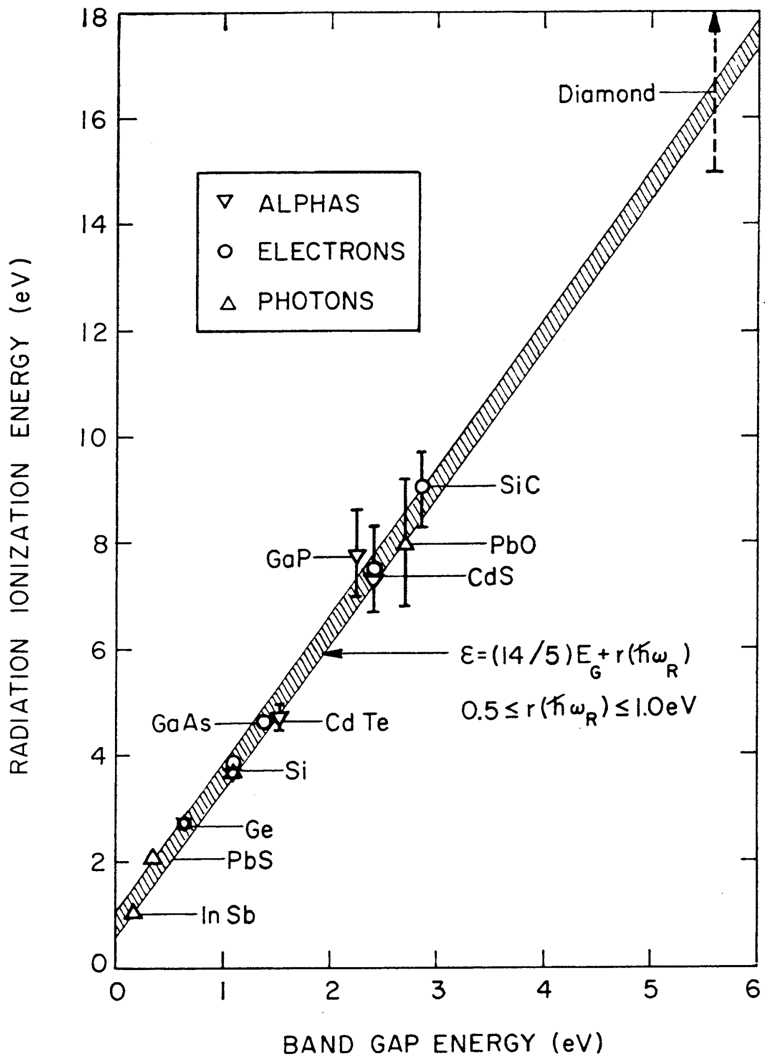
IX. Semiconductor Detectors – Part II

1. Fluctuations in the Signal Charge: the Fano Factor

It is experimentally observed that the energy required to form an electron-hole pair exceeds the bandgap.



C.A. Klein, J. Applied Physics **39** (1968) 2029

The mean ionization energy exceeds the bandgap for two reasons

1. Conservation of momentum requires excitation of lattice vibrations

2. Many modes are available for the energy transfer with an excitation energy less than the bandgap.

Two types of collisions are possible:

a) Lattice excitation, i.e. phonon production (with no formation of mobile charge).

b) Ionization, i.e. formation of a mobile charge pair.

Assume that in the course of energy deposition

N_x excitations produce N_P phonons and

N_i ionization interactions form N_Q charge pairs.

On the average, the sum of the energies going into excitation and ionization is equal to the energy deposited by the incident radiation

$$E_0 = E_i N_i + E_x N_x$$

where E_i and E_x are the energies required for a single excitation or ionization.

Assuming gaussian statistics, the variance in the number of excitations

$$\sigma_x = \sqrt{N_x}$$

and the variance in the number of ionizations

$$\sigma_i = \sqrt{N_i}$$

For a single event, the energy E_0 deposited in the detector is fixed (although this may vary from one event to the next).

If the energy required for excitation E_x is much smaller than required for ionization E_i, sufficient degrees of freedom will exist for some combination of ionization and excitation processes to dissipate precisely the total energy. Hence, for a given energy deposited in the sample a fluctuation in excitation must be balanced by an equivalent fluctuation in ionization.

$$E_x \Delta N_x + E_i \Delta N_i = 0$$

If for a given event more energy goes into charge formation, less energy will be available for excitation. Averaging over many events this means that the variances in the energy allocated to the two types of processes must be equal

$$E_i \sigma_i = E_x \sigma_x$$

$$\sigma_i = \frac{E_x}{E_i} \sqrt{N_x}$$

From the total energy $E_i N_i + E_x N_x = E_0$

$$N_x = \frac{E_0 - E_i N_i}{E_x}$$

yielding

$$\sigma_i = \frac{E_x}{E_i} \sqrt{\frac{E_0}{E_x} - \frac{E_i}{E_x} N_i}$$

Since each ionization leads to a charge pair that contributes to the signal

$$N_i = N_Q = \frac{E_0}{\varepsilon_i}$$

where ε_i is the average energy loss required to produce a charge pair,

$$\sigma_i = \frac{E_x}{E_i}\sqrt{\frac{E_0}{E_x} - \frac{E_i}{E_x}\frac{E_0}{\varepsilon_i}}$$

$$\sigma_i = \sqrt{\frac{E_0}{\varepsilon_i}} \cdot \sqrt{\frac{E_x}{E_i}\left(\frac{\varepsilon_i}{E_i} - 1\right)}$$

The second factor on the right hand side is called the Fano factor F.

Since σ_i is the variance in signal charge Q and the number of charge pairs is $N_Q = E_0/\varepsilon_i$

$$\sigma_Q = \sqrt{FN_Q}$$

In Silicon E_x = 0.037 eV
$E_i = E_g$ = 1.1 eV
ε_i = 3.6 eV

for which the above expression yields F= 0.08, in reasonable agreement with the measured value F= 0.1.

⇒ The variance of the signal charge is smaller than naively expected

$$\sigma_Q \approx 0.3\sqrt{N_Q}$$

A similar treatment can be applied if the degrees of freedom are much more limited and Poisson statistics are necessary.

However, when applying Poisson statistics to the situation of a fixed energy deposition, which imposes an upper bound on the variance, one can not use the usual expression for the variance

$$\mathrm{var}\, N = \overline{N}$$

Instead, the variance is

$$\overline{(N - \overline{N})^2} = F\,\overline{N}$$

as shown by Fano [1] in the original paper.

An accurate calculation of the Fano factor requires a detailed accounting of the energy dependent cross sections and the density of states of the phonon modes. This is discussed by Alkhazov [2] and van Roosbroeck [3].

References:

1. U. Fano, Phys. Rev. **72** (1947) 26

2. G.D. Alkhazov et al., NIM **48** (1967) 1

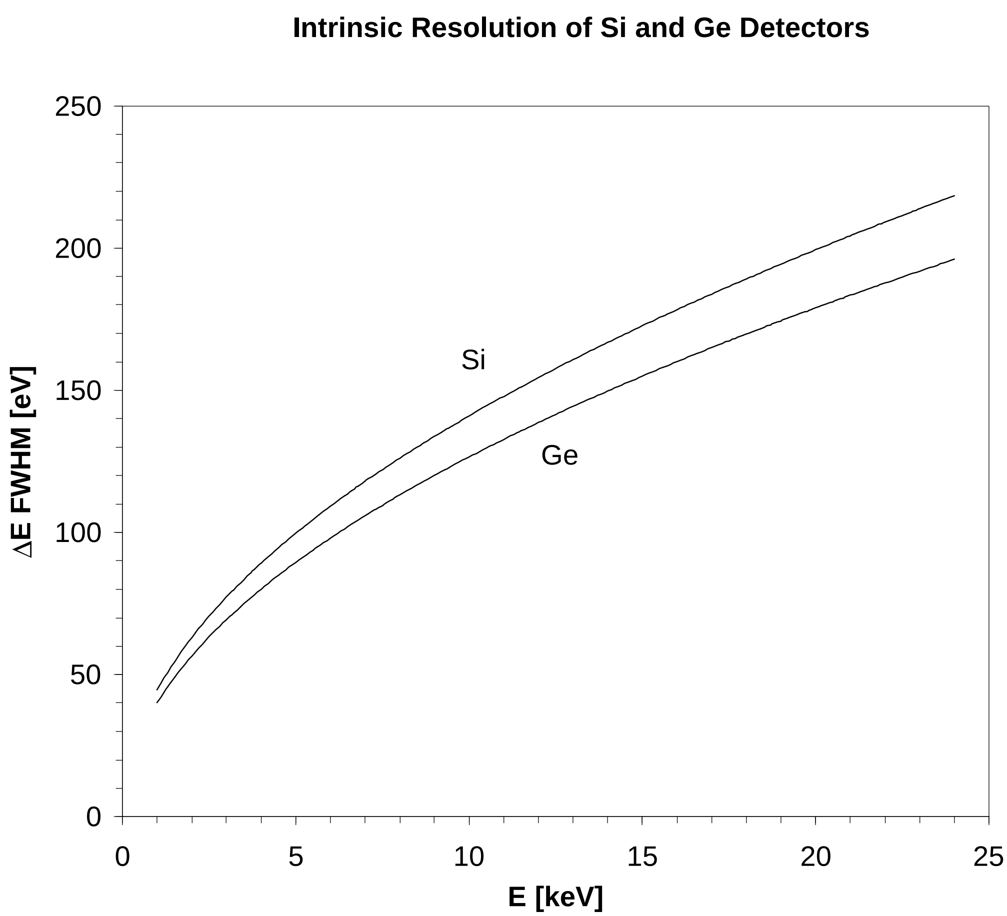
3. W. van Roosbroeck, Phys. Rev. **139** (1963) A1702

Intrinsic Resolution of Semiconductor Detectors

$$\Delta E = 2.35 \cdot \varepsilon_i \sqrt{F N_Q} = 2.35 \cdot \varepsilon_i \sqrt{F \frac{E}{w}} = 2.35 \cdot \sqrt{F E \varepsilon_i}$$

Si: ε_i= 3.6 eV F= 0.1

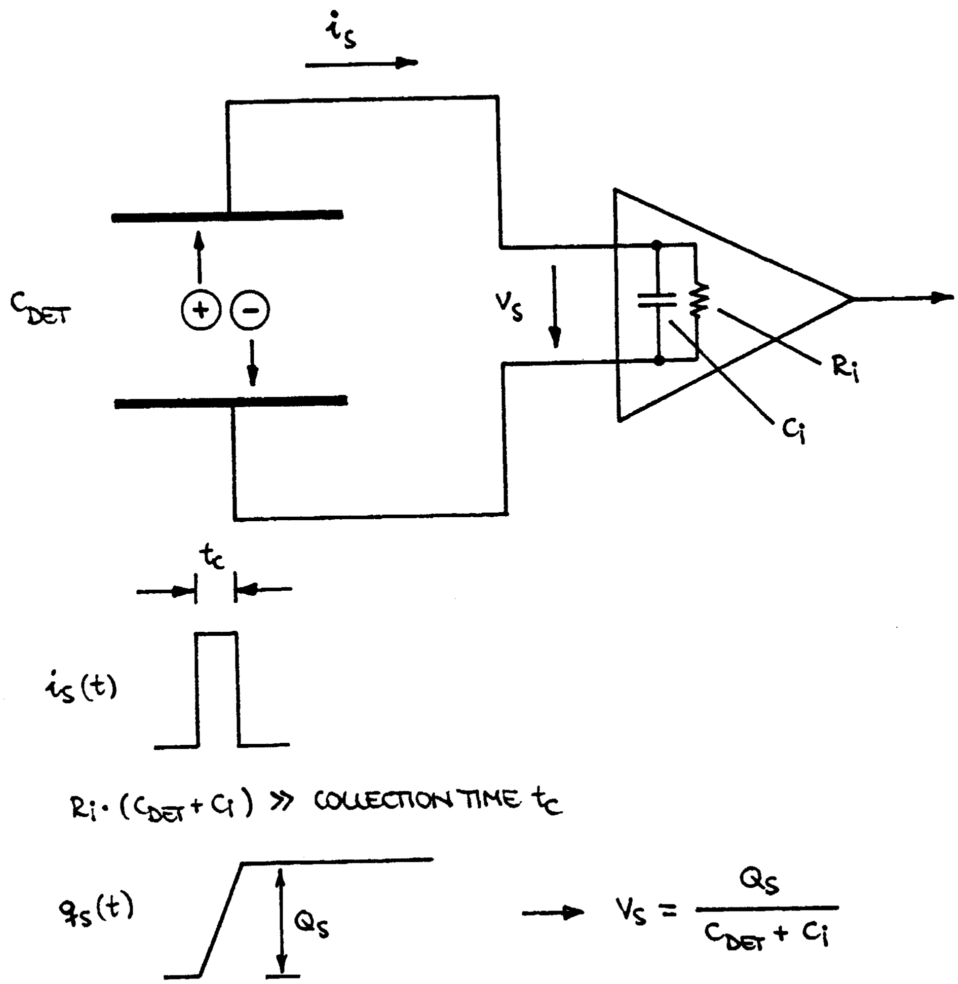
Ge: ε_i= 2.9 eV F= 0.1



Detectors with good efficiency for this energy range have sufficiently small capacitance to allow electronic noise of ~100 eV FWHM, so the variance of the detector signal is a significant contribution.

At energies >100 keV the detector sizes required tend to increase the electronic noise to dominant levels.

2. Induced Charge



When does the current pulse begin?

a) when the charge reaches the electrode?

or

b) when the charge begins to move?

Although the first answer is quite popular (encouraged by the phrase “charge collection”), the second is correct.

When a charge pair is created, both the positive and negative charges couple to the electrodes and induce mirror charges of equal magnitude.

As the positive charge moves toward the negative electrode, it couples more strongly to it and less to the positive electrode.

Conversely, the negative charge couples more to the positive electrode and less to the negative electrode.

The net effect is a negative current at the positive electrode and a positive current at the negative electrode, due to both the positive and negative charges.

Magnitude of the Induced Charge

(S. Ramo, Proc. IRE **27** (1939) 584)

Consider a mobile charge in the presence of any number of grounded electrodes.

Surround the charge q with a small equipotential sphere. Then, if V is the potential of the electrostatic field, in the region between conductors

$$\nabla^2 V = 0$$

Call V_q the potential of the small sphere and note that $V = 0$ on the conductors. Applying Gauss' law yields

$$\int\limits_{\substack{\text{sphere's}\\ \text{surface}}} \frac{\partial V}{\partial n}\, ds = 4\pi q$$

Next, consider the charge removed and one conductor A raised to unit potential.

Call the potential V_1, so that

$$\nabla^2 V_1 = 0$$

in the space between the conductors, including the site where the charge was situated. Call the new potential at this point V_{q1}.

Green's theorem states that

$$\int\limits_{\substack{\text{volume between}\\ \text{boundaries}}} (V_1 \nabla^2 V - V\, \nabla^2 V_1)\, dv = -\int\limits_{\substack{\text{boundary}\\ \text{surfaces}}} \left[V_1 \frac{\partial V}{\partial n} - V \frac{\partial V_1}{\partial n} \right] ds$$

Choose the volume to be bounded by the conductors and the tiny sphere.

Then the left hand side is 0 and the right hand side may be divided into three integrals:

1. Over the surfaces of all conductors except A. This integral is 0 since on these surfaces $V = V_1 = 0$.

2. Over the surface of A. As $V_1 = 1$ and $V = 0$ this reduces to

$$-\int_{\text{surface A}} \frac{\partial V}{\partial n} ds$$

3. Over the surface of the sphere.

$$-V_{q1} \int_{\substack{\text{sphere's}\\ \text{surface}}} \frac{\partial V}{\partial n} ds + V_q \int_{\substack{\text{sphere's}\\ \text{surface}}} \frac{\partial V_1}{\partial n} ds$$

   The second integral is 0 by Gauss' law, since in this case the charge is removed.

