TRIGGERLESS PARTICLE IDENTIFICATION SYSTEMS USING FPGA

by

Veli Uğur Güney

B.S., in Physics, Boğaziçi University, 2007

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in FBE Program for which the Thesis is Submitted

Boğaziçi University

2010

ACKNOWLEDGEMENTS

I am grateful to many people who directly or indirectly contributed to this work. In the presence of the millions who will read this thesis, I want to thank:

to all of the musicians I listened to for providing me an isolated place to concentrate,

to my dear instructors Alpar Sevgen and Yaşar Safkan for both their teaching efforts and introducing me with Taylan Akdoğan,

to my colleague assistants Cem Erol, Sırma Başak Yanardağ, Kazım Çamlıbel, Sina Türeli, Delalcan Kılıç, Emine Ertuğrul and Elif Demirbaş for helping me and replacing some of my tasks when I am not able to parallel process the work and the study at the same time,

to my Onur Albayrak for all his support and his presence in the lab. I have so much fun while being a labmate with you. Long live the fellowship of RLNP!

to my bandmates from Sakareller, Işık Barış Fidaner, Bahadır Maşa and Başar Uğur, who are exposed to my endless speeches about my fascination with physics topics. They also listened to so much my whining about I do not have enough time to do anything.

to my mom who prepared her son the required environment for a comfortable study, although he is nervous and constantly complaining.

to the defence jury, Alpar Sevgen, Serkant Çetin and Metin Arık, for sharing some of their precious time to evaluate our work.

to Haluk Beker, because I have no other option to declare my admiration to him before the public. His appreciation was so important for me. He helped me to have some self confidence in my physics abilities.

to the creators of FPGA technology, which let me feel that I witnessed a real Transformers technology.

to TÜBİTAK for breaking my false belief about no scientific work is encouraged hereabouts.

to Sezer Uğurdağ for Digital Design Lectures and introducing me a new and difficult field.

to Ali Başaran and his alter ego Bezgin Zor for all studies we made together, for our long deep and shallow conversations, for teaching me being more tolerable , and for ceaseless analyses of the "human condition".

to Fatih Uğurdağ for sharing his digital design super powers and enthusiasm for problem solving with us, and for interesting in my future plans, nice and long conversations before work.

And to Taylan Akdoğan for being a wonderful human being in every aspect. For being clever, kindhearted, knowing, considerate, indulgent, motivative, obliging at the same time. You patiently answered all of my questions about anything.

By being a part of this project, I know much more things than one year before on some subjects, and I used that knowledge in a serious and useful work. This is very satisfying.

It was an honour to study with you. I learned so much just by being near you and observing you. Your style of approaching to the problems (from physics to everyday ones) is the most efficient, reasonable, fast, structured and anti-depressing one I have

ever seen. That kind of education cannot be earned from books or anything without directly experiencing it.

I think if every M.S. and Ph.D. candidate had an adviser like you, the scientific world would be a much better place. I hope I will successfully grow the seed I get from you and use what I learned in my future works.

You are my hero and role model!

This work is dedicated to my father, who was the very reason that made me stay at B.U. and therefore, gave me the chance to participate this team work.

This project was funded by TÜBİTAK (No: 107T538).

ABSTRACT

TRIGGERLESS PARTICLE IDENTIFICATION SYSTEMS USING FPGA

As the beam energy and event rate in experimental particle physics setups increases, the acquisition and analysis of the data created at the experiments become more complex. In the beginning, counting the recorded events on a photo-plate by hand was enough. Then, computerized automated methods of event selection and offline analysis has been used. But in that kind of setup for every signal channel coming from a detector one needs to have some special purpose hardware to identify particles.

We made a next generation digital Data AQuisition (DAQ) system specialised to analyse pulse-based signals to replace the traditional setup. It is a triggerless particle identification system which uses Field Programmable Gate Array (FPGA) technology. It takes digitized electrical pulse signals coming from a scintillator-PhotoMultiplier Tube (PMT) system due to incoming charged particles as input and calculates the best fit parameters of a predefined modeling function to the pulse. All of the calculations are done on a specialized FPGA chip in real-time -our part of this work- with a deadtimeless manner and only important pulse parameters are sent to the computer.

ÖZET

FPGA KULLANAN TETİKLEYİCİSİZ PARÇACIK TANIMLAMA SİSTEMLERİ

Deneysel parçacık fiziği düzenekleri daha yüksek olay sıklıkları ve daha yüksek enerji seviyeleri kullandıkça, deneylerde yaratılan verilerin toplanması ve analizi gittikçe karmaşıklaşmaktadır. Baslangıçta fotoğraf kağıtlarına kaydedilen olayların elle sayılması yeterliydi. Sonra olayların seçimi ve çevrimdışı analizi için bilgisayarlı ve otomatikleştirilmiş yöntemler kullanıldı. Lakin bu tarz düzeneklerde dedektörlerden gelen her sinyal kanalı için bazı özel amaçlı donanımlar kullanmak gerekir.

Biz, geleneksel sistemlerin yerini alması niyetiyle atım-tabanlı sinyalleri analiz etmede uzmanlaşmış yeni nesil bir sayısal veri toplama sistemi yaptık. Bu FPGA teknolojisini kullanan tetikleyicisiz bir parçacık tanımlama sistemidir. Yüklü parçacıkların geçisi nedeniyle ışıldayıcı-ışık çoğaltıcı tüp sisteminden gelen sayısallaştırılmış elektrik atım sinyallerini girdi alan bu sistem, önceden belirlenmiş model fonksiyonunun atıma en iyi uyacağı parametreleri hesaplıyor. Tüm hesaplamalar bizim dizayn ettiğimiz uzmanlaşmış bir FPGA çipi tarafından ölu-zamansız bir şekilde gerçek zamanda yapılıyor ve sadece önemli atım parametreleri bilgisayara gönderiliyor. Bu sayede işlem gücünden, maliyetten, donanımdan ve enerjiden tasarruf yapılmış oluyor.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

a_0	Bohr radius
A	Atomic weight
c	Speed of light
e^-	Electron
h	Plank's constant
$\hbar$	Reduced Plank's constant
$\mathscr{H}$	Hamiltonian
m_e	Electron mass
N_A	Avogadro's number
r_e	Classical radius of the electron
Z	Atomic number
α	Fine structure constant
λ	Wavelength
ν	Frequency
ADC	Analog to Digital Converter
ANSI	American National Standards Institute
ASCII	American Standard Code for Information Interchange
CPU	Central Processing Unit
CRT	Cathode Ray Tube
DAQ	Data AcQuisition
DL	Dispatcher Logic
DUT	Device Under Test
FA	Fitting Agent
FIFO	First In First Out
FPGA	Field Programmable Gate Array
GP	Guess Pulse module
HDL	Hardware Description Language

HEP	High Energy Physics
IEEE	Institute of Electrical and Electronics Engineers
LUT	LookUp Table
OOP	Object-Oriented Programming
PCI	Peripheral Component Interconnect
PD	Pulse Detector module
PMT	PhotoMultiplier Tube
QED	Quantum ElectroDynamics
QCD	Quantum ChromoDynamics
RLC	Resistance inductance Capacitance
TR	Trigger module

1. INTRODUCTION

1.1. Problem Definition

In particle physics, the systems which are collecting, processing, filtering and recording the signals coming from the particle detectors due to the particle interactions happened inside are called Trigger and Data AcQuisition (TDAQ) systems. They are needed for many purposes such as creating and recording data packets which include information about the particle interactions into a digital medium, handling high event rates, differentiating important or rare events from background interactions, reducing the event rate below the recordable data transfer threshold determined by the technological limits, controlling and monitoring the experimental setup and its status, preparing the results of the experiment in a form which can be analyzed further etc.

The detector signals are usually in a form called a "pulse". A pulse has some characteristic properties such as the amplitude, the decay constants, the area under the curve, the timestamp. The amplitude is the height of the pulse. The decay constants are related how fast the pulse decays. The timestamp is a special point chosen on the pulse, such as the starting time or the time when the voltage is equal to some threshold value.

