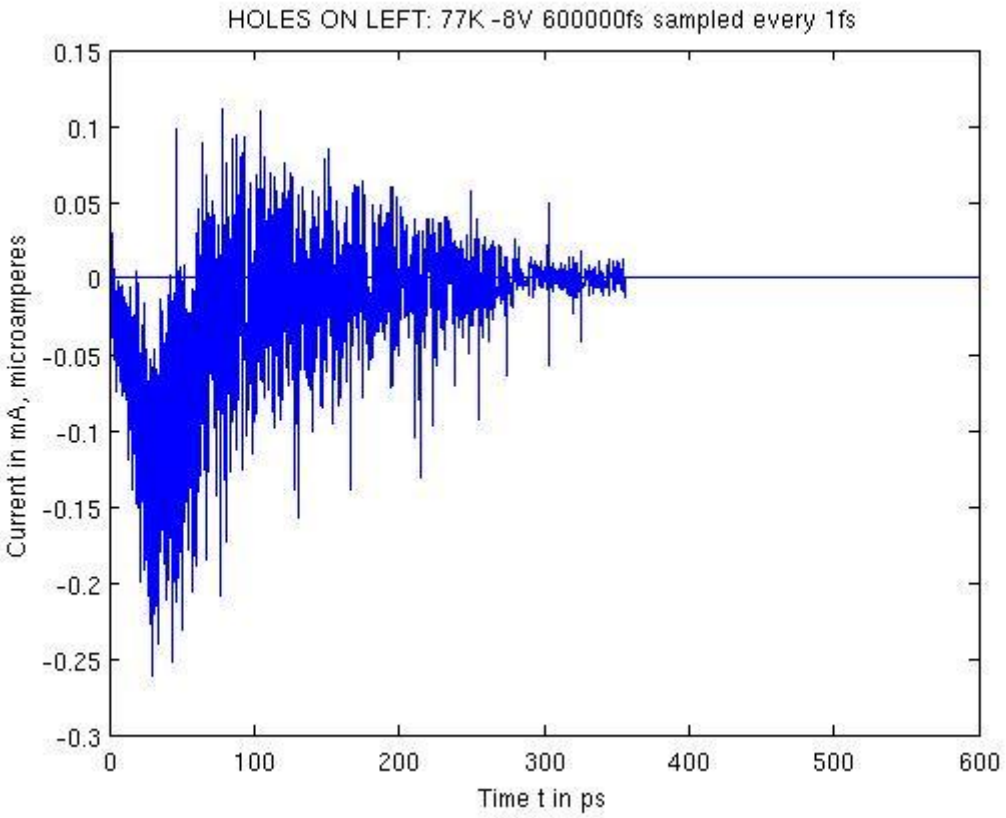


Figure 6-7. The current generated by holes on the left conductor edge. This is a different simulation, using only 600 ps length, as opposed to 1 500 ps which was used above, and sampled every fs, or every iteration pass.

An interesting observation to be made from the simulations is that the amount of carriers being stuck at a given time is very variable. In general, there is more carriers getting stuck in the 18V reverse bias, and in most runs all carriers tend to leave the APD device with and 8V reverse bias.

In the below plot, the exact same run parameters have been used, excepting different sampling rate for the APD current, and longer duration of the whole simulation.

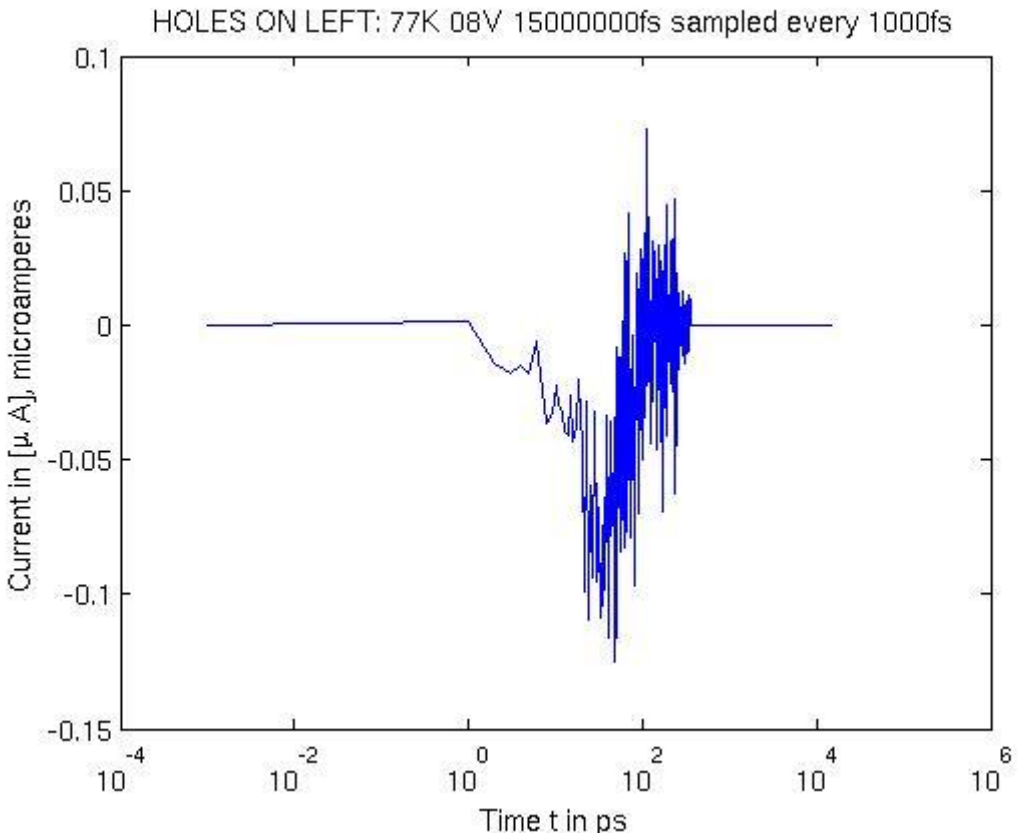


Figure 6-8. The current generated by holes on the left conductor edge. This is the same simulation as has been mentioned in the previous subchapters, using 15 000 ps total simulation length. This was also sampled every 1 000 fs/1 ps/1 000 iteration.

The initial burst is the main rush of photogenerated holes escaping the device while the latter parts after about 150 ps is dominated by stuck holes randomly moving in areas where they can be stuck, but slowly trickle out of those spots, and end up escaping. When the current hits zero, all carriers have escaped. It is not possible at temperatures above 0K to have zero current with any carriers still left inside the device.

The electron plots are similar in nature, only the peak comes much faster than for holes, as holes to a higher much higher degree loose energy to optical phonons

(Kinch, Beck et al. 2004). With the effective electron mass down to $\sim 7 \cdot 10^{-2} E_g$, where E_g is the band gap in electron volts, the conductivity for electrons is very high, and allows for a rapid acceleration towards the end conductor.