Combining these three integrals yields

$$0 = -\int_{\text{surface A}} \frac{\partial V}{\partial n} ds - V_{q1} \int_{\substack{\text{sphere's}\\ \text{surface}}} \frac{\partial V}{\partial n} ds = 4\pi\, Q_A - 4\pi\, qV_{q1}$$

or

$$Q_A = qV_{q1}$$

If the charge q moves in direction x, the current on electrode A is

$$i_A = \frac{dQ_A}{dt} = q\frac{dV_{q1}}{dt} = q\left(\frac{\partial V_{q1}}{\partial x}\frac{dx}{dt}\right)$$

Since the velocity of motion

$$\frac{dx}{dt} = v_x$$

the induced current on electrode A is

$$i_A = q\,v_x\frac{\partial V_{q1}}{dx}$$

where V_{q1} is the "weighting potential" that describes the coupling of a charge at any position to electrode A.

The weighting potential is for a specific electrode is obtained by setting the potential of the electrode to 1 and setting all other electrodes to potential 0.

- If a charge q moves along any path s from position 1 to position 2, the net induced charge on electrode k is

$$\Delta Q_k = q(V_{q1}(2) - V_{q1}(1)) \equiv q(\Phi_k(2) - \Phi_k(1))$$

- The instantaneous current can be expressed in terms of a weighting field

$$i_k = -q\,\vec{v}\cdot\overrightarrow{F_k}$$

  The weighting field is determined by applying unit potential to the measurement electrode and 0 to all others.

Note that the electric field and the weighting field are distinctly different.

- The electric field determines the charge trajectory and velocity

- The weighting field depends only on geometry and determines how charge motion couples to a specific electrode.

- Only in 2-electrode configurations are the electric field and the weighting field of the same form.

Example 1: Parallel plate geometry, disregarding space charge
(semiconductor detector with very large overbias)

Assume a voltage V_b applied to the detector. The distance between the two parallel electrodes is d.

The electric field that determines the motion of charge in the detector is

$$E = \frac{V_b}{d}$$

so the velocity of the charge

$$v = \mu E = \mu \frac{V_b}{d}$$

The weighting field is obtained by applying unit potential to the collection electrode and grounding the other.

$$E_Q = \frac{1}{d}$$

so the induced current

$$i = qvE_Q = q\mu \frac{V_b}{d} \frac{1}{d} = q\mu \frac{V_b}{d^2}$$

since both the electric field and the weighting field are uniform throughout the detector, the current is constant until the charge reaches its terminal electrode.

Assume that the charge is created at the opposite electrode and traverses the detector thickness d.

The required collection time, i.e. the time required to traverse the detector thickness d

$$t_c = \frac{d}{v} = \frac{d}{\mu \frac{V_b}{d}} = \frac{d^2}{\mu V_b}$$

The induced charge

$$Q = it_c = q\mu \frac{V_b}{d^2} \frac{d^2}{\mu V_b} = q$$

Next, assume an electron-hole pair formed at coordinate x from the positive electrode.

The collection time for the electron

$$t_{ce} = \frac{x}{v_e} = \frac{xd}{\mu_e V_b}$$

and the collection time for the hole

$$t_{ch} = \frac{d-x}{v_h} = \frac{(d-x)d}{\mu_h V_b}$$

Since electrons and holes move in opposite directions, they induce current of the same sign at a given electrode, despite their opposite charge.

The induced charge due to the motion of the electron

$$Q_e = q_e \mu_e \frac{V_b}{d^2} \frac{xd}{\mu_e V_b} = q_e \frac{x}{d}$$

whereas the hole contributes

$$Q_h = q_e \mu_h \frac{V_b}{d^2} \frac{(d-x)d}{\mu_h V_b} = q_e \left(1 - \frac{x}{d}\right)$$

Assume that $x = d/2$. After the collection time for the electron

$$t_{ce} = \frac{d^2}{2\mu_e V_b}$$

it has induced a charge $q_e/2$.

At this time the hole, due to its lower mobility $\mu_h \approx \mu_e/3$, has induced $q_e/6$, yielding a cumulative induced charge of $2q_e/3$.

After the additional time for the hole collection, the remaining charge $q_e/3$ is induced, yielding the total charge q_e.

In this configuration

- Electrons and holes contribute equally to the currents on both electrodes
- The instantaneous current at any time is the same (although of opposite sign) on both electrodes

The continuity equation (Kirchhoff's law) must be satisfied

$$\sum_k i_k = 0$$

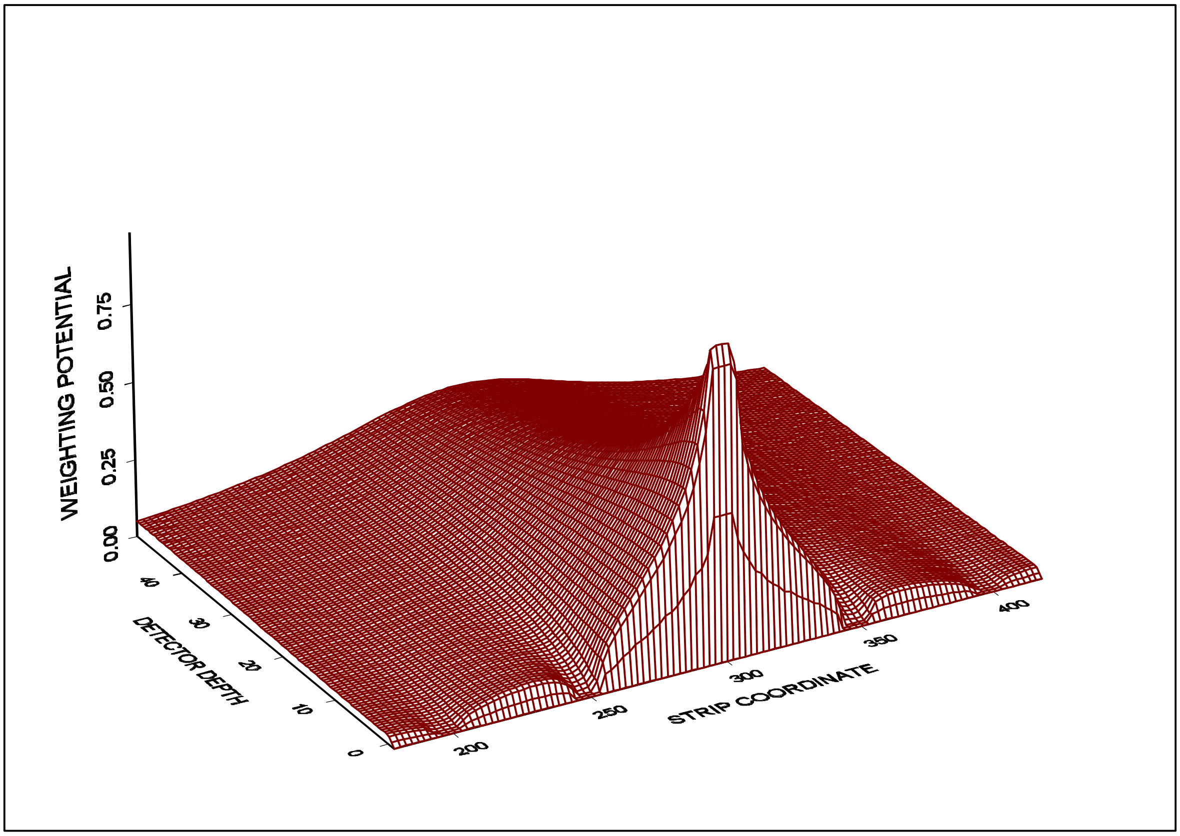
Since k=2: $i_1 = -i_2$

Example 2: Double-Sided Strip Detector

The strip pitch is assumed to be small compared to the thickness.

