Development of a Novel Readout System for Small Animal Positron Emission Tomography

Diplomarbeit

von

Jan-Frederik Pietschmann

Experimentelle Physik

Development of a Novel Readout System for Small Animal Positron Emission Tomography

Diplomarbeit

von

Jan-Frederik Pietschmann

Westfälische Wilhelms-Universität Münster

Institut für Kernphysik

— März 2008 —

Contents

CHAPTER I

Introduction

Positron Emission Tomography (PET) is a medical imaging method. Today, PET devices are present in almost every larger hospital. The great advantage of PET compared to Nuclear Magnetic Resonance (NMR) or Computed Tomography (CT) it its ability to visualize functional processes. This makes it popular in oncology and cardio vascular imaging [Tsui05].

As many imaging techniques, PET has also been adapted for small animals likes mice or rats as they are important models in clinical research. Due to their small size compared to humans, dedicated small animal PET scanner with increased spatial resolution and sensitivity have been developed.

At the Münster University Hospital, a quadHIDAC small animal PET scanner is installed. Unlike most other systems, this device is based on Multi Wire Proportional Chambers (MWPC) instead of scintillation crystals. It is used for research within the framework of the Sonderforschungsbereich 656 to examine the cardio vascular system.

However, as support is no longer available for the quadHIDAC device, it has been decided to develop a new small animal scanner prototype, called msPET, also based on MWPCs.

The main task of this thesis was to develop a readout system for this prototype, based on the n-XYTER readout chip. This chip has been developed within the framework of the European project "Detectors for Neutron Instrumentation" (DETNI). However, many of its properties are very suitable for PET. Furthermore, a Spartan 3 Field Programmable Gate Array (FPGA) has been used as a core of the readout board.

Finally, the performance of the existing MWPC prototypes for msPET has been evaluated. After examining the basic operation of the chambers, tests have been done to determine their efficiency and spatial resolution.

CHAPTER II

Positron Emission Tomography

Contents

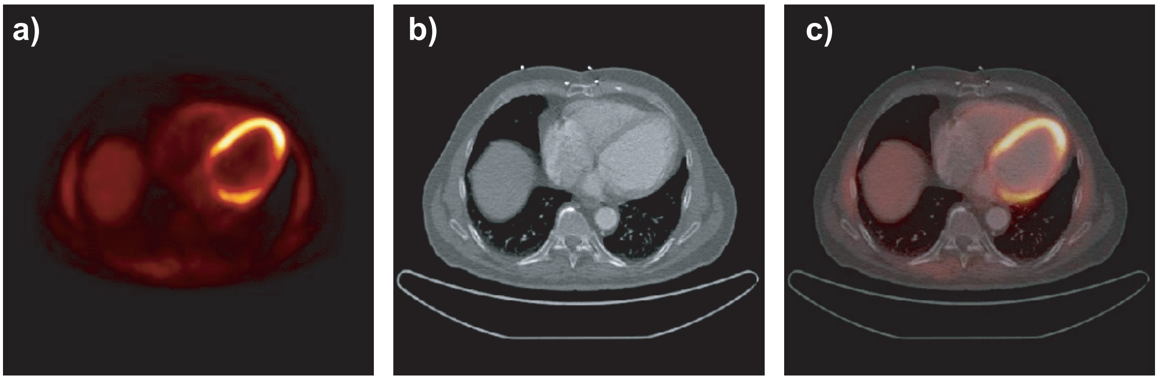


Figure II.1: Images obtained by a) PET only, b) CT only and c) their combination. [Büther07]

There is a great need in medicine for methods to non invasively image the interior of the human (or animal) body. Emission tomography (ET) is the generic term for all methods that are based on the detection of radiation emitted from radioactive material within the body. The two main representatives of ET are single photon emission computed tomography (SPECT) and positron emission tomography (PET). SPECT is a method which works with virtually every γ emitting substance. As shown in Figure II.2, it uses a collimator in front of a detector to determine the direction of an incoming photon. This leads to a cone on which the decay must have taken place, see Figure II.2 b). By rotating the detector and using appropriate algorithms, it is possible to reconstruct the three dimensional distribution of the radioactive emitter. The principle of PET will be explained in detail in the following sections.

The Tracer Principle

One important concept in nuclear medicine is the tracer principle, developed in about 1900 by George de Hevesy [deHevesy48]. It allows the examination of particular biological processes within animals and humans using PET and SPECT. De Hevesy recognized that integration of radioactive isotopes in organic compounds does not alter their biological behaviour. Therefore, radioactively marked biomolecules can be used to trace the concentration of specific substances within the body. This ability to examine functional processes distinguishes emission tomography from computed tomography (CT) or nuclear magnetic resonance tomography (NMR). These methods are only able to visualize the morphology of the tissue. However, as CT usually provides excellent spatial resolution, combinations of PET and CT, known as PET/CT, became very popular in recent years. Images created by PET, CT and their combinations are shown in Figure II.1.

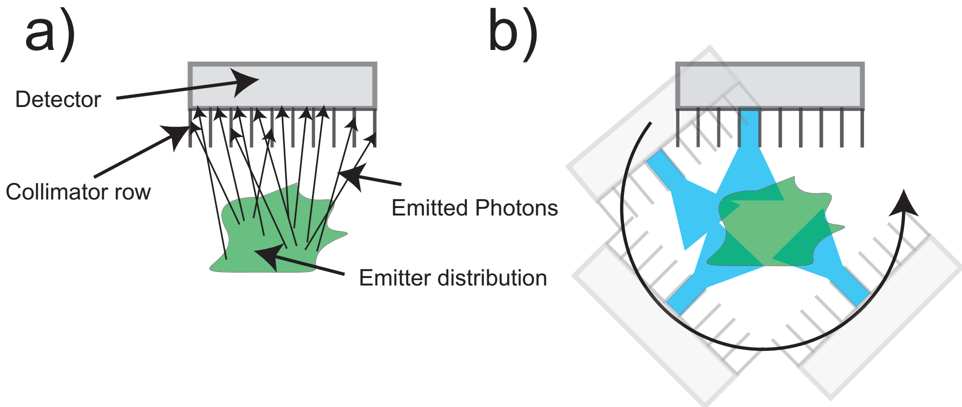


Figure II.2: The principle of single photon computed emission tomography (SPECT).

II.1 The Principle of PET

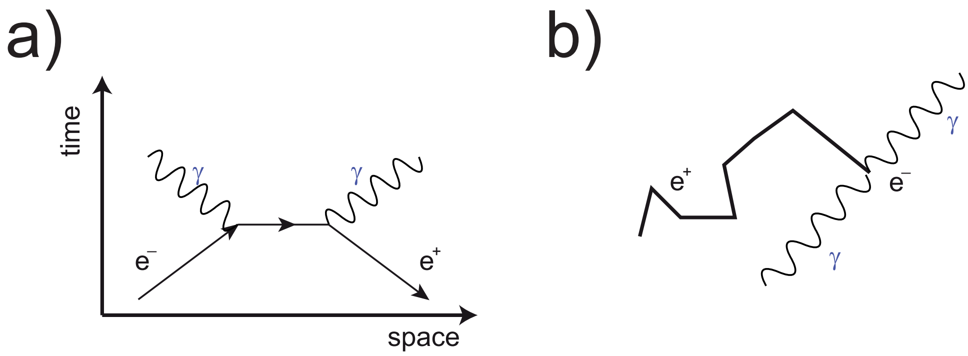


Figure II.3: a) One possible Feynman diagram for positron electron annihilation. b) A positron losing its enery in multiple scattering processes and finally performing positron electron annihilation.

Positron Emission Tomography uses photons that are emitted in positron electron annihilations. In this process, two photons are created in the reaction

$$e^+e^- \rightarrow \gamma\gamma. \tag{II.1}$$

This can be expressed in a Feynman diagram as shown in Figure II.3 a). As the reaction results in two identical bosons, another diagram with photons switched must be considered. Using these diagrams, it is possible to calculate the cross section of positron electron annihilation. As it is very tiny (at high kinetic energies T it behaves like $ln(T)/T$) [Lohrmann86], most positrons are slowed down until they practically have a velocity of zero. Then, they form a bound state with the electron which is called positronium before they finally annihilate. This is shown in Figure II.3 b). According to the law of energy conservation, the photons must have the energy of the electron and the positron. Therefore, they both most likely have a energy of 511keV. The important fact for PET is that the two 511 keV γ photons are emitted back to back. Detecting these photons in coincidence results in a line on which - neglecting the effects of the positron range - the decay must have taken place (Figure II.4, a). This line is commonly called *line of response* (LOR). It is obvious that the detection several of these lines increases the information about the distribution of the emitting element tremendously, as shown in Figure II.4 b). Using appropriate reconstruction

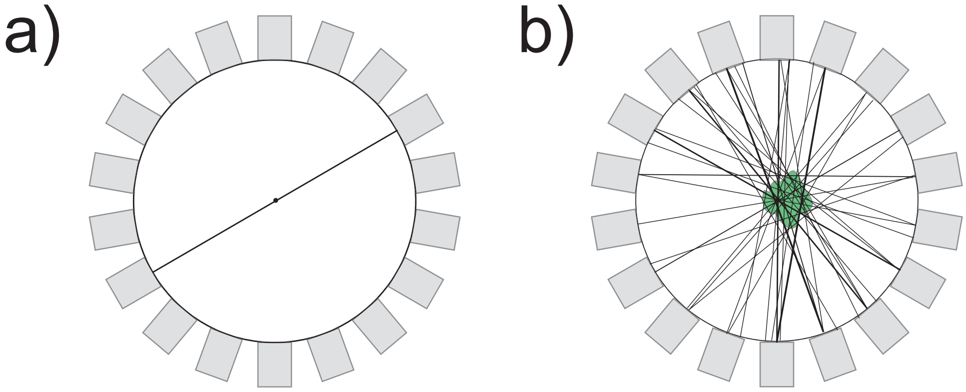


Figure II.4: a) A single line of response created by two 511 keV γ photons. b) Same situation for an extended source. Note that in practice, the coordinates of the detected photons are set to be in the middle of the szintillation crystals.

algorithms it is possible to reconstruct a three dimensional map of the sites of decay. In medical applications, the positrons needed for the annihilation are created by a β^+ isotope in the reaction

$$ {}^{A}_{Z}X \rightarrow {}^{A}_{Z-1}X + e^{+} + \nu_e. \qquad \text{(II.2)} $$

Compared to the detection of single photons in SPECT, coincidence measurements in PET allow a more accurate localisation of the decay. Furthermore, in PET far more of the emitted photons are used as it covers a greater solid angle. Each coincidence results in a line on which the decay must have taken place. In SPECT, each measurement only results in a cone. However this also introduces additional problems: If one of the photons is scattered (either on its way to or within the detector itself) the LOR does not intersect the real decay site. Events of this type are called “scattered events”, this phenomena is shown in Figure II.6 b). The second

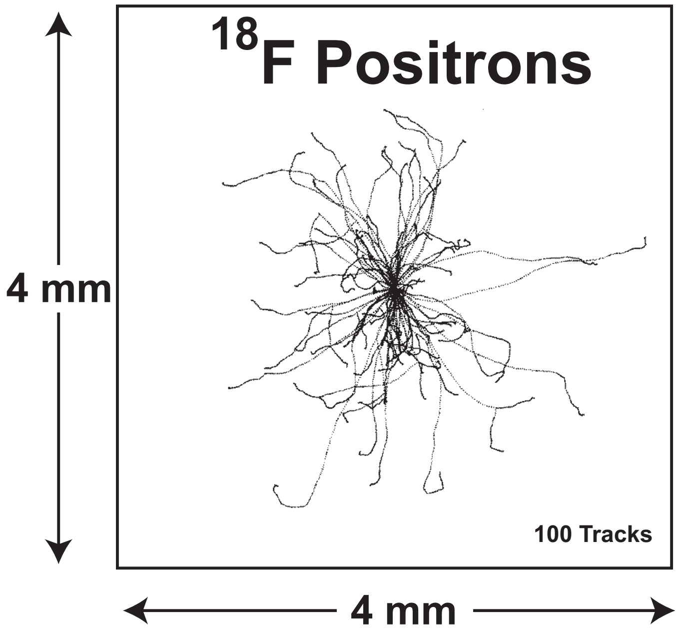


Figure II.5: The tracks of positrons emitted by ^{18}F in water. 100 tracks have been simulated. [Levin99]

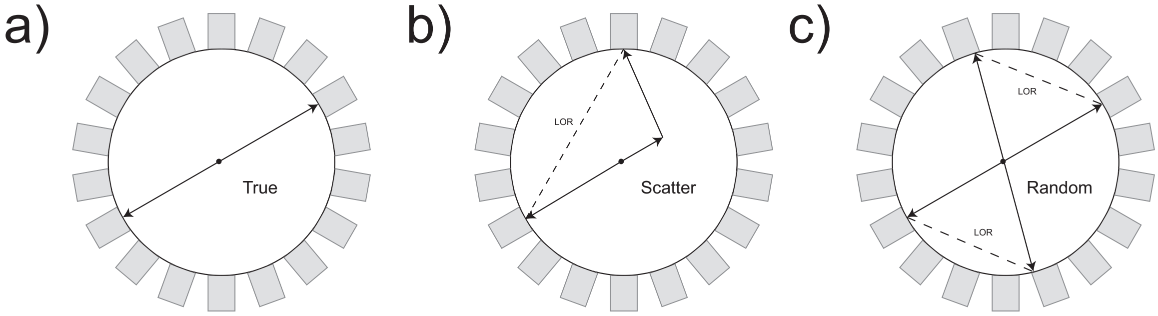


Figure II.6: a) A "true" coincidence. b) A "scattered" event. c) A "random" event. The dashed line indicates the measured line of response.

possible error occurs if two annihilations take place almost or completely simultaneously. Then, two photons emitted by different annihilations may accidentally be detected in coincidence. This also leads to a wrong LOR and is schematically shown in Figure II.6 c). Such events are commonly called “random events”. Both effects decrease the spacial resolution. They are also major problems for quantitative PET. Recent approaches try to estimate the scatter and random fraction by use of Monte Carlo simulations and adapt the reconstruction algorithms to correct them [Gottschlag07, Kösters07]. In PET, the set of possible tracers is of course limited

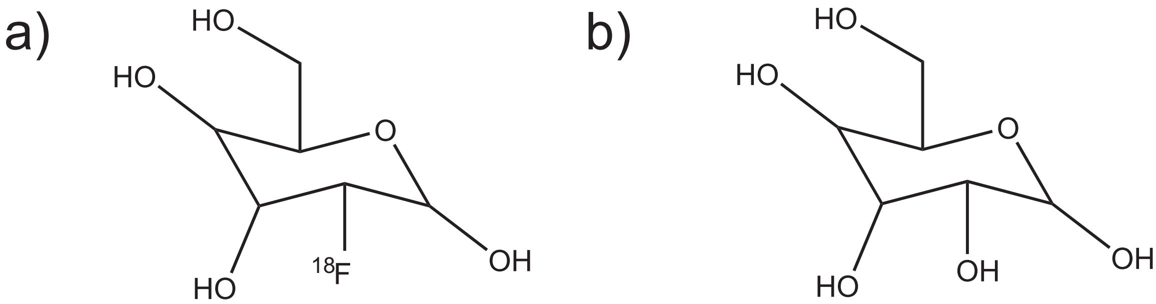


Figure II.7: a) ^{18}F-FDG. b) Ordinary glucose.

to β^+ emitter as only they create positrons. The most popular tracer in clinical PET applications today is ^{18}F-fluorodeoxyglucose ($^{18}\mathrm{F}-\mathrm{FDG}$). In $^{18}\mathrm{F}-\mathrm{FDG}$ one of the OH groups of ordinary glucose is replaced by a ^{18}F. $^{18}\mathrm{F}-\mathrm{FDG}$ enters a cell in the same way as usual glucose does. However, within the cell it is metabolized to a new compound that remains in its interior. This leads to an increasing concentration of ^{18}F, growing proportionally to the cells glucose metabolic rate [Wernick04]. Furthermore, ^{18}F has a half life of

$$t_{1/2} = 109.77 \text{ min},$$

which means that it exists long enough to perform an examination but does not remain very long in the patient’s body afterwards. The chemical structure of $^{18}\mathrm{F}-\mathrm{FDG}$ and ordinary glucose are shown in Figure II.7.

Commercially available devices

Commercial PET scanners are either based on scintillation crystals or on gaseous detectors. Scintillation scanners are more often used in practice. This is due to the fact that they are very reliable. For devices based on scintillation crystals, the spacial resolution basically depends on the size of the crystals. This means that decreasing the crystal size in order to obtain a higher spacial resolution, the number of crystals raises if the solid angle coverage is constant. As every crystal needs additional read-out electronics and the crystals themselves are expensive, this increases costs of such a scanner. On the other hand, scanners based on gaseous detectors have a greater spatial resolution and are much cheaper. They can provide an intrinsic resolution on the order of 1 mm or even better [Ott93].

II.2 A Short History of PET

The whole principle of PET is based on the discovery of the positron. This particle was postulated by P. Dirac[1] in 1928. In 1932, Carl Anderson[2] was able to experimentally find it in cosmic rays, using a cloud chamber [Anderson33]. He also gave it its name. Furthermore, he predicted the existence of positronium (also in 1932). It was experimentally discovered by Martin Deutsch[3] in 1951 [Deutsch51].

Even though the first imaging devices did not appear until 1940, there were a number of simpler applications beforehand. One of the earliest examples is the examination of the thyroid using radioactive substances. In the first simple approaches, a short lived iodine isotope was applied to the patient. Then, a map of the intensity distribution of the emitted γ rays was produced either by hand or using a dot tapper. An early image is shown in Figure II.8. In the mid- to late 1940s, the first clinical imaging devices were developed. They were based on Na(I)Tl scintillation crystals that were coupled to a motor driven system. Some of these devices already used a pair of opposite detectors. They contained an electronic coincidence circuit and were thus the first real PET devices. The first clinical PET scanner was built in 1952 by Brownell and Aranow [Brownell99].

In 1960, the first cyclotron dedicated to the production of isotopes needed for medical imaging was installed at the Hammersmith Hospital in London. Thenceforward, PET and other techniques based on emission tomography became more and more

[1]Paul Adrien Maurice Dirac was a British theoretical physicist. He was born August 8, 1902 in Bristol and died October 20, 1984 in Tallahassee. In 1933 he received the Nobel price “for the discovery of new productive forms of atomic theory”.

[2]Carl David Anderson was a US American physicist. He was born 3 September 1905 in New York and died 11 January 1991 in San Marino, California. In 1936, the received the Nobel price “for his discovery of the positron”.

[3]Martin Deutsch was an Austrian-American physicist. He was born on 29 January 1917 in Vienna and died 16 August 2002 in Cambridge, Massachusetts.

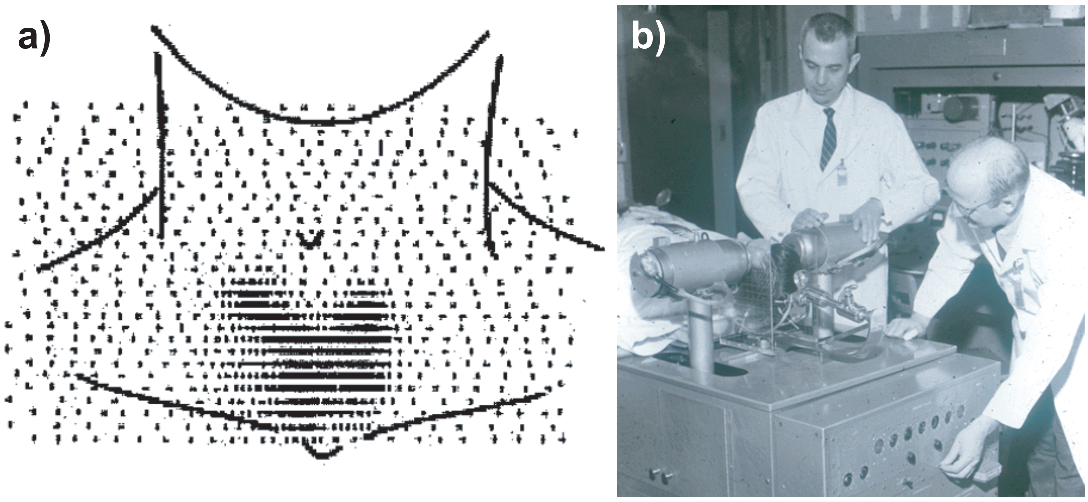


Figure II.8: a) Image produced with an experimental early imaging device. The picture has been recorded using a dot tapper. [Andrews62]. b) The first clinical PET device, developed in 1952 by Brownell and Aranow [Brownell99].

popular in clinical practice. Today, PET and SPECT are standard examinations in almost every larger hospital.

II.3 Small Animal PET

As other imaging methods like CT, PET has been adapted for the use in animal experiments. In case of PET there are some limitations concerning small animals such as mice or rats. Typical human scanners today provide a resolution of about 4 to 10 mm in clinical practice, specialized brain scanners achive up to 2 mm [Schäfers08]. The average mass of a mouse (rat) is 30 g (300 g). Compared to an average human with 70 kg, this is a factor of about 200 to 2000 in mass and thus volume. To achive comparable image quality and be able to examine the same biological questions, the spacial resolution of a dedicated small animal PET scanner must be about 1 mm. As the avarage range of positrons within tissue is also about 1 mm [Levin99], this sets a fundamental limit for the spacial resolution in PET.
Furthermore, small animal PET makes great demands on the sensitivity of the scanner. Sensitivity refers to the fraction of emitted γ photons that are detected by the scanner. This is an important quantity as the number of decays detected per voxel is proportional to the signal to noise ratio of the measurement and thus in the resulting image. As the voxelsize in small animal PET ($\sim \mu$l) is much smaller

than in human PET ($\sim$ ml), a better sensitivity is needed to compensate this effect. The relative reduction in sensitivity can partly be compensated by using a greater tracer concentration within the animals. However, the maximum concentration is still limited as it must not disturb the biological process that is examined in the experiment.
Considering all these criteria, a number of dedicated small animal scanners has been developed. Most of them are based on scintallation crystals. Their spacial resolution ranges from 1.8 to 2.1 mm. There is also one device available based on multi wire proportional chambers. It has a spatial resolution of about 1.2 mm.
For a short review on small animal scanners see [Wernick04].