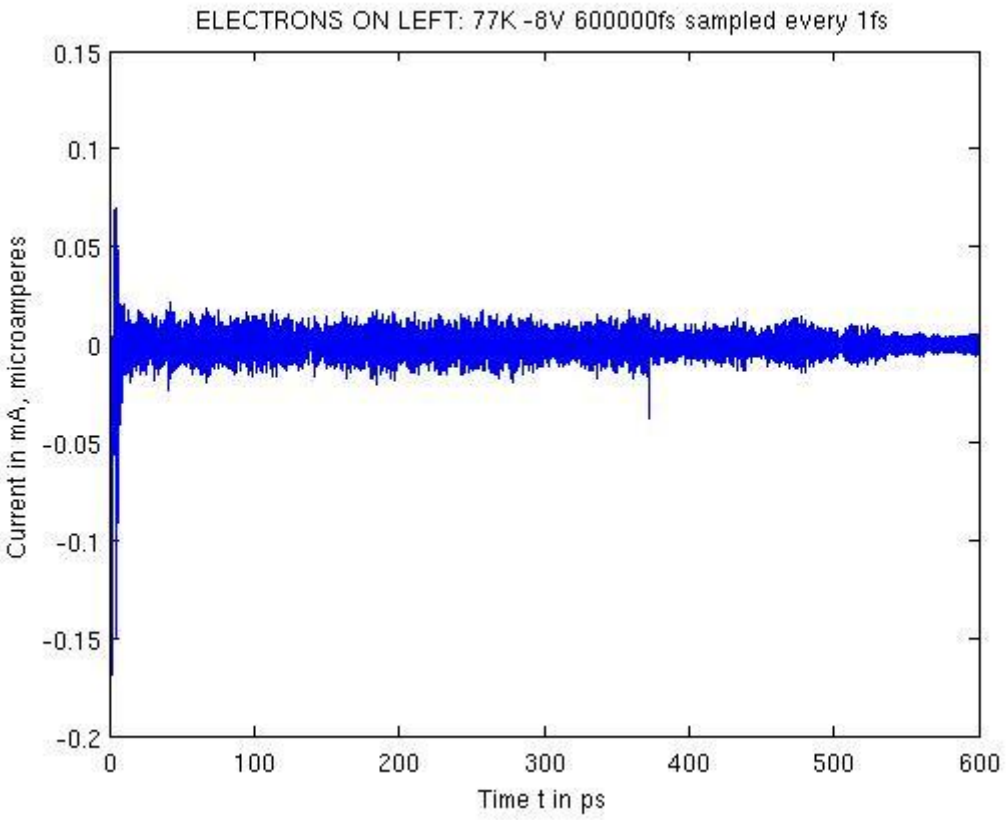


Figure 6-9. The current generated by electrons on the left conductor edge, for the 600 ps length simulation, sampled every 1 fs.

As predicted, this current is a significantly faster, and “spikier”. Whereas the hole avalanche takes between 100-150 ps, the electrons generate a shorter, but correspondingly higher current pulse, which appear to have a duration of between 25-40 ps. Note that both these estimates for spike burst duration is highly dependent on device size. As the electrons accelerate at such a rapid rate, they can be assumed to travel at a constant rate, thus making the travel time in which the carriers are generating current, increase linearly with APD device size.

Holes have a generally lower tendency to get stuck with an 8V reverse bias, but this effect is as mentioned earlier inherently somewhat random. Due to the randomness involved, two simulations with the same parameters rarely yield the exact same result, usually they vary in the number of impact ionization events, and the time at which the respective carriers exit the device.

As no relevant parameters were changed between the above and the below runs, yet they still yield significantly different results as to the end situation, it is not clear to the author any method of “helping” the carriers escape. Although, a way of testing this more thoroughly would be as already mentioned - using a higher point density mesh grid, if nothing else at least to have a more clear picture of the E-field in the highly doped corner regions.

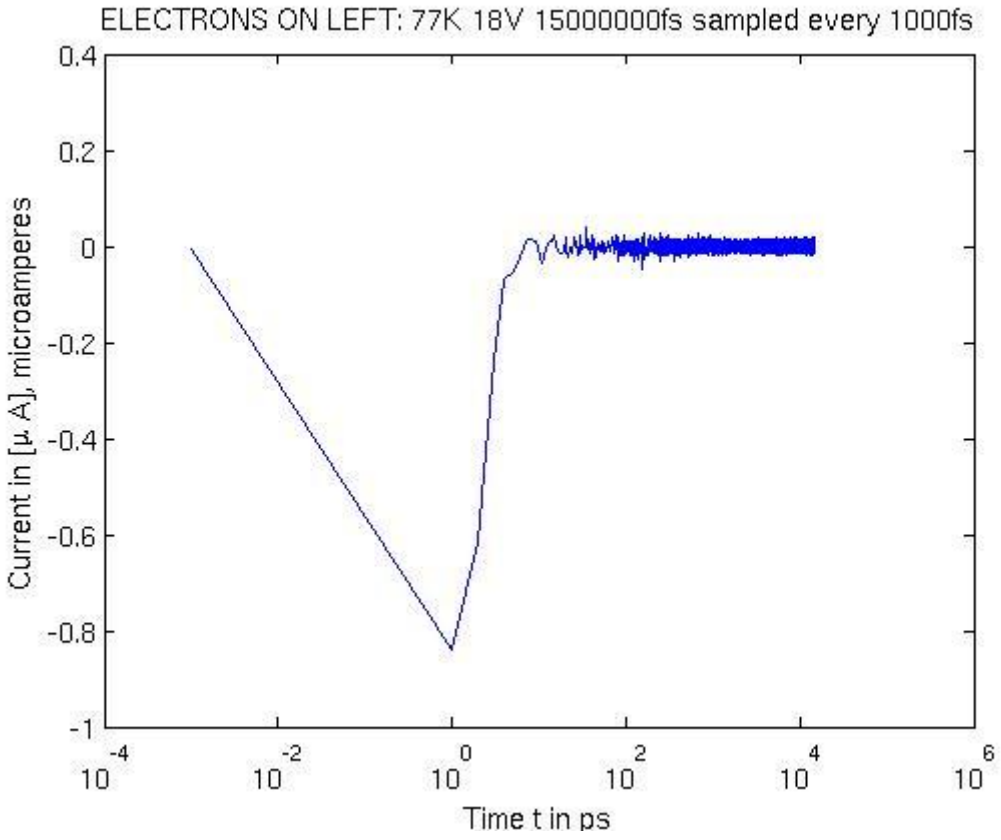


Figure 6-10. LOG-plot of the current generated by electrons on the left conductor edge, for the 15 000 ps length simulation, sampled every 1000 fs. The plot is cut off at 1000, since it was zero.

