Chapter 1

Signal formation in diamond

This chapter describes the fundamentals of signal formation in a diamond sensor, as well as its use as a particle detector. This is described in section 1.1 where energy deposition and signal formation mechanism are explained. Then some examples of ionisation are shown. Later, some of the internal lattice defects that effect the signal are described. The final section contains the description of the remaining part of the signal chain – signal amplifiers, digitisers and devices for signal processing. Noise contributions are discussed at every stage of the signal chain.

There are many types of radiation sensors existing, but in this chapter we will focus on semiconductors, in particular on diamond sensors. Diamond is a good insulator, but behaves as a semiconductor in certain cases. In fact, the main principle of operation is the same for diamond, silicon and other semiconducting materials – ionisation. An impinging highly energetic charged particle ionises the atoms in the lattice, freeing electrons and holes, which then drift towards positively and negatively charged electrodes, inducing an electrical signal. A sensor converts the energy deposited by a particle or a photon and to an electrical signal.

Silicon is currently considered as the industry standard for particle detection. However, there are some disadvantages of using silicon instead of diamond, due to significant differences in the material properties. In particular, the properties of silicon change significantly with radiation. For instance, the leakage current increases, which in turn increases shot noise and can lead to thermal runaway. In addition, due to induced lattice defects, which act as charge traps, its charge collection efficiency starts dropping quickly. Both are true for diamond as well, but on a much smaller scale.

Table 1.1 compares the properties of diamond and silicon. Some of these values will be revisited and used in the course of this thesis.

Property	Diamond	Silicon
Band gap energy E_g (eV)	5.5	1.12
Electron mobility μ_e (cm^2 V^{-1} s^{-1})	1800	1350
Hole mobility μ_h (cm^2 V^{-1} s^{-1})	1200	450
Breakdown field (V cm^{-1})	10^7	3×10^5
Resistivity (Ω cm)	$> 10^{11}$	2.3×10^5
Intrinsic carrier density (cm^{-1})	$< 10^3$	1.5×10^{10}
Mass density (g cm^{-1})	3.52	2.33
Atomic charge	6	14
Dielectric constant ε	5.7	11.9
Displacement energy (eV/atom)	43	$13 - 20$
Energy to create an e-h pair (eV)	13	3.6
Radiation length (cm)	12.2	9.6
Avg. signal created/μm (e)	36	89

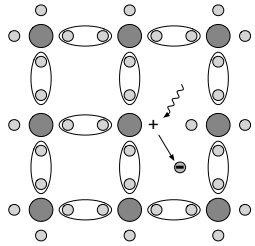
Table 1.1: Comparison diamond – silicon

1.1 Principles of signal formation in semiconductors

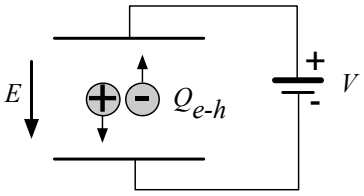
There are several ways the particles can interact with the sensor: via bremsstrahlung [], elastic or inelastic scattering (e-h pair production). Bremsstrahlung is radiation created when a particle is deflected from its original path due to attraction of the core of an atom. This is in principle an unwanted effect in semiconductors as it decreases the spatial resolution of the sensor. Elastic scattering is deflection of the particle's trajectory without energy loss. Inelastic scattering is the interaction through which the atom is ionised and an electron-hole pair is created. All these effects are competing and are dependent on the particle's mass, momentum etc.

Semiconductors are materials that are that are conductive only under specific conditions. They can be made up of atoms with four electrons in their valence band (e.g. silicon–Si, carbon–C or germanium–Ge) or as combinations of two or more different materials (e.g. gallium arsenide–GaAs). The atoms in the lattice form valence bonds with adjacent atoms, making solid crystal structures. These bonds can break up if sufficient external energy is applied. The electron that was forming the bond is kicked out, leaving behind a positively charged ion with a vacancy in its valence band (see figure 1.1a). A free electron-hole pair is thus created. The free electron travels through the crystal until it is caught by another hole. Similarly, the hole also "travels" through the material. Its positive charge attracts a bound electron in the vicinity, which breaks from the current bond and moves to the vacancy, leaving a new hole behind. The process continues, making it look like the vacancy – the hole – is traveling through the material.