II.4 Reconstruction Algorithms for PET

The standard method for reconstruction in computed tomography (CT) is the filtered back projection (FBP, for details see [Natterer01]). Even though this method is also applicable to PET, most of the newer devices use iterative algorithms. They produce better results because they take into account the statistical nature of the data.
In this model, the distribution of the tracer within the object is interpreted as a density function

$$f : \mathbb{R}^N \to \mathbb{R}^+, \qquad N = 2, 3. \tag{II.3}$$

Assuming a discrete set of possible lines of response, the number of events measured per line is the integral of f over this line:

$$\int_{L_i} f(x)dx \approx g_i, \qquad i = 1, \dots, m. \tag{II.4}$$

As the g_i are discrete numbers, this equation holds only approximately. The integral is known as Radon[4] transformation. In the next step, a grid is placed over the region of interest and it is assumed that f is constant in every pixel or voxel respectively. The value of f in pixel P_i is called f_i. Defining a_{ij} as the length of the intersection of line L_i with pixel P_j,

$$g_i = \sum_{j=1}^{n} a_{ij} f_j \tag{II.5}$$

[4] Johann Karl August Radon was an Austrian mathematician. He was born December 16, 1887 in Decin and died May 25, 1956 in Vienna.

holds. With the further definitions

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \quad g = \begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix},$$

equation II.4 can be rewritten as

$$Af = g. \tag{II.6}$$

In principle, every algorithm for solving linear systems of equation could be used to determine f. However as already mentioned, it is advantageous to consider the statistical nature of the g_i. As radioactive decay is a Poisson-distributed process, it is reasonable to assume that the number of photons that is emitted from pixel P_j and that are registrated on line L_i, X_{ij} are also Poisson-distributed. From this follows that the number of photons, emitted from every pixel that intersects with line L_i,

$$\gamma_i := \sum_{j=1}^{n} X_{ij} \quad i = 1, \ldots, m, \tag{II.7}$$

is Poison distributed as well. Therefore the measured data g_i can be interpreted as expectation of the random variable γ_i. To calculate the expectation value it is assumed the the matrix element a_{ij} corresponds to the probability that a photon is registrated on line L_i under the condition that is was emitted in pixel P_j. Then,

$$E(\gamma_i) = \sum_{j=1}^{n} a_{ij} f_j = (Af)_i, \quad i = 1, \ldots, m \tag{II.8}$$

holds. The probability for the measured values is then

$$P_f(g_i) = \frac{(Af)_i^{g_i}}{g_i!} e^{-(Af)_i}, \quad i = 1, \ldots, m. \tag{II.9}$$

The idea is now to variate f until the probability matches the measured values best. Formally, this is done by writing the probability as a function of f. This function is called the Likelihood function

$$L(f) := P_f(g) = \prod_{i=1}^{m} \frac{(Af)_i^{g_i}}{g_i!} e^{-(Af)_i} = g \cdot (Af) - (Af) \cdot 1. \tag{II.10}$$

To find the maximizer f, it is easier to consider the log-likelihood function. As the logarithm is a monotonic function, it has its maximum at the same place.

$$l(f) := \log L(f) = \sum_{i=1}^{m} (g_i \log(Af)_i - (Af)_i) \tag{II.11}$$

As the log-likelihood function is convex [Natterer01] is has a unique global maximum. f is this maximum, if the Kuhn-Tucker conditions are fulfilled. This is equal to the fact that f times the gradient of l vanishes, i.e.

$$f \nabla l(f) = f A^T \left(\frac{g}{Af} - 1 \right) = 0. \tag{II.12}$$

Here, 1 is a vector with only 1's and all operation are meant componentwise. Normalizing A so that each column sum is one (or $A^T 1 = 1$) one gets

$$f = f A^T \frac{g}{Af}. \tag{II.13}$$

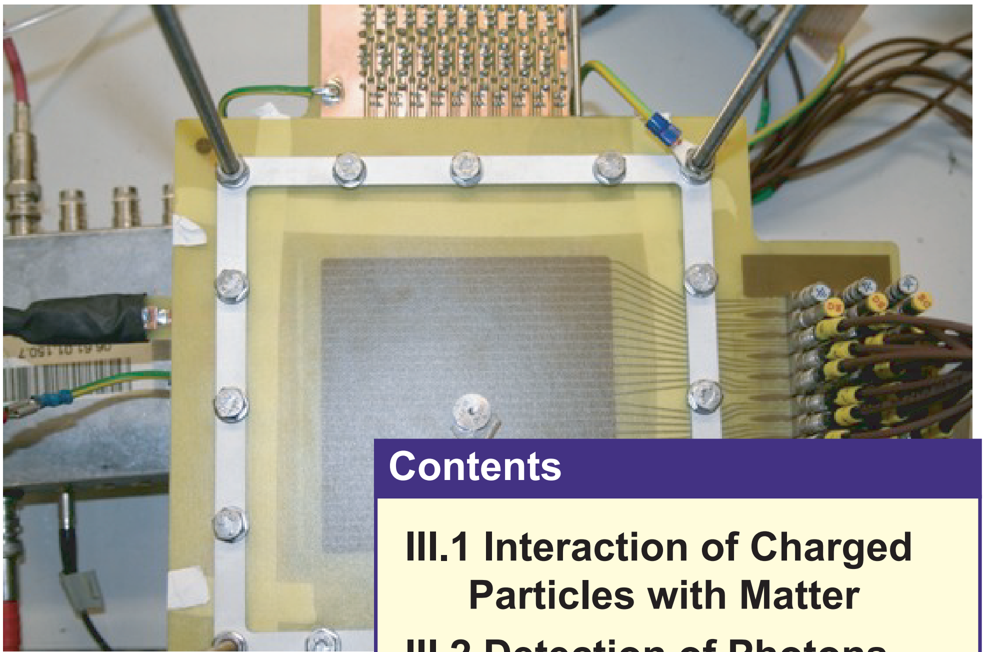
The easiest iterative algorithm to solve this equation would then be

$$f^{k+1} = f^k A^T \frac{g}{Af^k}, \quad k = 0, 1, \dots\ . \tag{II.14}$$

This algorithm is commonly called the EM algorithm. As its convergence is in general very slow, normally the events are subdivided into so-called subsets. This leads to the ordered subset expectation maximization (OSEM) algorithm that is commonly used today. For more detail on OSEM and its implementation see [Hudson94].

CHAPTER III

Gaseous detectors



Contents

Gaseous detectors are based on the interaction between an incident particle and a gas volume within the detector. There is an extended theoretical framework which provides cross sections and thus probabilities for such interactions. As this is the basis for every detector design, it will be considered first in this chapter. In the latter particular types of gaseous detectors will be described in detail. Beginning with simple proportional counters, multi wire proportional chambers will be explained, especially the Münster MWPC (msPET) prototype chamber .

III.1 Interaction of charged particles with Matter

If a charged particle traverses matter, it can interact with it via electromagnetic interactions. This principle it is commonly used for the detection of charged particles particles. In the following, we will only deal with the incoherent Coulomb scattering, as done in [Sauli77].

An important quantity in this context is the average energy loss within the traversed matter. This can be described by the famous Bethe[1]-Bloch[2] equation [Bethe30, Bloch33]. Approximately

$$-\frac{dE}{dX} = \frac{4\pi}{m_e c^2}\frac{nz^2}{\beta^2}\left(\frac{e^2}{4\pi\epsilon_0}\right)^2 \left\{\ln\frac{2m_e c^2\beta^2}{I\cdot(1-\beta^2)} - \beta^2\right\}, \qquad \beta = \frac{v}{c} \tag{III.1}$$

holds [Povh04]. In this formula, m_e and e are electron mass and charge respectively. ze is the charge of the particle, I is the medium's effective ionisation potential. It is obtained by measurements and, in most cases $I = I_0 Z$ is a good approximation. v is the velocity of the incoming charged particle. The first important thing to see is that the energy loss depends on the velocity and the charge of the transversing

[1] Hans Albrecht Bethe was a German-American physicist. We born on Juli 2nd, 1906 in Strasbourg, Alsace-Lorraine and died on March 6, 2005. In 1967, he obtained the Nobel price in physics for "for his contributions to the theory of nuclear reactions, especially his discoveries concerning the energy production in stars"

[2] Felix Bloch was a Swiss physicist. He was born on October 23, 1905 and died September 10, 1983. He obtained the Nobel price in 1952, together with Edward Mills Purcell "for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith"

particle but not on its mass. In general, the energy loss is normalized to the density ρ of the transversed matter, introducing a reduced length x so that

$$\frac{dE}{dx} = \frac{1}{\rho}\frac{dE}{dX} \tag{III.2}$$

holds. A qualitative look at equation III.1 shows that the energy loss is dominated by the β^{-2} term for small velocities. At relativistic energies, it increases due to the logarithmic term. These two competing terms lead to a minimum or an almost constant region for the energy loss as a function of projectile energy. This region is called the minimum ionizing region and is the probable case in particle physics [Sauli77].

III.2 Detection of Photons

As described above, in PET two γ photons with an energy of 511 keV each are emitted. The task of a PET camera is therefore to detect these photons with a maximum efficiency and spatial resolution. Unfortunately, these photons cannot be detected directly with a MWPC as their interaction probability with the gas volume is too low. Therefore, they are converted to electrons first. This is done in a so-called “converter” material that is located directly on the pad planes. There are three different processes of interaction of a photon with matter that lead to the liberation of a electron. These are photoelectric absorption, compton scattering and pair production (for details on these processes, see [Leo87]). However, if the cross-section of all these processes is put together a formula for the overall attenuation of a photon beam can be given as

$$I = I_0 e^{-\sigma N X} = I_0 e^{-\mu x}, \tag{III.3}$$

where σ is the cross-section, μ the mass attenuation coefficient and $x = \rho X$ is the reduced thickness of the medium. This formula can also be written as

$$I = I_0 e^{-(\frac{\mu}{\rho})\rho X}. \tag{III.4}$$

The quantity μ/ρ is called “mass attenuation coefficient”. Finally, defining the mean free path λ of the particle as $\lambda = (\mu\rho)^{-1}$, equation

$$I = I_0 e^{-\frac{X}{\lambda}} \tag{III.5}$$

holds. As the cross section increases with the atomic number, a good choise for the converter is gold ($Z = 79$) or lead ($Z = 82$).

III.3 Gaseous detectors

In the case discussed, the medium in which ionization takes place is always a gas volume. Therefore, diffusion and drift of electrons in gases are important processes. Of special interest is the case of a electric field applied over the gas volume that leads to a net movement of electrons and ions.

III.3.1 Strong Electric Field and Avalanche Formation

If a electric field is applied over the gas volume that exceeds a certain value, a electron can gather enough energy to ionize gas molecules. The resulting electrons can cause further ionization and thus a so-called "avalanche" is created. Knowing the mean free path of electrons for secondary ionization α, a formula for the number of electrons created per length can be derived. The inverse of α, the probability of a ionization per unit length, is often called the "first Townsend[3] coefficient". Starting with n electrons,

$$dn = n\alpha dx \tag{III.6}$$

further electrons will be created. Integration of this formula leads to

$$n = n_0 e^{\alpha x}. \tag{III.7}$$

From a given number of initial electrons, the multiplication factor or gas gain (after a way of length x) is[4]

$$M = \frac{n}{n_0} = e^{\alpha x}. \tag{III.8}$$

As this quantity is a very important parameter for a proportional counter, many models for α in different gases have been developed. For a review, see [Kowalski85].

III.3.2 Proportional Counter

Considering all the facts mentioned above, a simple gas detector setup could consist of two parallel, planar plates. In this case, the space between the plates is filled with gas and a electric field is applied. In a charged particle transverses the gas volume,

[3] John Sealy Edward Townsend was a Irish physicist. Be was born June 7, 1868 in County Galway, Irland and died February 16, 1957 in Oxford.

[4] In the more general case of a non-uniform electric field, α is a function of x and $M = exp\left(\int_{r_1}^{r_2} \alpha(x)dx\right)$ holds instead.

it causes ionization. Due to the electric field, the electrons produced are accelerated and gain enough energy to perform further ionization. An avalanche develops and creates a signal. However, there is a serious drawback of such a geometry. The amplitude of the signal (that is proportional the total ionization) depends on the point where the interaction of the incident particle within the detector took place. Thus, the correlation between the energy of the incident particle and the measured signal is lost. This problem can be solved by choosing an geometry in which the avalange formation takes place only in close proximity to the wire. The most popular geometr is a cylindrical setup with a anode wire in the center. This geometry can be described by the two parameters a which is the wire diameter and b the distance between the wire surface and the detector wall, see Figure III.1 a). The lateral electric field decreases with r^{-1} as a function of distance from the wire (Figure III.1 b)). The exact expression is

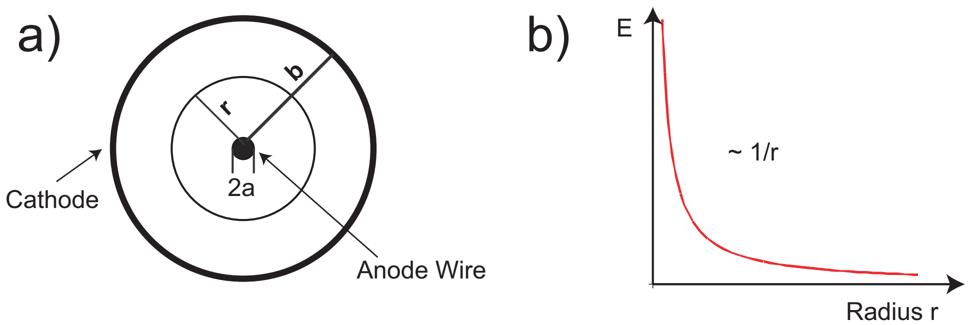


Figure III.1: a) A slice of a proportional counter perpendicular to the signaling wire. b) The electric field as a function of distance r from the center of the counter.[Sauli77]

$$E(r) = \frac{CV_0}{2\pi\epsilon}\frac{1}{r}, \tag{III.9}$$

which may also be written as a potential

$$\varphi(r) = -\frac{CV_0}{2\pi\epsilon} ln\left(\frac{r}{a}\right). \tag{III.10}$$

Here, r is the radial distance from the Wire, V_0 the applied voltage, ϵ the dielectric constant and

$$C = \frac{2\pi\epsilon}{ln(\frac{b}{a})} \tag{III.11}$$

the capacitance per length. If an incident particle liberates one or more electrons somewhere within the counter, the electric field accelerates them towards the anode wire, but the electrons do not gather enough energy to cause further ionization.

Only when they arrive at the very close proximity (a few radii in general) of the wire, the field strength is high enough to allow the creation of an electron avalanche. Because of lateral diffusion and the small dimensions the wire, the avalanche will completely surround the wire as shown in Figure III.2 a). If the number of electrons created in the avalanche is proportional to the number of electrons created in the primary ionization process, the signal is proportional to the energy of the incident particle. As shown in Figure III.2 b) this is only valid for a certain voltage range. If the device is operated within this range, it is called “proportional chamber”. In this mode, gains from 10^4 up to 10^6 can be obtained. If the applied voltage is too low, the electrons will still be moved to the anode wire, but no amplification will occur. Therefore, the signal will be too low to be measured. If, on the other hand, the voltage is higher the proportionality is lost gradually, as space charges build up around the anode. Finally, at even higher voltages, avalanches will spread out over the whole counter and the signal will become completely independent of the primary electrons. This is called the “Geiger-Müller” region.

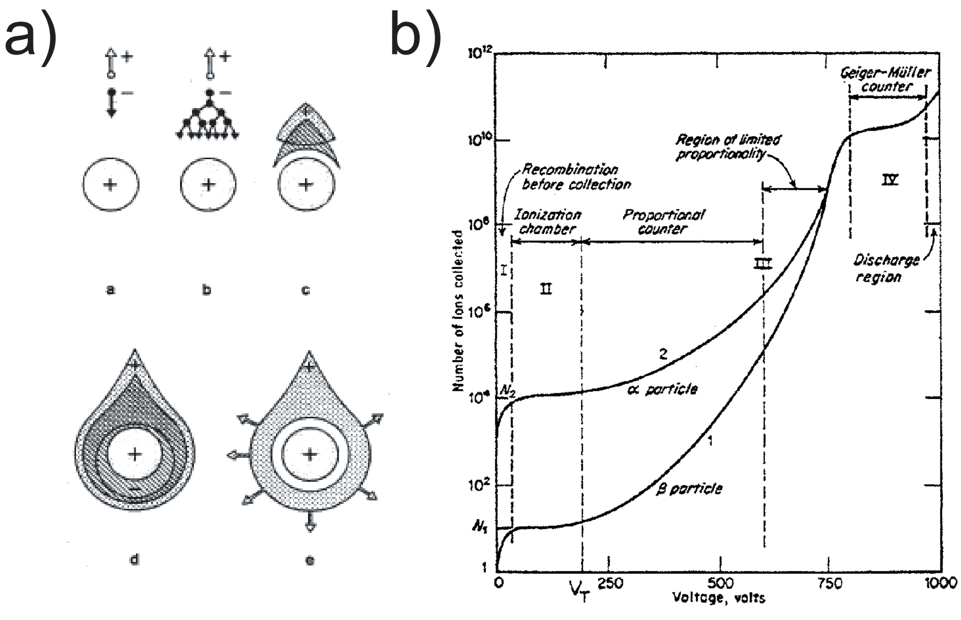


Figure III.2: a) Time development of an avalanche around an anode wire. b) The different modes of operation a proportional counter works depend strongly on the applied voltage. Here, the number of created ions is shown as a function of voltage. [Sauli77]

Signal Creation and Readout

It is important to realize that the creation of the signal is due to the movement of charges within the electric field rather than to the collection of charges themselves. From electrostatics, the induced signal can be calculated. For a charge q that is moved dr within a detector of (total) capacitance lC the induced voltage dv is given by the equation

$$dv = \frac{q}{lCV_0}\frac{d\varphi}{dr}dr. \tag{III.12}$$

In principle, the integration of this formula over the radius of the whole detector would lead to the induced voltage. However, as the electrons are liberated very close to the anode wire, their contribution to the signal is quite small. Assuming that all charges are created at the same distance d_C from the wire, the integral can be split up into two parts. Integration from a to d_C leads to the fraction of the signal created by the electrons (v^-) and from d_C to b that of the ions (v^+) respectively. The overall signal is of course given by

$$v = v^+ + v^-.$$

The ratio between the electron and the ion signal then is

$$\frac{v^-}{v^+} = \frac{\ln(a + d_C) - \ln a}{\ln b - \ln(a + d_C)}. \tag{III.13}$$

For typical parameters ($a = 0.02$ mm, $b = 10$ mm, $d_C = 0.001$ mm), the ratio is ≈ 0.8 %. To obtain a formula for the time dependence of the signal is possible by assuming that all ions are leaving the wire region with constant mobility. Then, integration reads to

$$v(t) = -\int_0^t dv = -\frac{Q}{2\pi\epsilon_0 l}\ln\frac{r(t)}{a}. \tag{III.14}$$

Using the definition of mobility gives

$$\frac{dr}{dt} = \mu^+\frac{E}{P} = \frac{\mu^+ CV_0}{2\pi\epsilon_0 P}\frac{1}{t}. \tag{III.15}$$

Integrating this formula and substituting the result in III.14 finally leads to

$$v(t) = -\frac{Q}{4\pi\epsilon_0 l}\ln\left(1 + \frac{\mu^+ CV_0}{\pi\epsilon_0 P a^2}t\right) = -\frac{Q}{4\pi\epsilon_0 l}\ln\left(1 + \frac{t}{t_0}\right). \tag{III.16}$$

The total drift time can of course be obtained by the condition $r(t) = b$ and is thus

$$T = \frac{\pi\epsilon_0 P(b^2 - a^2)}{\mu^+ CV_0}. \tag{III.17}$$

In Figure III.3 the function $-v(t)$ combined with the exponential decharge for different time constants is shown. The function has a very steep rise at the beginning which means that most of the signal develops within about the first one thousands of the total time [Sauli77]. Therefore, the signal normally is differentiated by a RC-circuit[5] with time constant $\tau = RC$.

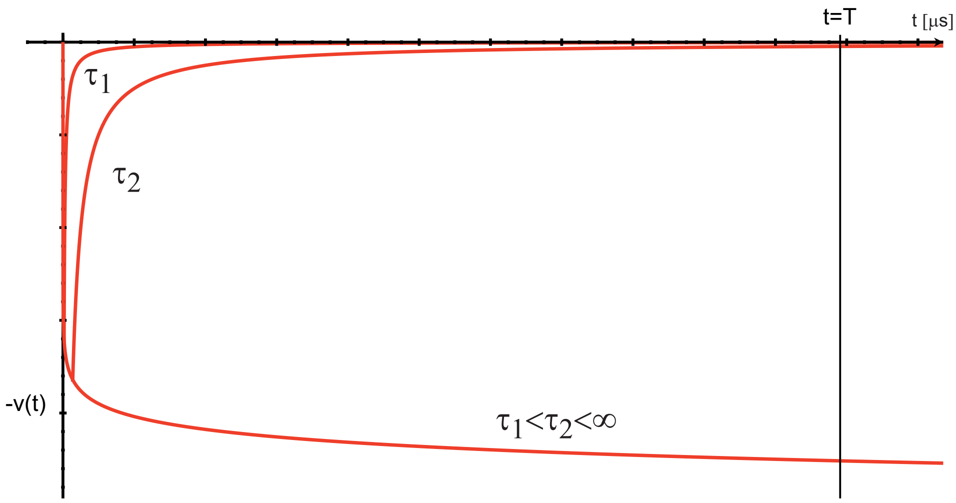


Figure III.3: Signals created by a MWPC. The sharp rise at the beginning is due to the drifting electrons. The later part is the exponential discharge of the chamber.

III.3.3 Filling Gas

As the principle of avalanche multiplication does not depend on the type of gas, generally every gas or gas mixture can be used as filling gas. However, miscellaneous mixtures also differ very much in their properties. For example, some allow a very high gain factors while others guarantee good proportionality. As these requirements often conflict with each other, the choice of gas is always a compromise. Experimental results have shown, that in noble gases multiplication occurs at a much lower voltage than in more complex molecules. This is due to the fact that molecules with more than one atom have a number of non-ionizing dissipation modes (e.q. vibrational or translational states). Within the group of noble gases, argon is in most[6] cases the first choice for it is much less expensive than xenon or krypton.
In practice however, counters filled with argon only cannot exceed a gain of 10^3

[5]For details on RC-circuits see [Horowitz89]
[6]The Alice TRD uses Xenon as it has a higher interaction probability for transition radiation.

to 10^4. This is because of secondary ionization that takes place if an argon atom is excited but not ionized. The atom can return to its ground state only through the emission of a photon. The minimum energy of such a photon is 11.6 eV for argon and is thus well above the ionizion potential of a copper cathode, which is only 7.7 eV. Therefore, these photons can ionize the copper atoms and create further electrons that start new avalanches. Argon ions on the other hand migrate to the cathode where they are neutralized with an electron from the copper. The balance energy of this process is also either radiated or a secondary emission takes place [Sauli77]. These secondary electrons cause additional avalanches shortly after the initial one. This puts the counter in the regime of permanent discharge. To suppress these effects, another component is added to the gas mixture. As already mentioned, polyatomic molecules posses a number of non ionizing absorption modes. Therefore, they can also absorb the photons created by the excited argon atoms without creating further electrons (methane, for example absorbs photons in a range from 7.9 up to 14.5 eV). This is a common property of most organic compounds but also of some inorganic molecules as CO_2, BF_3 or freon[7]. If an ionized polyatomic molecule neutralizes at the cathode, emission of a secondary electron is very unlikely. Instead, the molecule either dissociates or forms larger complexes (polymerisation). The second gas component is called "quencher", because it quenches the emitted photons and secondary electrons. Common mixtures are P10 which is 90% argon and 10% methan or a mixture of argon and $\mathrm{CO_2}$. Using a quencher with good photon absorption properties, gains larger than 10^6 are possible [Sauli77]. There are some more criteria for the choice of the gas, for a detailed discussion see [Sauli77].