An interesting detail to the graphs of the electrons is that they tend to have a positive peak at the end of the avalanche phase, which might suggest that a high number of electrons actually do get stuck for a short time, and/or are getting accelerated away from the corner region of the APD for a short time. Due to the limited resolution of the mesh it might be difficult to say for certain whether this is a physical plausible behaviour, or a consequence of the simulated electron hitting an unreasonable strong localized field accelerating it in the opposite direction.

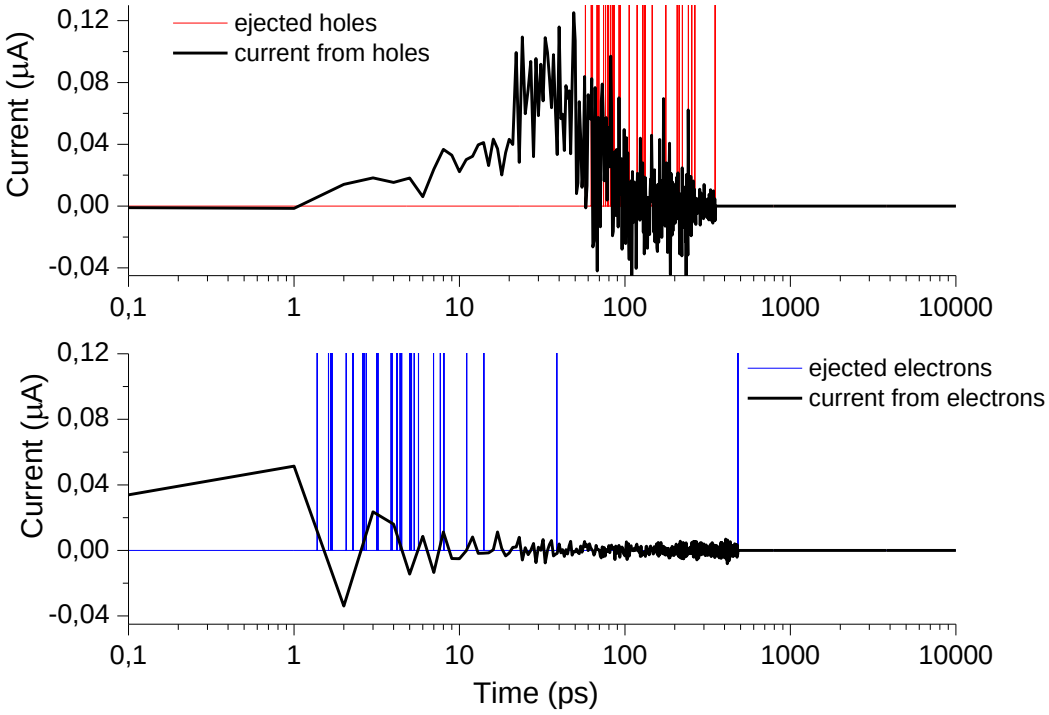


Figure 6-11. For illustration purposes, two LOG-plots of the currents generated by holes and electrons with ejected carriers in the same plot for 8V reverse bias.

6.5 Shockley-Ramo Calculated Currents for 18V

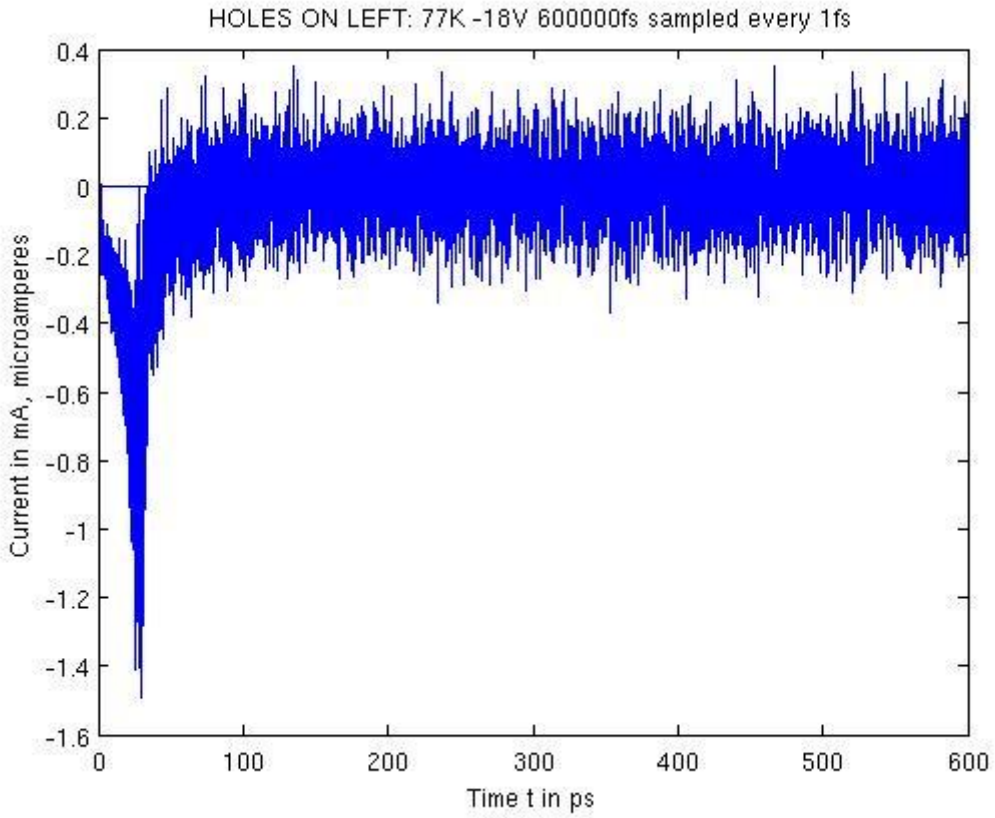


Figure 6-12. The current generated by holes on the left conductor edge, for the 600 ps length simulation, sampled every 1 fs. Note the significantly increased noise.

The peak is much clearer and sharper in the 18V reverse bias setup than in the 8V setup. The clearness is due to the stronger field, making the electrons less likely to stop and/or change direction away from the heavily doped corner and ultimately, the conducting edge of the APD device. On the other hand, the noise is much larger, which is due to the fact that more holes are getting stuck in the 18V reverse bias device. The reason why the “stuck” carriers generate such a noise current is explained in the above subchapters.

The increased sharpness is due to the higher acceleration each carrier experiences, both hole and electron. The higher acceleration means that after every impact ionization event the newly generated carriers, along with the generating electron (we are still modelling this as impact ionization can only be initialized by electrons) is almost instantly accelerated to their drift speed. This leads to a significantly larger

proportion of the carriers reaching the doped corner than with using an 8V external field.