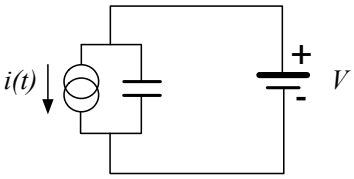
The electrons need to absorb a certain energy to get kicked out of the atomic bond – to get ionised. The minimal energy required to excite (ionise) an electron in a semiconductor is equal to the energy gap E_g. Typical widths of the forbidden gap are 0.7 eV in Ge, 1.12 eV in Si, 1.4 eV in GaAs and 5.5 eV in Di. Due to the small



(a) Valence bonds in the crystalline structure can be broken, producing a free electron-hole pair



(c) The freed electron-hole pair starts drifting in the externally applied electric field. The electron and hole drift in the opposite directions towards the oppositely charged electrods



(d) Equivalent electrical circuit. The moving charges act as a current source

Figure 1.1: In the equivalent electrical circuit diagram, electron-hole creation and drift can be modelled as a current source with a capacitor in parallel

band gap in semiconductors some electrons already occupy the conduction band at room temperature (RT). The intrinsic carrier concentration n_i in semiconductors is given as

$$n_i = T^{3/2} \cdot exp(-\frac{E_g}{2kT}) \tag{1.1}$$

wherein $k = 1.381 \times 10^{-23}$ m^2 kg s^{-2} K^{-1} is the Boltzmann constant and T is the temperature. The

If an external electric field is applied to the crystalline structure, the free electrons and holes drift toward the positive and negative potential, respectively (see figure 1.1c). While drifting, the charges couple with the electrodes, inducing current in the circuit, which is explained by the Shockley–Ramo theorem []. The charges recombine upon reaching the electrodes.

Energy deposition of α radiation and heavy ions

Energy deposition of β and γ radiation The mean energy loss of a particle traversing the detector with respect to its momentum is given with the the Bethe-Bloch equation []:

$$-\left\langle \frac{dE}{dx} \right\rangle = \frac{4\pi}{m_e c^2} \cdot \frac{nz^2}{\beta^2} \cdot \left(\frac{e^2}{4\pi\epsilon_0} \right)^2 \cdot \left[ln\left(\frac{2m_e c^2 \beta^2}{I \cdot (1-\beta^2)} \right) - \beta^2 \right] \tag{1.2}$$

The resulting function for a muon (a heavy electron) is shown in figure 1.2. At the momentum of around 300 MeV/c the particle deposits the lowest amount of energy. That is called a minimum ionising particle or a MIP.

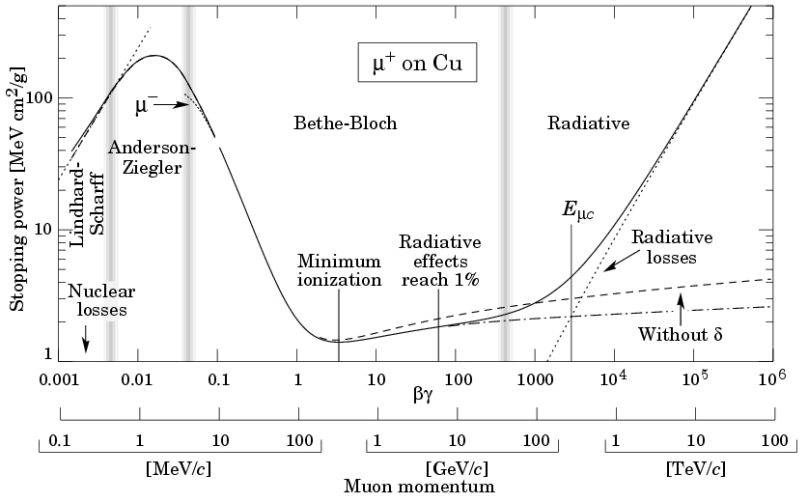
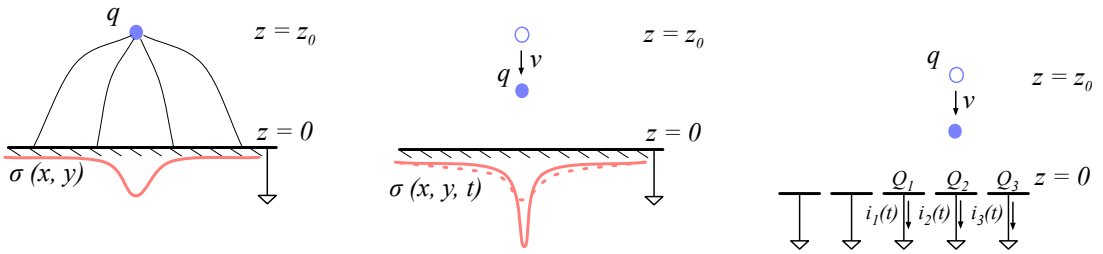