From these values some properties of the incident particle can be deduced. For example the amplitude and the area under the curve are related to the energy lost by the particle and deposited in the detector. The decay times may contain information about the type of particle, if the detector has different responses to different particles. From timestamps information about the interaction times and particle velocities can be found.

In traditional DAQ systems there is another specialized module for measuring a property of the pulse. The signal split into many lines and processed in different hardware. Generally, in the first stages there are triggers which make relatively rough

measurements, such as whether the signal has crossed a given threshold. Triggers decide whether or not to record the event. If allowed by the trigger, the rest of the DAQ system records the signal and measures some of its properties. Usually the raw data of the digitized pulse is also recorded in the event frame.

The need to have a special module to measure each of the pulse properties increases the complexity in general. Every connection introduces some risk of defect and noise, increases the cost of building and maintenance. The number of channels in recent high energy experiments is in the order of thousands. Hence trying to reduce the number of components in DAQ systems is an important task for the general maintainability of a high energy physics experiment.

1.2. Our proposal

We propose a next generation DAQ system we call Gigahertz Pulse Fitter (GPF) which is one smart module which replaces all of the traditional modules and finds all of the pulse parameters by fitting a model function of the pulse to the actual input. It works in real-time, thus outputs the parameters while the pulses are coming. It can handle high event rate in a deadtimeless manner. It runs on low cost hardware and is triggerless (has an autotriggering system). The DAQ systems using this smart module is scalable, because thenumber of required devices is equal to the number of channels.

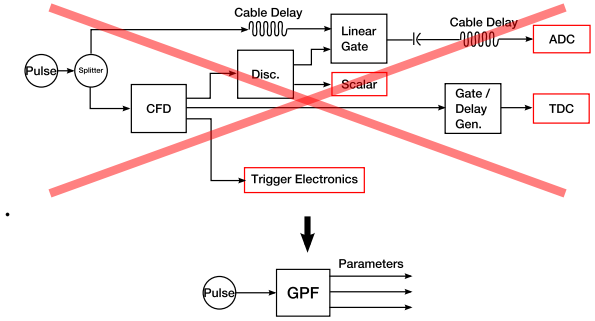


Figure 1.1: Traditional DAQ system vs. GPF

We are using fast digitizer and FPGA technology. The detector output signal is sampled with a sample rate of 1.5 GHz. That samples are processed according the firmware designed by us. Best fit parameters of a modeling function onto the actual

sample values are found. The parameters, such as the amplitude, decay time constant and timestamp, are sent to a computer memory. Every part of the system is placed on a PCI board which can fit in a generic computer.

The project contains works on physics, electronics and computer science.

- An analysis of particle detector interactions is made and the passage of a charged particle through a scintillator-photomultiplier tube system is simulated to help the determination of the fit function.
- Algorithms for pulse signal processing are studied.
- Digital design for programming the processing chip is done. The digital design includes the design of the system on the paper, the simulations of the hardware in the software domain from pulse generation to controlling the operational correctness, the actual design in a hardware description language (HDL) and the synthesis and implementation of the hardware into a programmable chip.
- Device drivers are written to communicate and control the DAQ system with the computer.

Our team consists of Prof. Taylan Akdoğan, Uğur Güney and Berkin Bilgiç from Boğaziçi University and Prof. Fatih Uğurdağ and Ali Başaran from Bahçeşehir University.

Although the distribution of the tasks is blurred due to multipurpose approach of the members, the author of this thesis is mainly contributed

- to the fit function determination by choosing the physical phenomena to be taken into account and writing the simulation code,
- to the development of hardware simulations by writing SystemC code and implementing the the new algorithm ideas in the simulation domain,
- to HDL writing process by providing input output bit streams of the ports of the modules from the simulation for testbench tests to the HDL writers.
- to the intrinsic time measurement by writing special control software of the equip-

ment.

1.3. Abstract of the thesis

In the second chapter a general survey of the particle physics is made. The perspective of the particle physics, the approach of experimental particle physics, the methods of measurements in the atomic scale, the reason of the need for high energy and particle acceleration, the main components of an experimental particle physics setup such as particle sources, accelerators, detectors and DAQ systems, the main interactions between the incident particle and the detector and the working principles of the particle detectors are explained. A literature survey of the recent DAQ systems using FPGA and a summary of our work included.

The third chapter is about how we chose the fit function to use in our system. The stochastic processes which are responsible for the output signal from a scintillator and photomultiplier tube system are explained. Outputs of a computer simulation of an electron passing through the detecting medium and the resulting electrical signal are plotted.

Fourth chapter is about the hardware we are using. A short introduction to digital hardware design and digitization is made. FPGA technology is introduced. The inner modules of the system, their tasks and structure is explained.

Chapter five includes information about what the SystemC hardware simulation library is and how we used it, results of the simulations and the Downhill simplex algorithm which is used to find the best fit parameters.

The last chapter is about the intrinsic time resolution of the system. The experimental setup and results of the measurement are expressed. The thesis is finished with future work suggestions and conclusion.

2. A SURVEY OF PARTICLE PHYSICS

Plato believed that the fundamental building blocks of matter are tiny perfect geometric objects, which we now call Platonic Solids [1]. For him, earth was made of cubes, air octahedrons, water icosahedrons, and fire tetrahedrons (Figure 2.1). We came a long way since Plato wrote Timaeus. Today, we believe that elementary particles are dimensionless, point-like objects. (Figure 2.2) (Although some string theorists are trying to bring in more exotic shapes into this old metaphysical debate.)

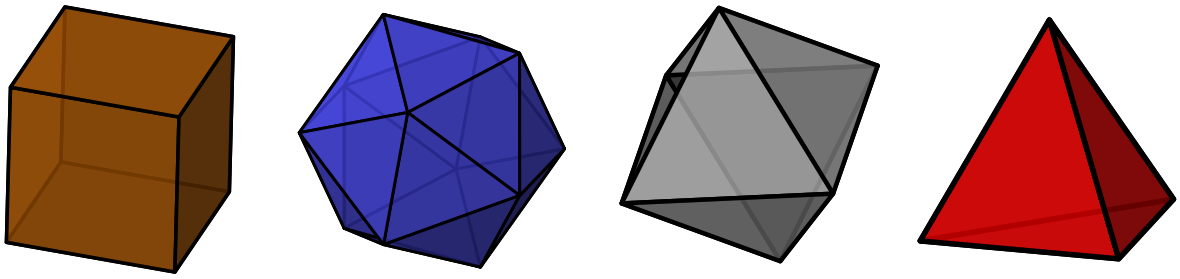


Figure 2.1: Platonic solids for earth, water, air and fire

But the point where we differ from the ancient philosophers who also claimed that all kind of matter is composed of a finite set of tiny elementary particles, such as Anaxagoras, Democritos etc., is not our envisagement about the exact shape of the atoms, the smallest indivisible units, but that we are now able to make experiments on these objects and test our reasoning. Physics is an experimental science and particle physics is the branch of physics which investigates the nature at its lowest level.

•

Figure 2.2: Standard Model's conceptualization of particles

Particle physicists study the properties of particles. They ask what kind of particles are there with what kind of properties, and how they interact with each other. The main idea is that all of the matter and energy that exist is an organization of

elementary particles. They interact in different ways so that higher level structures emerge.

Elementary particles build up composite particles which are bound states of elementary particles, ie. quarks build up hadrons like protons and neutrons. They, together with leptons, electrons, create different kinds of atoms, chemical elements. They come together as molecules. Organisms, planets, stars, galaxies are made of them. From bottom to top it is a journey which covers many orders of magnitudes.

At every stage different structures and natural laws to explain and govern them emerge, but basically all of this beauty is a consequence of the games played between these small agents. Nuclear reactions and decays are due to the interactions of quarks and bosons, and chemical properties of matter are due to the electromagnetic forces atoms and molecules, and chemical reactions are elementary interactions of higher-level organizational structures and so on.