It will be more noise simply because of the fact that the stronger field leads to higher kinetic energy, and thus speed for both types of carriers. Thus the stuck carriers also “bounce” more vigorously back and forth, leading directly to stronger currents, as the generated current is based on the speed of the respective carrier.

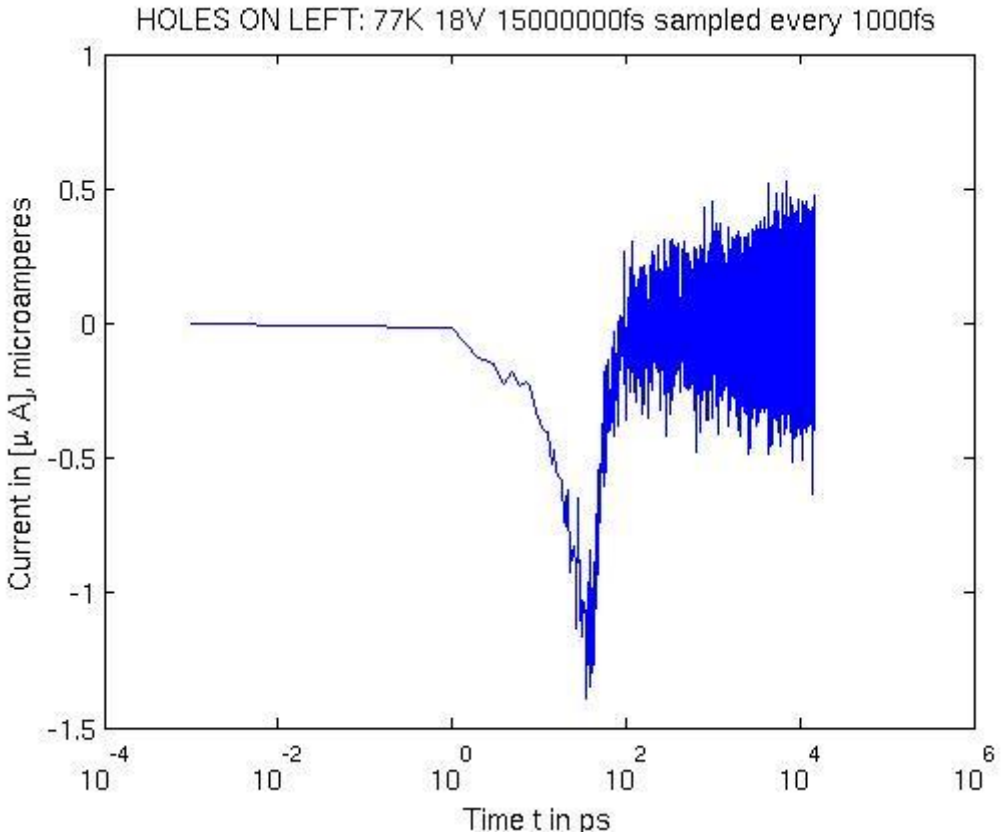


Figure 6-13. The current generated by holes on the left conductor edge, for the 15 000 ps length simulation, sampled every 1000 fs. Note the significantly increased noise.

In the above graph, the general characteristic can be recognized fig 6-12, except the graph appears thinner. This is partly explained by the different scale on the x-axis, but mainly due to the reduced sampling rate. It is apparent that this sampling rate is good enough to capture the current generated from the holes. The electrons will prove slightly more different to detect, as will be discussed later on.

Also note the scale on the y-axis for the current. It is over 10 times higher than for 8V. This significantly increased current is due to the nonlinear scaling of carrier multiplication with increasing external field.

The noise for the holes is stronger with 18V reverse bias field, approximately around double magnitude.

Below is the corresponding plot for the 600 ps duration simulation for electrons. The very sharp peak in generated current is a consequence of the strong electric field. The noise is around the same order of magnitude as for the 8V field, contrasting the noise current level for holes, which is marginally higher.

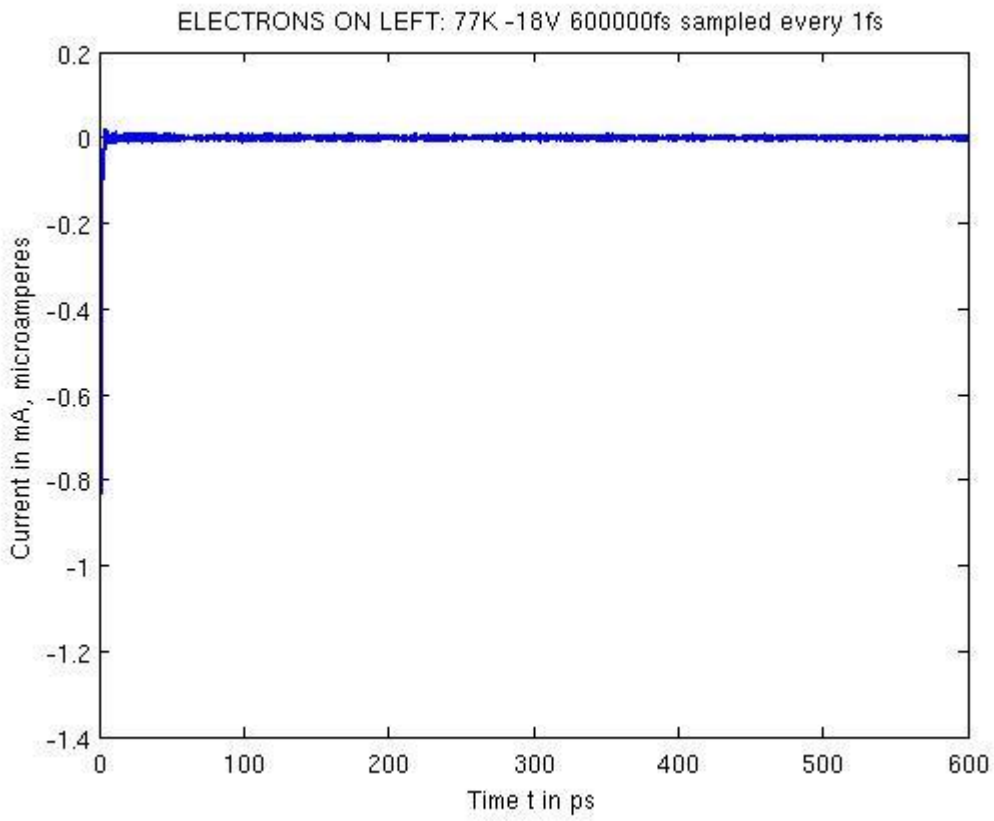


Figure 6-14. The current generated by electrons on the left conductor edge, for the 600 ps length simulation, sampled every 1 fs. Note the significantly increased noise.

In both the above plot and the below plot it is obvious that even though there are many carrier particles “stuck” in these few dead end regions that never escape, they

do not seem to disturb the functionality of the device, as the noise is random static which can be filtered away, either with a frequency filter, or a magnitude filter.