Figure 1.2: Stopping power for muons according to the Bethe-Bloch formula []



(a) Newly created point charge couples with the conductive plane

(b) When the charge drifts, the charge density in the plane changes

(c) The changing charge density in the small regions of the plane induces current

Figure 1.3: A point-like charge inducing current in a conductive plane

1.1.1 Signal induction by moving charges

The book [3] gives a simple introduction to understanding signal induction in a conducting plane by a point-like charge. The idea behind it lies in the coupling of the charge with the electrode. The electrode can be in this case modelled as an infinite conducting plane. When the point charge q is created (e.g. an electron-hole pair created via ionisation), its electrostatic field lines immediately couple with the electrode, as seen in figure 1.3a. The electric field on the metal surface due to a point-like charge q at the distance z_0 equals

$$E_z(x,y) = \frac{qz_0}{2\pi\epsilon_0(x^2+y^2+z_0^2)^{\frac{3}{2}}} \qquad E_y = E_z = 0. \tag{1.3}$$

A mirror charge appears on the conducting plane, with a charge density distribution

$$\sigma(x,y) = \epsilon_0 E_z(x,y) = \frac{qz_0}{2\pi(x^2+y^2+z_0^2)^{\frac{3}{2}}}. \tag{1.4}$$

The charge density integrated over the whole plane gives the mirror charge Q, which has the opposite value of the point charge q:

$$Q = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sigma(x,y)dxdy = -q. \tag{1.5}$$

Now we segment the plane into infinitely long strips with a width w whereby each of the strips is grounded (figure 1.3c). With the charge density distribution 1.4, the resulting mirror charge on a single strip Q_2 directly below the point charge ($x = 0, y = 0$) will be equal to

$$Q_2(z_0) = \int_{-\infty}^{\infty}\int_{-w/2}^{w/2} \sigma(x,y)dxdy = -\frac{2q}{\pi} arctan\left(\frac{w}{2z_0}\right) \tag{1.6}$$

If the charge starts moving towards the conducting plane, the mirror charge density distribution also changes (see figure 1.3b). This results in the $Q_2[z_0(t)]$ to change with time, inducing an electric current $i_n(t)$:

$$i_n(t) = -\frac{d}{dt}Q_2[z_0(t)] = -\frac{\partial Q_2[z_0(t)]}{\partial z_0}\frac{\partial z_0(t)}{\partial t} = \frac{4qw}{\pi[4z_0(t)^2 + w^2]}v. \tag{1.7}$$

The movement of the point-like charge therefore induces current in the conducting plane. The induced current is linearly dependent on the velocity of the point-like charge.

W. Shockley [5] and S. Ramo [4] independently proposed a theory which explains how a moving point charge induces current in a conductor. The Shockley-Ramo theorem can therefore be used to calculate the instantaneous electric current induced by the charge carrier or a group of charge carriers. It can be used for any number of electrodes. It states that the current $I_n^{ind}(t)$ induced on the grounded electrode n by a point charge q moving along a trajectory $\mathbf{x}(t)$ equals

$$I_n^{ind}(t) = -\frac{dQ_n(t)}{dt} = -\frac{q}{V_w}\nabla\Psi_n[\mathbf{x}(t)]v(t) = -\frac{q}{V_w}E_n[\mathbf{x}(t)]v(t), \tag{1.8}$$

where $\mathbf{E}_n(\mathbf{x})$ is the electric field in the case where the charge q is removed, electrode n is set to voltage $V_w = 1$ and all other electrodes are grounded. $\mathbf{E}_n(\mathbf{x})$ is also called the *weighting field* of electrode n and is defined as the spatial differential of the *weighting potential*: $\mathbf{E}_n(\mathbf{x}) = \nabla\Psi_n(\mathbf{x})$. In the case of two parallel electrodes, the weighting field is $E_w = -\frac{d\Psi}{dx} = -1/d$, where d is the distance between the electrodes. The resulting induced current is therefore

$$i(t) = \frac{q}{d}v_{drift}(x,t), \tag{1.9}$$

whhereby v_{drift} is the drift velocity of the point-like charge and d is the distance between the electrodes.