The main idea is that all of the matter and energy that exist is an organization of elementary particles. They interact in different ways so that higher level structures emerge. Of course, the whole is more than the sum of its parts. The *interactions* between the parts create some constraints for their collective behavior. And these constraints in the lower-level become the patterns and governing rules in the higher level. So, we are not just a collection of particles in hustle and bustle, but are irreducible identities with identifiable patterns which become entitled to be called "being"s.

What we have learned so far are collected in the quantum field theoretical model called "the standard model of particles". Our octahedrons are quarks and leptons which are matter particles and our tetrahedrons are photons, gluons, Z and W bosons which are interaction particles. Our information about the nature is not stored in the structure of the particles but in the physical properties attached to them and in mathematically defined interactions between them.

In the physical model which is used to describe the phenomena, these interactions are point interactions. They have no extension in space and no duration in time. [1] Particles meet at some space-time points, called vertecies, where and when the interaction occurs. The existence of a particle or a vertex at a space-time points is called an event. An interaction can be represented by Feynman diagrams, which can be used to calculate the occurrence probabilities of the interaction. See Figure 2.3. Particle physicists try to unfold the (casual) relations between events.

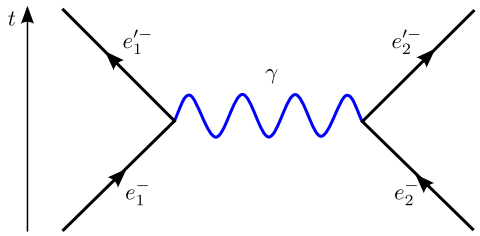


Figure 2.3: A description of an interaction in terms of a Feynman diagram. In this Feynman diagram two electrons are interacting with a virtual photon exchange.

Standard model is the composition of quantum electrodynamics (QED), interactions of electrically charged particles, weak force, the force responsible for nuclear decay and proton's transmutation to neutron, quantum chromodynamics (QCD), interactions between quarks, theory of strong force. But it does not include dark matter and dark energy, about which new cosmological research says that they occupy 23% and 72% of the mass-energy content of the universe respectively, and does not include gravitational attraction and neutrinos, etc. However, its predictions are in agreement with the experiment for the remaining 4% with more than enough precision for every day observations.

So we should make further experiments to test its predictions, correct its deficiencies, develop new models beyond standard model, which is not theoretically satisfying due to the lack of the explanations about its parameters and about why that kind of

[1] In classical formalisms there are continuous distributions of mass and charges which are interacting with action at a distance. Electrodynamics introduced the concept of the field which is the mediator of an interaction by propagating the disturbance. This mediation has a speed limit c. The discovery of atoms (matter particles) leads to discrete distributions of sources. (mass/charge) Quantum mechanics introduces the quantization of energy and discritization of the fields, and the interactions are explained in terms of other kind of particles, such as photons. Hence, in the course of history we returned to the idea that every physical interaction is made via side by side objects.

particle with that properties exist.

Furthermore, the elementariness of the particles are due to the technical limitations of the current experimental technologies. It is normal to suspect whether they have an internal structure or not. Just like once we thought that atoms are the elementary particles but on the contrary they consist of nuclei and electronic shell, what we know today as dimensionless particles may be bound states of other structures. Even new forces or classes of particles (SUSY) may be found.

2.1. Experimental Particle Physics

Experimental particle physicists are trying to

- Measure the known particle properties with better precisions. For example life time, charge distribution of composite particles, and masses, spins and charges of elementary ones etc.,
- Discover new types of particles, new forces or matter particles,
- Measure the probability of a certain interaction to happen,
- Discover certain symmetries and their exceptions

As humans beings we can not see particles or experience what happened between them in any direct way. Even the light itself is a kind of particle, called photon. What we perceive with our eyes is the statistical average of many interaction between the incoming photons and our sensory mechanisms.

Experimental setups amplify what happened between them to macroscopic scales. In the end, what we get is an indirect result of some "chain reaction"s. What happened, and how it was happened is investigated using the byproducts of the reactions which leave traces on the detectors. Today's technology makes our devices be able to react even to single particles.

Our knowledge comes from three sources [2]:

- Scattering state of the interacting system, where interacted particles scatter and their trajectories are deflected. Quantities like the type, distribution, direction and velocity of the scattered particles are the clues which lead us to educated guesses about the nature of the interaction,
- Decay events where a particle spontaneously disintegrates into another one and releases some other particles, which can be thought of as a change from a bound state to a scattering state,
- Bound states where interacted particles stick together and properties of the composite object are studied.

2.2. Measurements in the Microscopic World

2.2.1. Diffraction and Distance Resolution

In the atomic world the objects are very small. Therefore it becomes important to know the experimental limit of the smallest distance which can be measured. It is related to the distinguishability of two points in different positions. This may due to the quality of the equipment or due to the nature of light.

The simplest example of distance measurement is the measurement of the size of a narrow slit from which we will learn the concept of spatial resolution.

The setup consist of a light source, an opaque surface which has an rectangular slit on it, one side of the length d is comparable to the wavelength of the light, λ, and other side is larger than λ and a screen on which the shadow of the surface and the intensity pattern of the incoming light falls. (Figure 2.5a)

According to Huygens-Fresnel principle every point on a wavefront can be thought as a source of spherical wave. Their superposition makes the wavefront and let it keep its shape and direction. If the wavefront contacts with an obstacle the point sources contacting with the obstacle can not contribute the superposition and the wavefront scatters. (Figure 2.4) This phenomena is called diffraction. (See the appendix for a

detailed explanation of the diffraction phenomena)

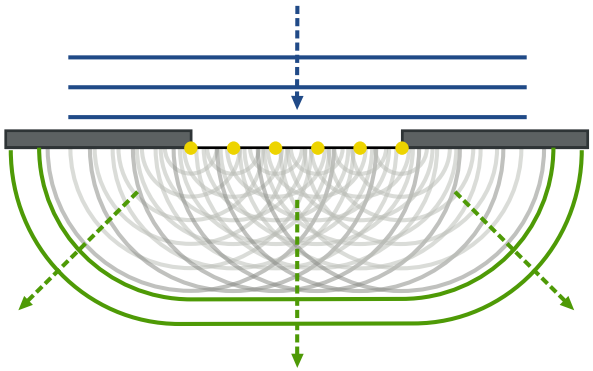
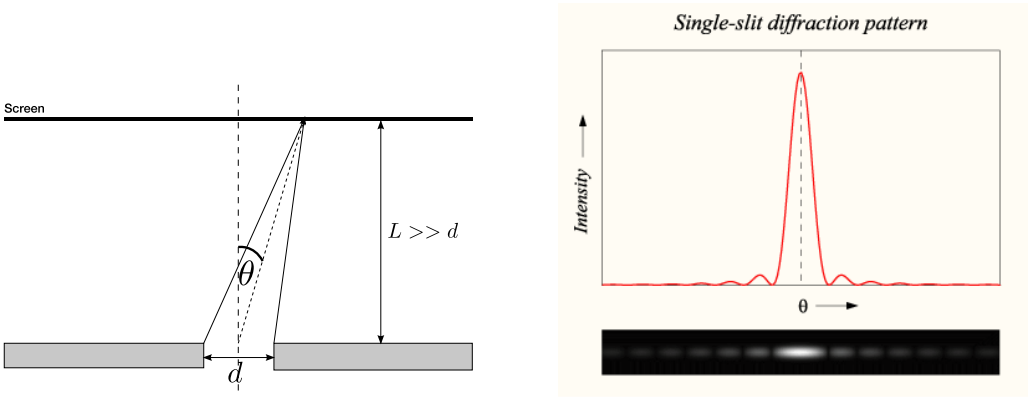


Figure 2.4: Huygens-Fresnel principle [3]

In the single slit experiment the light source is far away from the opaque surface, hence the incoming wave is a planar wave. When the wavefront hits the surface only the points on the slit can continue to being spherical wave sources and their superposition scatters and hence the incoming light diffracts.