The current peak might seem surprisingly low, compared to the one generated by the holes, but closer inspection of the hole data reveals that it is much more “spiky” in the avalanche phase than the electron graphs. The hole graphs are highly fluctuating, giving the impression of a thick line, while it in reality has an average closer to what can be expected, and closer to the electron-generated graphs.

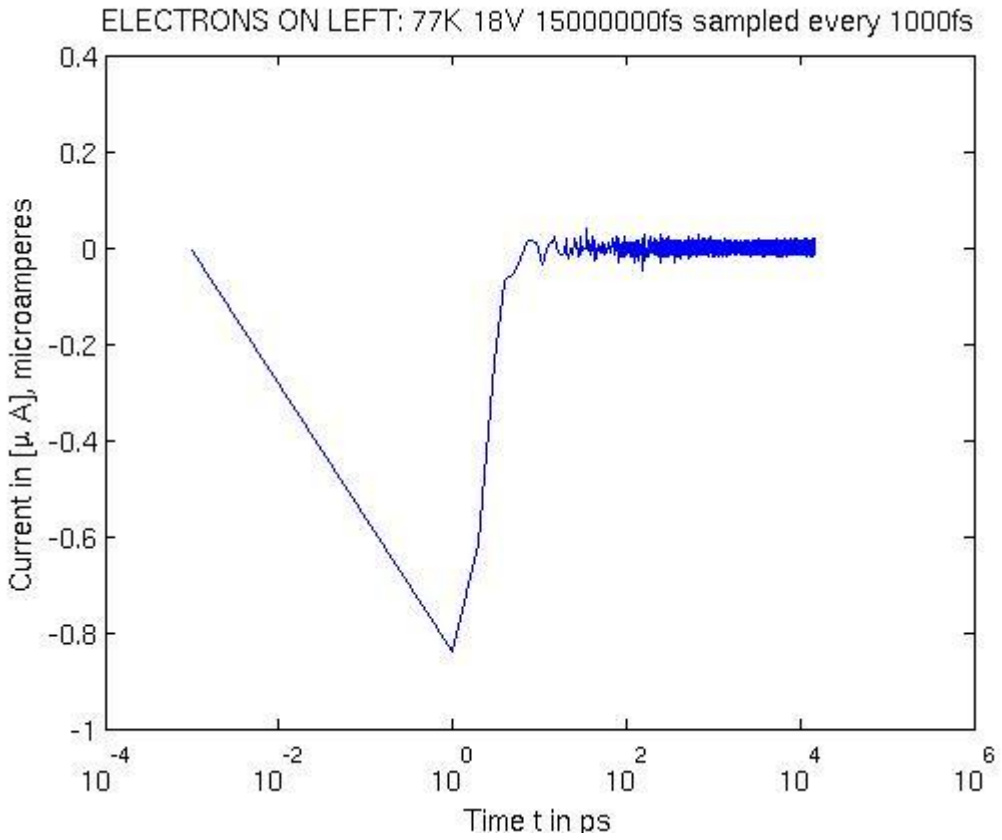


Figure 6-15. The current generated by electrons on the left conductor edge, for the 15 000 ps length simulation, sampled every 1000 fs. Note the static noise.

The hole graphs are naturally more stretched out in time as compared to the electron graphs, for two reasons. Firstly they are moving slower, meaning that the time it will take the avalanche of holes to stop moving (whether by exiting the APD, or by getting “stuck”) will be longer. Secondly, due to the (reasonable) assumption in the code that

holes do not initialize impact ionisation events, all new carriers will be generated on the right side of the initial electron/hole pair. This means that all holes will have to travel longer than the initial one, and due to their relatively large effective mass their acceleration towards the left corner region will be significantly slower than that of the electrons.

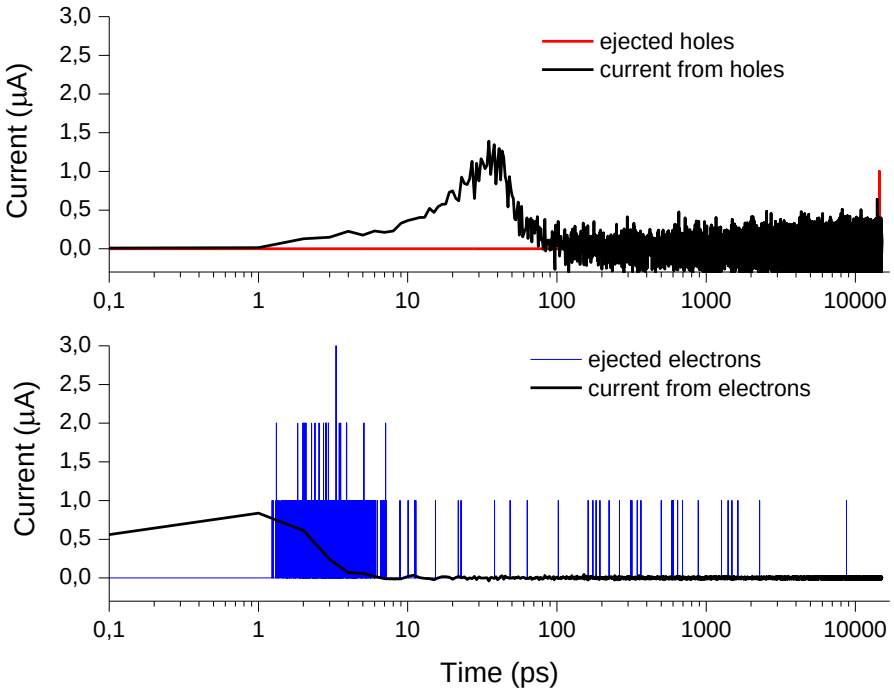


Figure 6-16. For illustration purposes, two LOG-plots of the currents generated by holes and electrons with ejected carriers in the same plot for 18V reverse bias.

7 Discussion and Summary

The Monte Carlo simulator which has been developed as a student project at FFI has during this work been made capable of modelling devices. A fast Poisson solver which relies on the successive overrelaxation method has been implemented and an impact ionisation model has been implemented in order to make the simulator capable of modelling APDs.

During this work, two simple pn-junction devices has been simulated, a pn-diode and an APD. The device characteristics, including current-voltage characteristics and switching times have been obtained. For the APD, we have studied multiplication and noise and obtained the gain-voltage characteristics.

We simulated simple devices in order to discuss a few problematic aspects of the Monte Carlo method which arise when simulating CMT pn-junctions. Within the framework of the Monte Carlo method, the dark current needs to be neglected for CMT APDs. We have pointed to the fact that the current impulse response may turn out inaccurate due to violation of the continuity equation in the surrounding circuit. Furthermore, we believe that the self-consistent solution of the transport equations and Poisson's equation using the Monte Carlo method is well suited for studying large photo currents in APDs because the frozen field assumption is easily relaxed within the Monte Carlo framework.