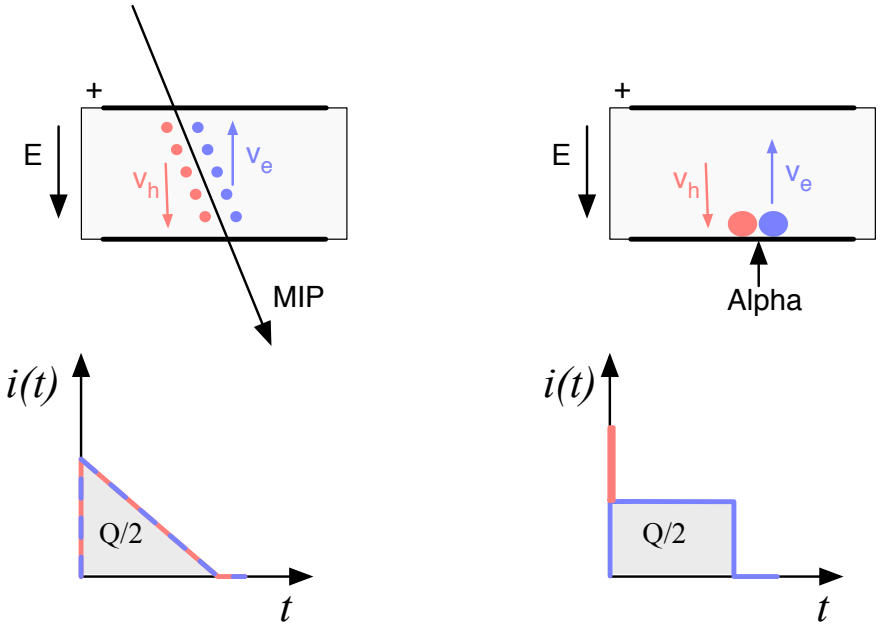


Figure 1.4: Charge carrier drift in diamond for β/γ and for α particles

1.1.2 Radiation-induced electrical pulses

When a highly-energetic particle travels through the sensor, it interacts with atoms in the lattice. It ionises the valence electrons, creating electron-hole (e-h) pairs on its way. It can either deposit only a fraction of its energy and fly exit the sensor on the other side or it can get stopped in the bulk, depositing all of its energy. A special case is when it interacts with the core of the atom in the middle of the sensor via a nuclear interaction. All these various types interactions produce different amounts and different spatial distributions of e-h pairs. The induced electrical current will therefore differ for different types of interaction. Two most frequent types are shown in figure 1.4. The first diagram shows the interaction of a minimum ionising particle (an electron or a proton) or in some cases a photon, if it is energetic enough. The electrons and holes are created all along the trajectory of the particle and immediately start drifting towards the positive and negative electrode, respectively. At the beginning, all charges drift and contribute to the induced current. Those closest to the electrodes have a very short drift path and recombine quickly, reducing the induced current. Gradually all the charge carriers recombine. The resulting current signal is a triangular pulse with a sharp rising edge and a linear falling edge. The accumulated charge Q_s equals to the sum of the contributions of the positive and negative charge carriers. The second type of interaction happens when the particle is stopped in the diamond close to the point of entry. Most of its energy is deposited in one point. A cloud of charge carriers is created and the charges with the shorter path to the electrode recombine almost instantly. The carriers of the opposite charge, however, start

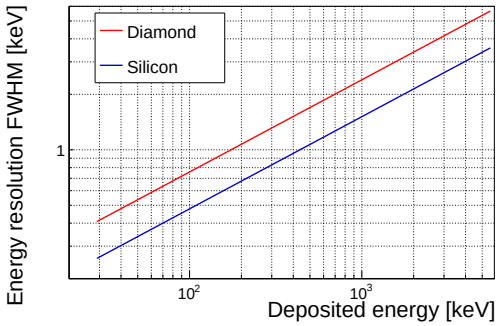


Figure 1.5: Calculated intrinsic energy resolution for silicon and diamond

drifting through the sensor to the other electrode. In an ideal diamond sensor, their velocity is constant throughout the drift up until they recombine on the other side. The contribution of the first charge cloud is a peak with a short time. The cloud drifting through the sensor, on the other side, induces a current signal with a flat top. The resulting signal is rectangular shaped, with a spike in the beginning. This spike is filtered out in a real device because it is too fast for the electronics existing currently. The accumulated charge Q_s is equal to a half of the deposited charge by the stopped particle.