(a) Setup (b) Intensity Pattern

Figure 2.5: Diffraction

The result of the diffraction can be seen on the screen as the intensity pattern. (Figure 2.5b). The intensity at the point θ is the sum of infinitesimal intensities due to point sources on the slit. When the path length difference, ΔL, of different rays is half of the wavelength they interfere destructively.

$$\Delta L = \frac{\lambda_{\text{light}}}{2} \tag{2.1}$$

Then the relation between the angle at which minimum intensity occurs, θ_{min}, and ΔL

is

$$\Delta L = \frac{d}{2} \sin \theta_{\text{min}} \tag{2.2}$$

Hence

$$\lambda_{\text{light}} = d \sin \theta_{\text{min}} \tag{2.3}$$

For $\lambda = d$, $\sin \theta_{\text{min}}$ becomes $\frac{\pi}{2}$ which means that destructive interference does not occur. After $\lambda > d$ one can only deduce that the size of the object is smaller than λ and d can not be measured from the intensity pattern. Hence two points separated by a distance d can be resolved only if λ_{light} does not exceed d. This is the spatial resolution.

Visible light is in the range 400 nm $< \lambda_{\text{visible light}} <$ 700 nm which is much larger than the size of an atom ($\approx$ 0.3 nm).

$$d_{\text{atom}} \ll \lambda_{\text{visible light}}$$

So, visible light can not be used to measure the atomic sizes. This is not the fault of the microscope or its quality but it is due to the nature of light. Therefore we need smaller wavelength to probe smaller distances.

2.2.2. High Energy Physics

Today a part of particle physics is called "High Energy Physics" (HEP). Why do we need higher energies lies in the relation

$$E = h\nu \tag{2.4}$$

This expression due to Einstein gives the relation between energy, E, of a particle and its corresponding frequency, ν. The constant of proportionality is Planck's constant,

h. $\nu\lambda = c$ for light particles, where c is the speed of light. Then $\nu = \frac{c}{\lambda}$ Hence

$$E = \frac{hc}{\lambda} \tag{2.5}$$

So, wavelength and energy are inverse proportional. Particles with higher energies have smaller wavelengths.

$hc = 1240$ eV nm $\approx 10^{-6}$ eV m. Which means that with an energy E = 1 eV, one can probe 10^{-6} meters. (Table 2.1)

A wave scatters from a target if its wavelength is comparable or smaller than the size of the target. (By this means radio waves, $\lambda \approx 10^3$ m, can travel long distances without being scattered from the obstacles on its way like buildings etc.) Hence distances $\Delta x < \lambda$ is indistinguishable for waves. λ has to be smaller than the dimensions of the inspected object. Therefore for probing smaller distances higher energies, are needed.

In the ultraviolet domain electromagnetic waves with shorter wavelengths than visible light exist. For example x-rays are in the nanometer range and their λ is similar to the adjacent planes in a crystal structure. Hence using the technique called "X-ray crystallography" the form of molecules can be deduced, by analyzing their diffraction patterns. Gamma rays have even shorter wavelengths, but they are hard to produce. We do not have a gamma-ray laser today. Hence it is the natural limit for a light microscope.

Fortunately matter particles show wave behavior too and waves with shorter λ can be created by giving them enough kinetic energy/momentum. The needed energy can be calculated from the relativistic energy relation:

$$E^2 = p^2c^2 + m^2c^4 \tag{2.6}$$

But one needs to accelerate them to the relativistic speeds to obtain length resolutions

in the order of nucleus etc.

Because, unlike photons, we can accelerate charged particles using electric and magnetic fields, the technological limit of λ for matter waves is shorter than of light waves hence matter particles are used in modern "microscopes". For example electron microscopes, which can achieve $10^6\times$ magnification, has greater resolution than light microscopes, which are limited to $1000\times$.

While these lines are written, the highest energy loaded to a proton in an artificial accelerator is 3.5 TeV, in LHC. The aim is to achieve 7 TeV and collide two opposing beams with a total energy of 14 TeV.

Although this is a huge amount of energy for a single particle, it is very small for macroscopic scales. For example, 7 TeV is equal to 1.12×10^{-6} Joules, which is roughly the kinetic energy of a mosquito ($m = 0.25$ mg, $v = 1$ m/s).

2.2.3. Energy, Temperature and Time

The need of high energy can be seen from another perspective which is related to the temperature and age of the universe.

The relationship between energy and temperature for a bunch of colliding particles is given by:

$$E = \frac{3}{2}kT \tag{2.7}$$

where, k is the Boltzmann constant, T is temperature. Approximately the room temperature corresponds to 35 meV, 1 eV corresponds to 10^4 K. Which means air molecules have an average energy of 35 meV when colliding with each other.

Just like when we throw something very fast (higher than the escape velocity), namely with enough kinetic energy, it can escape the gravitational pull of Earth, an

electron around the nucleus can be freed, if enough energy is given. The amount of the needed energy is called the binding energy. If the colliding particles have kinetic energies above the binding energy, they may break down and their internal structure scatters.

For many molecules the binding energy of outer electron is in the order of 1 eV. Hence chemical structures and reactions can exist at room temperature, which is well below 10^4 K. For example, the binding energy of hydrogen atom is 13.6 eV, which means that it can survive on Earth. But the inner temperature of the Sun 10^7 K, which corresponds to an energy of 1 keV, do not let atoms to survive. The sun is made of free electron gas and a plasma of hydrogen and helium nuclei.

Today the average temperature of the universe is about 2.7 K. It is in the form of Cosmic Microwave Background, the remnant of Big Bang. Also we know that the galaxies are moving away. By rewinding this process it is calculated that the universe had began 13 billion years ago in a very hot state, the Big Bang. It is cooling down since then, while expanding. At the beginning no particle pair can be bound (or "in bound state"), but today it is cold enough for atoms and molecules to exist.

A relation between time and temperature, $T(t)$, can be made. It can be seen from this considerations that, higher energies corresponds to higher temperatures, which exist in earlier stages of the universe. Hence higher energies are needed to simulating the early conditions of the universe. (See Table 2.1.)

According to special relativity particles can be created from energy. The relation between the mass of the particle and corresponding energy is: $E = mc^2$. For example electron has a mass 0.5 MeV in terms of energy. During a collision, if the kinetic energy of the colliding particles is high enough, a pair of particles with that rest mass energy can be created. 1 MeV is needed for creating an electron, anti-electron (positron) pair.

Therefore, by creating conditions similar to the early universe, collisions with higher energies produces more exotic particles (with higher mass) which were found

Table 2.1: Approximate relations between particle energy and its wavelength, its temperature and the time when average temperature of the universe is equal to it.

Energy	Wavelength (m)	Temperature (K)	Time	Allowed Processes
1 meV	10^{-3}	10^{1}	now	Galaxies
1 eV	10^{-6}	10^{4} (Earth)	300000 years	Molecules
1 keV	10^{-9}	10^{7} (Sun)		
1 MeV	10^{-12}	10^{10}	100 sec	Nuclei synthesis
1 GeV	10^{-15}	10^{13}	10^{-9} sec	Nucleon synthesis
1 TeV	10^{-18}	10^{16}	early universe	Plasma

more widely in early universe, like muons, heavier quarks etc. "In so doing we are creating for a brief moment in a small region of space an intense concentration of energy, which replicates the nature of the universe as it was within a split second of the original Big Bang. Thus we are learning about our origins" [4].

Also the energy which will be loaded to particles can be tuned finely to produce exactly a kind of particle. This process is called resonating. For example, 1232 MeV, for Δ.

Hence the main idea of high energy physics is more sophisticated than smashing something to see how it works.

2.3. Experimental Setup of Particle Physics

Particle physicists do their inquiries in the micro world using experimental setups consist of radiation sources, detectors, accelerators, colliders, data acquisition systems and analysis.