7.1 On the Nyquist Theorem and Sampling Rate

If we look at the two above plots, we can see that any device measuring the output current of this APD will need to be remarkably quick. So much so, in fact, that it might be unfeasible to achieve such a high degree of precision and sampling rate. Nyquist's Theorem states that we must have a measuring frequency of

$$f_s > 2 f_e .$$

f_e represents the frequency of the electron current pulse. For simplicity, define the avalanche current spike as our whole signal. Assume the whole signal takes 10 ps,

which is a decent estimate for the electron avalanche, just to get a proper feeling with the magnitudes involved. Since the frequency is given by $f_e = 1/T_e$, this means $f_e = 10^{11}$ Hz or 100 GHz for the relevant signal. This means the Nyquist's sampling frequency is

$$f_s > 2 f_e$$
$$f_s > 2 \cdot 10^{11}\,\text{hz}$$

This is an immense requirement, at least 200 GHz, especially given that most ordinary computers these days have around 3-4 GHz processors.

This frequency can and will be lowered if the APD were to increase in size. It was mentioned in chapter 0, but it is worth summarizing, that the electron travel time will increase linearly with the size of the device. As the APD we have simulated in this context measures 600 nm x 200 nm, this means that there are still some potential for increasing the dimensions without introducing other effects or errors.

The corresponding frequency f_h for holes is much more lenient. The current pulse is now closer to $T_h = 85 - 90$ ps long.

$$f_h = 1/T_h \approx 1.176 \cdot 10^{10}\,\text{Hz},$$

or 11.76 GHz. It is still very fast, but significantly slower than the electron burst.

It should be noted that these numbers are the result of *one* photon generating the first electron/hole pair. In practise, it can be expected in any practical setup[1] single photons are a rare occurrence, and it is highly likely that several will appear within a reasonably short amount of time, meaning that this pulse will be extended the more photons that are hitting the APD device.

[1] Except in a device specifically intended as a single photon detector, but then other devices might be more suited.

7.2 On the Frozen Field and Number of Mesh Points

As has been mentioned in the Results chapter, the basis for the simulations of the APD and carrier transport within it, relies on a solver for the Poisson equation

$$\nabla^2 V = \rho ,$$

given the charge density ρ. The PDE is then solved to yield a potential throughout the mesh point lattice, and this potential V is used to determine the field at any point. However, with strongly varying, or highly non-uniform potentials and thus, E-fields, and the number of mesh points too limited, there is a high risk of not calculating the correct field. In general, the more non-uniform the field is, the higher number of mesh points is needed. Note that the error in precision in the calculations, measured by the error estimate, is separate from errors introduced by a low number of mesh points. Assessing the quality of the field estimate and whether or not enough mesh points has been used in general is no easy task, and requires lengthy experience in the area. Thus, it is hard for the author to precisely pinpoint any weird behaviour to a potential low number of mesh points.

There is also the option of doing a geometric smoothing of the raw field data to reduce the effect of erroneous values in a few mesh points. This is not technically providing any new information, but the method can be employed to simply smooth the graph. Naturally, this works best the more uniform the field is, as this lessens the chance of the method smoothing out spikes that are actually present, and is not a result of an error somewhere.

Still, based on the appearance of the field, it appear it would be optimal to perform a more thorough calculation of the frozen field in both cases, using a higher number of mesh points only in the highly doped corner regions.

7.3 General Effects of Increased Reverse Bias Voltage

- Increasing the reverse bias potential increases multiplication degree of the charge carriers.

As explained earlier in the thesis, the stronger field allows the electrons to accumulate more kinetic energy, thus leading to a higher rate of impact ionization events. Keep in mind, in reality, this happens for holes too, but the rate at which hole-initiated impact ionization occurs, is much rarer. Hole-initiated impact ionization is ignored completely in our simulation, and thus does not occur during simulations. Earlier work (Kirkemo 2011) also indicated this trend, although this was only done with reverse bias voltage up to -7V. In one simulation for -8V the multiplication was 48. This means that 48 impact ionization events occurred. In a corresponding simulation for -18V reverse bias the multiplication was found to be 966.

- Increasing the reverse bias potential increases peak magnitude, and reduces avalanche duration.

This means that the avalanche current peak will become more compact and intense. Some consequences of this are that the signal might be easier to detect in equipment with low sensitivity, but also making it harder to detect a single photon, due to the duration of the signal decreases. If the APD device can be guaranteed to be fed a pack of photons, and cooling is of little concern, it appears a strong field is to be preferred, as that will make a signal stronger, while the fact that several it is several photons will increase the likelihood of detection due to several current peaks travelling. Using several photons has not been a focus of this thesis, but there might be issues with saturation

Another possibility is the use on an integrator for the current, thus essentially focusing on total charge generated. This method will be unaffected by a potentially short duration of the signal.

This is, of course, only valid up to a certain point, where the amount of electrons generated start having an effect on the frozen field, but that is beyond the scope of this thesis.

8 Appendices

8.1 Proof of Green's Reciprocation Theorem

This theorem is central in Shockley's proof for the Induced Electrode Currents, recited in chapter 4. In its full glory, the theorem states

$$\int_v \left(\rho\Phi' - \rho'\Phi\right) dV = \oint_s \left(\sigma'\Phi - \sigma\Phi'\right) dA$$

Here, ρ is charge density in a volume V, σ is charge distribution over a surface s, and Φ is an electromagnetic potential function. The ' marks for two different non-coexistent situations. Green's Theorem is an excellent starting position:

$$\int_v \left(\Phi\nabla^2\Psi - \Psi\nabla^2\Phi\right) dV = \oint_s \left(\Phi\frac{\partial\Psi}{\partial n} - \Psi\frac{\partial\Phi}{\partial n}\right) dA$$

Start by substituting $\Psi = \Phi'$. Furthermore, we have these two equations linking ρ and σ with the derivatives.

$$\nabla^2\phi = -\rho/\epsilon_0 \text{ and } \sigma = \epsilon_0\frac{\partial\Phi}{\partial n} \Rightarrow \sigma/\epsilon_0 = \frac{\partial\Phi}{\partial n}.$$

which obviously also holds for the ' marked situation. Using these two equations, we get

$$\int_v \left(\Phi\rho' - \Phi'\rho\right) dV = \oint_s \left(\Phi\sigma' - \Phi'\sigma\right) dA,$$

after we cancel out the $1/\epsilon_0$ -factors.

Now we are almost done. In his work, Shockley uses a slightly modified version of this equation. Firstly, Shockley assumes no surface charge, $\sigma = 0, \sigma' = 0$, which basically renders the right hand side to 0.

$$\int_v \left(\Phi\rho' - \Phi'\rho\right) dV = 0$$

Secondly, he is using a discrete model for the spatial charge distribution ρ, turning the left hand side integration into a summation, and the continuous ρ into Q_i.