III.4 Multi Wire Proportional Chambers

Even though proportional counters are well suited for the detection of the energy loss of radiation, a serious drawback is their lack of spatial resolution. It can only be determined whether a particle transversed a chamber or not. A straight forward approach to overcome this limitation would be a device with a multiwire structure. However, it was longtime believed that such a device would not work due to capaci-

[7]Freon is a commercially available refrigerant. It is produced by the DuPont company. See http://refrigerants.dupont.com for details.

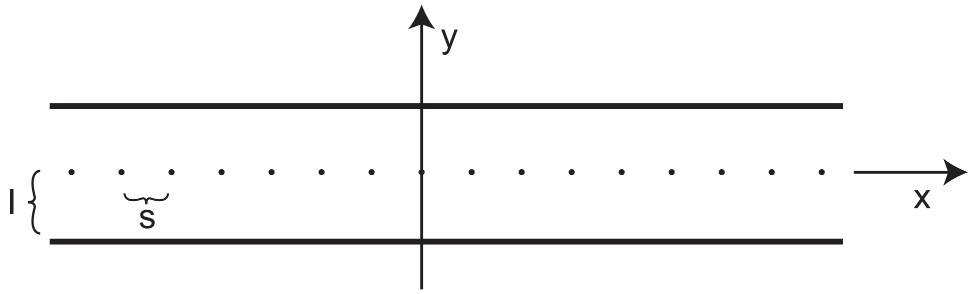


Figure III.4: Important geometry parameters of a typical MWPC. [Sauli77]

tive coupling between the individual wires. In 1967/1968, George Charpack[8] and his coworkers proved that black is white by constructing the first Multi Wire Proportional Chamber or MWPC short [Charpack68]. They recognized that the negative signals produced by capacitive coupling are largely compensated by the induced positive signals in all wires surrounding the anode of interest. In fact, every wire acts as a single, independent detector.

Soon after their invention, MWPCs became extremely popular in high energy physics. One reason is that until then, most detectors were optical in nature (e.g. cloud- or spark-chambers). There was a need for purely electronic devices to acquire data at higher rates and improve data processing. On the other hand, the developments in semiconductor electronics had just made it possible to handle large numbers of channels.

The basic geometry of MWPCs is almost everywhere the same[9]. The MWPC consists of a anode wire layer sandwiched between two cathode planes. Theses cathodes can either be wires as well or a continuous plane of some conductive material. In most cases, this material is subdivided into so-called “pads”. This allows a special readout technique that will be explained in the following. In Figure III.4, the geometry and common names of important geometrical parameters are shown. Because the electric field within the chamber is very sensitive to the position of the wires, they are kept under mechanical tension to assure that they stay in their positions. This requires a very stable frame which is one of the main mechanical defiances in the construction of MWPCs. If the chamber’s dimensions exceed a certain limit (about 1 m), there even have to be supportive elements for the wires.

[8] Georges Charpak is a Polish-French physicist. We was born on August 1, 1924. In 1992, he obtained the Nobel price in physics “for his invention and development of particle detectors, in particular the multiwire proportional chamber”

[9] One common extension are driftchambers, for details see [Blum93] or cylindric chambers.

III.4.1 Electrostatics within a MWPC

The electric field or the electric potential respectively can be obtained by solving Laplace's[10] equation

$$\Delta V(x, y, z) = 0 \tag{III.18}$$

with appropriate boundary conditions. Assuming an infinite wire plane and a zero wire diameter, the problem is invariant under any translation in z and only the two dimensional Laplace equation for a potential $V(x, y)$ has to be solved. As to be expected, the field is constant and the field lines are parallel to each other and orthogonal to the cathode planes in most of the chamber. They are similar to the field lines in a plate capacitor with the same dimensions. However, in proximity to the wires, the field is deformed as shown in Figure III.5 a). Even though an analytical formula for the potential can be obtained [Blum93], today simulations are the appropriate tool to get detailed information on the electric field. These are based on numerical solutions of the Laplace equation. In the simplest case, a cartesian grid is placed over the whole chamber. The differential operator Δ is then replaced by finite differences on the grid. A simple forward Euler[11] algorithm

$$V(i, j) = \frac{1}{4}\left[V(i+1, j) + V(i-1, j) + V(i, j+1) + V(i, j-1)\right] \tag{III.19}$$

already produces some results. In this equation $V(i, j) = V(x + i\Delta x, y + j\Delta y)$ if Δx and Δy are the cell sizes of the cartesian grid. However, more sophisticated (i.e. finite element) methods provides a convergence of higher order. For details see [Burger06] or [Giordino97] .

Furthermore, simulations can give information of field disturbance if for example one wire is slightly displaced. This is shown in Figure III.5 b)

III.4.2 Readout and Resolution Recovery

In principle, there are several possibilities to readout a MWPC. The simplest would be to consider each wire to be a single detector with its own electronics. This approach has several drawbacks. It results in a large number of channels to be read (a chamber with a 20 cm wire plane, a wire radius of 20 μm and a wire distance of

[10] Pierre-Simon Laplace was french mathematician. He was born in March 23, 1749 in Beaumont-en-Auge (Normandie) and died March 5, 1827 in Paris.

[11] Leonhard Paul Euler was a swiss mathematician. He was born April 15 1707in Basel and died September 18 1783 in St. Petersburg.

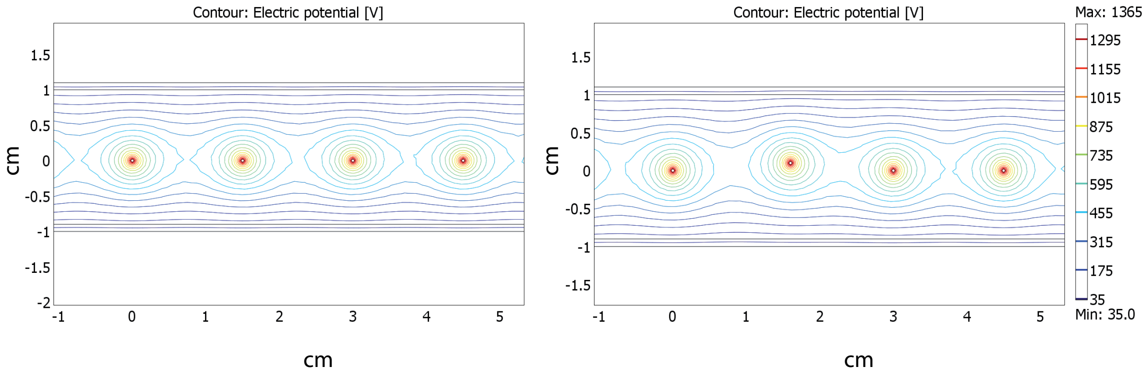


Figure III.5: a) Equipotential lines for a chamber with 20 μm wire diameter, 1.5 mm distance between the wires and 1 mm between wire- and pad-plane. b) Equipotential lines for the same geometry. One wire is displaced by 10 % in x- and y-direction. All plots have been created with COMSOL using a finite element method.

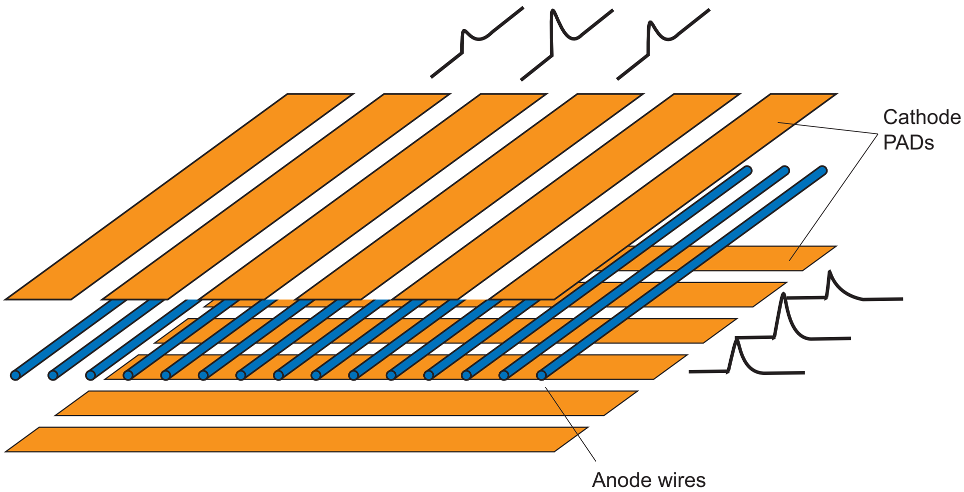


Figure III.6: Schematic setup of a MWPC with pad readout. [Sauli77]

1.5 mm contains $\approx$ 130 wires), every wire must be soldered to a single output pad and has to be decoupled from the high voltage. Although this is possible with modern electronics, it would provide spacial resolution only perpendicular to the direction of the wires, i.e. in one dimension. To overcome these limitations, the cathode plaines are subdivided into so-called "pads". These are strips that are electrical insulated from each other. Therefore the charge induced by the moving ions is shared by a number of pads and can be measured as a function of pad number. Using the center of gravity method (i.e. weighted average) of several pads gives a first estimate for the coordinate of the incident particle. The avalanche on the anode wire is also highly localized along the wire direction [Leo87]. This allows, using two perpendicular pad planes, to determine both x and y coordinate. A setup of such a MWPC is shown in Figure III.6.

To understand how the signal on the pads is created, we consider a setup in which only one of the two cathode plaines has one pad which means that it is split up into three pieces. Furthermore, we assume a series of avalanches that come from a straight particle track and neglect the width of the avalanches. The track is then reduced to a infinite thin line of charge with density λ, located in the middle of a capacitor. The goal is to calculate the surface charge density $\sigma(x)$ induced on the pad. From Gauß[12] law

$$\sigma(x) = \epsilon_0 E(x) \tag{III.20}$$

holds. To obtain an expression for the electric field E, we use the *Method of images*[13]. The images of a line of positive charge are alternating positive and negative lines at the position

$$z_k = \pm(2k+1)\frac{D}{2}, \qquad k = 1, 2, \dots \text{ and } z_0 = -\frac{D}{2}. \tag{III.21}$$

Each pair of these lines at $\pm z_k$ creates an electric field which is normal to the surface at $z = 0$ and is equal to

$$E_k^{(n)} = \frac{\lambda}{\pi\epsilon_0}\frac{z_k}{x^2 + z_k^2}. \tag{III.22}$$

As the total surface charge is the sum of all these parts, this leads to

$$\sigma(x) = -\frac{\lambda}{\pi}\sum_{k=0}^{\infty}(-1)^k\frac{(2k+1)\frac{D}{2}}{x^2 + (2k+1)^2\frac{D^2}{4}} = -\frac{\lambda}{2D}\frac{1}{cos\left(\frac{\pi x}{D}\right)}. \tag{III.23}$$

This function is shown in Figure III.7. Note that the integral

$$\int_{-\infty}^{\infty}\sigma(x)dx \tag{III.24}$$

is $-\lambda/2$. This means that half of the original charge is transfered to a cathode plain of infinite dimension.

[12] Johann Carl Friedrich Gauß was born on 30 April 1777 in Braunschweig and died 23 February 1855 in Göttingen. He was a german mathematician and scientist and contributed to many field, e.g. number theory, magnetism, etc.

[13] The principle of this method is explained in [Nolting07] or every other book on electrostatics.

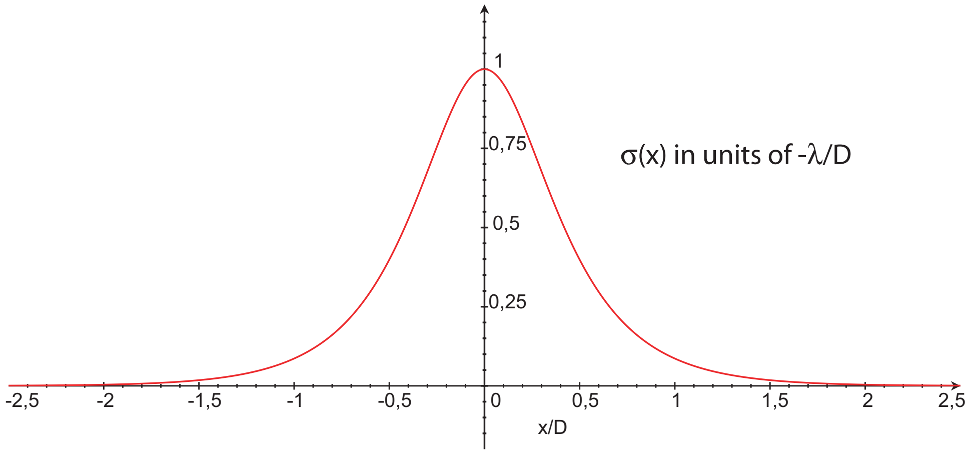


Figure III.7: Charge density induced on a pad by a line of charge with density λ [Blum93].

Spacial Resolution and PAD Response Function

To reconstruct the coordinate of an incident track from the signals measured on the pads, the amount of charge deposited on a pad as a function of the coordinate of the incident particle is needed. If W is the width of a pad strip, and x the distance from the center of the strip, this is given by

$$P(x) = \int_{x-\frac{W}{2}}^{x+\frac{W}{2}} \sigma(x^{'})dx^{'}. \tag{III.25}$$

$P(x)$ is called "pad response function". The important fact for reconstruction of the coordinate is, that this function can be approximated by a Gaussian curve,

$$P_0(x) \approx Ae^{-\frac{x^2}{2s_0^2}}. \tag{III.26}$$

This was determined experimentally [Blum93]. Measuring the signals on three consecutive pads results in the system of equations

$$p_{i-1} = A\exp\left(-(x-(x_i-W))^2/2s^2\right) \tag{III.27}$$
$$p_i = A\exp\left(-(x-x_i)^2/2s^2\right) \tag{III.28}$$
$$p_{i+1} = A\exp\left(-(x-(x_i+W))^2/2s^2\right), \tag{III.29}$$

with p_i the pulse height on the ith pad and $x_i - W$, x_i, $x_i + W$ the centers of three adjacent strips. A is a constant proportional to the overall charge. In principle, the three unknowns A, x and s can be obtained by solving this system if p_i, p_{i-1} and p_{i+1} are measured. However, solving it for s first,

$$s^2 = -W^2 \ln\left(\frac{p_{i-1}p_{i+1}}{p_i^2}\right), \tag{III.30}$$

makes it possible to measure the pad response function and thus determining s in advance. The coordinate x is then given by[14]

$$x = \frac{1}{w_1 + w_2}\left[w_1\left(x_i - \frac{W}{2} + \frac{s^2}{W}\ln\frac{p_i}{p_{i-1}}\right) + w_2\left(x_i + \frac{W}{2} + \frac{s^2}{W}\ln\frac{p_{i+1}}{p_i}\right)\right]. \quad \text{(III.31)}$$

The weights w_1 and w_2 are usually chosen as p_{i-1}^2 and p_{i+1}^2 as the measurement error is roughly proportional to the recorded pulse height on the strips with the smaller heights, cf. [Blum93].

III.5 The Münster Prototype

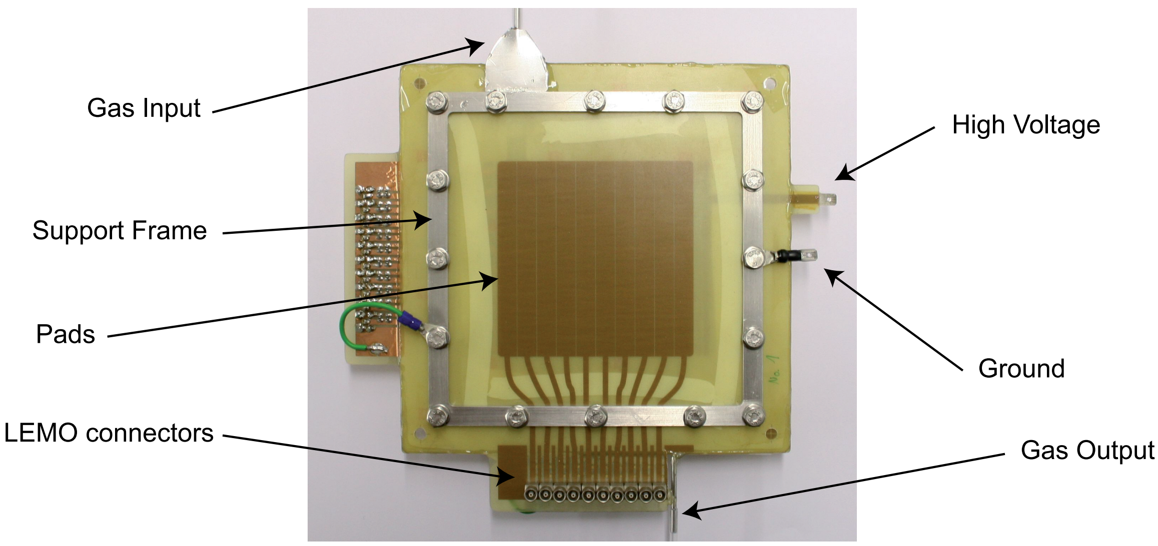


Figure III.8: Photo of the 3 mm wire-to-pad prototype with 10 pad strips.

With the goal of building a complete small animal PET prototype, the construction of MWPC prototypes for this purpose started in Münster in 2005. As mentioned above, the idea of a wire chamber based small animal PET has already been realized in the quadHIDAC scanner. However, as these devices (and service quantities) are no longer commercially available, it was decided to start the design of msPET.
The first step was the design of small MWPCs needed for this purpose. The prototype consists mostly of F10 (or G10 which is the halogen free version of F10). This is a compound of woven glass and epoxy as usually used for printed circuit boards

[14]This equation is easily calculated by combining eqn. XX a) and XX b) first, then XX b) and XX c) and finally add these two terms for x with the weights w_1 and w_2.

(PCBs). Therefore, it is commercially available with a laminated copper layer. It has been chosen because of it good insulation properties, does not outgase and has a small radiation length. Furthermore, it provides enough mechanical stability to act as a frame for wires under tension. As shown in Figure III.9 a), the chamber is composed of four layers: a frame supporting the wires sandwiched between two pad planes. The pads are made of copper. They are etched using a negative pattern (Figure III.9, b), exactly as done for regular PCBs. At the end of the plane, every pad is connected to a LEMO connector. Two stainless steel pipes on two opposite sides allow a continuous flow of the gas mixture. To apply high voltage to the wires, they are all soldered onto a strip of copper that is connected to an external pin. For additional stability, two stainless steel frames are screwed together with the whole chamber.

The initial prototype had a wire-to-pad distance of 3 mm and a gas volume of $100 \times 100 \times 3$ mm^3. However, measurements with lead foils showed that electrons created within these foils by γ photons leave them under a certain angular distribution with mean exit angle

$$\alpha_{mean} = 36.0^\circ \pm 2.8^\circ$$

[Hünteler07]. This has a negative effect of the spacial resolution of the chamber that scales with the height of the chamber. To preserve its good intrinsic resolution, a thiner prototype (1 mm wire-to-pad distance) has been built. For further details, see [Hünteler07].

In addition, two versions of the 1 mm prototype with different pad sizes were built. In the first version, the pads have a width of 9.7 mm. In the second version, the width has been reduced to 3.03 mm to optain a higher spatial resolution.

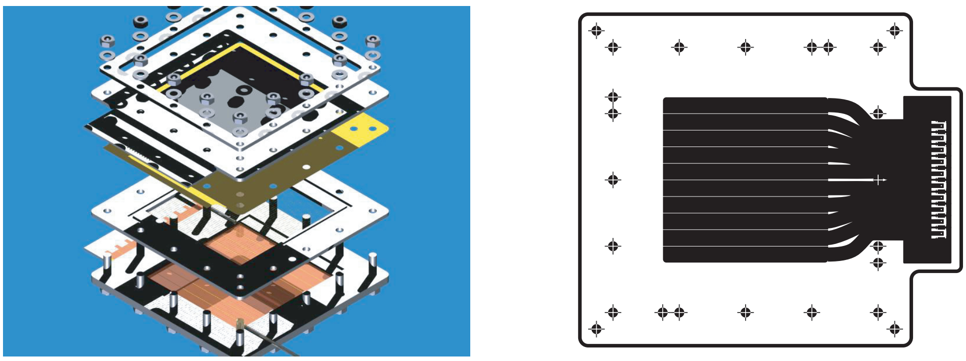
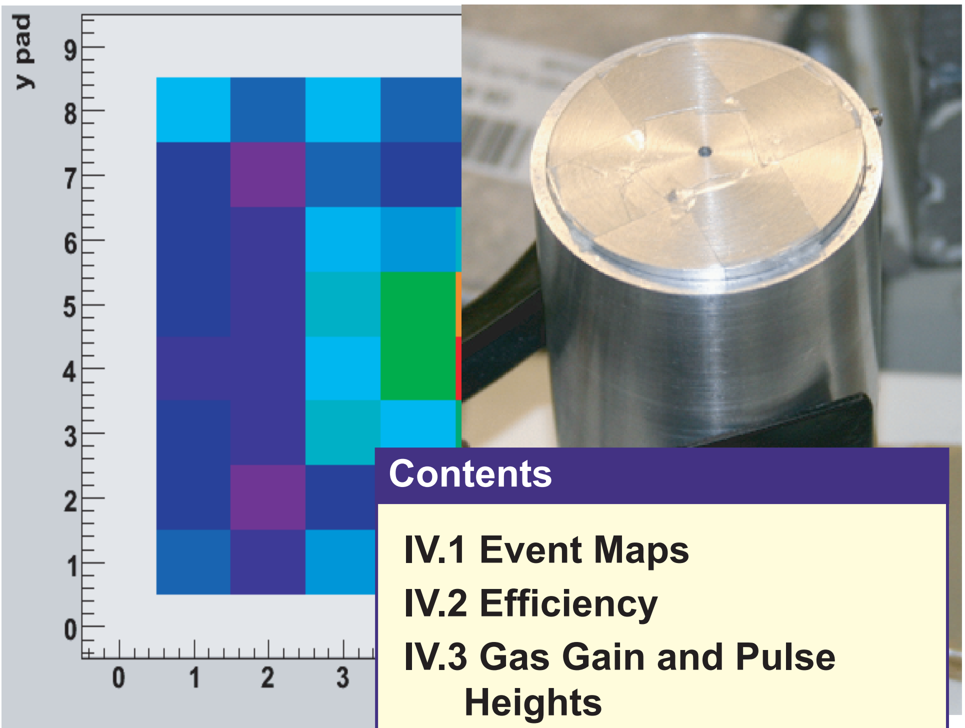


Figure III.9: a) Engineering plans of the first MünsterPET prototype. b) The negative used to etch the pads. [Gottschlag07]

CHAPTER IV

Experimental Evaluation of the msPET MWPC Prototypes



Contents

Before building a complete small animal PET prototype, it is important to assure that the MWPC prototypes work properly. Therefore, a number of tests have been performed. Beginning with simple event maps it has been tried to examine the efficiency and spatial resolution of the prototypes.