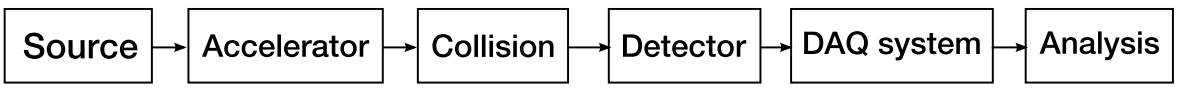


Figure 2.6: A diagram of experimental particle physics setup

Particle beams that are produced in the sources are loaded with energy in accel-

erators. Then they are aimed to fixed targets or collided with an other beam. The outcomes of their interaction are captured in the detectors by transforming the information about the position, momentum, energy etc. of the incoming particles to electrical signals. These signals are sent to DAQ systems where significant ones are sorted out from physically less important ones using trigger systems. The events that pass the trigger filters are recorded in computer media and then analyzed utilizing analysis software mostly by PhD students. At the end, theories are checked or new ones are formed upon these observations.

Let us discuss the main components of the experimental setup in the following sections.

2.4. Particle Sources and Beams

2.4.1. Cosmic Rays

There are different kinds of radiation sources. One of them is cosmic rays whose existence was first postulated by Victor Hess in 1912. The reactions in stars, supernovae and other astronomical objects eject free particles to the space which sometimes hit the Earth's atmosphere. The incoming rays are composed of 90% hydrogen nucleus (proton), 9% helium nucleus and 1% other elements and electrons. Only 0.01% of them are antimatter, which is an indication of matter-antimatter asymmetry. It should be noted that lifetime of proton is long enough to have a long journey.

In the atmosphere they interact with the nuclei of air molecules. When they react, a shower of secondary particles falling down to Earth occur. For example after the collision pions and kaons are produced which decay to muons very fast. These muons can be detected at the sea level with an average intensity of 100 $m^{-2}sec^{-1}$.

The cosmic rays may produce very high energetic collisions which can not be gained in laboratory conditions, but the incoming particles are random and their intensity is very low. This means that the chance of getting an intended type of particle

is very low. Between 1930 and 1950 they were the main source of the HEP research. Positron and muon are discovered during this era. But because today one needs constant high intensity beams to probe some interactions with very low occurrence probability (cross sections), cosmic rays are not useful sources for controlled experiments. Except neutrinos produced in the Sun while nuclear reactions, which are another kind of cosmic rays with high intensity but have extremely low interaction rates.

2.4.2. Radioactive Decay

Another set of sources are nuclear reactors and some nuclear reactions. While disintegrating a nucleus emits a variety of particles: neutrino, neutron, alpha, beta and gamma particles etc.

These names -alpha (α), beta ($\beta^{\pm}$), gamma (γ)- are given when the categorization of the particles were not well defined. Alpha particles are bound states of two neutrons and two protons (helium nucleus), beta particles are electrons or positrons and gamma particles are high energetic photons. The radioactive elements ^{90}Sr and ^{106}Ru are beta (e^{-}) sources. ^{241}Am is an alpha and ^{252}Cf is a neutron source.

Alpha rays were used by Rutherford while examining the atomic nucleus. But particles produced by available natural sources do not have enough high energy demanded by recent HEP experiments. Hence they are used for calibration purposes only. Elements have a determined energy spectrum of emitted particles and detectors are adjusted to give correct values. The monoenergetic spectrum of natural decays make them appropriate reference points. And they are also used to exercise detectors before using accelerated beams.

2.4.3. Ionization

Protons are produced by ionization of hydrogen atom.

Electrons are produced easily by heating a metal. When a filament is heated

while there is an electrical current through it, it releases some electrons. This filament is put in an electric field so that the released electrons are directed to the next stage of the accelerator. This is the starting point of both cathode ray tube (CRT) monitors and electron accelerators.

2.4.4. Accelerators (Secondary Beams)

The particle accelerators themselves are used to create certain kinds of high energetic particles. First, particles easier to produce, like electrons and protons, are accelerated and directed to a target, called "spullation target". These collisions create secondary beams which can be investigated with detectors or can be sorted out and used as beams for successive collisions. While these collisions, pairs of particles and antiparticles may be created via virtual photons and gluons. Magnetic fields are used to separate them.

2.5. Accelerators

Accelerators are facilities which are used to load kinetic energy to particles.

2.5.1. Rutherford and the Need for Particle Acceleration

The idea of accelerating particles is first expressed by Ernest Rutherford in 1927. In a speech before the Royal Society of London he drew the attention of the scientific community to the need of higher energies.

He had previously used the natural α-sources to probe the nucleus, by firing α-particles to a thin gold foil. Some of the α's goes easily through the foil, because they have enough kinetic energy. But some of them bounce back by the gold nuclei. This leads to a conclusion about the structure of the atom, which is that the nucleus is much smaller than the atom and concentrated in its center and has nearly the whole mass.

Rutherford wanted to investigate further the structure of the atomic nucleus. He

again used α-particles with 5 MeV to disintegrate the nitrogen nuclei [5]. But the energies due to the natural α-decays are not enough to disintegrate nuclei with higher binding energies. Therefore projectile supplies with higher energies than naturally occurred α and β decays is needed for the future of HEP experiments [6].

2.5.2. J. J. Thomson and Using EM Fields to Manipulate Charged Particles

Particles are so small that they can not be held even with the tiniest tweezers. Actually the concept of "holding" itself is meaningless at the atomic scale. As an approximation, a free particle is in the middle of the empty space and only affected by the fields around it. Hence only the particles which have electrical charge can be accelerated by manipulating them using artificial electrical and magnetic fields.

The idea of using EM fields for manipulating particles goes back to J. J. Thomson, who bent cathode rays in vacuumed tubes in 1895. He created a constant electric field between metal plates and let the beam go through and bent it. That way electrons and hence the existence of subatomic particles are discovered. Thomson also used a combination of both electric and magnetic fields to calculate the e/m ratio of the electron.

This technology was used in TV sets and computer monitor until LCD systems became widespread. CRT monitors are simple accelerator-detector systems. The beam source is produced in an electron gun. The electrons forming the beam are speed up and their direction is changed using electromagnets and deflector plates in a vacuum and hit the phosphor target, the screen. At the point where the incoming electrons collide with the phosphor molecules on the screen light is emitted. They work at 10 keV.

2.5.3. Cockcroft & Walton and the First Accelerator

The first and simplest way for accelerating particles is to put them in a static electrical field created by a constant voltage difference. The equation for a static field

is:

$$\mathbf{E} = -\nabla V \tag{2.8}$$

Hence the energy gained from point $\mathbf{r_1}$ to $\mathbf{r_2}$ is

$$\Delta E = q\left[V(\mathbf{r_2}) - V(\mathbf{r_1})\right] \tag{2.9}$$

This basic principle is the underlying mechanism of the first device intended to accelerate particles.

The electrostatic devices at the time when Rutherford announced the need of the higher energies, could not create enough voltage difference. In 1928 Gamov had taken the quantum tunnelling effects into account and calculated that 500 keV suffice to split the atom. In 1932 Cockcroft and Walton designed a voltage multiplier and made a device capable of producing 800 KV steady potential. Using it they linearly accelerated protons (H^+ ions) and disintegrated lithium atom into two α particles at 400 keV [7]. (Equation 2.10

$$p + {}^7\mathrm{Li} \rightarrow \alpha + \alpha \tag{2.10}$$

Today Cockcroft-Walton accelerator is used as the first stage of many HEP experimental setups.

2.5.4. Van de Graaff Generator

Another static high voltage device developed at about the same time was Van de Graaff Generator. Robert J. Van de Graaff used electrostatic charge belts which carry charges to collect them in insulated conductors instead of voltage multiplier rectifiers

design by Cockcraft [8].

He filled two conducting spheres with opposite charges until they discharge by creating arcs and sparks between them. This simple idea is similar to loading a glass rod with charge by rubbing it with a wool cloth. In Van de Graaff generators this "rubbing" is done by motorized silk belts. Charges are loaded to silk belts and transferred from the belt to the spheres by point discharges from needle points and metal brushes. In 1932 Van de Graaff succeed to reach ±750 KV, in total 1.5 MV difference between spheres.