Tertiary and lastly, Shockley's generic potential Φ is replaced with an electronic potential V_i.

Then, the form of the GRT used by Shockley looks like this:

$$\sum_{\text{all charges}} \left(V_i Q_i' - V_i' Q_i \right) = 0$$

QED.

8.2 The FMM Method

It is also worth mentioning the method of FMM, or fast multipole moment calculations. This is a method of calculating interaction potential between charges. It is an approximate solution to the N-body problem of interacting particles, less precise than its unaccelerated version, that of multipole moment. Let us first take a quick recap of the derivation of this method.

We start with looking at a vector from a fixed reference point to a point in the system, called r. Then, let r_1 be the vector from the fixed reference starting point to the observation point

$$d^2 = r_1^2 + r^2 - 2r_1 r\cos\phi = r_1^2\left(1 + \frac{r^2}{r_1^2} - 2\frac{r}{r_1}\cos\phi\right)$$

Taking the square root and inverting the expression gives us

$$\frac{1}{d} = \frac{1}{r_1}\left(1 - 2\xi y + \xi^2\right)^{-1/2}$$

Here we also substituted $\xi = \frac{r}{r_1}$ and $y = \cos\phi$. The key now to recognize this expression as the generating function of Legendre polynomials given below

$$\left(1 - 2\xi y + \xi^2\right)^{-1/2} = \sum_{i=0}^{\infty} \xi^i P_i(y)$$

Substituting this back into the main equation for the inverse distance 1/d yield

$$\frac{1}{d} = \frac{1}{r_1}\sum_{i=0}^{\infty}\left(\frac{r}{r_1}\right)^i P_i(\cos\phi) = \sum_{i=0}^{\infty}\frac{1}{r_1^{i+1}} r^i P_i(\cos\phi),$$

Thus proving that any physical potential that is proportional with $1/d$ can be expressed as a multipole expansion.

This method, however, has a runtime of $O(n^2)$, which means that the execution time of a program can escalate drastically as problem size increases. Thus, it is relevant to optimize this method to reduce execution time of such an algorithm. This is the FMM, or fast multipole method. This method revolves around separating between

local and distant interactions, and by greatly reducing the number of execution time to $O(N + M \log M)$. N represents particle number, and M is the number of mesh points in problem and Is usually chosen to be smaller but proportional to N. This further simplifies to $O(N \log N)$, however the computational cost is often found in practise to be proportional to $O(N)$. The algorithm uses some simplifications, and is very vulnerable to a highly varying or non-uniform electric field (technically, this sensitivity of the algorithm is bound to the Poisson solver used within). The impact on performance might be either in the form of significant increase in running time, or a corresponding decrease in precision of the final answer.

The algorithm is fairly sophisticated, and any short reproduction besides what is already given here would likely not do it justice. For more information on this subject, refer to (D. Vasileska 2010).

9 References

Bellotti, E; Moresco, M; Bertazzi, F. (2011). A 2D Full-Band Monte Carlo Study of HgCdTe- Based Avalanche Photodiodes. Journal of Electronic Materials DOI: 10.1007/s11664-011-1635-8.

Bertazzi, F.; Moresco, M.; Penna, M.; Goano, M.; Bellotti, E. (2010). Full-Band Monte Carlo Simulation of HgCdTe APDs. Journal of Electronic Materials DOI: 10.1007/s11664-010-1198-0.

Blanco-Filgueira, B.; Lopez, P.; Roldan, J.B. (2012). Analytical modelling of size effects on the lateral photoresponse of CMOS photodiodes. Solid-State Electronics Vol. 73, 15.

Cavalleri, G.; Fabri, G.; Gatti, E.; and Svelto, V. (1963). On the induced charge in semiconductor detectors. Nuclear Instruments and Methods in Physics Research Vol. 21, 177.

Cavalleri, G.; Gatti, E.; Fabri, G., and Svelto, V. (1971). Extension of Ramo's theorem as applied to induced charge in semiconductor detectors. Nuclear Instruments and Methods in Physics Research Vol. 92, 137.

Dabrowski, W. (1987). Comments on the collective and corpuscular approach of generation-recombination noise in a p-n junction. Solid-State Electronics Vol. 30, 205.

Dabrowski, W. (1989). Transport equations and Ramo's theorem. Progress in Quantum Electronics Vol. 13, 233.

de Visschere, P. (1990). The validity of Ramo's theorem, Solid-State Electronics Vol. 3, 455.

Djuric, Z; Smiljanic, M., and Radjenovic, B. (1984). The application of Ramo's theorem to the impulse response calculation of a reach-through avalanche photodiode. Solid-State Electronics Vol. 27, 833.

Dubaric, E.; Nilsson, H.E.; Frøjdh, C., and Norlin, B. (2002). Monte Carlo simulation of the response of a pixellated 3D photo-detector in silicon. Nuclear Instruments and Methods in Physics Research A Vol. 487, 136.

Ettenauer, S. (2008). Considerations of the application of Ramo's theorem for segmented
HPGe detector pulse-shape calculations. Nuclear Instruments and Methods in Physics Research A Vol. 588, 380.

Fabjan, C.W., and Riegler, W. (2004). Trends and highlights of VCI 2004. Nuclear Instruments and Methods in Physics Research A Vol. 535, 79.

Gatti, E., and Geraci, A. (2004). Considerations about Ramo's theorem extension to conductor media with variable dielectric constant. Nuclear Instruments and Methods in Physics Research A Vol. 525, 623.

Gatti, E., and Padovini, G. (1982). Signal evaluation in multielectrode radiation detectors by means of a time dependent weighting vector. Nuclear Instruments and Methods in Physics Research Vol. 193, 651.

Gook, A; Hambsch, F.J.; Oberstedt, A., and Oberstedt, S. (2012). Application of the Shockley-Ramo theorem on the grid inefficiency of Frisch grid ionization chambers. Nuclear Instruments and Methods in Physics Research A Vol. 664, 289.

Gruber, L.; Riegler, W., and Schmidt, B. (2011). Time resolution limits of the MWPCs for the LHCb muon system. Nuclear Instruments and Methods in Physics Research A Vol 632, 69.

Gunn, J.B. (1964). A general expression for electrostatic induction and its application to semiconductor devices. Solid-State Electronics Vol. 7, 739.

Hamel, L.A., and Julien, M. (2008). Generalized demonstration of Ramo's theorem with space charge and polarization effects. Nuclear Instruments and Methods in Physics Research A Vol. 597, 207.

Haus, H.A. and Melcher, J.R. (1989). Electromagnetic Fields and Energy, Prentice-Hall, Englewood Cliffs, NJ.