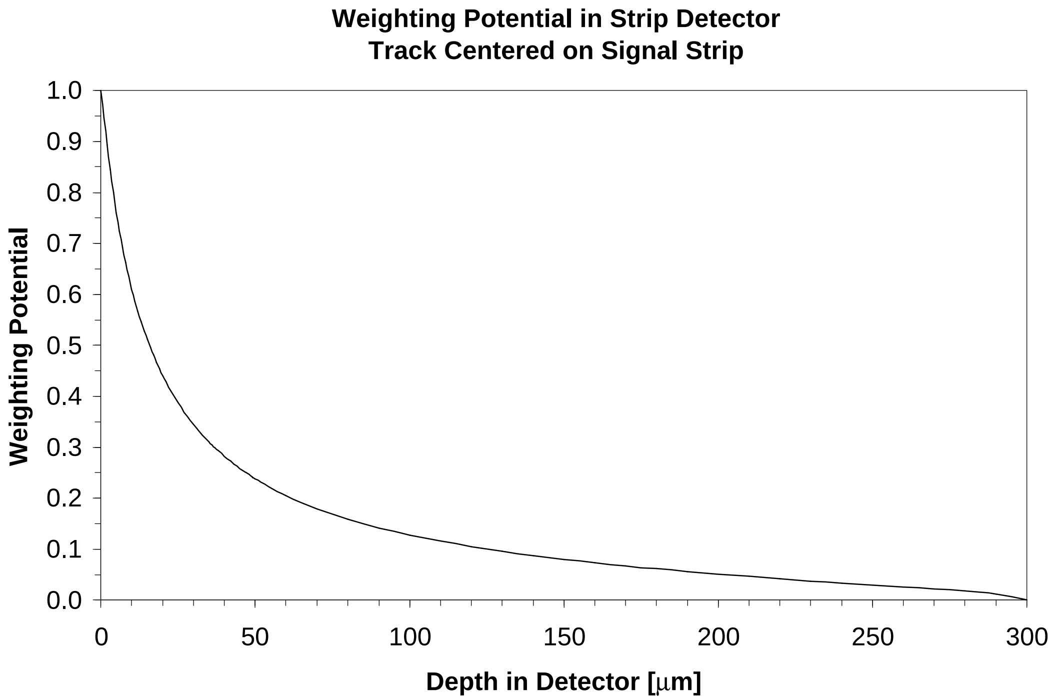
The electric field is similar to a parallel-plate geometry, except in the immediate vicinity of the strips.

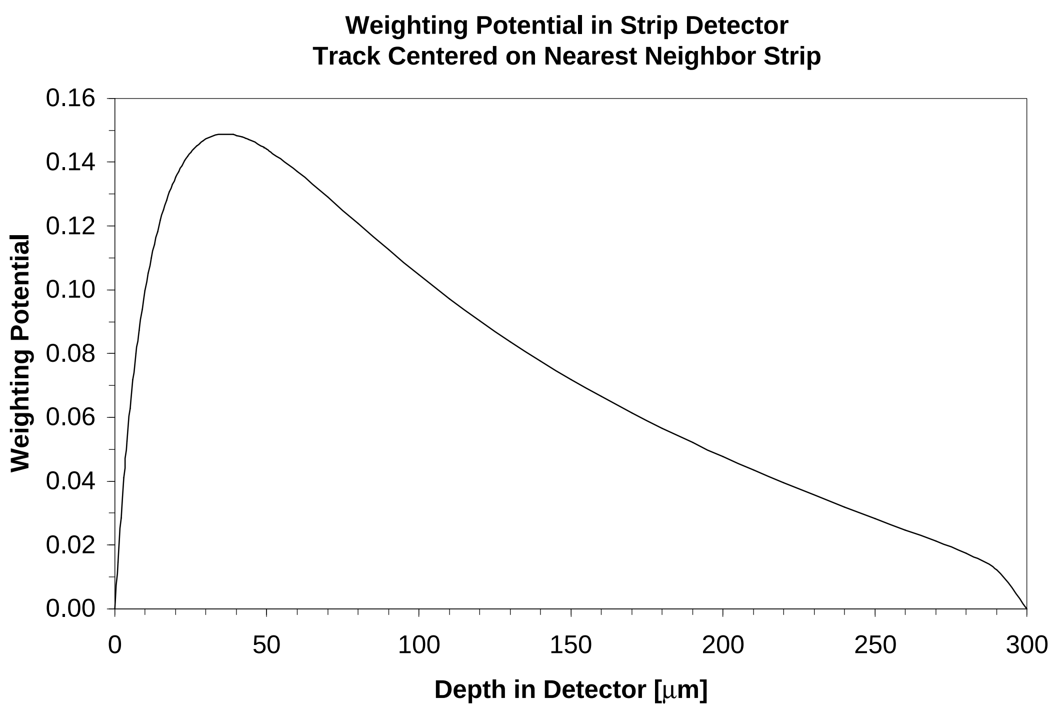
The signal weighting potential, however is very different.



Weighting potential for a 300 µm thick strip detector with strips on a pitch of 50 µm. Only 50 µm of depth are shown.

Cuts through the weighting potential





Consider an electron-hole pair q_n, q_p originating on a point x_0 on the center-line of two opposite strips of a double-sided strip detector. The motion of the electron towards the n-electrode x_n is equivalent to the motion of a hole in the opposite direction to the p-electrode x_p. The total induced charge on electrode k after the charges have traversed the detector is

$$Q_k = q_p[\Phi_{Qk}(x_p) - \Phi_{Qk}(x_0)] + q_n[\Phi_{Qk}(x_n) - \Phi_{Qk}(x_0)]$$

since the hole charge $q_p = q_e$ and $q_n = -q_e$

$$Q_k = q_e[\Phi_{Qk}(x_p) - \Phi_{Qk}(x_0)] - q_e[\Phi_{Qk}(x_n) - \Phi_{Qk}(x_0)]$$

$$Q_k = q_e[\Phi_{Qk}(x_p) - \Phi_{Qk}(x_n)]$$

If the signal is measured on the p-electrode, collecting the holes,

$$\Phi_{Qk}(x_p) = 1,$$
$$\Phi_{Qk}(x_n) = 0$$

and $Q_k = q_e$.

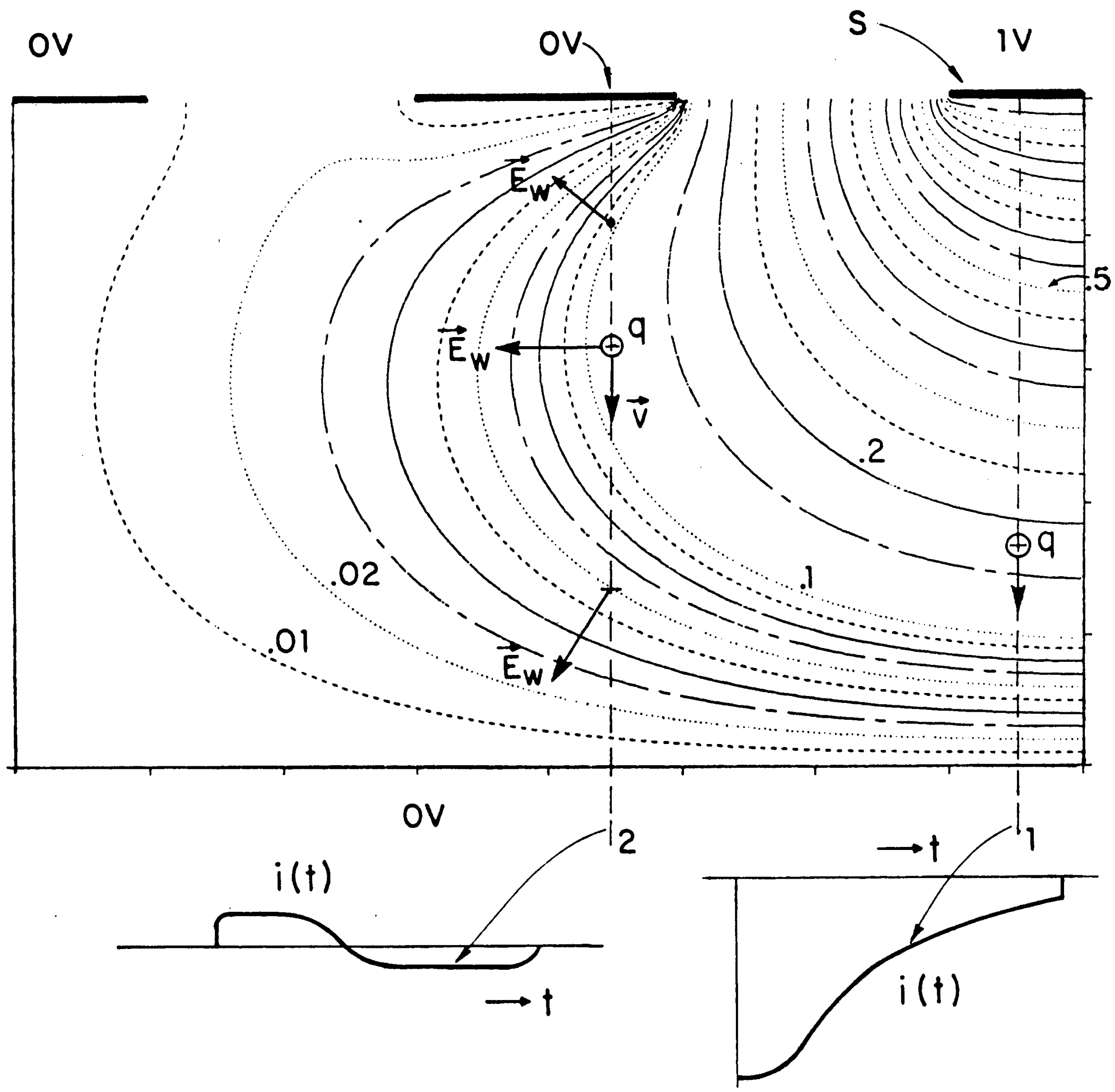
If, however, the charge is collected on the neighboring strip k+1, then