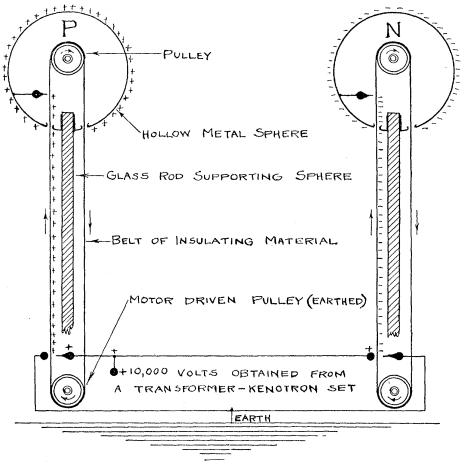


Figure 2.7: Schematic of the Van de Graaf Generator [8]

The spark formation is reduced by increasing the sphere sizes and putting them in inert gases. Todays generators can make voltage differences up to 15 MV.

2.5.5. Wideröe and Acceleration with Alternating Voltage

In the direct-voltage systems the limit of the maximum energy to be gained is equal to the maximum voltage be able to produced. To overcome this limitation a design based on alternating-voltage is used.

The idea of applying the same voltage repeatedly using an alternating field is first developed by Ising in 1924 [9]. (Figure 2.8) In 1928 Rolf Wideröe made a simple system using this technique, which is also called resonant acceleration [10].

There are gaps between conducting evacuated tubes through which electrons are flowing. An oscillating voltage is applied to the tubes at a fixed radio frequency (RF) in MHz region. Because tubes are conducting there is no field inside the tubes and electrons are accelerated along the gaps between the tubes. The oscillation frequency and phase is adjusted according to get the maximum voltage across each gap when electrons passes through it. The energy, E, they gain at each gap is $E = eV_{\max}$

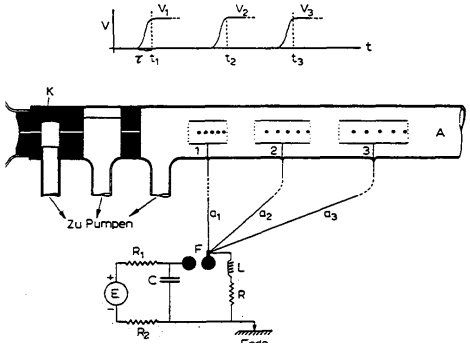


Figure 2.8: An alternating voltage model, schematic from [9]

Wideröe's device had two gaps, hence the energy applied to potassium ions using 1 MHz oscillations is 50 keV which is twice of the amplitude of the oscillating voltage, 25 KV. But there is no limit for the number of gaps. Hence,

$$E = N_{\text{gaps}} e V_{\max}$$

Only, the distance between the gaps should adjusted according to the oscillation frequency, voltage, mass of the particles etc. The tubes should gets longer while the particle gets energy and speed. Resonant acceleration technique makes it possible to accelerate a charged particle to a large energy using a small amplitude.

The phenomenon which makes RF acceleration feasible is called the longitudinal phase stability. A particle beam consist of discrete bunches. Because particles do not have the same exact velocity, every bunch is dispersed in space hence occupies some volume. Therefore some particles in the bunch come earlier to the acceleration gap and some become late. As can be seen in the Figure 2.9, if one chooses a phase value, ϕ_0, for the oscillation, such that bunches arrive to the gap when voltage value is V_0 which

is smaller than $V_{\max}$ the dispersion decreases. Because a greater voltage is applied to the late ones and let them accelerate more and a lesser voltage to the earlier ones and accelerate them less.

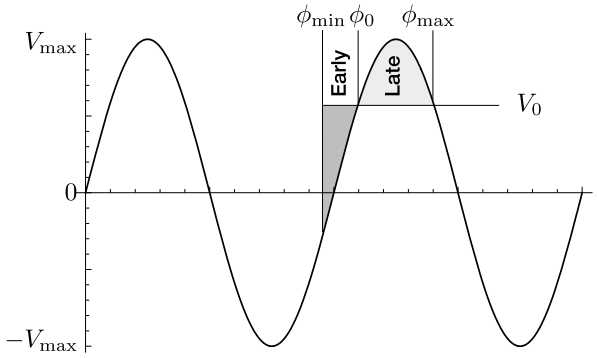


Figure 2.9: Longitudinal phase stability in RF acceleration

2.5.6. Cyclotron, Betatron and Synchrocyclotron

Wideröe's model is a linear accelerator and at that time they are useful for accelerating heavier nuclei. But they are not that useful for accelerating smaller particles, because they need longer vacuum tubes and faster RF circuits. And the amount of energies loaded to heavier nuclei is not enough to produce significant amount of nuclear disintegration like Cockcraft's device.

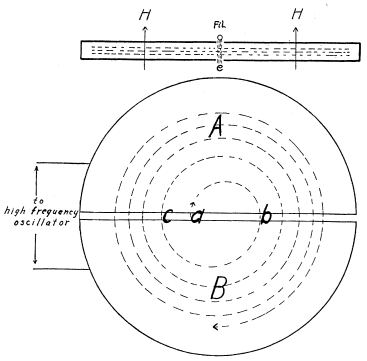


Figure 2.10: Cyclotron [11]

In 1931, Lawrance made a device, named cyclotron, in which particles have a circular trajectory due to the applied constant magnetic field [11]. They pass through the same gap between the "D"s repeatedly. An oscillating voltage is applied on the gap of which frequency is adjusted such that protons are accelerated each time they cross

the gap. At every half turn their energy is increased hence the radius of the trajectory also increases and the particle spirals out.

The time it takes for a half revolution is, $T = \pi r/v$. Therefore, the voltage oscillation frequency, which is also called the "cyclotron frequency", is chosen such that, $\omega_{\text{osc}} = \pi/T = v/r$. Under constant magnetic field B, the momentum of the particle is $p = eBr$. Hence for nonrelativistic speed, the orbital frequency is $\omega_{\text{orb}} = v/r = veB/p = eB/m$. In order to accelerate the particles each time they cross a gap ωs must be equal, $\omega_{\text{osc}} = \omega_{\text{orb}}$.

In relativistic region $p = \gamma mv$, and orbital frequency becomes $\omega_{\text{orb}} = eB/\gamma m$, which depends on the particle velocity. Hence if
$omega_{\text{osc}}$ is kept constant, particles may not be accelerated to highly relativistic speeds.

Circular accelerator for electrons is invented by D. W. Kerst, and is called betatron. Because electron is a light particle the radiation due to acceleration (curved motion), which is called the synchrotron radiation becomes important.

The more advanced circular accelerators on which the oscillation frequency is changed synchronously with the inverse of the particle energy, are called synchrocyclotron. Because there is no need to have to apply a great voltage at every turn, synchrocyclotrons works with relatively small voltages at acceleration gaps such as 10 keV. The limitation of the Synchrocyclotron depends on the highest available magnetic field or largest radius of the accelerator. An increase in one of them also increases the highest energy to be reached. Their practical limit is about 700 MeV.

2.5.7. Synchrotron

In a synchrotron the particles are moving inside vacuumed tubes on a fixed radius circle, on which at some places RF cavities are placed. At every passage through a cavity particle gets a small energy increase. Because the radius is fixed, the applied magnetic field has to be increased synchronously with the increase in the loaded energy.

This technology is possible due to (transverse) phase stability, which is achieved by an arrangement of magnets which is called "weak focusing", found in 1945. Say, for example, the energy of a particle is slightly higher than the average, E_0, it will have a slightly greater radius $r = r_0 + \Delta r$. In weak focusing the magnetic fields are adjusted such that that oscillates transversely around r_0.

Other technological advance in synchrotron acceleration is the "strong focusing", where an array of magnets alternatingly focuses and defocuses the beam. The result is damping in the oscillations. In higher energy devices the focusing is done by quadrupole magnets. First strong focusing devices did reach to 30 GeV.

The cyclotrons and synchrotrons are also used for medical purposes, ie. producing radioactive isotopes for PET scans or in material science, chemistry, molecular biology etc.

2.5.8. Storage Rings

Particles can be accelerated to the intended speed in just a few seconds. But before the actual collision they may have to wait for some hours. Hence they need to be kept at some place while maintaining their speeds. These places are called storage rings.