As seen the two types of interactions have well defined signal responses. Nuclear interactions on the other hand yield various results. The resulting signal shape depends on the decay products of the interaction – they can be α, β or γ quanta, inducing a mixed shaped signal.

1.1.3 Signal charge fluctuations

Two of the important sensor characteristics are the magnitude of the signal and the fluctuations of the signal at a given absorbed energy. They determine the relative resolution ΔE/E. For semiconductors the signal fluctuations are smaller than the simple statistical variance $\sigma_Q = \sqrt{N_Q}$, where N_Q is the number of released charge pairs (ratio between the total deposited energy E_0 and the average emergy deposition E_i required to produce a charge pair). In fact, [] shows that the variance is $\sigma_Q = \sqrt{FN_Q}$, where F is the Fano factor [] (0.08 for diamond and 0.115 for silicon []). Thus, the variance of the signal charge is smaller than expected, $\sigma_Q \approx 0.3\sqrt{N_Q}$. The resulting intrinsic resolution of semiconductor detectors is

$$\Delta E_{FWHM} = 2.35\sqrt{FEE_i} \tag{1.10}$$

wherein $E_i(Si)$= 3.6 eV and $E_i(C)$= 13 eV. E.g., for an α particle with energy E_α = 5.486 MeV the calculated resolution in diamond is equal to $\Delta E_{FWHM} = 5.6$ keV. This defines the minimum achievable resolution for energy spectroscopy with diamond. Figure 1.5 shows the calculated energy resolution function for silicon and diamond.

1.2 Carrier transport in a diamond sensor

This section describes the carrier transport phenomena in diamond. This theory provides the basis for discussion about the measurements in chapter **??**.

Free charge carriers in a semiconductor get thermally excited and scatter in random directions with a thermal velocity v_{th} []. Their integral movement due to thermal excitation equals zero. Their transport is instead by means of drift and diffusion. Diffusion is caused by the concentration gradient. In its presence the carriers tend to scatter in the direction of the lower concentration. Drift on the other hand is caused by an externally applied electrical field. In that case the carriers move in parallel to to the field lines. In a sensor with a high applied field the diffusion contribution is negligible.

Diffusion The concentration profile dissolves with time forming a Gaussian distribution with variance $\sigma(t) = \sqrt{Dt}$ [] .

Drift velocity and mobility The charge carriers drift through the diamond bulk with a drift velocity $v_{drift}(E)$ [], which is proportional to the electric field E at low electric fields: $v_{drift} = \mu E$. The proportionality factor μ is defined as the mobility in $\mathrm{cm^2V^{-1}s^{-1}}$. For higher fields, however, the velocity saturates. The final equation for v_{drift} is therefore

$$v_{drift}(E) = \mu(E)E = \frac{\mu_0 E}{1 + \frac{\mu_o E}{v_{sat}}} \tag{1.11}$$

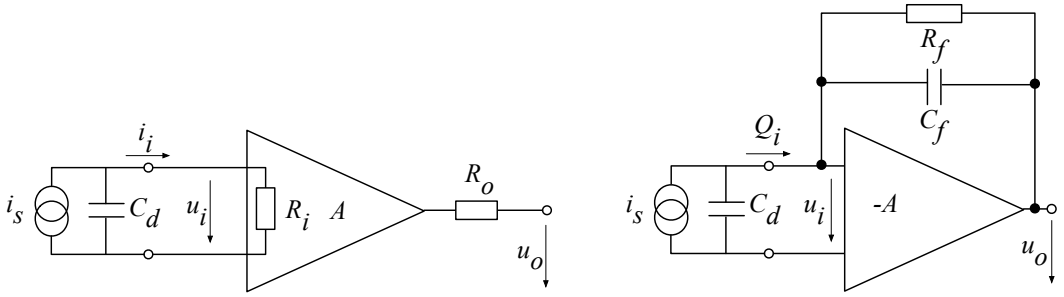
where μ_0 is the low drift mobility and v_{sat} is saturation velocity. The drift velocity can be retrieved experimentally via the transit time measured with the Transient Current Technique (TCT). This technique enables the measurement of transit time t_t of the carriers through the sensor with the thickness d.

$$v_{drift}(E) = \frac{d}{t_t(E)}. \tag{1.12}$$

The velocities for holes and electrons usually differ. In diamond, the holes travel 30 % faster than electrons [].