IV.1 Event Maps

Here, the term event map refers to a plot in which a histogram is filled with the number of the pad carrying the highest charge in each single event. Combining the information of the two perpendicular pad planes, the result is a checker board like pattern.
To create such maps, it is necessary to be able to read the voltages on all pads. For the first 1 mm prototype, this means 20 values. However, as the peak sensing ADCs[1] used feature only 16 channels per module, only the eight inner pads per plane have been considered. As the ADCs need a gate, the signal from the wires plane of the MWPC has been used as a trigger. This setup in schematically shown in Figure IV.1. The first test that has been done was the measurement of cosmic radiation. As

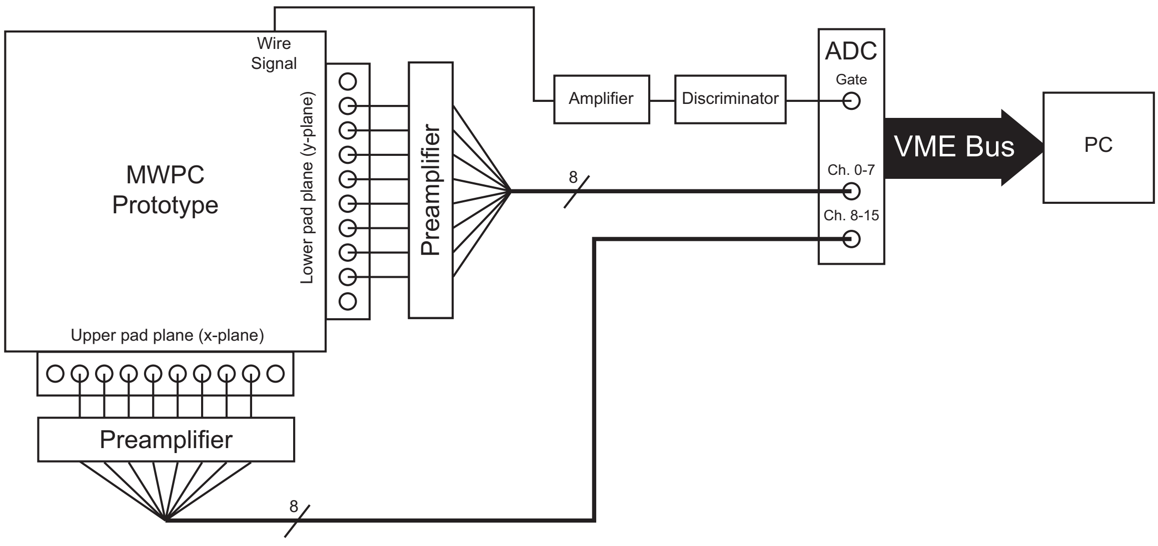


Figure IV.1: Setup for acquiring event maps.

this radiation is equally distributed over the surface of the whole chamber, the event map should, neglecting statistical variations, essentially be homogeneous. However, as shown in Figure IV.2 a) this is not the case. There are fluctuations, especially

[1] ADC stands for **a**nalog **d**igital **c**onverter.

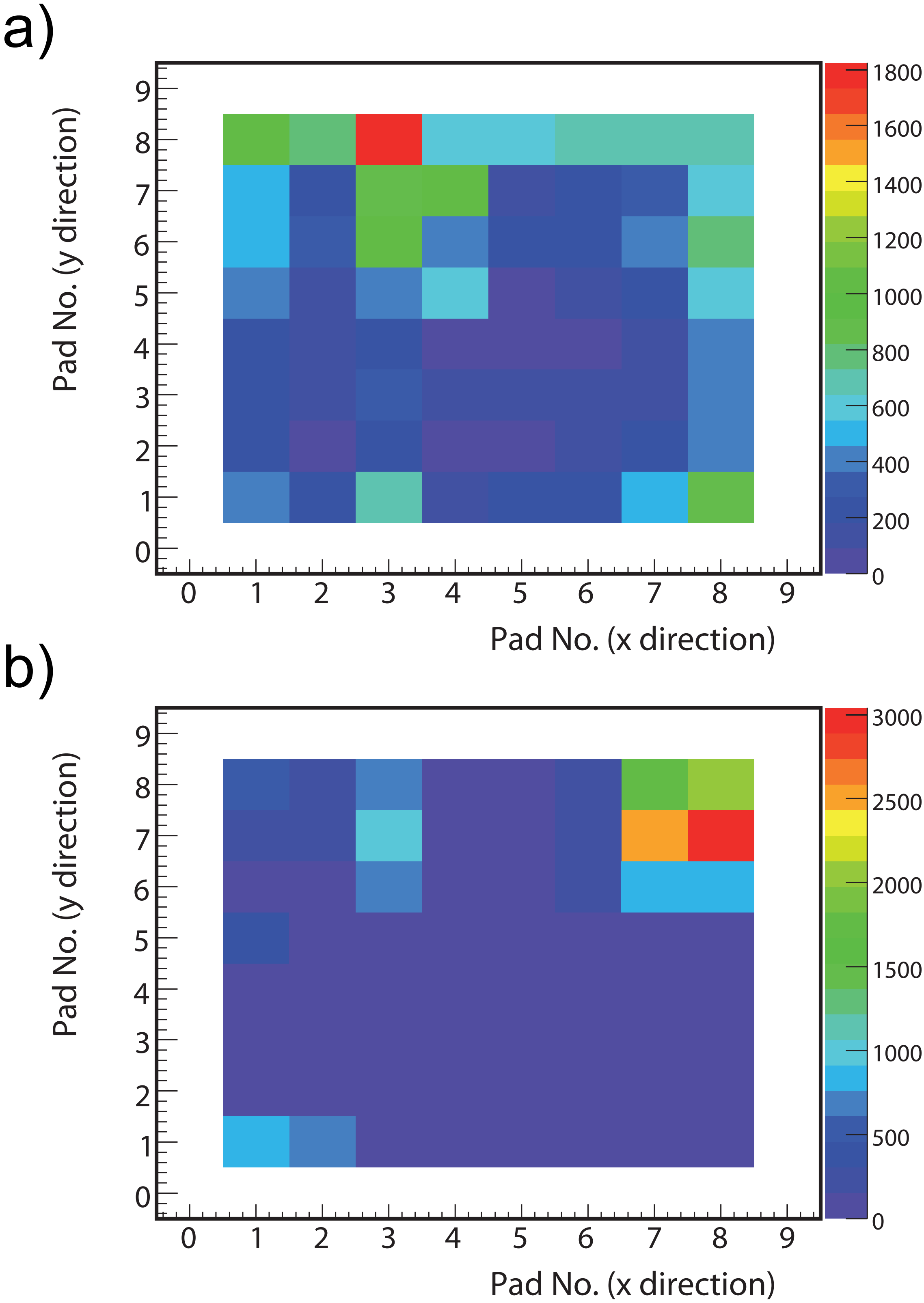


Figure IV.2: a) Shows a event map created using cosmic radiation as source. 25,000 events were taken. b) Event map created with a ^{55}Fe source positioned on the top right corner of the chamber. 25,000 events were taken.

at the border of the map. There is a number of possible explanations for this. For example, different wire tensions could lead to a deformation of the electric field within the chamber and thus a different gain in different regions (cf. Figure III.5). Furthermore the area of increased activity coincidences with the gas inlet. Due to perturbations, a higher density of the gas could arise. This also would lead to more events.
In the next step a X-ray source has been placed on the chamber. A ^{55}Fe source with an activity of 37 MBq has been used. It emittes photons at a energy of 231 keV in electron capture. In Figure IV.2 b), a event map with the source placed on the chamber is shown.
These event map suggests that the chamber is working as desired because the position of the source on the chamber can be seen in the plot. However, more detailed measurements were performed to study the chamber in more detail and determine the important parameters efficiency and spatial resolution.

IV.2 Efficiency

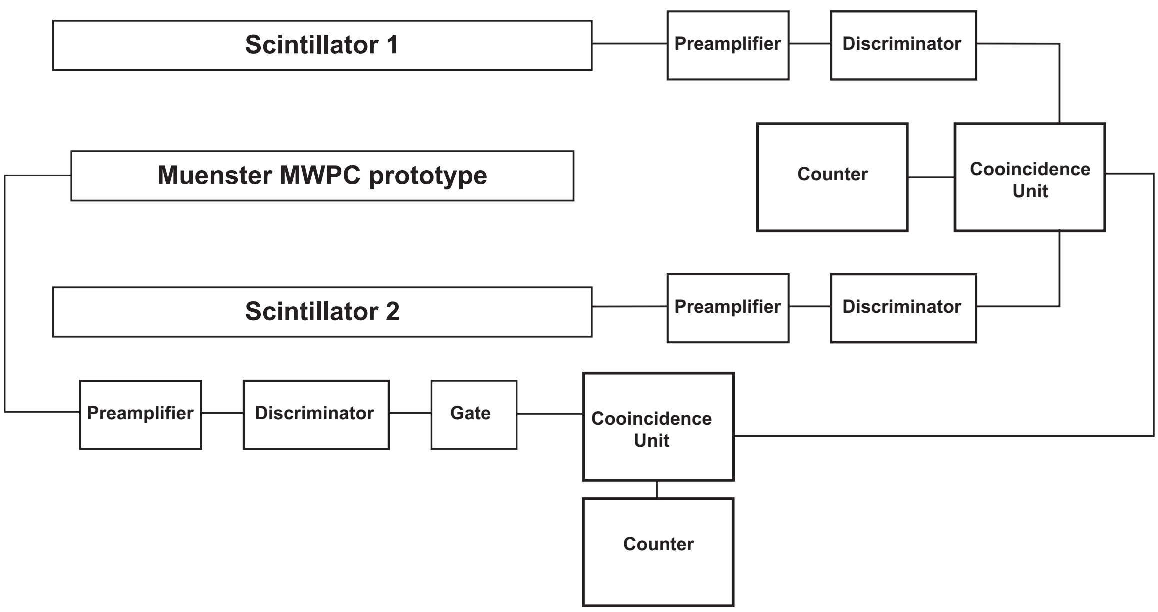


Figure IV.3: Schematical view of the efficiency setup.

Efficiency is defined as the number of charged particles that are detected by the chamber over the total number of particles traversing it. To test the prototype, cosmic rays were used as source. In the lower atmosphere, they consist mostly of muons and electrons. The flux of the myons crossing a unit horizontal area is $\approx 130\ \mathrm{m}^{-2}\mathrm{s}^{-1}$ or $\approx 50\ \mathrm{m}^{-2}\mathrm{s}^{-1}$ for electrons respectively. The mean energy of the myons is about 4 GeV [PDG06]. For Argon, the number of primary ions created by minimum ionizing particles is 22.9 per cm [Kleinknecht84] and thus 2.3 per mm. As in principle, every electron released in the gas starts an avalanche, theoretically an efficiency of 100% is expected and values up to 95% have been obtained for other chambers, see [Nettebrock91].

To experimentally determine the efficiency, the chamber has been "sandwiched" between two scintillators, operating in coincidence. This is necessary to consider only particles that really cross the chamber. Furthermore, it reduces the noise of the photomultipliers. Assuming that they are able to detect every charged particle that transverses them, the efficiency of the chamber is simply the number of coincidences of the photomultipliers and the chamber over the number of photomultipier coincidences only. The complete setup is schematically shown in Figure IV.3.

Efficiency measurements have been done for both P10 and a 82% Argon/18% CO_2 mixture using the 1 mm prototype. The results are shown in Figures IV.4. For P10, above 1520 Volt the chamber got into a mode of spontaneous discharge and could no longer be operated. Using the Argon/CO_2, the chamber could be operated up to 1640 Volt due to the better quenching properties of this mixture. As the rate of events became very low for the Argon/CO_2 mixture at lower voltages, the measurement was started at 1540 Volt. The measurements show that the chamber achives a maximum efficiency of about 40% at 1520 Volt using P10 and of about 35% at 1640 Volt using Argon/CO_2. There are two possible explanations for these results. Either, the chamber has to be operated at a higher voltage and thus at a higher electrical field. This would allow the chamber to create a measurable signal even when only very few electrons are created. The other possible explanation is that for a chamber with such a small gas volume, the number of electrons created is not always high enough to start an avalanche.

Furthermore, it would be worthwhile to repeat the measurements with other gas mixtures. In fact, a system allowing the flexible mixing of Argon with CO_2 or other quenching gases is currently beeing installed in Münster.

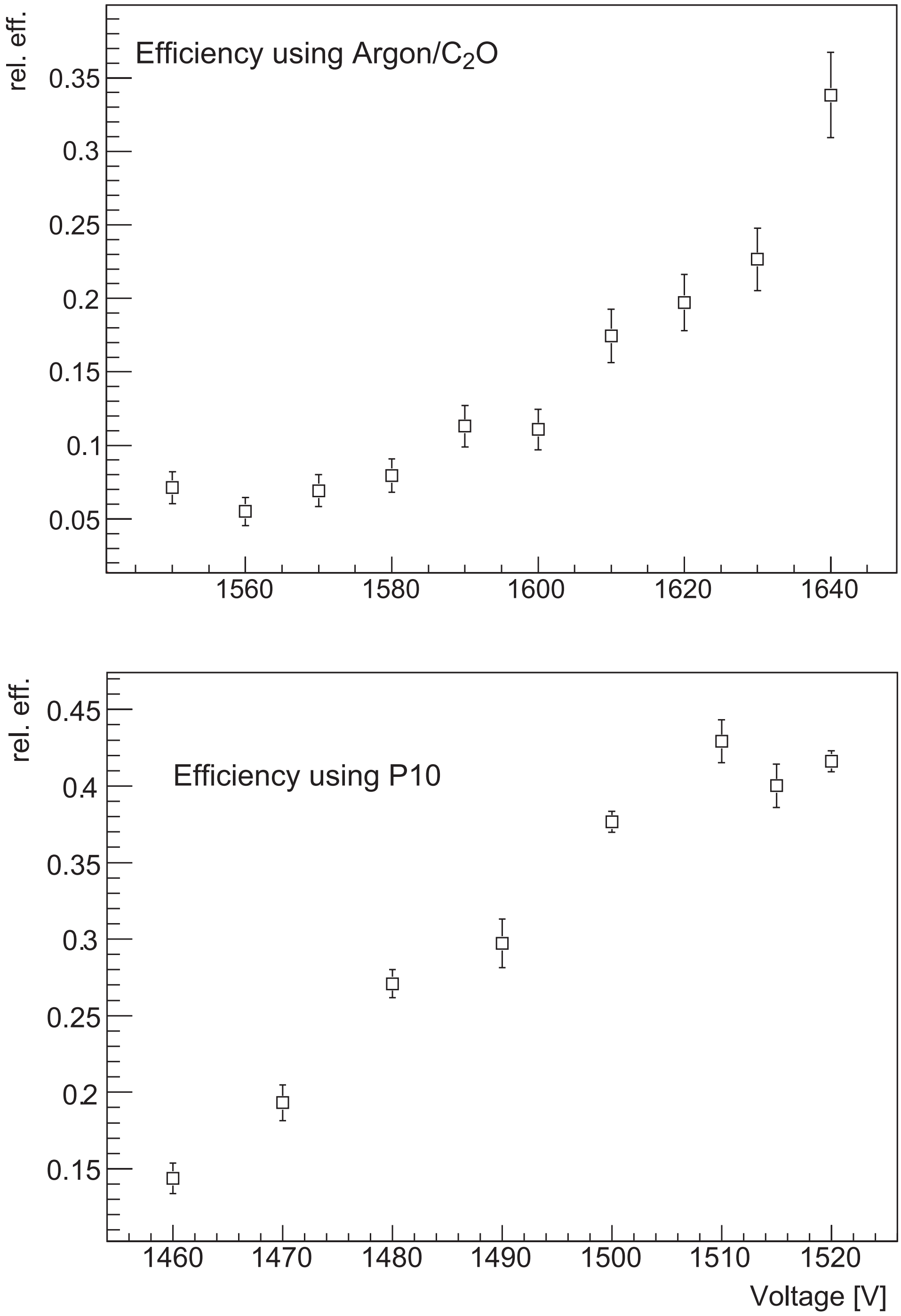


Figure IV.4: Results for the efficiency measurements. a) Using P10 as a filling gas b) Using Argon/CO_2 as a filling gas. Both measurements were taken using a 1 mm prototype.

IV.3 Gas Gain and Pulse Heights

As derived in a previous chapter (Equation III.8), the gas gain of a MWPC is defined as

$$M = \frac{n}{n_0} = e^{\alpha x}.$$

Furthermore, it has been shown that the amount of charge deposited on a pad is proportional to the number of secondary electrons n. This means that using a charge sensitive preamplifier, it is possible to determine M from the height of the measured signal. Here a Ortec Model 124 amplifier with a gain of 275 mV/pC has been used. The output of the amplifier has been further analysed with a LabView[2] based readout system. This system is able to automatically determine the pulse height. Finally, a histogram of the pulse height distribution of 5000 pulses at a given voltage was created. This has been done a range from 1500 to 1620 volts, in steps of 10 volts. Figure IV.5 shows every second histogram. As it is not clear, how many electrons are created in the primary process (most likely the interaction of a myon with the gas volume), only the measured charge is shown. Note that using only one pad, it has to be assumed that in most cases only a fraction of the charge induced is really measured. This is due to the fact that the charge is normally distributed to two or three adjacent pads. However, according to theory, a shift of the mean of the histogram is expected as the gain should increase with the voltage. The mean value versus the voltage is shown in Figure IV.6. The plot confirms the increase of the mean charge as a function of voltage.

IV.3.1 Simulations with Garfield

The measurements have also been compared to simulation made with Garfield. Garfield is a program dedicated to the simulation of MWPCs. It has been developed in 1984 by Rob Veenhof at CERN[3]. Garfield allows the simulation of the electric potential within the MWPC as well as the gain factor. For the msPET geometry, the gain has been simulated using a standard numerical method as well as a Monte Carlo simulation. This has been done for different gas mixtures. The results are displayed in Figure IV.7. Unfortunately, the gain predicted by Garfield is much higher than the measured one. Above a certain voltage, the predicted values become as

[2] LabView is a graphical data acquisition system. For details see [Baumann05, Gottschlag05].
[3] CERN stands for European Organization for Nuclear Research.

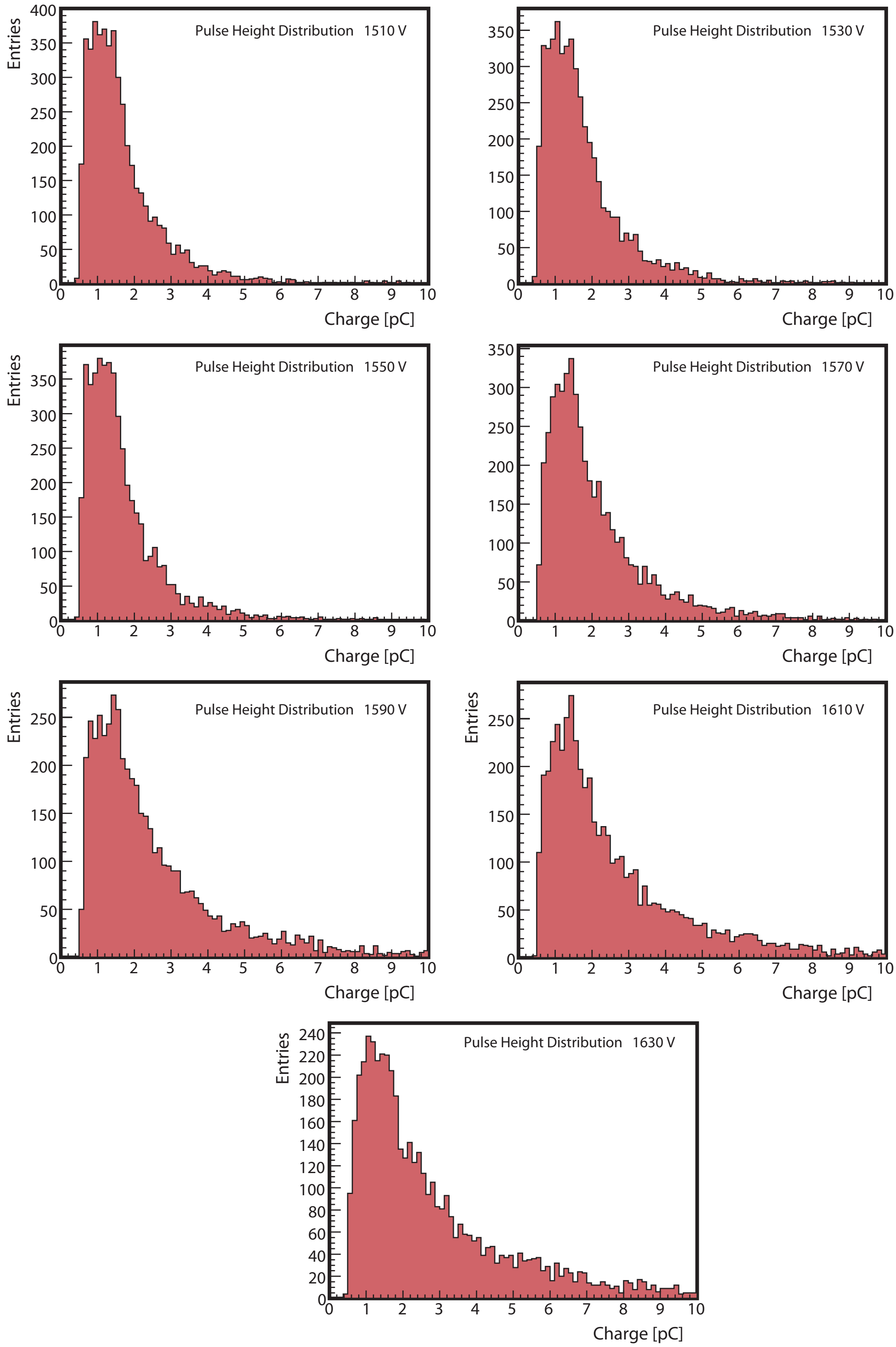


Figure IV.5: Histograms showing the charge distribution on a single pad for different voltages.

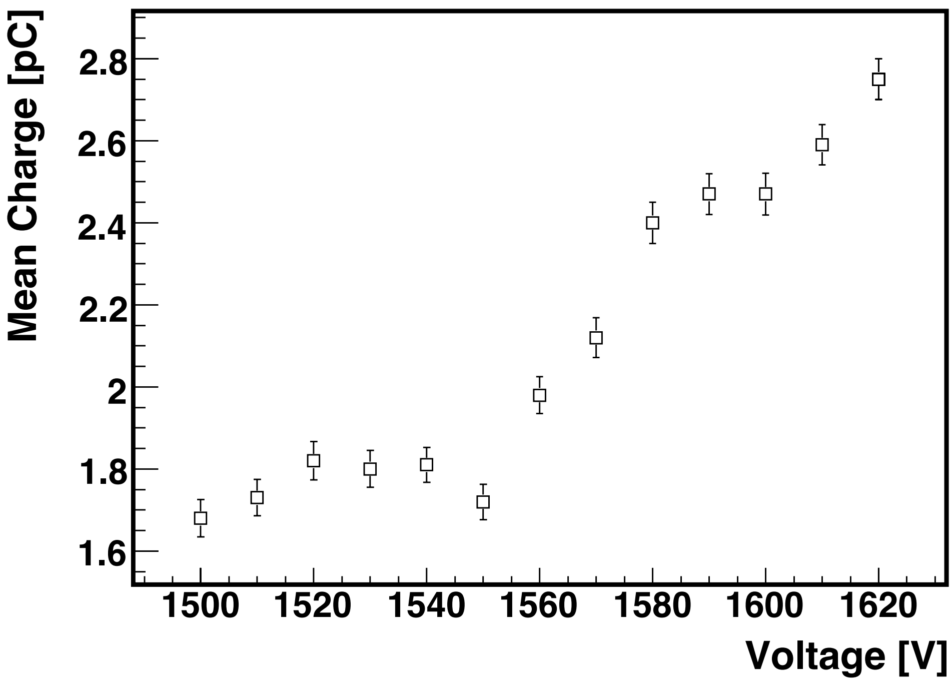


Figure IV.6: The Mean Charge is increasing as a function of voltage.

high as 10^{19} and stay constant. This most likely means that the algorithm used in Garfield does no longer works properly for the given set parameters.