Storage rings are circular vacuumed pipes on which a constant magnetic field is maintained with magnets for keeping the particle beams circulating continuously with constant frequency. Because of the long storage time they need high vacuum and constant application of magnetic field.

Two storage rings may be intersected to study the collisions of stored beams or particle and anti-particle beams can be collided on one storage ring. Every time the beams cross each other some of the particles collide and interactions happen.

2.5.9. Livingston Plot

The accelerator technology shows an exponential growth behavior in energy similar to Moore's Law in the developments of computation technologies. The graph which shows the timelines of different acceleration technologies and their energies is called Livingston plot. See Figure 2.11 and 2.12.

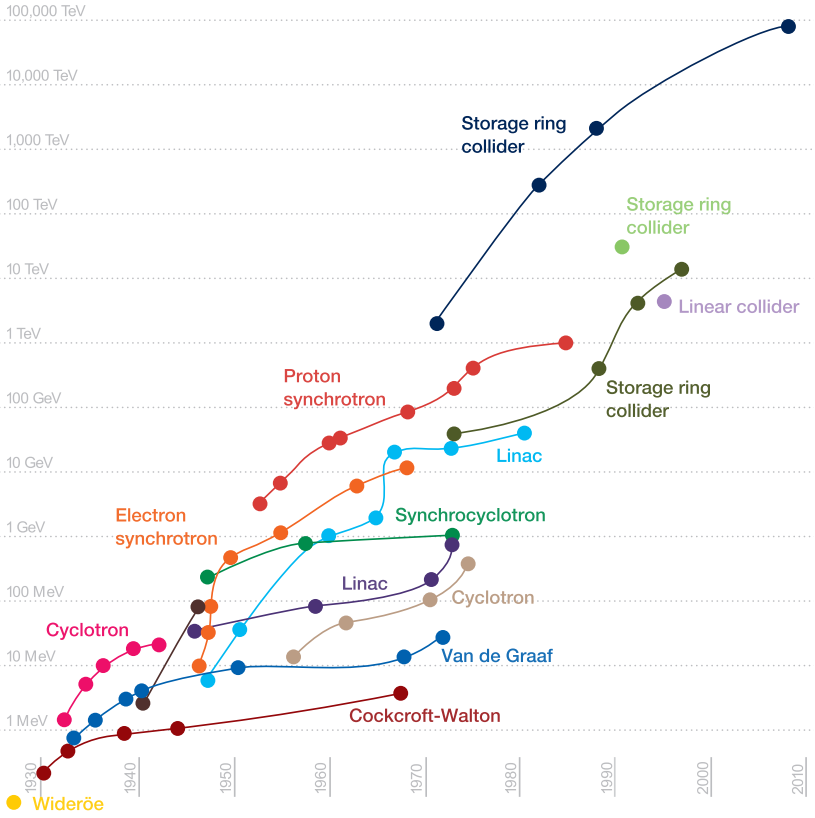


Figure 2.11: Livingston plot. Adaptation of the Livingston plot from the 2001 Snowmass Accelerator R&D report. Energy of colliders is plotted in terms of the laboratory energy of particles colliding with a proton at rest to reach the same center of mass energy. Using these units, the energy of collisions at the Large Hadron Collider is nearly 100000 TeV. Adapted from [12]

2.5.10. Other Important Aspects of Accelerators

Although the main attention of the Livingston plot is on energy, which defines the limit to observe new physics, there are other important factors which affects the accelerator operation. These may be the beam intensity (luminosity), duty cycle,

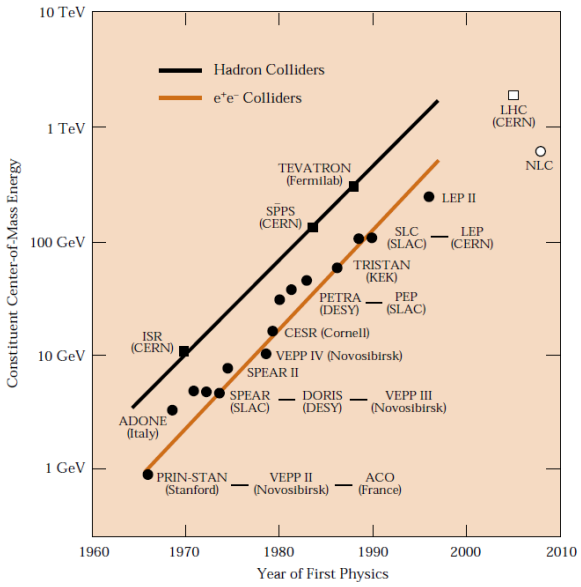


Figure 2.12: The energy in the constituent frame of electron-positron and hadron colliders constructed [13]

amount of background radiation, the ability to focus the beam, vacuum etc.

The particle flux which is also called the luminosity, L, is the amount of particles that are let to interact in unit time and area. The occurrence rate of a certain interaction, R, is the product of L and cross section, σ. $R = L\sigma$.

There is a trade-off between luminosity and energy. In reality they can not have both the maximum possible value at the same time. Because of the wave nature of the matter and probabilistic aspects of quantum mechanics interactions can occur at lower energies with a smaller probability. Hence higher densities are needed. They can compensate each other.

Duty cycle is the percentage of time the beam is actually on. In HEP experiments the beams do not have a steady flow but it is pulsed on and off. If the duty cycle is small many of the interactions occur in a close time which makes it harder to isolate individual ones.

Background radiation, interactions that are already known and hence contain no

new information.

Because the cross sections of rare events are small beams must be focused using magnets to increase the interaction probability.

2.5.11. Accelerator Complexes

The above mentioned accelerator types are used together in accelerator complexes such as the Conseil Européen pour la Recherche Nucléaire (CERN, European Organization for Nuclear Research).

"CERN's accelerator complex is a succession of particle accelerators that can reach increasingly higher energies. Each accelerator boosts the speed of a beam of particles, before injecting it into the next one in the sequence".

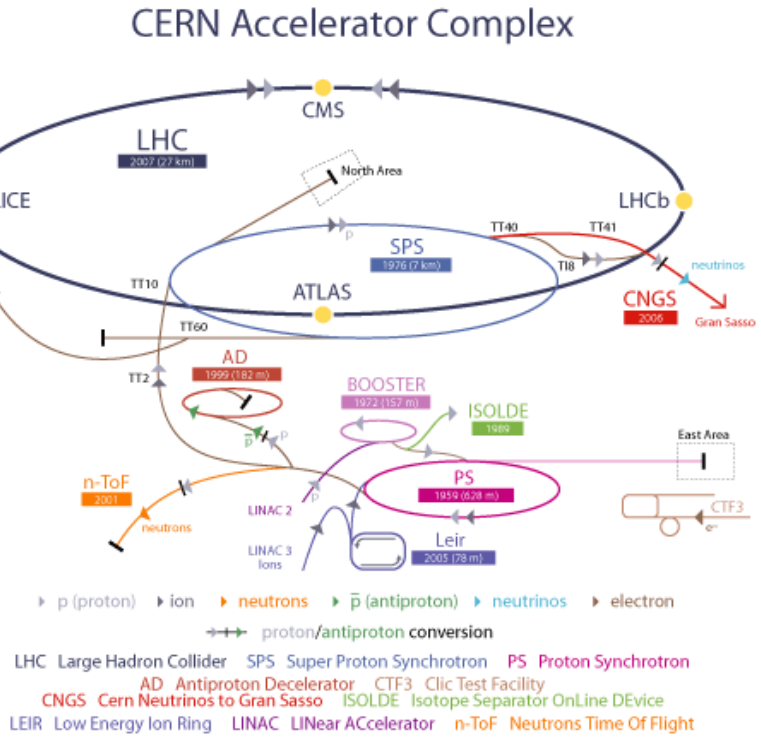


Figure 2.13: CERN accelerator complex

"Protons are obtained by removing electrons from hydrogen atoms. They are injected from the linear accelerator (LINAC2) into the PS Booster, then the Proton Synchrotron (PS), followed by the Super Proton Synchrotron (SPS), before finally

reaching the Large Hadron Collider (LHC). Protons will circulate in the LHC for 20 minutes before reaching the maximum speed and energy" [14].