He, Z. (2001). Review of the Shockley-Ramo theorem and its application in semiconductor gamma-ray detectors. Nuclear Instruments and Methods in Physics Research A Vol. 463, 250.

He, Z.; Knoll, G.F.; Wehe, D.K., and Miyamoto, J. (1997). Position-sensitive single carrier CdZnTe detectors. Nuclear Instruments and Methods in Physics Research A Vol. 388, 180.

Heubrandtner, T.; Schnizer, B. (2002a). The quasi-static electromagnetic approximation
for weakly conducting media. Nuclear Instruments and Methods in Physics Research A Vol. 478, 444.

Heubrandtner, T.; Schnizer, B.; Lippmann,C.; Riegler, W.; (2002b). Static electric fields in an infinite plane condenser with one or three homogeneous layers. Nuclear Instruments and Methods in Physics Research A Vol. 489, 439.

Heubrandtner, T.; Schnizer, B., Schöpf, H. (1998). A simple theory for signals induced by a point charge moving in a resistive plate chamber. Nuclear Instruments and Methods in Physics Research A Vol. 419, 721.

Jackson, J.D. (1975). Classical Electrodynamics, 2nd ed., Wiley, New York.

Kim, K.O.; Kwon, T.J.; Kim, J.K. (2011). A new approach for evaluating the mobility-lifetime products of electron-hole pairs in semiconductor detectors. Journal of the Korean Physical Society Vol. 59, 20.

Kim, H.; Min, H.S.; Tang, T.W.; Park, Y.J. (1991). An extended proof of the Shockley-Ramo theorem. Solid-State Electronics Vol. 34, 1251.

Kinch, M. A.; Beck, J.D.; et al. (2004). "HgCdTe electron avalanche photodiodes." Journal of Electronic Materials **33**(6): 630-639.

Kittel, C. (2005). Introduction to solid state physics 8th ed. New York Wiley.

Kirkemo, C (2011). Monte Carlo Simulation of PN-junction, Master Thesis, UiO.

Kotov, I.V. (2005). Currents induced by charges moving in semiconductor. Nuclear Instruments and Methods in Physics Research A Vol. 539, 267.

Laurenti, J. P.; Bouhemadou, J. C. A.; Toulouse, B.; Legros, R.; Lusson, A. (1990). Temperature dependence of the fundamental absorption edge of mercury cadmium telluride. Journal of Applied Physics 67, 6454

Lippmann, C.; Vincke, H.; Riegler, W. (2009). Simulation of RPC performance for 511 keV photon detection. Nuclear Instruments and Methods in Physics Research A Vol. 602, 735.

Neyts, K.; Beeckman, J., Beunis, F. (2007). Quasistationary current contributions in electronic devices. Opto-Electronics Review Vol. 15, 41.

Nilsson, H.E.; Dubaric, E.; Hjelm, M.; Bertilsson, K. (2002). Simulation of photon and charge transport in X-ray imaging semiconductor sensors. Nuclear Instruments and Methods in Physics Research A Vol. 487, 151.

Norum, O. C. (2009). Monte Carlo Simulation of Semiconductors – Program Structure and Physical Phenomena. MSc, NTNU.

Olivero, P.; Forneris, J.; Gamarra, P. Jakšic, M.; Lo Giudice, A.: Manfredotti, C.; Pastuovic Z.; Skukan, N.; Vittone, E. (2011). Monte Carlo analysis of a lateral IBIC experiment on a 4H-SiC Schottky diode. Nuclear Instruments and Methods in Physics Research B Vol. 269, 2350.

Owens, A., and Kozorezov, A.G. (2006). Single carrier sensing techniques in compound semiconductor detectors. Nuclear Instruments and Methods in Physics Research A Vol. 563, 31.

Pellegrini, B. (1986). Electric charge motion, induced current, energy balance, and noise. Phys. Rev. B Vol. 34, 5921.

Petrovici, M.; Herrmann, N.; Hildenbrand, K.D.; Augustinski, G.; Ciobanu, M; Cruceru, I.; Duma, M.; Hartmann, O.; Koczon, P.; Kress, T.; Marquardt, M.; Moisa, D.; Petris, M.; Schroeder, C.; Simion, V.; Stoicea, G. and Weinert, J. (2002). A large-

area glass-resistive plate chamber with multistrip readout. Nuclear Instruments and Methods in Physics Research A Vol. 487, 337.

Ramo, S. (1939). Currents induced by electron motion. Proceedings of the IRE Vol. 27, 584.

Riegler, W. (2002). Induced signals in resistive plate chambers. Nuclear Instruments and Methods in Physics Research A Vol. 491, 258.

Riegler, W. (2004). Extended theorems for signal induction in particle detectors VCI 2004, Nuclear Instruments and Methods in Physics Research A Vol. 535, 287.

Riegler, W. (2009). Time response functions and avalanche fluctuations in resistive plate chambers. Nuclear Instruments and Methods in Physics Research A, Vol. 602, 377.

Samedov V.V. (2009). Radial dependence of induced current density and small pixel effect in parallel-plate detectors. IEEE Conf. Proceedings.

Schuttauf, A.; Ciobanu, M.; Hildenbrand, K.D.; Kis, M.; Deppner, I.; Herrmann, N.; Kim, Y.J.; Kang, T.I.; Koczon, P.; Lopez, X.; Leifels, Y.; Petrovici, M.; Piasecki, K.; Reischl, A.; Reisdorf, W.; Ryu, M.S.; Simion, V.; Zernezki, N.; Zhang, X. (2009). Multi-strip MRPCs for FOPI. Nuclear Instruments and Methods in Physics Research A Vol. 602, 679.

Shockley, W. (1938). Currents to conductors induced by a moving point charge. J. Appl. Phys. Vol. 9, 635.

Svaasand, L. O. (1981). Elektrisitet og magnetisme del 2. Magnetostatikk og elektrodynamikk, Tapir.

Vasileska, D; Goodnick, S.M.; Klimeck, G. (2010). Computational Electronics - Semiclassical and Quantum Device Modeling and Simulation, CRC Press.

Vittone, E.;. Fizzotti, F.; Lo Giudice, A.; Paolini C.; Manfredotti, C. (2000). Theory of ion beam induced charge collection in detectors based on the extended Shockley-Ramo theorem. Nuclear Instruments and Methods in Physics Research B Vols. 161-163, 446.

Yoder, P. D.; Gartner, K., and Fichtner, W. (1996). A generalized Ramo-Shockley theorem for classical to quantum transport at arbitrary frequencies. J. Appl. Phys. Vol .79, 1951.

Yoder, P. D.; Gartner, K; Krumbein, U., and Fichtner, W. (1997). Optimized terminal current calculation for Monte Carlo device simulation. IEEE Transactions on Computer Aided Design of Integrated Circuits Vol 16, 1087.