Velocity saturation At higher velocities the carriers lose more energy to the lattice (phonon transport).

Space carge



(a) A capacitive source and a current amplifier

(b) A capacitive source and a charge amplifier

Figure 1.6: Simplified equivalent circuits of a current and charge amplifier

1.3 Electronics for signal processing

This section describes the electronics of a detector, starting with a description of signal amplifiers and then discussing the digitisation and signal processing. All these stages are necessary to extract information from the sensor. First, the signal has to be amplified. Then it is digitised and finally processed in a specially designed processor or a logic unit.

1.3.1 Signal preamplifiers

The signal charge generated in the sensor by a single highly energetic particle or photon is of the order of fC. The induced current is ranging between 10^{-8} A (β, γ radiation) and 3×10^{-7} A (α radiation). Signals as low as these have to be pre-amplified before processing. Depending on the measurement, several types of signal amplifiers can be used. The preamplifiers have to be designed carefully to minimise electronic noise while maximising gain – thus maximising the signal-to-noise ratio (SNR). A critical parameter is the total capacitance, i.e. sensor capacitance and input capacitance of the preamplifier. The SNR improves with a lower capacitance. Several types of amplifiers can be used, all of which affect the measured pulse shape. They behave differently for resistive or capacitive sources. Given that semiconductors are capacitive sources, we will focus on these. Two preamplifiers are used most commonly, a current and a charge amplifier. Both are described below in detail.

Current-sensitive amplifier

Figure 1.6a shows the equivalent circuit of a capacitive source and a current amplifier. An amplifier operates in current mode if the source has a low charge collection time t_c with respect to the $R_i C_d$ time constant of the circuit. In this case the sensor capacitance discharges rapidly and the output current i_o is proportional to the instantaneous current i_i. The amplifier is providing a voltage gain, so the output signal

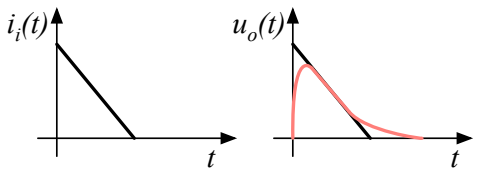


Figure 1.7: Input and output signal of the current amplifier

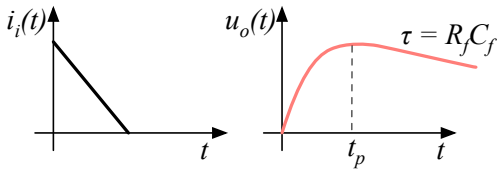


Figure 1.8: Input and output signal of the charge amplifier

voltage u_o is directly proportional to the input voltage u_i:

$$u_o(t) = A \cdot R_i \cdot i_s(t). \tag{1.13}$$

The detector capacitance C_{det} together with the input resistance of the amplifier R_i defines the time constant of the signal (see figure 1.7). The higher the C_{det} is, the slower will be the response of the amplifier. For the case of the diamond sensor, which has the capacitance of the order of 2 pF and the input resistance of 50 Ω, the resulting time constant is $\tau = 10^{-10}$ s. This yields the signal rise time $t_r \sim 2.2\tau = 0.22\ ns$.

Charge-sensitive amplifier

(0.5 pg) In order to measure integrated charge in the sensor, a feedback loop is added to the amplifier (see figure 1.6b). The feedback can be used to control the gain and input resistance, as well as integrating the input signal. The charge amplifier is in principle an inverting voltage amplifier with a high input resistance.

In an ideal amplifier the output voltage u_o equals $-Au_i$. Therefore voltage difference across the capacitor C_f is $u_f = (A+1)u_i$ and the charge deposited on the capacitor $Q_f = C_f u_f = C_f(A+1)u_i$. Since no current can flow into the amplifier, all of the signal current must charge up the feedback capacitance, so $Q_f = Q_i$.