$$\Phi_{Qk+1}(x_p) = 0,$$
$$\Phi_{Qk+1}(x_n) = 0$$

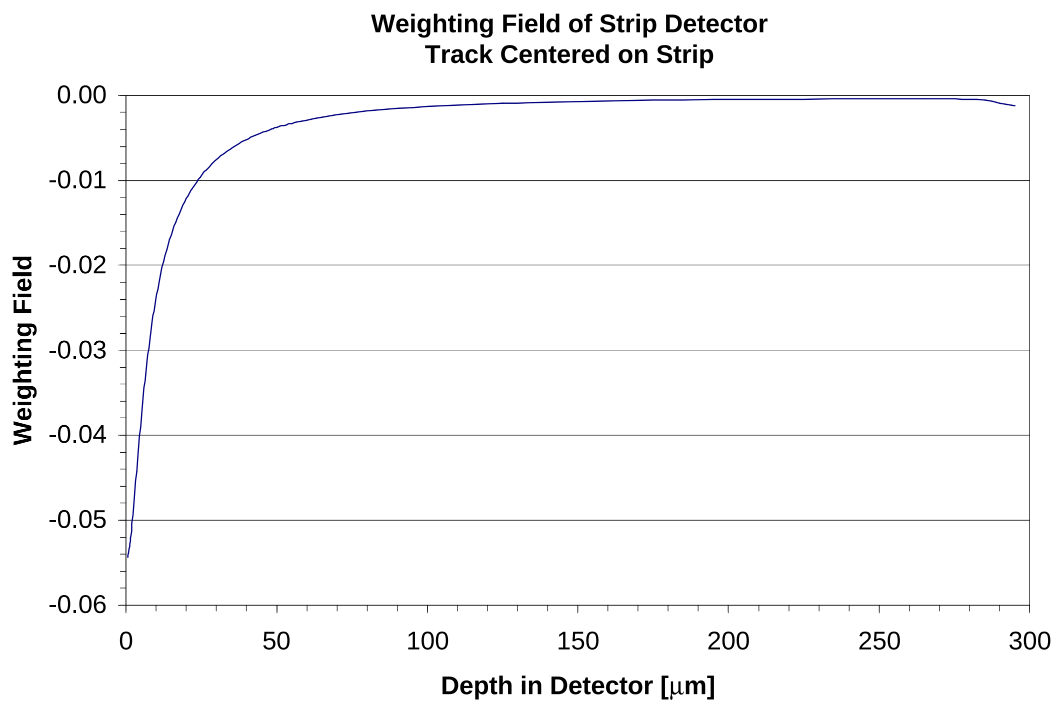
and $Q_{k+1} = 0$.

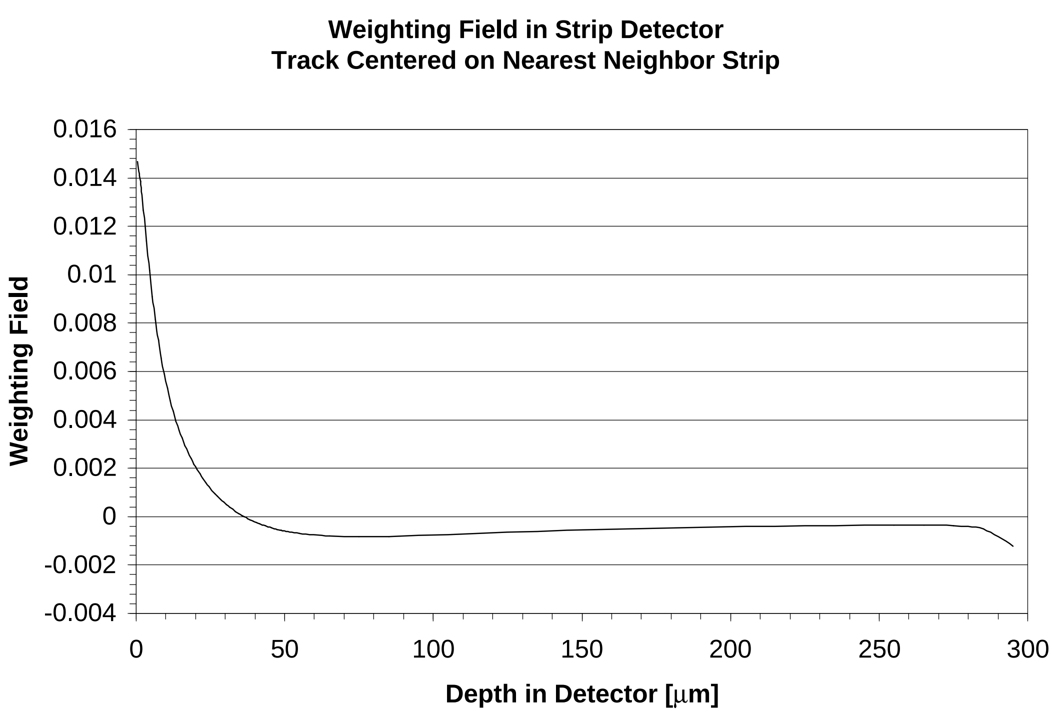
In general, if moving charge does not terminate on the measurement electrode, signal current will be induced, but the current changes sign and integrates to zero.

This is illustrated in the following schematic plot of the weighting field in a strip detector (from Radeka)



Cuts through the Weighting Field in a Strip Detector
(d= 300 µm, p= 50 µm)





Note, however that this charge cancellation on “non-collecting” electrodes relies on the motion of both electrons and holes.

Assume, for example, that the holes are stationary, so they don't induce a signal. Then the first term of the first equation above vanishes, which leaves a residual charge

$$Q_k = q_e[\Phi_{Qk}(x_0) - \Phi_{Qk}(x_n)]$$

since for any coordinate not on an electrode

$$Q_k(x_0) \neq 0,$$

although it may be very small.