IV.4 Spatial Resolution

Spatial resolution is the ability of the MWPC to resolve the position where it has been traversed by a charged particle. It is usually defined as the minimum distance at which two events can still be separated. To determine the spatial resolution of our prototype, we used the ^{55}Fe source partly collimated by a mask. The mask consists of pairs of holes with different distances between them as shown in Figure IV.8. To reconstruct the position of the incident particle from the signal measured on the pads, the pad response method that is described in III.4.2 has been used. Here, the approach to approximate the PRF with a exponential function

$$P_0(x) \approx e^{-\frac{x^2}{2s^2}}$$

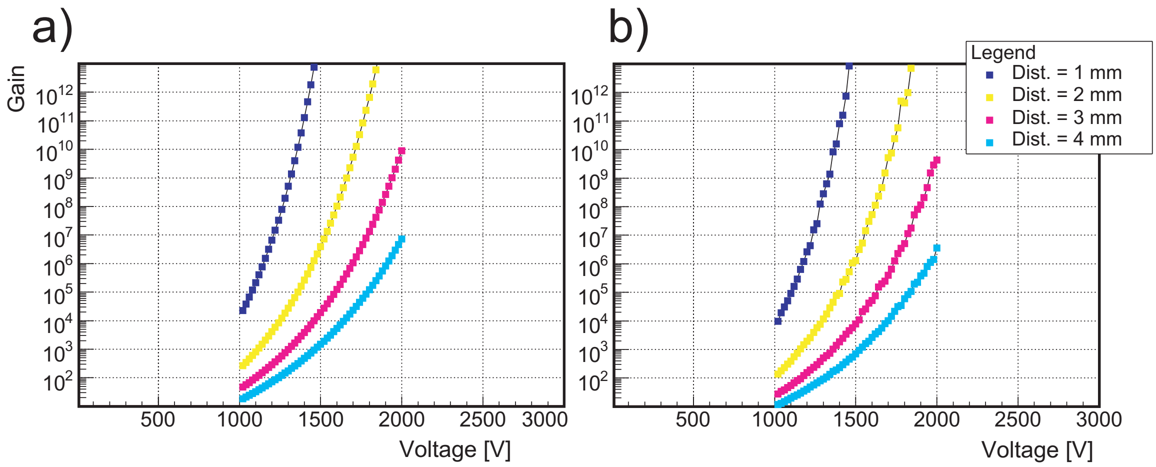


Figure IV.7: Simulated gains for a) P10 and b) 82% Argon/18% CO_2 mixture for different pad to wire distances.

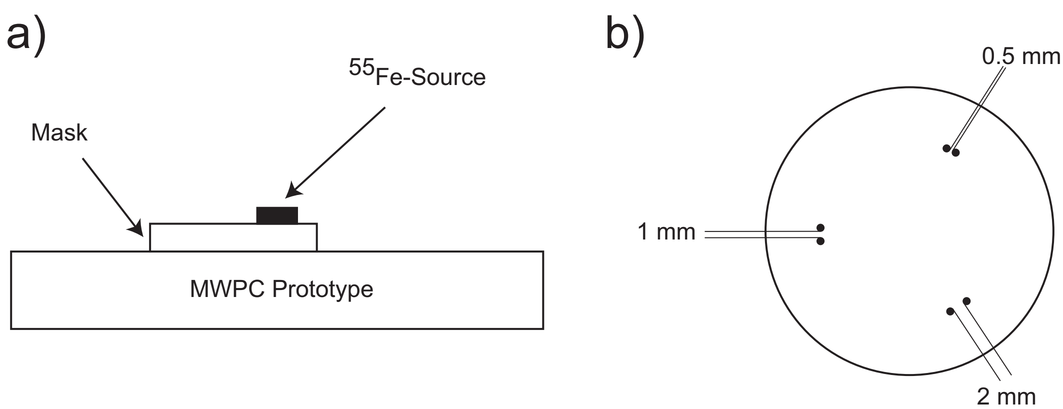


Figure IV.8: a) Determining the spatial resolution using a discriminator mask and a ^{55}Fe source. b) Detailed view of the discriminator mask.

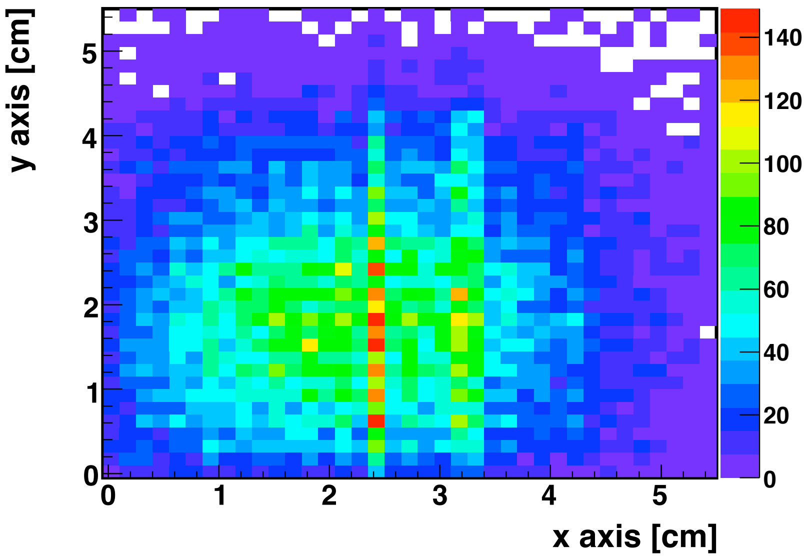


Figure IV.9: Reconstructed image using the pad response function of a ^{55}Fe source collimated by two wholes with a diameter of 1 mm each. The distance between the center of the holes is 2 mm. It is not possible to distinguish between the contribution of the two holes. The lines in the middle is regarded as some measurement artefact.

---

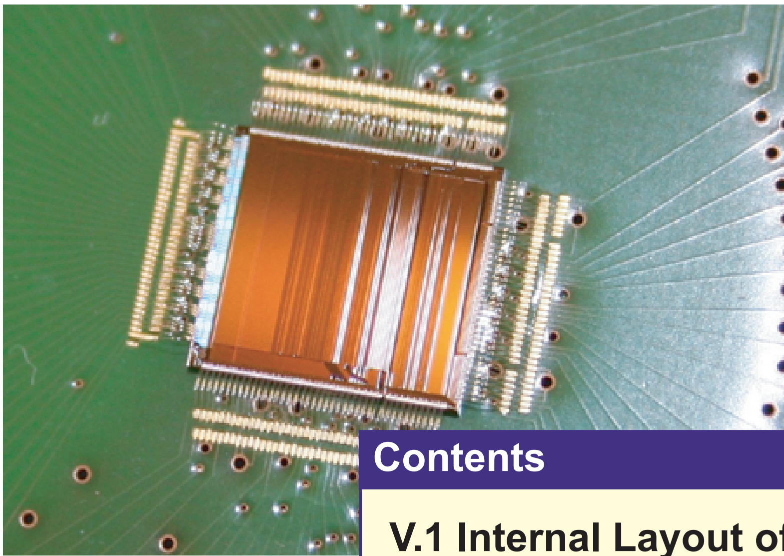
has been used. Thus, the first step in this method is to determine the parameter s. This has been done as described in III.4.2, using a cosmic ray data set with 120,000 events. The results are (for a pad width of $W = 0.3$ cm)

$$s_x = 0.33 \text{ cm} \qquad s_y = 0.30 \text{ cm} \tag{IV.1}$$

for the x- or y-plane respectively. The error of these quantities is due to the measured signals on the pads and thus depends on the gain of the preamplifiers and on the accuracy of the ADCs. As it is very hard if not impossible to obtain quantitative results for these errors, the overall error of s_x and s_y are neglected here. However it is clear that this error has a negative effect on the spatial resolution. In the next step, a measurement with the ^{55}Fe source discriminated by two holes with distance 2 mm and 55,000 events has been reconstructed. The result is shown in Figure IV.9. The lines in the middle of the plot is regarded as an error of the data acquisition system. However, neglecting this, it is still not possible to determine the position of the two holes in the collimation mask. Thus, the spatial resolution of the msPET prototype could not be determined.

CHAPTER V

The n-XYTER Readout Chip



Contents

n-XYTER stands for **n**eutron-**X**-**Y**-**T**ime-**E**nergy **R**ead-out. The n-XYTER is an ASIC[1] readout chip which has been developed within the framework of the european project DETNI. Although it has been intended to use for the detection of neutrons, its specifications perfectly match the needs of a readout chip for multi wire proportional chambers. This is because neutrons as well as gamma rays emitted by electron positron annihilation are statistical in nature. All 128 channels on the chip are purely data driven and therefore no external trigger is needed. Containing an integrated timestamp generator with a resolution up to 1 ns [Brogna06] the detection of coincidences in PET is easy to realize.
Still, as there is no readout board available at the moment, a prototype of such a board has been developed based on the DL701 LogicBox created in Heidelberg. This box contains a Xilinx Spartan 3 FPGA[2] which is able to process the n-XYTER's data and perform slow control via I^2C[3] as well as a USB interface to communicate with any standard PC. To convert the analog peak height output signal of the n-XYTER, a MAX106 flash ADC[4] has been used.

V.1 Internal Layout of the n-XYTER

The n-XYTER consists of an analog front-end part that performs pulse height detection and time stamp creation. It is followed by a clock synchronous back end that is responsible for the derandomization of the recorded data. The complete analog part as well as some fraction of the back end are repeated for every of the 128 channels. Furthermore, the n-XYTER contains several units for slow control via I^2C, digital to analog converters (DACs) to set trigger thresholds as well as a clock driven timestamp generator and a token manager. An overview the n-XYTER's architecture is given in Figure V.1.
The n-XYTER can be driven by two independent clocks, *clk256A* and *clk256B*. They are applied via different input pads. *clk256A* is the main clock for the chip. It drives the counter which creates the first 12 MSBs[5] of the timestamp (MSB first) and is used for the synchronisation of test signals. By default, it also generates the readout-clocks *clk32o* and *clk128o*. The secondary clock, *clk256B* can be used for

[1]ASIC stands for **A**pplication **S**pecific **I**ntegrated **C**ircuit.
[2]FPGA stands for **F**ield **P**rogrammable **G**ate **A**rray
[3]I^2C stands for **I**nter **I**ntegrated **C**ircuit Bus
[4]ADC stands for **A**nalog **D**igital **C**converter.
[5]MSB stands for **M**ost **S**ignificant **B**it

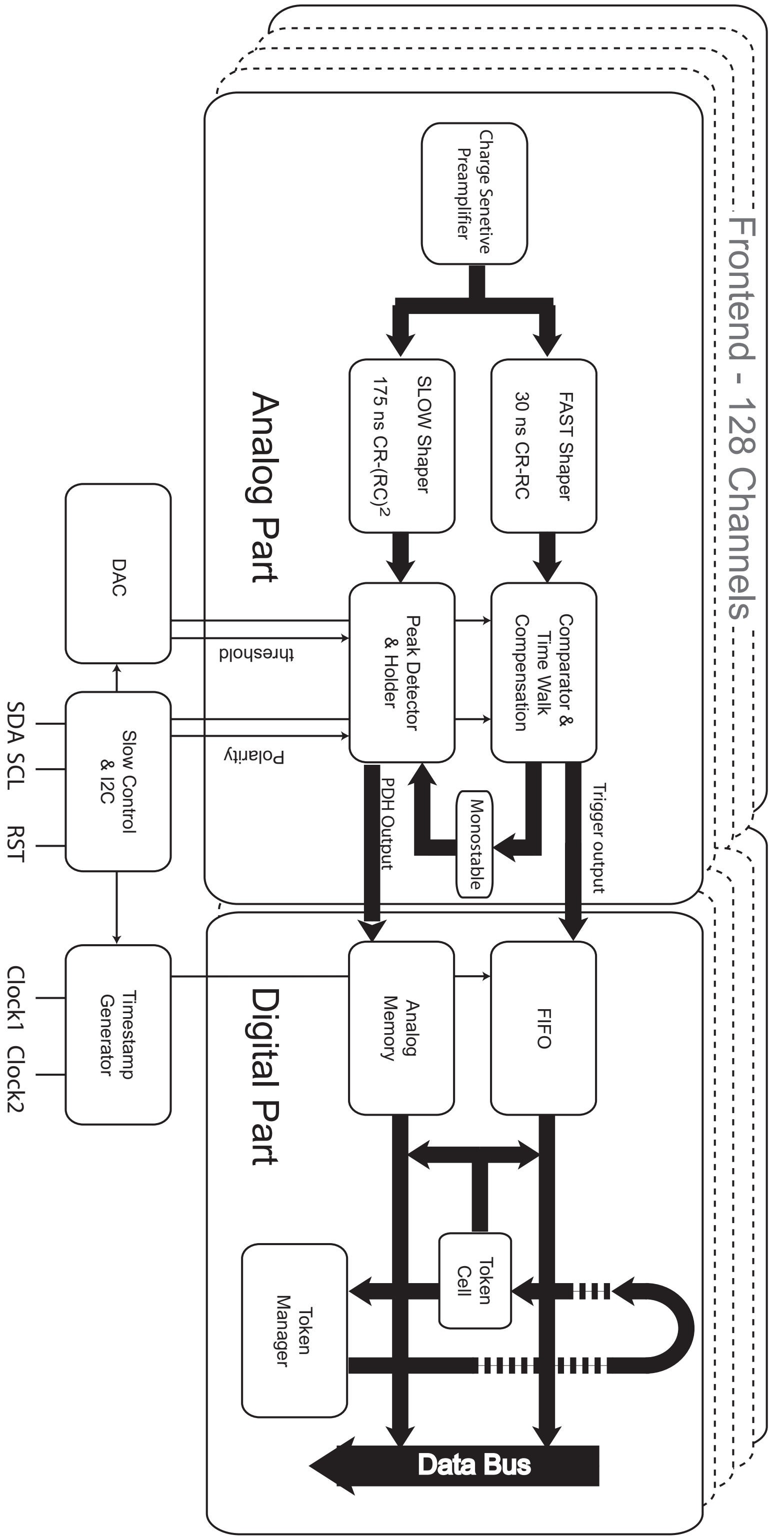


Figure V.1: The n-XYTER's chip architecture [Brogna07].

readout clock generation as well. Furthermore, it can be used to generate the LSB[6] of the timestamp. Originally, *clk256B* has been indended as a failback for the first design. In case *clk256A* failed, it could take over its function [Schmidt07].

V.1.1 Analog Frontend

The layout of the analog front-end is shown in Figure V.2. As each channel features its own analog front-end, this circuit is present 128 times within the chip. The front-end begins with a charge sensitive preamplifier which is based on a classical folded-CASCADE architecture [Brogna06]. To keep noise low, a n-tpye MOSFET is used. Finally, the signal is differentiated (by a Cr-Rd passive network) within the preamplifier.
In the next step, the signal is split up into two parts. One is fed into the fast- the other one into the slow-shaper. Within the fast-shaper, a fast uni-polar pulse is generated by a 30 ns peaking-time shaper. This shaper is followed by a comparator and a time-walk compensation circuit. The comparator compares the signal to a fixed value generated by DACs[7] and thus acts as a trigger for the signal. The time-walk compensation is needed to obtain an amplitude independent timestamp. The second part of the signal is processed in the slow-shaper, which has a peaking time of 175 ns. The pulse is then transferred to a peak detector and holder circuit to determine the pulse height.
As the n-XYTER allows direct access to the fast- and slow-shaper outputs of the test channel via external pads, these signals can be visualized with an oscilloscope.

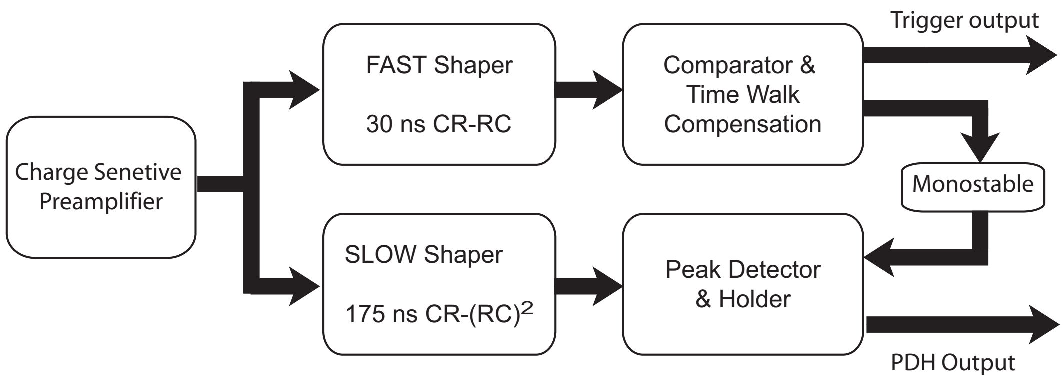


Figure V.2: The Analog Front-End Part of the n-XYTER.

[6] LSB stands for **L**east **S**ignificant **B**it
[7] DAC stands for **D**igital **A**nalog **C**onverter

V.1.2 Digital Part

Following the analog front-end, the output of the fast shaper is used to generate a timestamp. This is done using a 12 bit counter, called timestamp generator, driven by *clk256A*. In addition *clk256B* can be used to create the LSB of the timestamp (according to register settings). The timestamp is then stored in a FIFO[8] memory, triggered by the fast-shaper signal via a SR latch. Meanwhile, the slow signal, i.e. the output of the peak detection and holder circuit is stored within an analog memory[9]. The further processing is organised as a token ring readout. This means, a token manager generates a so-called token, which asynchronously passes from channel to channel. The token is generated only if at least one of the FIFOs has data in it. The output of the FIFO and the analog memory for each channel are controlled by a token cell. If there is data available on a channel when the token arrives, the data readout is initiated and the stored analog and digital values are transferred to the data bus. In the next readout clock cycle, the token continues to the next channel containing data and finally returns to the token manager. At the end of the data bus, the digital data is demultiplexed and converted to the LVDS[10] signaling standard. The analog signal is also converted to a differential standard. Finally, these signals are set on the output pads.

V.1.3 Slow Control and Test Channel

To control the overall behaviour of the n-XYTER (e.g. DAC settings, mask channels, etc.) the chip possesses 45 registers, which can all be addressed via I^2C. The first 16 registers (i.e. register 0 to 15) are used to mask every of the 128 input channels. Registers 16 to 31 set up the front-end DAC values used by the comparator circuit. Register Vth (18) is of special interest. Vth sets the comparator threshold voltage thus defining the trigger threshold for the input channels. Registers 32 and 33 are actual configuration registers and will be treated in more detail in the following. There are also some status registers that contain information of the missing token counter (register 36 to 37) and the overflow counter (34 to 35). Theses registers are read-only and automatically reset after each readout. Moreover, there are some

[8] FIFO stands for **F**irst **I**n **F**irst **O**ut. This means that the element first stored in the memory is also the first to appear on the output, like in a queue. The opposite principle is called LIFO meaning Last in First Out.

[9] An analog memory is basically made of a capacitors and switches. For details, see [Leo87]

[10] LVDS stands for **L**ow **V**oltage **D**ifferential **S**ignaling.

registers to control the clock (43 to 45), the test pulse (38) and the test trigger delay (39) as well as one register (42) to trim the DAC. Table V.1 gives an overview of all registers.

The n-XYTER provides several test features for the analog and digital part. In

Table V.1: The n-XYTERs I^2C registers. [Brogna07]

I^2C Address	Register
0 - 15	mask registers (external)
16 -31	reserved for front-end DACs (external)
32 - 33	configuration register
34 - 35	overflow counter, readout only, reset after readout
36 - 37	missing token counter, readout only, reset after readout
38	test pulse delay register
39	test trigger delay register
40 - 41	Spare
42	trim DAC register (bit 0-4)
43	clock delay register No. 1
44	clock delay register No. 2
45	clock delay register No. 3

test mode 0 (full test mode), a pulse is generated within the calibration block and submitted to the analog channels via a 100 pF capacitor. The pulse is applied to groups of every 4th channel, according to the settings in register 32. However, in a chosen group, it still can be decided if the test pulse is applied to a single channel using the mask register. The pulse height - and thus the amount of injected charge - is set in a register called 'cal' (24). Furthermore, the polarity of the test pulse can also be set in register 32.

In test mode 1, a test trigger is transferred directly to the digital part of the n-XYTER allowing to test the timestamp generation. The test trigger has to be applied externally as shown in Figure V.3. Furthermore, the test pulse and test trigger signals can be delayed if needed.

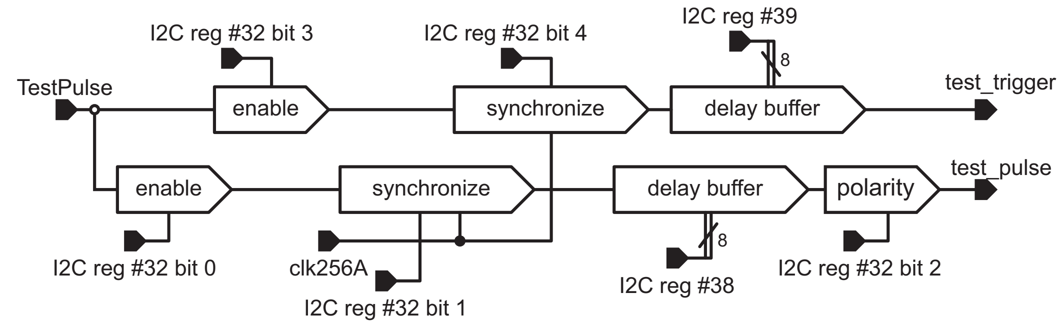


Figure V.3: The creation of the test_pulse and test_trigger signals from the TestPulse input pad [Brogna07].