In reality, however, charge-sensitive amplifiers respond much slower to the than the time duration of the current pulse from the sensor. In addition, a resistor is added to the feedback line in parallel to the capacitor. The resistor and capacitor define the decay time constant of the pulse (see figure 1.8). This is necessary to return the signal to its initial state and ready for a new measurement.

Analogue electronic noise

(2 pg) Electronic noise determines the ability of a system to distinguish signal levels. The analogue signal contains a lot of information, which can quickly be erased or altered if the signal properties change. It is therefore instrumental to understand the noise contributions to the signal to qualify the information it carries. There are several noise contributions, of which the important ones are listed below. The thermal noise is the dominant noise contribution in the use case for diamond detector signal amplification and therefore defines the limitations of the detector system. Thermal noise or Johnson–Nyquist noise is generated by the random thermal motion of charge carriers in the conductor. The frequency range of the thermal noise is from 0 to ∞ with a more or less uniform distribution. Therefore this is nearly white noise. The resulting signal amplitude has a Gaussian distribution. The RMS of the noise amplitude is defined as

$$u_{RMS} = \sqrt{4k_B RT\Delta f}\ [] \tag{1.14}$$

where k_B is the Boltzmann constant, R is the resistance of the conductor, T its temperature and Δf the frequency range. This equation shows that it is possible to reduce the noise RMS by either (1) reducing the frequency range, (2) reducing the resistance of the conductor or (3) cooling the conductor.

Contributions of shot noise, flicker noise and burst noise and other types are not significant relative to the thermal noise. However, the contributions of external factors can severely deteriorate the signal. This means the noise produced by capacitive or inductive coupling with an external source, which causes interference in the signal. These effects can be reduced by shielding the electronics and avoiding ground loops.

1.3.2 Analogue-to-digital converters

An analogue-to-digital converter (ADC) is a device that converts the analogue electrical signal on the input to its digital representation - a series of digital values. This involves a quantisation – *sampling* of the signal at a defined sampling period, resulting in a sequence of samples at a discrete time period and with discrete amplitude values. The resolution of the ADC is the number of output levels the ADC can quantise to and is expressed in bits. For instance, an ADC with a resolution $n = 8$ bit will have the dynamic range $N = 2^n = 256$ steps. The resulting voltage resolution Q_{ADC} at the input voltage range of $V_{ADC} = \pm 50$ mV is then equal to

$$Q_{ADC} = \frac{V_{ADC}}{2^n} = \frac{100\ mV}{2^8\ steps} = 0.39\ mV/step. \tag{1.15}$$

With a sampling period of t_s =1 ns it will produce the sampling rate of f_s =1 GSPS (gigasample per second).

Quantisation error and quantisation noise (or a round-off error) is a contribution to the overall measurement error due to digitisation (rounding). It is defined as a difference between the actual analog value and a digitised representation of this

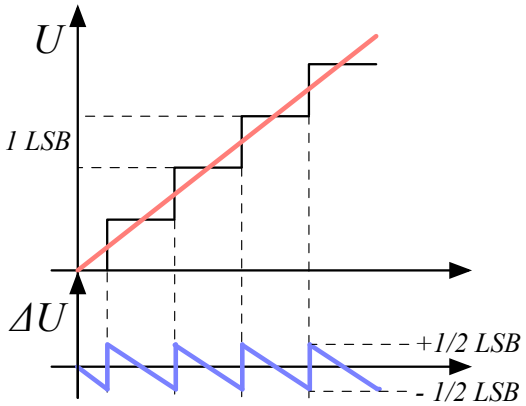


Figure 1.9: Input signal digitisation and quantisation error

value. The error is defined by the least significant bit (LSB), as seen in figure 1.9. Typically, the input signal amplitude is much larger than than the voltage resolution. Therefore the quantisation error is not directly correlated with the signal and has an approximately uniform distribution []:

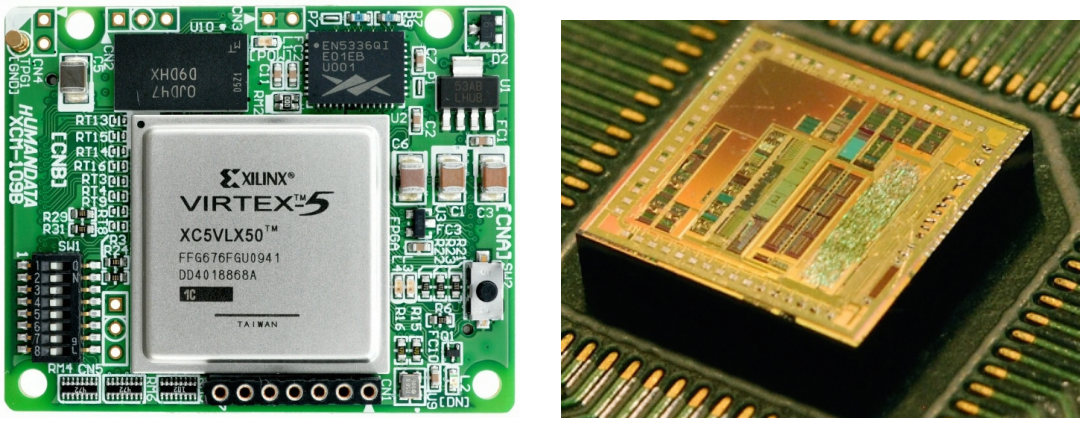
$$\Delta Q_{ADC} = \frac{1}{\sqrt{12}} LSB \sim 0.289\ LSB. \tag{1.16}$$

For the example above the quantisation error will be $\Delta Q_{ADC} = 0.289 \cdot 0.39$ mV=0.11 mV. The error depends strongly on the linearity of the ADC, but this will not be discussed in this document as the devices used have ADCs with a linear response.

1.3.3 Digital signal processing

The digitised signal can be processed to extract useful information. Therefore after the signal amplification and digitisation the signal is routed in a device which handles the analysis. The signal can either be processed immediately (in real time) or it can be saved to a data storage for analysis at a later stage (offline). The devices carrying out the processing can be multipurpose (e.g. Field programmable gate arrays) or dedicated (e.g. application-specific integrated circuits). Each of the two has its advantages and disadvantages, which are listed below.

Field programmable gate array (FPGA) is an integrated circuit designed to be reprogrammable and configured after manufacturing. It consists of a set of logic gates that can be interconnected in numerous combinations to carry out a logic operation. Many such logic operations can take place in parallel, making the FPGA a powerful tool for signal processing. FPGAs are often used during system development or in systems in which the requirements might change with time. They can be reprogrammed in the order of seconds. In addition, the logic design only needs minor changes when



(a) Xilinx Virtex 5 FPGA [2]

(b) ASIC [1]

Figure 1.10: An example of an FPGA and an ASIC chip

migrating to a newer version of the FPGA chip of the same vendor. They also offer faster time-to-market with comparison to application-specific solutions, which have to be developed. On the other hand, the price per part can be significantly higher than for the application-specific solutions. Also, their other major disadvantages are a high power consumption and relatively low speed. However, today's solutions are capable of speeds of the order of 500 MHz. Together with the integrated digital signal processing blocks, embedded processors and other modules, they are already very powerful and versatile. All in all, FPGAs are a good choice for prototyping and limited production, for projects with a limited requirements for speed and complexity.

Application-specific integrated circuit (ASIC) is an integrated circuit designed for a specific use. The design cannot be modified after chip production, as compared to FPGAs. On the other hand, the ASICs can be optimised to perform a required operation at a high speed and low power. In addition, due to the specific design the size of the chip can be much smaller. ASICs can be designed as hybrid chips, containing both a digital and an analog part. To update the chip, the design has to be submitted to a foundry, which produces the new chips with a turnover time of 4—6 weeks. The costs of a submission start at $ 50 000, but the price per part can be reduced significantly with a high volume. To sum up, ASICs are used for high volume designs with well defined requirements where some stringent constraints in terms of power consumption and speed have to be met.

Bibliography

[1] *ASIC*.

[2] *Xilinx Virtex 5 FPGA*.

[3] W. Blum, W. Riegler, and L. Rolandi. Particle Detection with Drift Chambers, volume 2 of *2*. Springer-Verlag, Berlin Heidelberg, 2008.

[4] S. Ramo. *Currents Induced by Electron Motion. Proceedings of the IRE*, 27:584–585, 1939.

[5] W. Shockley. *Currents to Conductors Induced by a Moving Point Charge. Journal of applied Physics*, 9:635, 1938.