An important consequence of this analysis is that one cannot simply derive pulse shapes by analogy with a detector with contiguous electrodes (i.e. a parallel plate detector of the same overall dimensions as a strip detector). Specifically,

1. the shape of the current pulses can be quite different,

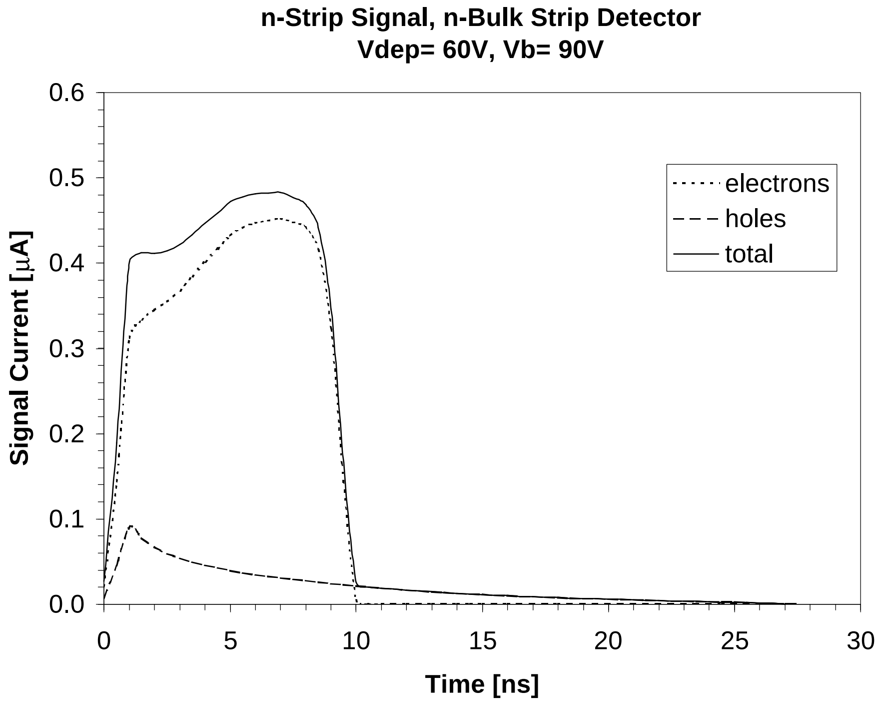
2. the signals seen on opposite strips of a double-sided detector are not the same (although opposite in sign), and

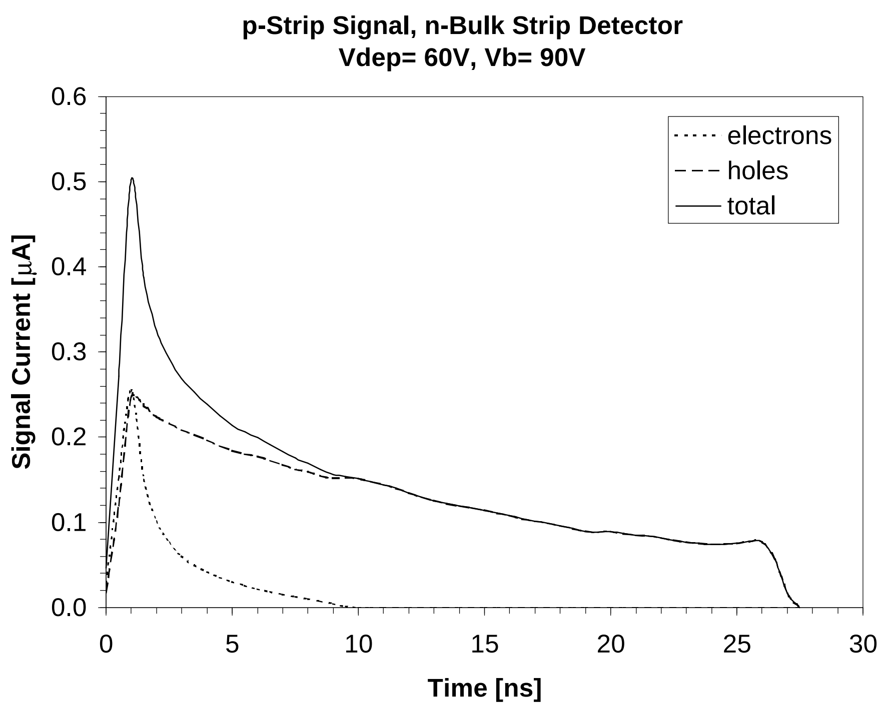
3. the net induced charge on the *p*- or *n*-side is not split evenly between electrons and holes.

   - Because the weighting potential is strongly peaked near the signal electrode, most of the charge is induced when the moving charge is near the signal electrode.

   - As a result, most of the signal charge is due to the charge terminating on the signal electrode.

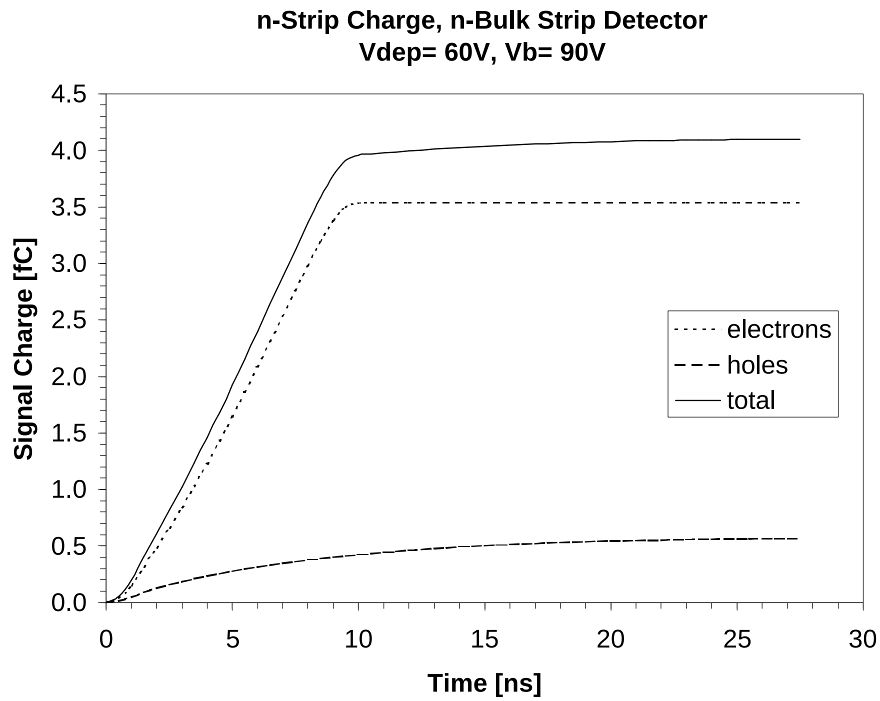
Current pulses in strip detectors (track traversing the detector)

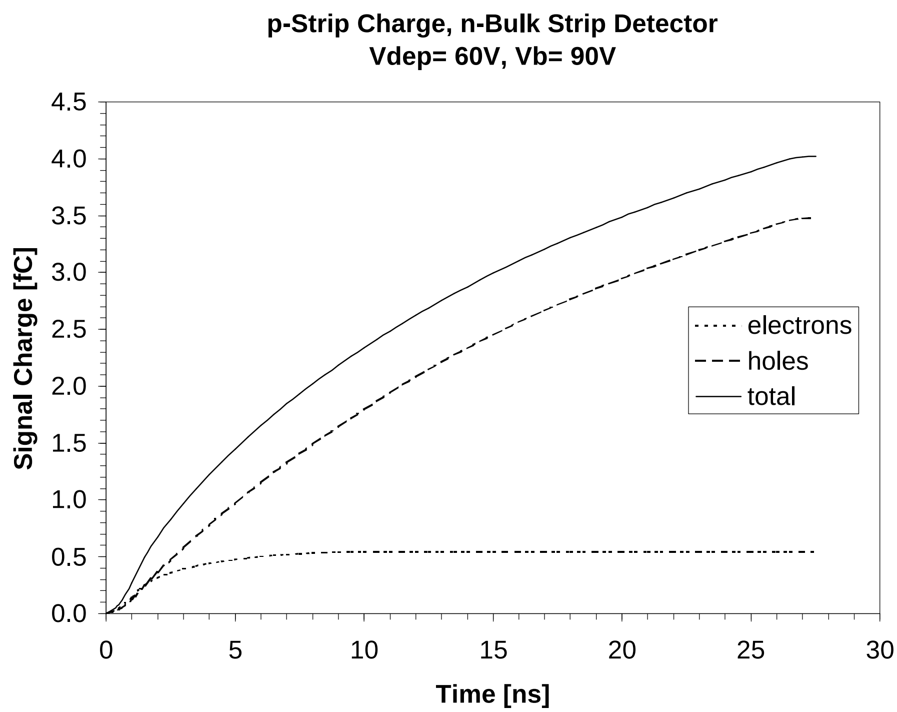




The duration of the electron and hole pulses is determined by the time required to traverse the detector as in a parallel-plate detector, but the shapes are very different.