V.2 The n-XYTER's Digital Data Output Format

The n-XYTER generates two readout clocks either from *clk256A* or *clk256B* (according to the settings of bit 3 in I^{2}C-register No. 33). They are called *clk32o* and *clk128o*. These clocks are generated by dividing the chips main clock. If the chip is operated at the standard frequency of 256 Mhz, they are running at 32 Mhz (*clk32o*) or 128 Mhz (*clk128o*). Each pulse recorded is encoded in a 25 bit package (see Figure V.2). The timestamp is 14 bits, the channel id 7 bits wide. The package also contains a data-valid bit, as well as a pile-up, an overflow and a parity bit. These 25 bits are split up into 4 packages of 8 bit width which are transmitted on consecutive cycles of *clk128o*. Furthermore, they are transmitted in gray code[11]. The analog pulse

Table V.2: n-XYTER digital data output. The four packages of 8 bits are transmitted consequitively. dv means data-valid. [Brogna07]

	7	6	5	4	3	2	1	0
0	DV	TS13	TS12	TS11	TS10	TS9	TS8	TS7
1	0	TS6	TS5	TS4	TS3	TS2	TS1	TS0
2	0	ID6	ID5	ID4	ID3	ID2	ID1	ID0
3	0	0	0	0	0	PileUp	OverF	Parity

[11] Gray code is an alternative binary decoding scheme for numbers. In gray code, two successive values differ by only one bit. This is especially useful for error correction.

height information is provided on the output simultaneous to the transmission of the 3rd data package. The detailed timing can be seen in Figure V.4. The packages are transmitted synchronous to *clk128o*. *clk32o* is used to determine the start of a each package (V.2).
However, actual measurements at the GSI[12] have shown that *clk128o* is shifted by 180 degrees with respect to the Figure V.4 [Lalik07]. This is important to assure proper timing when merging digital and analog data.

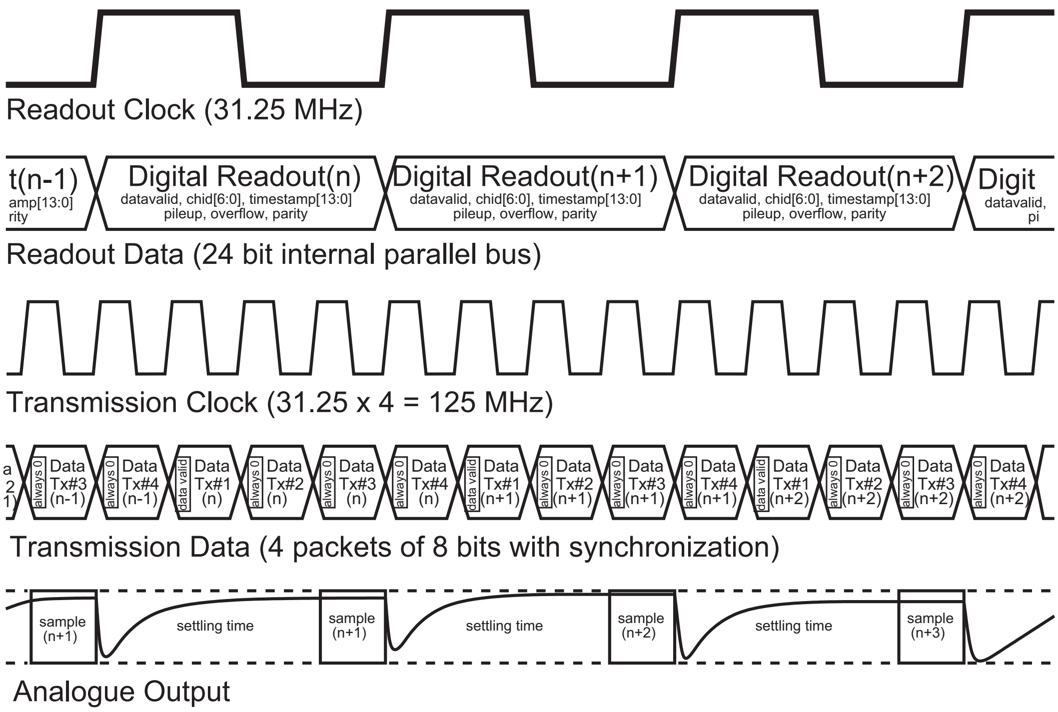
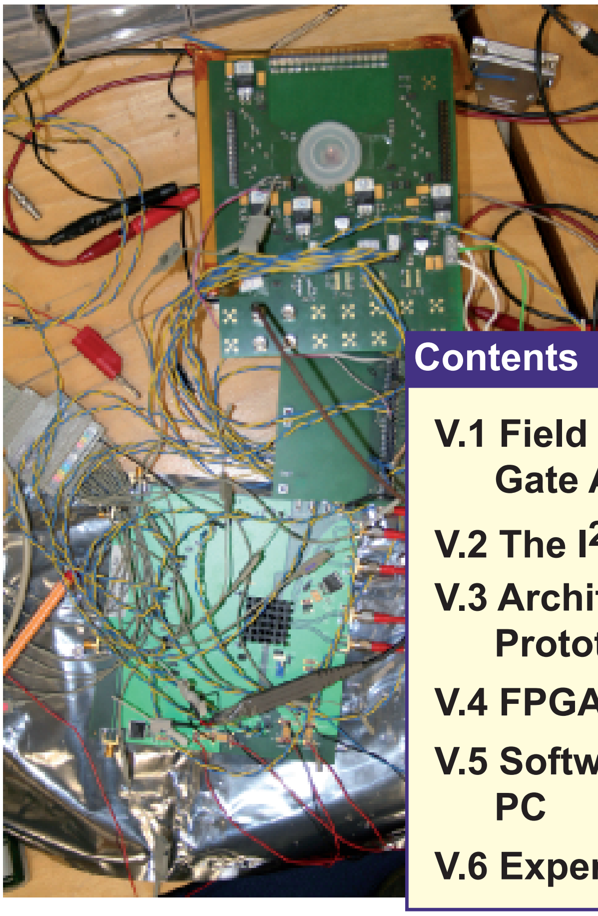


Figure V.4: Data scheme of the n-XYTER's digital output. Readout clock refers to *clk32o*, transmission clock to *clk128o* [Brogna07].

[12]GSI stands for **G**esellschaft für **S**chwer**i**onenphysik and is located near Darmstadt.

CHAPTER VI

A Readout Board Prototype



Contents

VI.1 Field Programmable Gate Arrays

Field Programmable Gate Arrays are user programmable devices that allow the implementation of complex logic functions. They are able to process multiple signals at a time and are thus often used for data processing at high rates [Maxfield03]. This makes them an ideal core for a dedicated n-XYTER readout board. In the following sections, their historical development, architecture as well as the design process will be described. Special attention is payed to the Xilinx Spartan 3 device that is used here.

VI.1.1 A Short History of Programmable Logic Devices

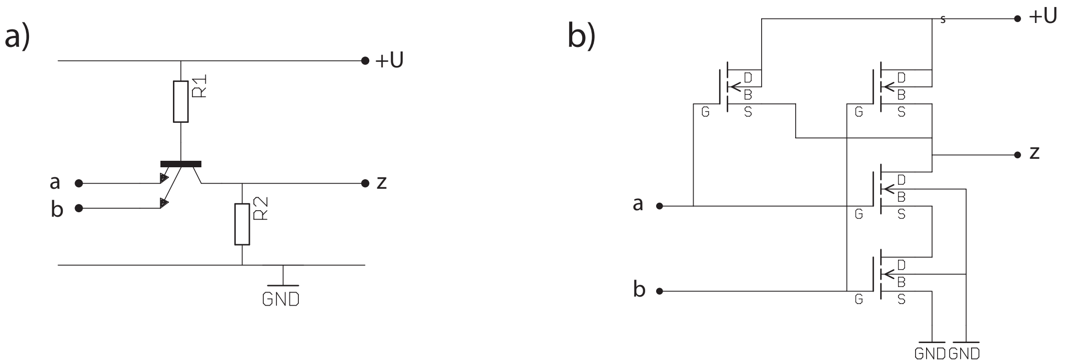


Figure VI.1: a) shows a TTL AND gate implemented using a multi emitter bipolar junction transistor. b) shows a NAND gate implemented in CMOS logic.

The history of modern digital electronics started with the invention of the transistor in 1947[1]. Only a short time later, the bipolar junction transistor was invented. Connecting several of these transistors, it is possible to create digital logic gates [Tietze83]. According to the way, in which the transistors are connected, the gates are classified either transistor-transistor logic (TTL) or emitter coupled logic (ECL). Logic gates build in ECL are faster than their TTL counterparts but also consume more power [Horowitz89].

In 1962, the metal-oxide semiconductor field-effect transistor (MOSFET) was invented. As ordinary transistors, MOSFETs are available in two polarities: p- and n-type. Combining these two in a logic gate, the resulting circuit is classified as

[1]It is assumed that the reader is familiar with some basic electronic elements such as transistors, MOSFETs, diodes, etc. If not, [Horowitz89] is a good reference.

complementary metal-oxide semiconductor (CMOS). The main advantage of CMOS is its low power consumption in non-switching states [Tietze83]. In Figure VI.1 the implementation of a TTL AND and a CMOS NAND gate are shown as examples. With the invention of lithographic techniques it became possible to create integrated circuit (IC) chips containing a number of logic gates. For example, the 7400 device by Texas Instruments contained four two-input NAND gates in a single chip. Combining different chips, it is possible to realize more "complex" logic functions.
The next step towards a more sophisticated, user programmable logic device was the development of the programmable read-only memory (PROM) in the 1970s. Although these devices were initially intended to act as computer memories, they can be used to implement simple logic functions such as lookup tables or state machines. To understand how a PROM acts as a logic device, it can be considered as combination of a fixed AND and a programmable OR gate [Maxfield03] as shown in Figure VI.2. The address lines act as inputs of the logic function whereas the

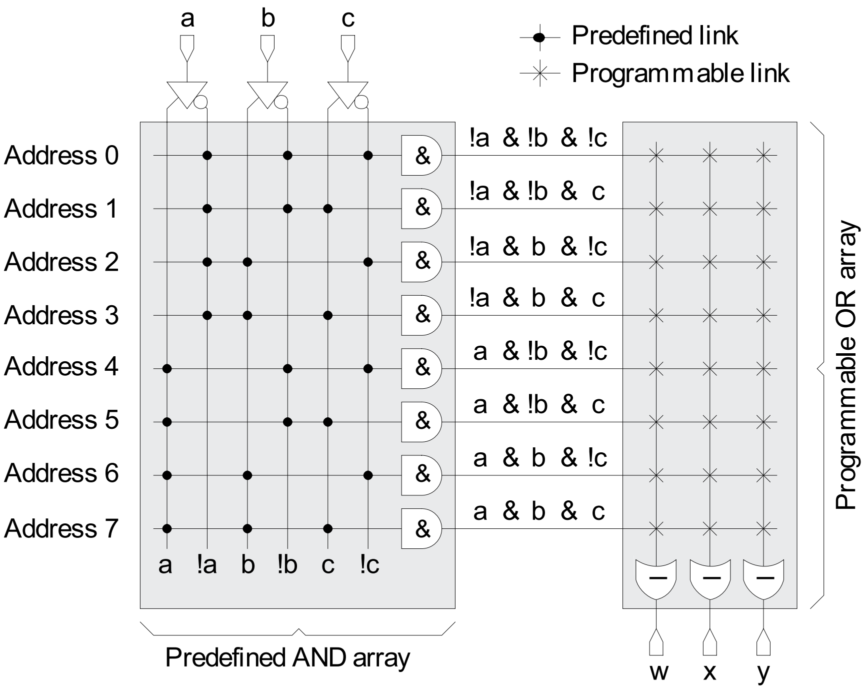
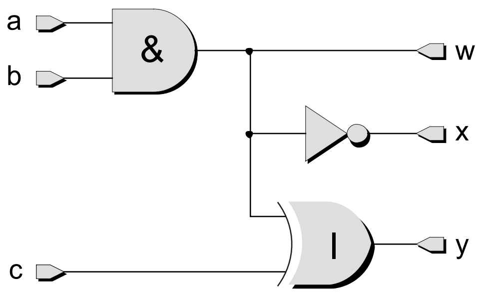


Figure VI.2: The principle of a PROM acting as logic device. Note that this is only a model and not the real physical architecture of a PROM. [Maxfield03]

data lines represent the result. Thus, a 3-input, 3-output PROM for example can be used to implement every combinatorial function with 3 inputs and 3 outputs. A simple example is given in VI.3. However, PROMs are very slow devices. Typical



a	b	c	w	x	y
0	0	0	0	1	0
0	0	1	0	1	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	0	1	0
1	0	1	0	1	1
1	1	0	1	0	1
1	1	1	1	0	0

Figure VI.3: Example of a simple logic circuit expressed as a lookup table that could be implemented using a PROM. [Maxfield03]

access times are in the order of 40 ns [Zeidman02].
Subsequently, more devices based on the same principle were developed. Programmable logic arrays (PLAs) have both a programmable AND and OR gate. However, as signals take longer passing programmed links than fixed ones, these devices still were relatively slow. To overcome these limitations, programmable array logic (PAL) are made of a programmable AND but fixed OR gate.
As all these devices only allow the implementation of more or less simple logic function (and are thus called simple programmable logic devices or SPLD short), several of these SPLDs were combined using programmable links. The programmable links can be realized in several ways. One possibility is to use fusible-link technology. Here, in the initial state, all connections are made using fuses. In the programming process, the fuses that are not needed in the design are burnt out. The drawback of this technology is, that once programmed, the design cannot be changed any more. Using other techniques like SRAM[2] or EEPROM[3] technology, it is possible to create devices that can be programmed multiple times. For more details on this see [Zeidman02]. The resulting devices are then called complex PLD or CPLD short. The fact that all PLDs are user programmable makes prototyping very easy. However, compared to the complexity of application specific integrated circuits (ASIC), they are still very simple devices. ASICs are built up of defined cells containing a number of transistors. The user is free to interconnect these in almost every combination. This allows the implementation of very complex circuits. The final device is produced in a factory using the same lithographic techniques as for integrated cir-

[2]SRAM stands for **S**tatic **R**andom **A**ccess **M**emory
[3]EEPROM stands for **E**lectrically **E**rasable **PROM**

cuits. Thus, ASICs are very fast but are not reconfigurable and their development is expensive.[4]
To close the gap between these two technologies, Xilinx Inc. developed the first FPGA which was commercially available in 1982.

VI.1.2 FPGA Architecture

A FPGA consists of three basic[5] elements: Input Output Buffer (IOB), Configurable Logic Blocks (CLB) and structures to interconnect these elements.
The Input Output Buffer connect the FPGA to the outside world. They are usually able to convert a wide range of input signals (single- as well as double-ended) to the FPGA's internal signalling standard and vice versa. Configurable Logic Blocks are units with a given number of input and output lines that are able to perform a given logic operation. They are realized in different ways, for example by use of a lookup table (LUT). Using the internal interconnection structure it is possible to to connect IOBs and one or more CLBs to generate complex logic functions. The important feature of FPGAs is that the interconnection as well as the LUTs can be programmed by the user, on most FGPAs even multiple times. This makes it possible to test different designs very easily and was in fact a revolution for product design.

VI.1.3 Hardware Description Languages

Early programmable logic devices only contained a small number of logic gates. To program them, engineers created schematics of the logic function they wanted to implement, mostly by using pencil and paper. From these schematics, they created text files which described the connections to be cut or made within the PLD. Finally these textfiles were entered manually into the programming device. As PLDs became more complex, this was no longer possible and more sophisticated design tools were needed. Hardware description languages are the solutions for these problems. These languages allow the description of the hardware in text form and thus independently from the real components. This makes it possible to describe even very large design

[4]Especially, the so-called non-recurring engineering costs are important. These are the costs that are needed to produce the ASIC in a factory and are independent of the design.

[5]There are many other elements, but as all of them are manufacturer dependent they are not mentioned in this general discussion.

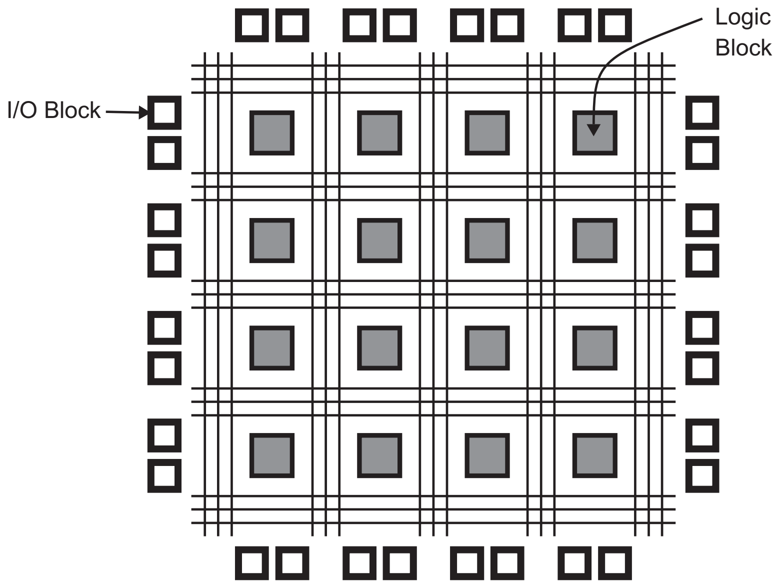


Figure VI.4: Typical architecture of a FPGA device. The spacial arrangement of the component may differ from company to company. [Xilinx07]

with a maintainable expenditure of time. One of the first HDL was the advanced boolean expression language (ABEL). It allows the user to enter the desired behaviour in form of boolean equations and truth tables. Furthermore, ABEL already contained minimization algorithms to reduce the number of logic gates needed to realize the design.
The two most common HDLs today are the VHSIC[6] hardware description language (VHDL) and Verilog. VHDL was developed on behalf of the US Department of Defense in the late 1980s. Verilog was developed by Phil Moorby and Prabhu Goel in 1985 at Automated Integrated Design Systems. Today, it is very popular in north America whereas VHDL is wide spread in Europe. There are also other approaches, e.g. SystemC. For details see [Maxfield03].

Abstraction Layers

Compared to the first, manual design process, ABEL already provided some kind of abstraction from the structural (i.e the logic gates) to a functional (i.e. boolean equations) level. Increasing this level even further leads to more complex functional or even behaviorial (i.e. algorithmic) descriptions of the design. Clearly, the more

[6]VHSIC stands for **V**ery-**H**igh-**S**peed **I**ntegrated **C**ircuits.

abstract the description becomes, the easier it is to describe complex designs. The next step of the functional level is the so-called register transfer level (RTL). On this level, the design is described by means of registers that are linked together by combinatorical logic. For example, consider a register $REGA$ that is either set to the value of a register $REGB$ or to zero, according to the value on a third register, $REGC$. Furthermore, we assume that this process is controlled by a global clock signal CLK. The RTL description of such a scenario is

```
process(CLK)
begin
        if( REGC == '1') then
                REGA = REGB;
        else
                REGA = '0';
        end if;
end;
```

The highest level of abstraction is the representation of the design on a behaviorial level. To illustrate this, a simple example is the equation

$$A = B * C + 4 * D,$$

if B, C, D are input and A is a output line of a logic circuit and all these lines are 4 bit wide. On a structural level, a number of gates is needed to perform such arithmetic, but on the behavioural level it can be described by only one line of text.

HDL Design Flow

Using HDLs, there is a standard design flow consisting of four different steps. In the first step, the designated functionality is implemented using the HDL. This step is followed by a behavioural simulation (also known as RTL simulation). The simulation is an important step since it allows to find errors by looking at every signal that is present in the design. Once the design is downloaded onto the chip, this can only be done using specialized tools. The simulation step is repeated until the design works properly which normally takes several iterations. The next step in the design flow is to synthesise the design. This produces a net list or gate-level net list. This is a file containing information about the logic elements used and their interconnections. The net list is FPGA specific as every FPGA has different components that may be used in the design. Using the net list, a functional simulation of the design is possible. Thereafter, the place and route step creates an actual implementation of the design. In this step, the components get their location within the FPGA and the wires between them are set. Therefore, a timing simulation is possible as all wires (and their lengths) are fixed. If the timing simulation is successful, a binary bitfile is created in a final step. This bitfile is then downloaded into the FPGA. The whole design process is summarized in fig. VI.5.

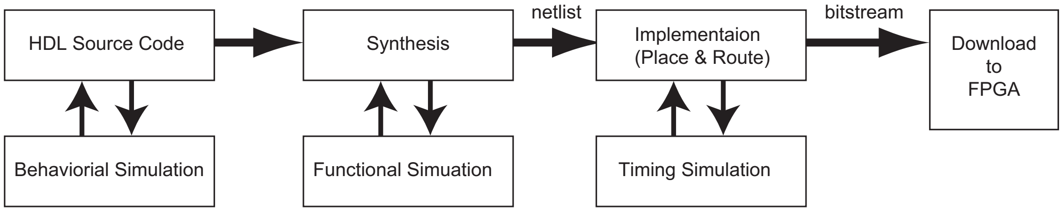


Figure VI.5: A Typical HDL design flow.

VI.1.4 Xilinx Spartan 3

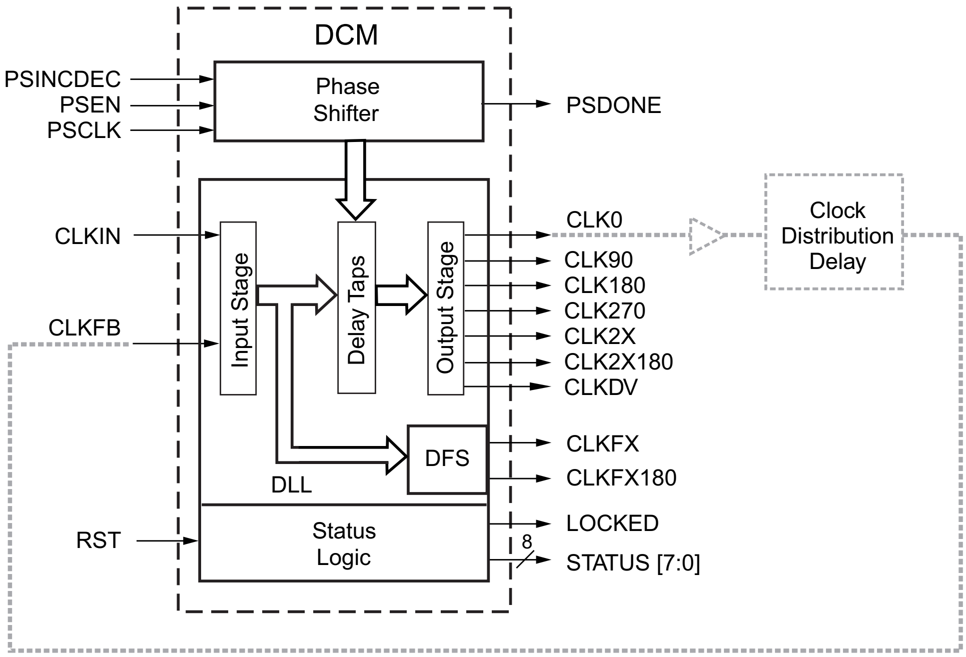


Figure VI.6: A digital clock manager within the Spartan 3.[Xilinx07].

The Spartan 3 is a FPGA designed by Xilinx Inc. and was introduced in 2003. In addition to logic blocks and input output buffer, it contains some additional components. There is block RAM distributed over the chip that is able to store up to 18 kbit of data. The memory can be grouped together and used in several configurations, for example as FIFO. Furthermore, the Spartan 3 possesses four digital clock managers (DCM). These units are able to create different clock frequencies from the chip's main clock. Figure VI.6 shows an overview. DCMs can also shift the clock's phase by an arbitrary angle and can be used to perform clock skew elimination. This means that they compensate random phase shifts. This can happen because clock signals arrive at different times at distant points in the FPGA due to small variations in the path lengths. Here, the DCM is important to create the clock signals for the n-XYTER. A schematic view of a Spartan 3 DCM is shown in Figure VI.7.