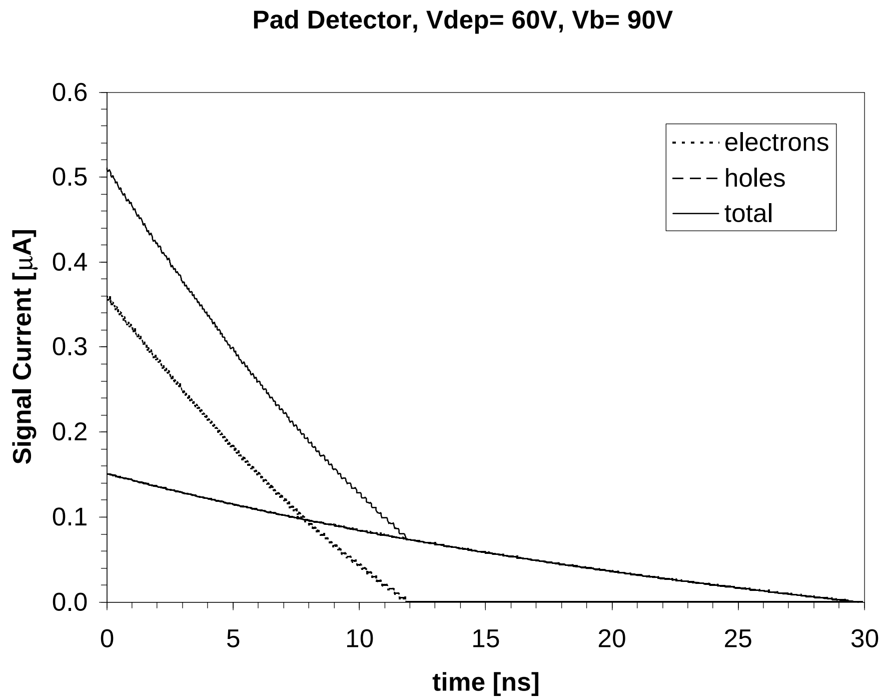
Strip Detector Signal Charge Pulses



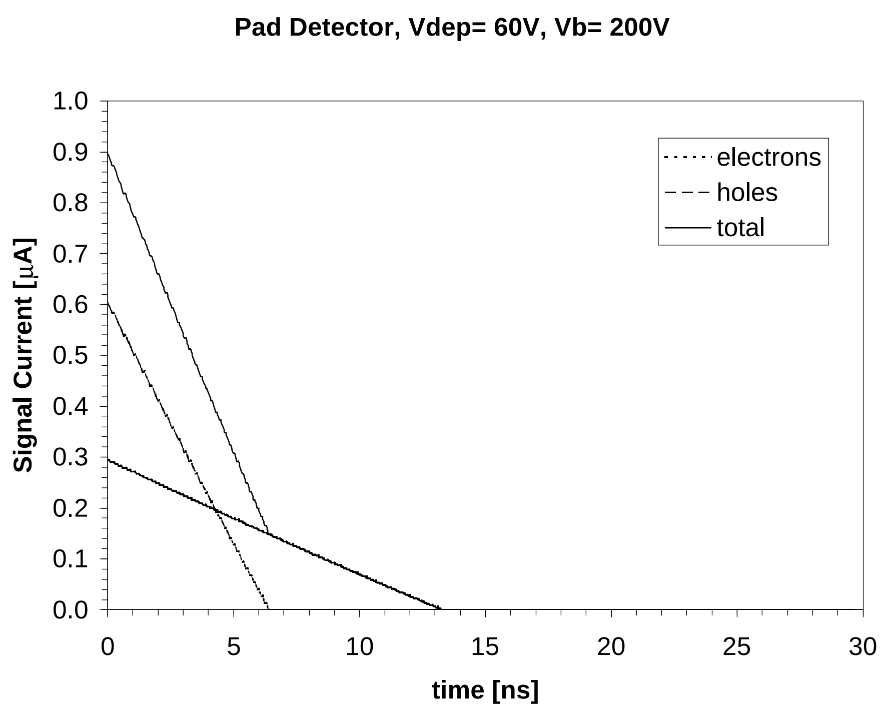


For comparison:

Current pulses in pad detectors (track traversing the detector)

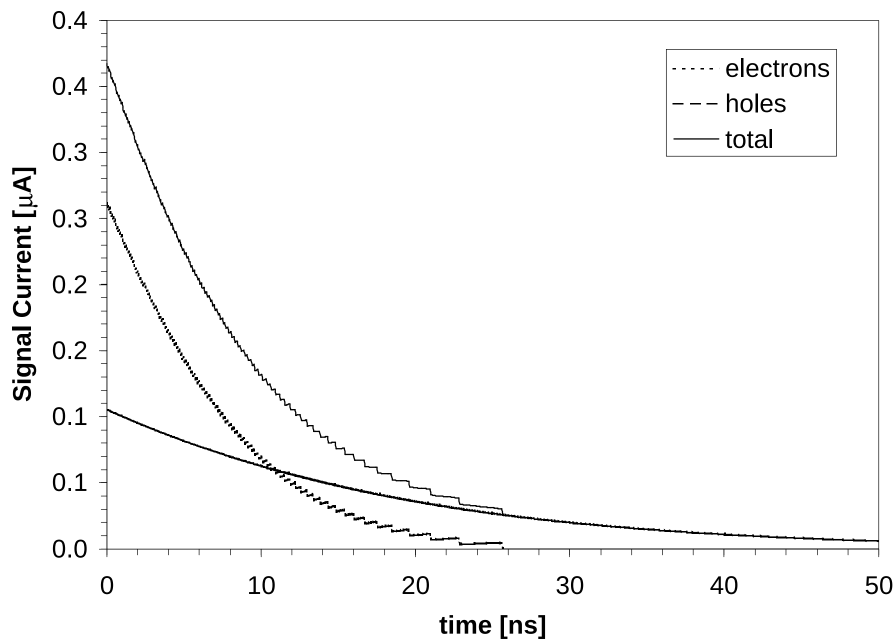


For the same depletion and bias voltages the pulse durations are the same as in strip detectors. Overbias decreases the collection time.

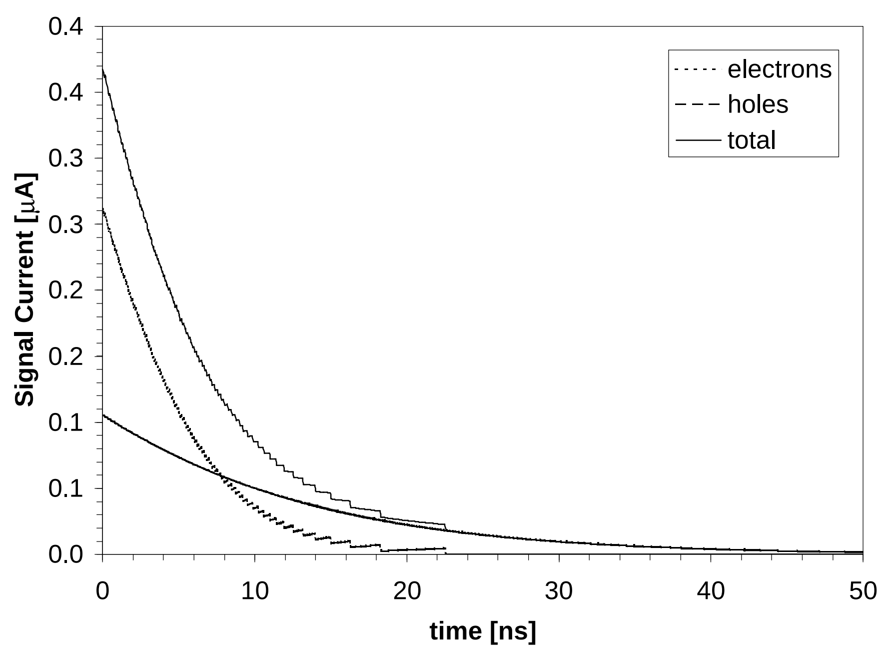


Operation at or under full depletion leads to long “tails” from the low-field region.