The input output blocks support 18 single ended and 8 double ended signaling

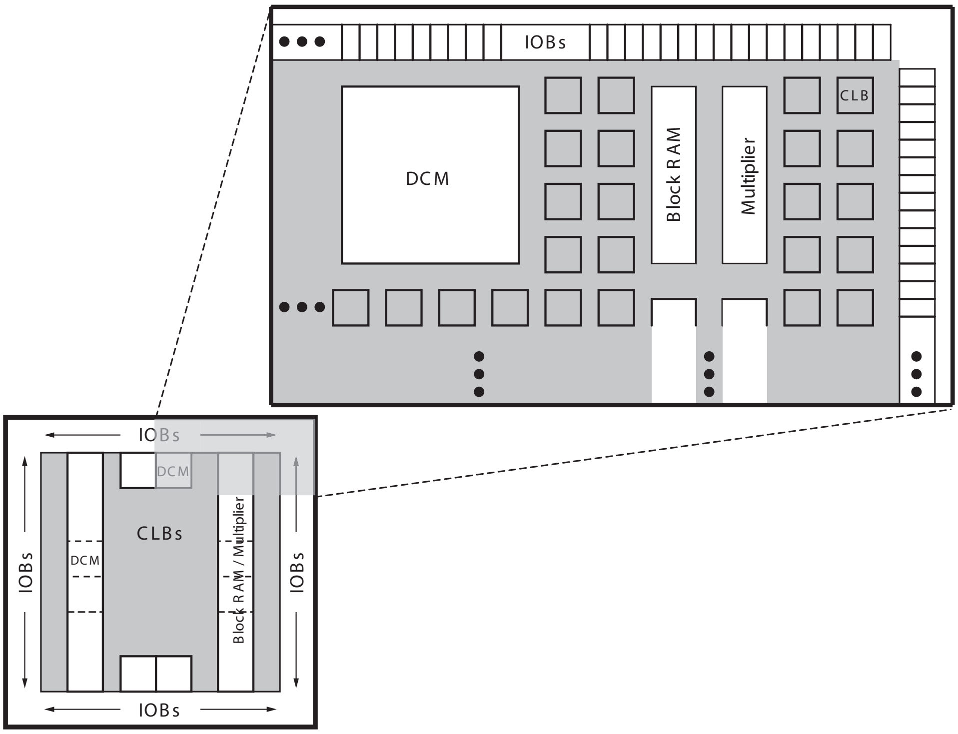


Figure VI.7: The architecture of the Xilinx Spartan 3 FPGA device. Beside IOBs and CLBs it also contains digital clock managers (DCMs) and 18 kbit of block ram distributed over the device. Furthermore, it contains some specialized units such as dedicated multipliers. [Xilinx07].

standards. Most important, the Spartan 3 supports LVDS to interfere with the n-XYTER and LVPECL to address the ADC needed for the readout board.

The CLBs of the Spartan 3 are already quite complex. Each of them comprises four interconnected slices within the FPGA. They contain a number of elements such as two logic function generators, lookup tables, two storage elements, etc. This makes them able to realize a number of logic functions. In the prototype board, the Spartan 3 runs with a clock frequency of 100 MHz.

VI.2 The I^{2}C Bus

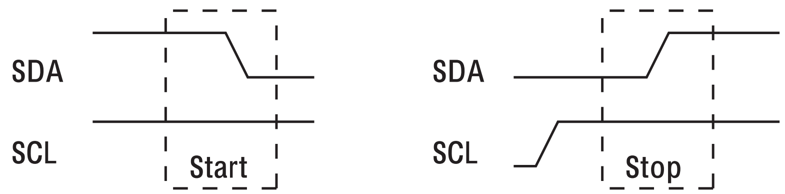


Figure VI.8: I^{2}C start and stop signals

The I^{2}C-Bus is a two wire, bidirectional serial bus that is used for slow control of the n-XYTER. It was developed by Phillips in the early 1980s [NXP07], originally to allow the communication of several integrated circuits within customer electronics. Thus its name **I**nter **I**ntegrated **C**ircuit Bus. The first standardized specification, version 1.0 was published in 1992. Beside the normal transmission speed (100 kbit/s, standard mode) the I^{2}C-Bus can also operate in a fast mode (400 kbit/s). Since version 2.0 even a high-speed mode which can transfer data with up to 3.4 Mbit/s is available.
The bus itself consists of a clock (SCL) and a data (SDA) line. In its idle state, both lines are pulled up to the (positive) supply voltage. This is defined as the logic '1' state. Devices connected to the bus can act either as transmitter, receiver or both. Furthermore they are divided into 'master' and 'slave' devices. A master is a device that initiates data transfer and generates a clock signal to permit data during the transfer. Slave is the general description for every device that is addressed by a master. The I^{2}C-Bus is a multi-master bus which means that more than one master device can be connected to the bus at a time. To actually start a data transfer, the master generates a special condition on the bus, a so-called 'start condition'. It is defined as high to low transition on the SDA line while the SCL line remains high. There also exists a stop condition to terminate the transfer, defined by a low to high transition on SDA while SCL is high[7], see Figure VI.8. After the start condition, the master sends a seven bit slave address. This address is followed by the read/write bit which allows to master to indicate whether it wants to read ('1') or write ('0')

[7]However, a new transfer following a previous one can be initiated by simply generating another start condition, a so-called restart, without sending a stop signal before.

data. The slave has to acknowledge the reception of these information. This is done by simple keeping the SDA line low for one clock cycle. After the acknowledge, the master can start transmitting or receiving data. The whole transfer is ended by a stop condition generated by the master. Figure VI.9 shows an example of a data transfer on a single master bus.

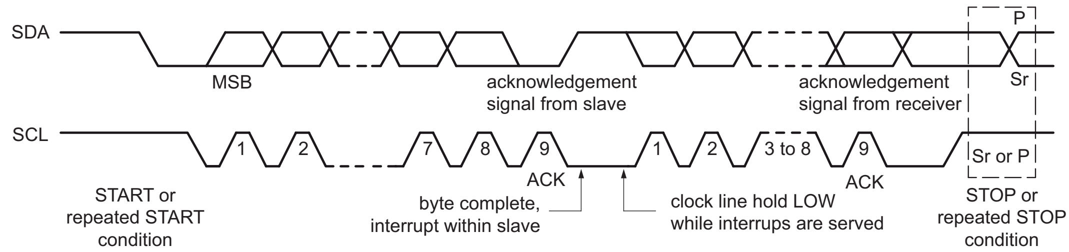


Figure VI.9: Example of a data transfer on the I^2C bus. [NXP07]

VI.3 Architecture of the Prototype

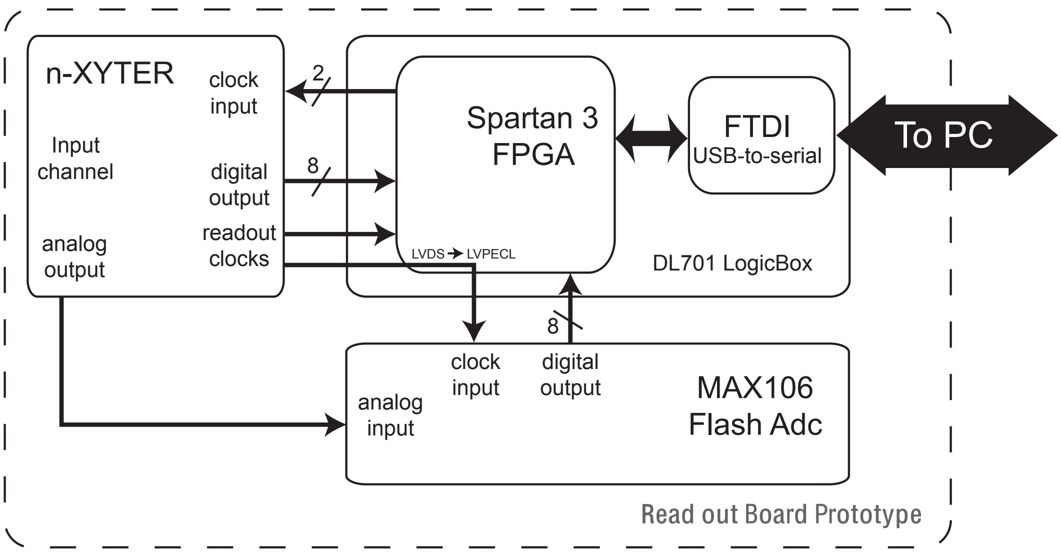


Figure VI.10: Schematic architecture of the read-out board including the connections of its several components.

One of the main tasks that was done in this thesis was the construction of a prototype read-out board for the n-XYTER chip. The schematic architecture of this board is shown in Figure VI.10. The board consists of two main components: The DL701

LogicBox (that contains a Spartan 3 FPGA) and a MAX106 ADC.
The ADC converts the differential analog output of the n-XYTER to an eight bit digital value. The Spartan 3 concanates this information with the n-XYTER's digital output. Finally, the data is transfered to a personal computer using a FTDI[8] parallel to USB converter chip. There, it is further analyzed using a C++ program.

VI.3.1 MAX106 Analog To Digital Converter

The MAX106 is a flash ADC which means that it continuously samples values from its analog input. This input is differential which is important, as the n-XYTER's analog output is also differential. The signalling standard of the MAX106's digital output is low voltage differential PECL[9]. The clock and reset inputs also operate with this standard. In the simplest mode of operation, the data is presented on the digital output exactly nine clock cycles after it appeared on the analog input. Using *clk32o* to clock the ADC, this easily allows to match the analog with the digital data. The clock signal is converted from LVDS to LVPECL using the Spartan 3.

VI.3.2 The DL701 LogicBox

The DL701 logic box has been designed at the electronics workshop in Heidelberg. It contains the Spartan 3 FPGA as well as a FTDI USB to parallel converter chip. The FTDI chip converts the serial USB data to an eight bit wide parallel stream. Along with the DL701, a VHDL module is provided which handles all the communication with the FTDI chip. It is called "DL701_CBUS" and allocates an address as well as a data register to the FPGA. A schematic of the DL701 and the internal data lines is shown in Figure VI.11. Within the n-XYTER readout board, the DL701 acts as central data processing unit.
The Spartan 3 onside the DL701 is able to handle the n-XYTER's LVDS signals using specified input and output buffer respectively. These buffers are called I/OBUFDS and are part of the Xilinx design library. As LVDS is a differential signalling standard, two pins are needed for each signal.

[8]FTDI stands for **F**uture **T**echnology **D**evices **I**nternational Ltd, see http://www.ftdi.com.
[9]PECL stands for **p**ositive **e**mitter **c**oupled **l**ogic

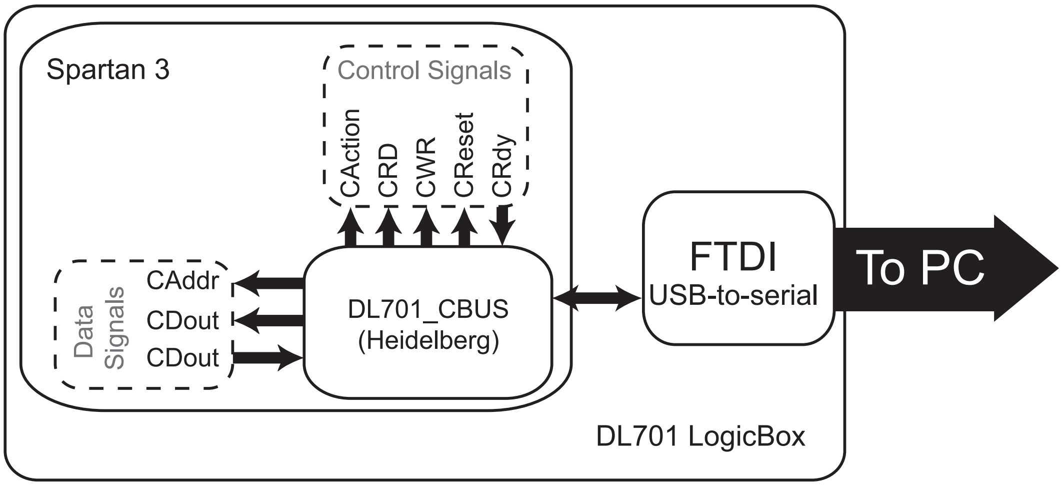


Figure VI.11: Overview of the DL701 logic box created in Heidelberg.

VI.4 Spartan 3 Program Design

The central tasks when building a read-out board from the components mentioned above, is to program the FPGA onside the DL701 LogicBox to process the (digital) signals from the n-XYTER as well as the output of the ADC. The FPGA has to decode the n-XYTER's data scheme and has to assure that the timing is right according to the n-XYTER's readout clocks. Furthermore, it has to decode the timestamp and channel information and gray-decode them. Finally, all information has to be sent to a PC using the FTDI USB chip. An I^2C-Controller also has to be implemented for slow control. An overview of the complete design is given in Figure VI.14.

VI.4.1 I^2C-Controller

The I^2C-Controller implemented here is based on a simple I^2C-Controller available as open source IP[10]-core at Opencores.org[11]. This controller is called *simple_i2c* and consists of two components, both containing a state-machine. The unit called *i2c_core* is responsible for creating the basic I^2C commands in terms of transitions of the SDA/SCL lines. Therefore a state machine is implemented that translates the input of the *cmd* line to the corresponding signals i.e. start, restart, stop, write and read. The state-machine is driven by the FPGAs main clock and each command is

[10]IP stands for intellectual Property. IP cores are pieces of HDL code that implement a certain functionality, e.g. VGA controller or encryption algorithms.

[11]Opencore.org is a website that collects IP cores which are under an open source license.

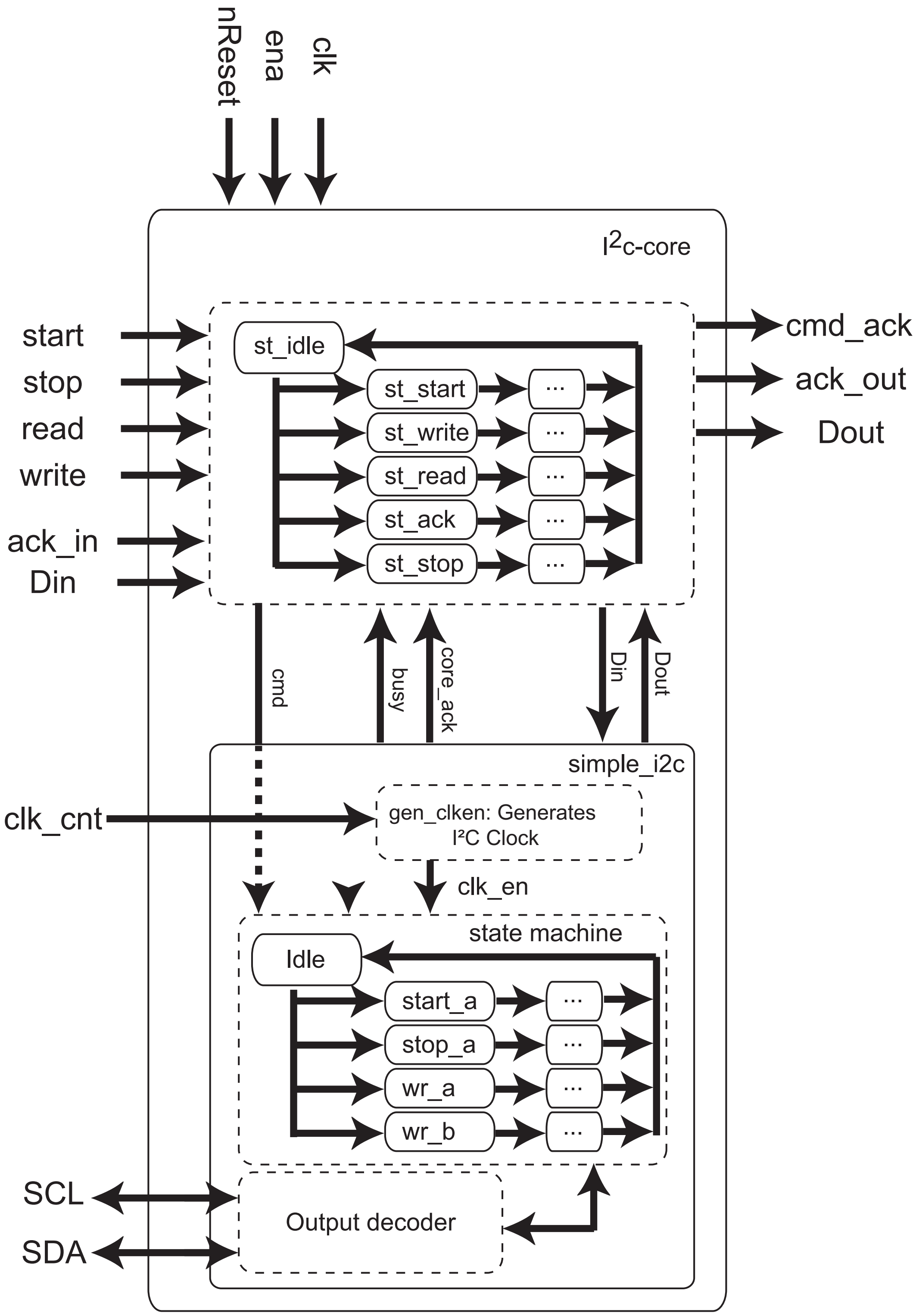


Figure VI.12: The complete architecture of the I^{2}C controller.

processed in four steps (A to D). However, to create I^2C-Signals at a lower frequency, the unit also contains a counter that acts as a clock divider.
The *i2c_core* unit is controlled by the *simple_i2c* unit itself. This unit contains input signals for start, stop, read and write as well as parallel data input and output ports (8 bits wide). Before the data is sent to the *i2c_core* it is serialized and vice versa for data read from the I^2C bus. The complete architecture of the controller is shown in Figure VI.12.
Finally, in a third layer, another state machine is needed to actually send the right combination of I^2C commands expected by the n-XYTER to read and write registers. This also is done in a state machine called "main_statemachine" within the file **I2C_Bus.vhd** (see Figure VI.13 b). As usual, at the beginning, the state machine is in the "idle" state. According to the input line *cmd*, the state machine branches to either state "rd1" or "wr1". In both cases, the first step is to send the n-XYTER's slave address followed by a write byte. Then, the address given at the "reg_addr" line is transmitted to the n-XYTER. If the task is to write data to the register, the next step simply is to write the 8 bit data given on the "Din" line and send a stop condition to complete the transfer. If a register value is to be read, the state machine will again send the slave address, followed by a read-bit. Then, the data send from the n-XYTER is read and put onto the "Dout" line. In the last step, the state machine branches to a "done" state and puts a "1" on the "clr_cmd_out" line to indicate that the task is done. In case of a read command, data is available on "Dout". If, during this process, the I^2C controller does not receive all acknowledge bits from the slave the "error" line is also set to "1". The whole read and write cycle is schematically shown in Figure VI.13 a).

VI.4.2 The DL701_CBus unit

The "DL701_CBus" unit is a precompiled NGC[12] module which manages the communication with the FTDI USB to parallel converter. It has generously been provided by Peter von Walter[13]. On FPGA side, the unit has a "CAction" bit, which is set high for one clock cycle after each valid USB command. The "CWR" and "CRD" line are high (also for one clock cyle) if a read or write command is sent. The "CAddr", "CDout" and "CDin" lines are all 32 bits wide. "CAddr" and "CD-

[12]NGC stands for **N**etlist **G**ates **C**onstants.

[13]Peter von Walter is the head of the electronic workshop at the physikalisches Institut in Heidelberg. His email address is vwalter@physi.uni-heidelberg.de

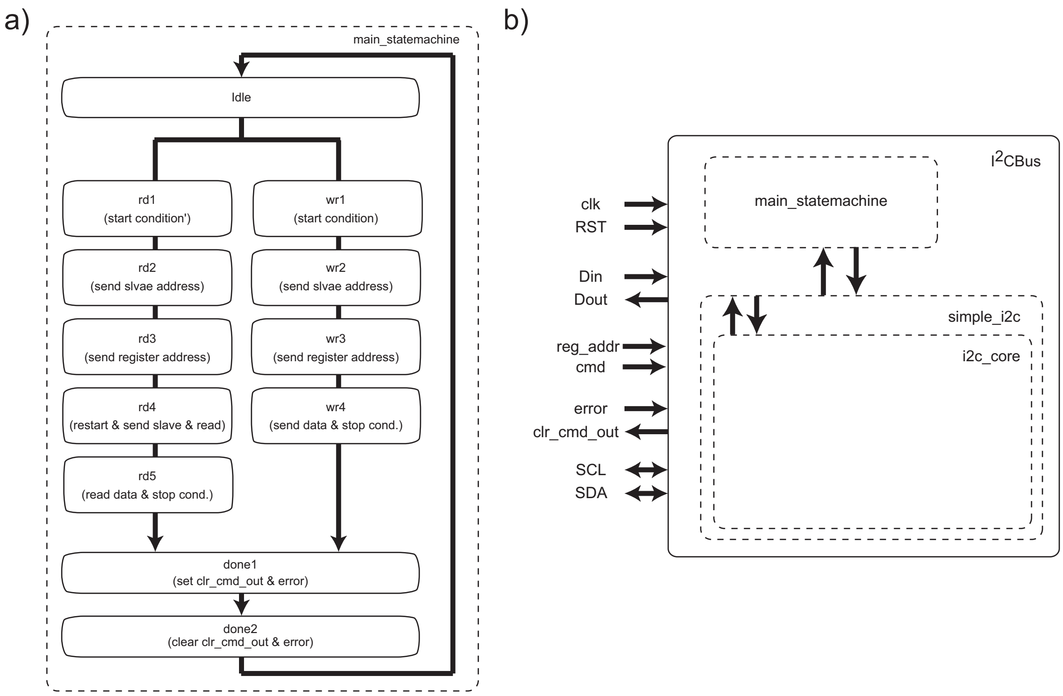


Figure VI.13: a) shows the state machine controlling the read and write operations of the I^{2}C controller. b) shows the data lines and internal architecture of the I^{2}C controller.

out" can be set from PC side. "CDin" is used to send data off the FPGA to the readout PC. The "CRdy" line can be used to pause the data transfer if the FPGA needs some time after a read or write request. Then, the FGPA can set "CRdy" to low for as long as it needs to complete its operation. All connection lines of the DL701_CBus module are also shown in Figure VI.11.

VI.4.3 Data Decoding

As mentioned above, the n-XYTER uses two read-out clocks to transmit data off the chip (see Figure V.4). In the FPGA, the decoding of these data packages is done in the unit "nXYTER_decode". As input, *clk128o*, *clk32o* and the 8 data lines are needed. The unit saves every data package transmitted with *clk128o*. When a low to high transition of *clk32o* is sensed, another single data package is saved and in the consecutive cycle of *clk128o* the last four data packages are put together. The timestamp as well as channel information are extracted. These two values are decoded from gray code to standard binary by a subunit called "gray_decode"[14].

[14]This unit works exactly as is recommended in the n-XYTER Reference Manual, Chapter 7

Alignment of Analog and Digital Data

When the digital data is merged with the information from the ADC, it is essential to assure that it is matched with the proper analog value. The ADC used here provides the converted value nine clock cycles after it was available at the analog input. Therefore, the digital output from the n-XYTER must be delayed. This is done using a shift register. Every cycle of *clk128o*, the values are shifted by one position. This is done no matter whether new valid data is available or not. After nine cycles, the analog and digital data are conflated. Finally the information is put on the output lines "ts_out", "id_out" and "adc_out". If the data was valid, the output line "dv" is set to "high" for one cycle of *clk128o*. This allows the main program to distinguish whether valid data is available or not.

VI.4.4 The Main Program

The term "main program" refers here to the top most layer of the VHDL code which interconnects all units described above. This means in particular that the data flow from and off the n-XYTER and the readout PC respectively has to be managed. For slow control the I^2C controller is connected to the DL701_CBus unit, which manages USB data transfer. The "reg_addr" input is connected to the 8 LSBs of the "CAddr" line. The same is done with "Din" (connected to CDout) and "Dout" (connected to CDin). The read or write process is started by low to high transitions on the "RWD" or "CWD" line[15]. Furthermore, the I^2C controller only starts its work, if CAddr < 255 to assure that there is no conflict with the data readout.
For data readout from the n-XYTER things are unfortunately not that easy. As the rate of data from the n-XYTER can be very high (they are transmitted at the frequency of *clk128o*), it is necessary to implement some kind of buffer. Otherwise, data would get lost because of the limited speed of the USB link to the read-out PC. Here, this buffer is realized as FIFO memory which is put between the nXYTER_decode and the DL701_CBus unit. It is 29 bits wide and has a depth of 4098 bits. The "almost-full" output of the FIFO is connected to an external LED to indicate the danger of data loss. An overview of all components in the VHDL program and how they are connected is shown in VI.14.

[15] Actually, the is another small state machine to manage data transfer with the I^2C module.

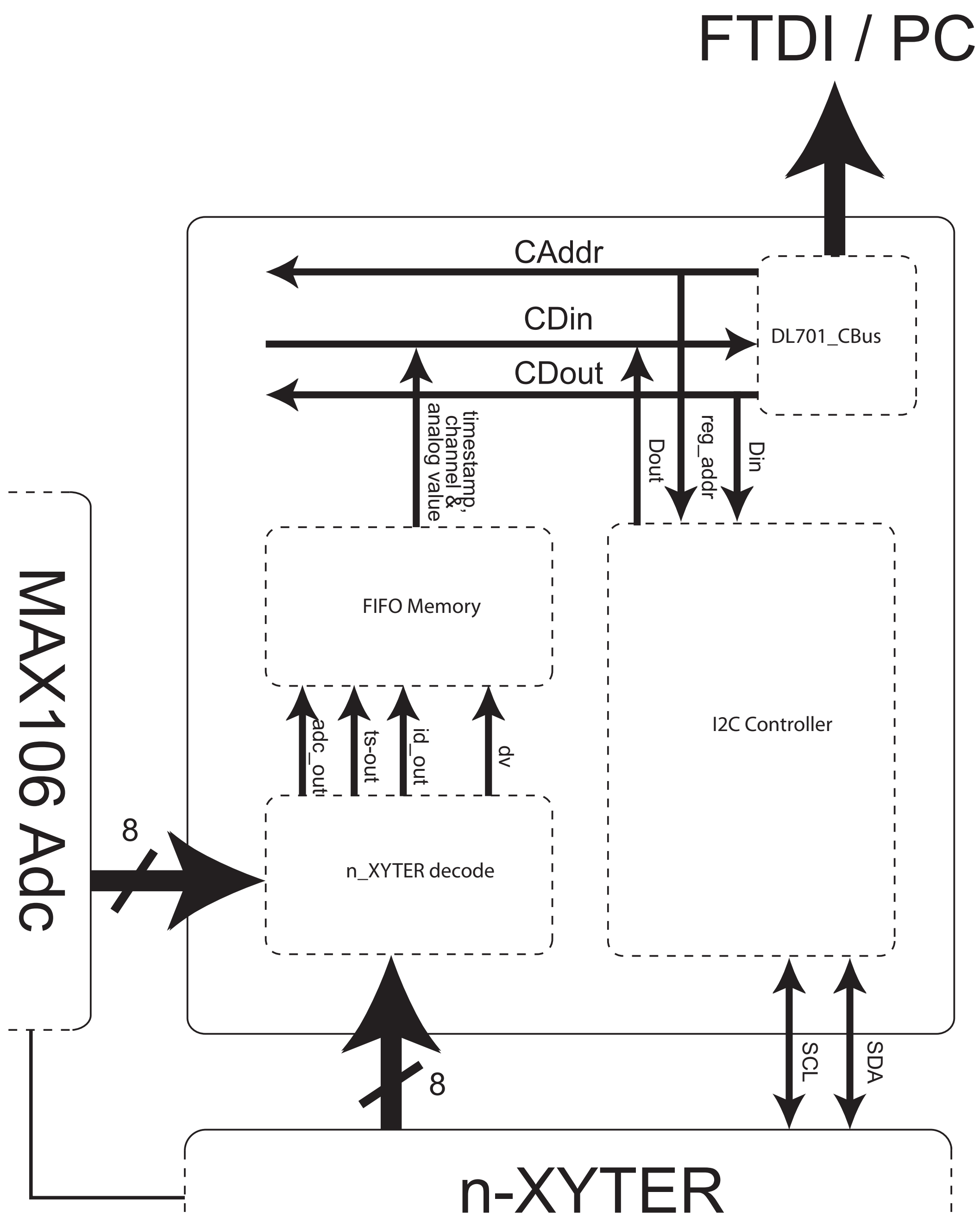


Figure VI.14: The topmost layer of the VHDL design. The arrows indicate the connections between the single components as well as with the n-XYTER and the ADC.

VI.5 Software on the Readout PC

Within the FPGA, the communication with the FTDI chip is handled by the "DL701_CBUS" unit. It responds to a set of commands that can be sent either via a virtual COM port or directly via a DLL[16] to the FTDI. A C++ program has been created to send and read data from the FTDI using the DLL library functions. This allows slightly higher data rates than using the virtual COM port.
The PC interface allows to read and write the n-XYTER's I^2C registers (either one by one or using a textfile that contains all values). Furthermore, it allows to read the data from the n-XYTER and save it into a textfile for further processing. A screenshot of the program is shown in VI.15.

VI.6 Experimental Test: Timestamp Accuracy Using the Test Trigger Input

A first test to control the readout board is to test the the accuracy of the time stamp using an external test trigger impuls. This test would also assure that the transfer of digital data from the n-XYTER via the FPGA to the PC works correctly. The principle of this test is as follows. Test trigger signals with different frequencies are applied to the n-XYTER. The timestamp creation circuit is mainly a counter driven by *clk256a*. Therefore the ratio between the test trigger signal and *clk256a* determines the differences in the timestamp from trigger pulse to trigger pulse. For example, if a trigger pulse of 1 MHz is applied and *clk256a* has a frequency of 100 Mhz, the counter increases by 100 each pulse. As the counter only counts the first 12 MSBs of the timestamp, two zeros must be added to the binary value of 100. This results in a difference of 400 in the timestamp per pulse. This has been confirmed experimentally. In VI.16, the difference in the timestamp for several trigger frequency to *clk256a* ratios is shown. The graph is linear, as expected, up to a certain frequency. This change in the form of the graph means that the frequency is too high so that the n-XYTER no longer works correctly. However, as this behaviour only appears for frequencies greater than ≈ 1.5 MHz, this is no restriction in practical cases.

[16] Drivers for the COM port as well as the DLL are provided by Future Technology Devices International Ltd, see http://www.ftdi.com for more details and code samples

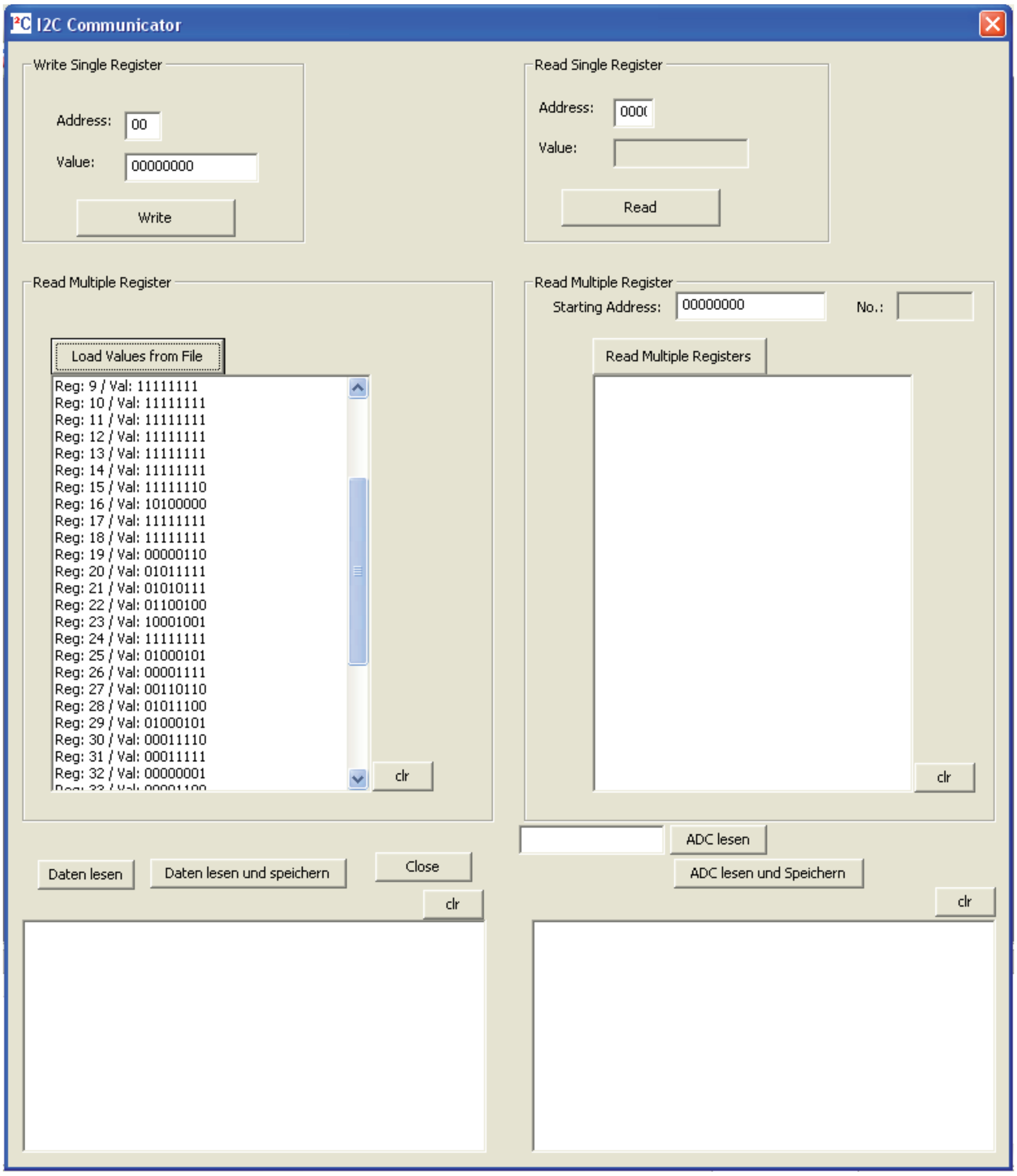


Figure VI.15: This figure shows the interface of the PC readout software.

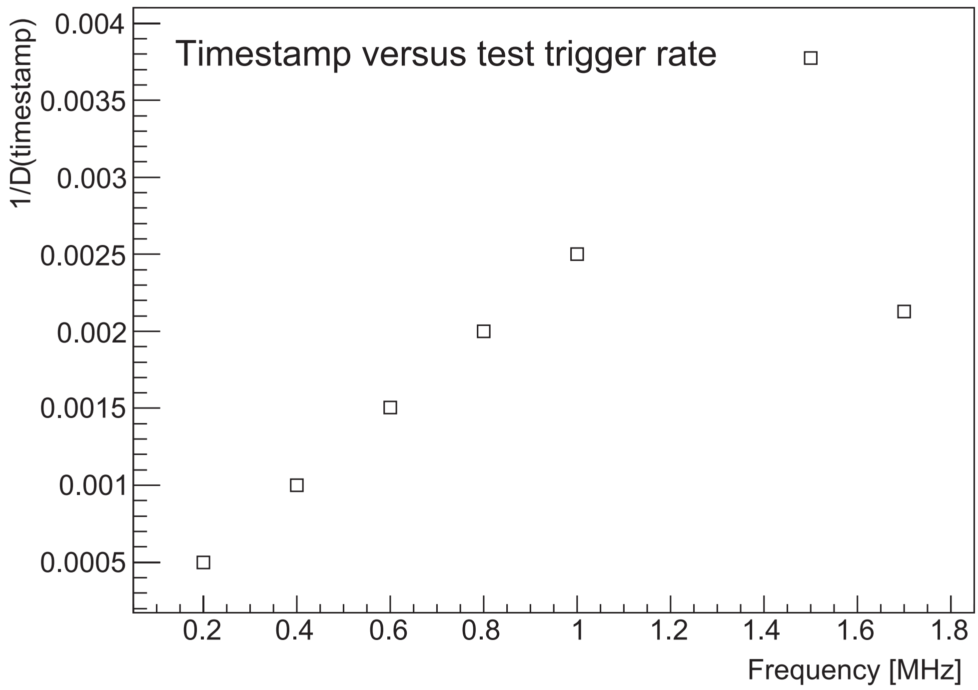


Figure VI.16: Test trigger frequency versus one over the change in the timestamp value. If the n-XYTER works properly, a linear correlation is expected. This is the case up to about 1.5 MHz

CHAPTER VII

Summary

In this thesis, two subjects have been addressed. The first one is the evaluation of the existing msPET MWPC prototypes. The properties of interest were spatial resolution, gas gain and efficiency of the chambers. To determine the spatial resolution, it was necessary to record the pulses induced on the pad strips of the chambers. This has been done using peak detecting ADCs and a program was developed to interface with these ADCs via the VME bus. The gas gain has been figured out using a charge sensitive preamplifier with known gain and a LabView program to record the pulses from the preamplifier. For the efficiency, a combination of hardware coincidence units and another LabView program have been used. The program counted the number of events on both scintillators and MWPC and was furthermore able to control the high voltage. Thus, a completely automatic measurement is possible.
Unfortunately, the results obtained for the msPET prototype chamber are not optimal. Even though it has been shown, that the chamber is in principle working, we were not able to determine the spacial resolution. The efficiency was successfully determined but was lower than expected. A number of possible reasons for these results have been discussed. However, it is clear that further prototype have to be developed.
The second subject of this thesis was the development of a novel readout system for PET using the n-XYTER readout chip. To complete this task, a FPGA design has been developed to interface with the n-XYTER. This design includes a I^2C controller to fully configure the n-XYTER and read status information. Also, a unit to decode the n-XYTER's internal data scheme was implemented. This setup has been

tested using the test trigger mode and has been shown to work fine. Furthermore, an ADC to process the nXYTER's analog output was incorporated in the readout board design. The VHDL code has been extended to combine both analog and digital data and to assure that their relative timing is correct.
The next step in the development of a complete readout board for the msPET prototype would be the design of a PCB containing the FPGA, the ADC and the n-XYTER to reduce noise and make the system more compact. Finally, the readout board has to be tested with a real MWPC.

Bibliography

[Anderson33] C. D. Anderson. The Positive Electron. *Phys. Rev.*, 43(6):491–494, Mar 1933. doi:10.1103/PhysRev.43.491.

[Andrews62] G.A. et al. Andrews. The use of thyroid scanning in thyroid cancer. *Progress in Medical Radioisotope Scanning*, pages 312–343, 1962.

[Baumann05] C. Baumann. X-ray Transmission of Radiators for the ALICE-TRD. diploma thesis, 2005.

[Bethe30] H. A. Bethe. Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie. *Annalen der Physik*, 397:325–400, 1930.

[Bloch33] F. Bloch. Zur Bremsung rasch bewegter Teilchen beim Durchgang durch Materie. *Annalen der Physik*, 408:285–320, 1933.

[Blum93] W. Blum and L. Rolandi. *Particle Detection with Drift Chambers.* Springer, 1993.

[Brogna06] A.S. Brogna. N-XYTER, a CMOS read-out ASIC for high resolution time and amplitude measurements on high reade multi-channel counting mode neutron detectors. 2006.

[Brogna07] A.S. Brogna. The n-XYTER Reference Manual. 2007.

[Brownell99] G. L. Brownell. A history of positron imaging. Presentation given in celebration of the 50^{th} year of services by the author to the Massachusetts General Hospital on October 15^{th}, 1999., 1999.

[Burger06] M. Burger. Scriptum: Numerik partieller Differentialgleichungen. 2006.

[Büther07] F. Büther. private communication, 2007.

[Charpack68] G. Charpack. The Use of Multiwire Proportional Counters to Select and Localize Charged Particles. *Nucl. Instrum. Methods, Vol. 62, p. 235*, 1968.

[deHevesy48] G. de Hevesy. *Radioactive indicators: their application in biochemistry, animal physiology, and pathology.* Interscience Publishers, New York, 1948.

[Deutsch51] M. Deutsch. Evidence for the Formation of Positronium in Gases. *Phys. Rev.*, 82(3):455–456, May 1951. doi:10.1103/PhysRev.82.455.

[Giordino97] N. J. Giordino. *Computational Physics.* Simon and Schuster, 1997.

[Gottschlag05] H. Gottschlag. Entwicklung einer Apparatur zur automatisierten Positions- und Spannungsbestimmung von Drähten in Vieldrahtproportionalkammern. diploma thesis, 2005.

[Gottschlag07] H. Gottschlag. private communication, 2007.

[Horowitz89] P. Horowitz and W. Hill. *The Art of Electronics.* Cambridge University Press, 1989.

[Hudson94] H.M. Hudson. Accelerated Image Reconstruction using Ordered Subsets of Projection Data. *IEEE Trans. Med. Img.*, 13(4):594–601, 1994.

[Hünteler07] H. Hünteler. Development and Evaluation of a Multiwire Proportional Chamber for a High Resolution Small Animal PET Scanner. diploma thesis, 2007.

[Kleinknecht84] K. Kleinknecht. *Detektoren für Teilchenstrahlung.* Teubner, 1984.

[Kösters07] T. Kösters. private communication, 2007.

[Kowalski85] T. Z. Kowalski. Measurement and parametrisation of the gas gain in proportional counters. *Nucl. Instrum. Methods, Vol. 234, pp. 521-526*, 1985.

[Lalik07] R. Lalik. private communication, 2007.

[Leo87] W. R. Leo. *Techniques for Nuclear and Particle Physics Experiments*. Springer, 1987.

[Levin99] C. S. Levin. Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution. *Phys. Med. Biol.*, 44:781–799, 1999.

[Lohrmann86] E. Lohrmann. *Hochenergyphysik*. Teubner, 1986.

[Maxfield03] C. Maxfield. *The Design Warriors's Guide to FPGA*. Elsevier Academic Press, 2003.

[Natterer01] F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction*. SIAM, 2001.

[Nettebrock91] P. Nettebrock. Konstruktion und Test einer Vieldrahtproportionalkammer. diploma thesis, 1991.

[Nolting07] W. Nolting. *Grundkurs Theoretische Physik 3*. Springer, 2007.

[NXP07] NXP. I^2C-bus specification and user manual (available at http://www.nxp.com). 2007.

[Ott93] R. J. Ott. Wire chambers revisited. *J. Nucl. Med.*, 20(4), 1993.

[PDG06] PDG. *Particle Physics Booklet*. BNL and CERN, 2006.

[Povh04] B. Povh. *Teilchen und Kerne*. Springer, 2004.

[Sauli77] F. Sauli. Principles of Operation of Multiwire Proportional and Drift Chambers. *Nucl. Instrum. Methods, Vol. 62, p. 235*, 1977.

[Schäfers08] K. Schäfers. private communication, 2008.

[Schmidt07] C. J. Schmidt. private communication, 2007.

[Tietze83] U. Tietze and Ch. Schenk. *Halbleiter-Schaltungs-Technik*. Springer, 1983.

[Tsui05] B. M. W. Tsui. High-resolution molecular imaging techniques for cardiovascular research. *J. Nucl. Cardiology*, 12(3), 2005.

[Wernick04] M. N. Wernick and J. N. Aarsvold. *Emission Tomography. The Fundamentals of PET and SPECT*. Elsevier Academic Press, 2004.

[Xilinx07] Inc Xilinx. Spartan-3 Generation FPGA User Guide, 2007. Available at http://www.xilinx.com.

[Zeidman02] B. Zeidman. *Designing with FPGAs and CPLDs*. CMP Books, 2002.

APPENDIX A

Acknowledgements

Hier möchte ich allen danken, die zur Entstehung dieser Arbeit beigetragen haben. Herrn Prof. Johannes Wessels danke ich für Möglichkeit in diesem spannenden Gebiet arbeiten zu dürfen, in dem Grundlagenforschung mit konkreten medizinischen Anwendungen verknüpft ist. Außerdem danke ich ihm für die Möglichkeit mehrere lehrreiche Aufenthalte an der GSI verbringen zu können.
Holger Gottschlag danke ich für die Betreung dieser Arbeit. Insbesondere seine Offenheit jeder Frage gegenüber, egal wie sinnlos sie auch war, hat mir bei der Erstellung dieser Arbeit oft geholfen. PD Dr. Klaus Reygers danke ich für viele nützliche Diskussionen. Christoph Baumann und Bastian Bathen danke ich für das gute Arbeitsklima im 'Kompetenzzentrum'. Roland Berendes danke ich für die vielen Hilfestellungen bei allen elektronischen Problemen und vor allem dafür, dass er alle Fragen geduldig beantwortet hat. Norbert Heine danke ich für die Hilfe bei vielen mechanischen Problemen und bei der Erstellung von Grafiken. Christian Schmidt, Knut Steffen Solva und Rafal Lalik danke ich für die freundliche Aufnahme an der GSI und die vielen Informationen über den n-XYTER. Peter von Walter am KIP in Heidelberg danke ich für die Hilfe mit der DL701 LogicBox der USB Kommunikation.
Weiterhin danke ich Dr. Tom Dietel, Helge Grimm, Markus Heide, Norbert Heine, Melanie Klein-Bösing, Stefan Korsten, Robert Lüchtenborg, Markus Rammler, Baldo Sahlmüller, Eva Sicking, Wolfgang Verhoeven, Uwe Westerhoff, Alexaner Wilk, Svenja Wulff und Dr. Oliver Zaudtke für das gute und freundschaftliche Arbeitsklima in der Arbeitsgruppe Wessels.

Meinen Eltern danke ich für die (materielle und vor allem ideelle) Unterstützung während meines Studiums, Maja dafür, dass ich, was auch immer mich beschäftigt mit ihr besprechen kann.

Eigenständigkeitserklärung

Ich versichere, diese Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.
Los Angeles, 13. März 2008

Jan-Frederik Pietschmann