Pad Detector, Vdep= 60V, Vb= 60V



Pad Detector, Vdep= 60V, Vb= 30V



Note: The “steps” in the curves are artifacts of the calculation resolution.

Application of Induced Charge Concept:
Charge Collection in the Presence of Trapping

Practical semiconductor crystals suffer from imperfections introduced during crystal growth, during device fabrication, or by radiation damage.

Defects in the crystal

- impurity atoms
- vacancies
- structural irregularities (e.g. dislocations)

introduce states into the crystal that can trap charge.

Charge trapping is characterized by a carrier lifetime τ, the time a charge carrier can "survive" in a crystal before trapping or recombination with a hole.

Trapping removes mobile charge available for signal formation.

Depending on the nature of the trap, thermal excitation or the externally applied field can release the carrier from the trap, leading to delayed charge collection.

Given a lifetime τ, a packet of charge Q_0 will decay

$$Q(t) = Q_0 e^{-t/\tau}$$

In an electric field the charge will drift. The time required to traverse a distance x is

$$t = \frac{x}{v} = \frac{x}{\mu E}$$

after which the remaining charge is

$$Q(x) = Q_0 e^{-x/\mu E \tau} \equiv Q_0 e^{-x/L}$$

Since the drift length is proportional to the mobility-lifetime product, $\mu\tau$ is often used as a figure of merit.

Assume a detector with a simple parallel-plate geometry. For a charge traversing the increment dx of the detector thickness d, the induced signal charge is

$$dQ_s = Q(x)\frac{dx}{d}$$

so the total induced charge

$$Q_s = \frac{1}{d}\int_0^d Q(x)dx = \frac{1}{d}\int_0^d Q_0 e^{-x/L} dx$$

$$Q_s = Q_0 \frac{L}{d} \left(1 - e^{-d/L}\right)$$

$$d >> L: \qquad \frac{Q_s}{Q_0} \approx \frac{L}{d}$$

In high quality silicon detectors:

$\tau \approx$ 10 ms
μ_e= 1350 V/cm·s^2

E=10^4 V/cm $\quad\Rightarrow\quad$ $L \approx 10^4$ cm

In amorphous silicon $L \approx$ 10 μm (short lifetime, low mobility)
In diamond, however, $L \approx$ 100 – 200 μm (despite high mobility)
In CdZnTe at 1 kV/cm, $L \approx$ 3 cm for electrons, 0.1 cm for holes

Although there is no “magic” solution to improving the charge yield of materials with marginal drift lifetimes, it is possible to mitigate the variations of measured signal.

For example, take a material with a long electron lifetime and a short hole lifetime used as a gamma ray detector (CdZnTe, for example). The interaction is point-like, so the charge originates from only one coordinate x_0.

In a parallel plate geometry, as shown previously

$$Q_s = Q_e + Q_h = Q_0 \frac{\Delta x_e}{d} + Q_0 \frac{\Delta x_h}{d} = Q_0 \frac{\Delta x_e + \Delta x_h}{d}$$

If the electron drift length Δx_e and the hole drift length Δx_h are sufficient so that

$$\Delta x_e + \Delta x_h = d$$

both terms add to yield the full signal charge $Q_s = Q_0$.

If, however, the hole lifetime is short, so that holes fail to reach their collection electrode before being trapped

$$\Delta x_e + \Delta x_h < d$$

and $Q_s < Q_0$.

Furthermore, the charge yield will depend on the position of the interaction. If the interaction is at the negative electrode, the signal will be exclusively due to electrons, so for a long electron lifetime the full charge will be measured.

If the interaction is at the positive electrode, only holes will contribute and

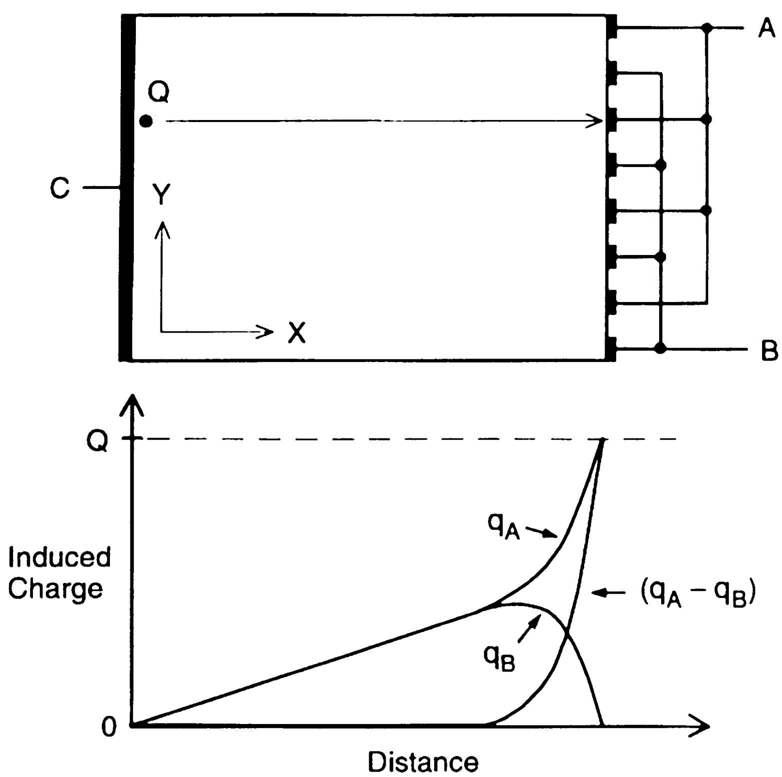
$$Q_s = Q_0 \frac{L}{d} \left(1 - e^{-d/L}\right)$$

For intermediate interaction sites the signal will vary between these extremes.

The charge response can be made more uniform by adopting a strip detector configuration.

Since the induced signal is formed predominantly in the vicinity of the strips, one can minimize the effect of the trapped charge carrier.

First applied by Paul Luke (IEEE Trans. Nucl. Sci. **NS-43** (1996) 1481) the configuration combines every other strip into two readout channels A and B.



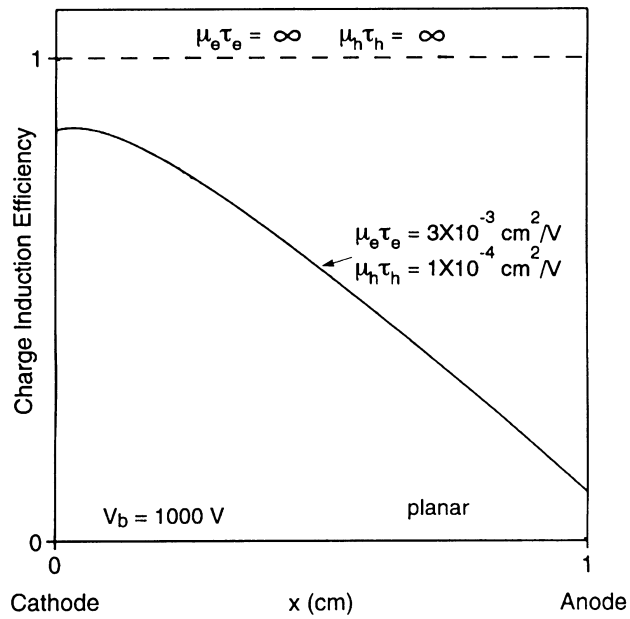
Channel A measures the sum of the electron and hole signals q_A. Channel B measures the signal q_B from the holes drifting towards electrode C (cf. discussion on p. 18).

The difference signal is predominantly the electron signal.

The simulated charge measurement efficiency vs. position for a CdZnTe detector using a simple parallel plate geometry is shown below.

In CdZnTe at 1 kV/cm,
$L \approx$ 3 cm for electrons,
0.1 cm for holes

The 1 cm thick detector is operated at 1 kV.



The strip readout provides a substantial improvement in uniformity.

Reducing the gain G of the non-collecting grid allows compensation of electron trapping and yields a nearly flat response.