2.6. Detectors

A detector is a kind of matter or setup which creates a macroscopic reaction when it contacts with an incoming radiation, or if it has enough sensitivity, with a single particle.

The role of detectors can be visualized by assuming that an interesting interaction occurs at a point in space and time. From that point several secondary particles of different masses are emitted with various angles and momenta as shown in Figure 2.14. It is the job of the experimentalist to measure the time of interaction, t, and the vector momenta, p_i, and masses, M_i, of those emitted particles.

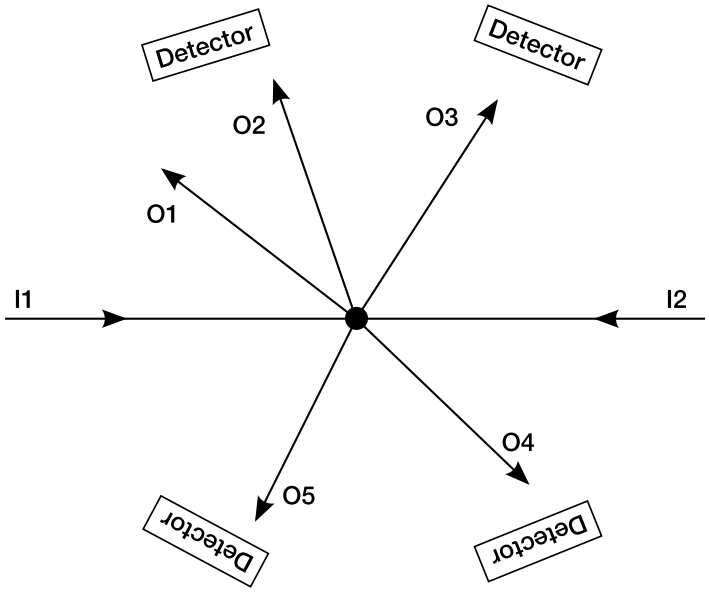


Figure 2.14: Interaction at a vertex

Detectors, their means of interactions and common types are investigated in the next section.

2.7. Particle Detection Methods

In this chapter we will summarize the methods and physical principles of particle detectors.

First of all, we said that Standard Model conceptualizes particles as point particles which do not have a volume and do not span in space. Then, how it is possible that two particles meet at absolutely the same coordinate in space and interact at that vertex, although there are infinitely many points between any two coordinates.

The solution arises from the quantum mechanical properties such as indeterminacy principles and the ability of the systems to be in superposition states. For example a free photon, which is going to interact with an electron in bound state around an atom of the detector, do not have definite energy and position values. It is in a superposition of different states corresponding to energies distributed around a mean value.

The linear combination of the states causes a dispersion in the position, momentum, energy distributions of the photon. Hence the state of the photon in the position domain is not a dirac delta function, $\delta(\mathbf{r} - \mathbf{r_0})$, but is a function, called wavepacket, which occupies some volume in space. In this manner, both the free photon and the bounded electron has different possible positions which continuously span some space, and hence the possibility to being at the same point, and hence interact at that vertex, is non zero.

2.7.1. Cross-section

Cross-section is a quantity which is used to describe interaction of two particles. It is the scattering rate, R_s over the incident flux, Φ_i.

$$\sigma = \frac{R_s}{\Phi_i} \left(\frac{1/[T]}{1/[L]^2[T]} \right) \tag{2.11}$$

Hence it has the same physical dimensions as the area, usually measured in barns. (1 barn = 10^{-28} m^2)

In the classical sense (where a particle or flux hits an object which has a volume in space) the total cross-section, σ, is the area of the projection of the target object on an infinite plane perpendicular to the direction of the incoming flux. It is how the incident particle "sees" the object, in other words it is the amount of the area upon where the projectiles may hit the target.

If σ is calculated by taking quantum mechanical effects into acount, the idea of projection onto a plane collapses, and the interaction probability somehow increases. Projectile may see the back of the target, due to the wave property of matter.

As an example, if the σ of a sphere with radius R is calculated using classical mechanics, one gets $\sigma = \pi R^2$, which is the area of a circle with the same radius. σ depends to the energy of the incident particle. If its energy is high, it is fast and spends less time around the target hence the interaction probability drops.

$$\sigma = \iint d\Omega \frac{d\sigma}{d\Omega} \tag{2.12}$$

The differential cross-section, $\frac{d\sigma}{d\Omega}$, gives finer detail about the interaction. It is the directional dependence of the total cross-section and proportional to the amount of the particles reflected to the given direction (solid angle) hence gives the angular distribution of scattered particles. (See Figure 2.15)

$$\frac{d\sigma}{d\Omega}(E, \Omega) = \frac{1}{\Phi_i} \frac{dN_s}{d\Omega} \tag{2.13}$$

where N_s is the number of scattered particles. Of course these are statistical quantities which are get by assuming the incident flux is homogeneous, and its area is larger

than the characteristic area of the target. Although the number of scattered particles fluctuate over time, N_s and σ values are time averages.

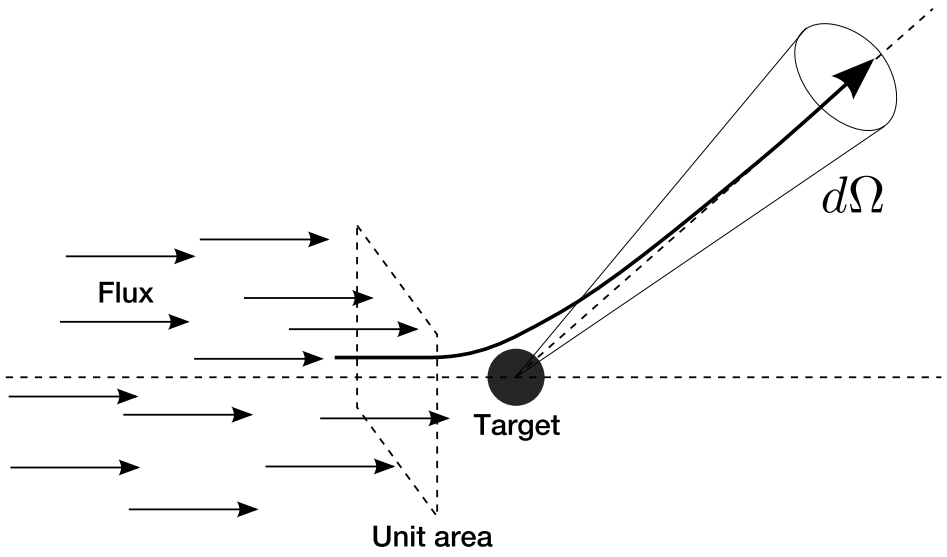


Figure 2.15: Differential cross-section

The cross-sections can be calculated from the first principles, if the form of the interaction is known mathematically. Then they can be measured in experiments and by comparison can be used to verify or reject theories. This approach is used by Rutherford when he studied the atomic nucleus. He calculated $\frac{d\sigma}{d\Omega}$ by assuming that the positive charges, which incident α particles are interacting, are point-like and in the center of the atom and compared the calculated value with the results of his Rutherford scattering experiment.

2.7.2. Passage of Charged Particles Through Matter

The passage of particles through generic matter is a very complex process because of the variety of interaction types and dependence on the compound of the matter and the type and the velocity of the incident particle. The main idea is that the particle goes along some trajectory, which may include multiple scatterings through small angles hence deflections from the incident direction, and on the way it looses energy by transferring some (or all) of its kinetic energy to the matter via electromagnetic and strong interactions, by changing the configuration of the matter such as heating it, taking apart electrons from atoms, putting molecules to other energy/vibrational

states etc.

The most famous everyday example is the passage of light particles through glass and air, where, essentially, the trajectories are bent or reflected. These phenomena are used for many technologies such as lenses etc.

Ordinary matter such as a living tissue is a complex compound of which analysis is hard due to its heterogeneous nature. Fortunately, the detectors used in particle physics (such as scintillators) are human constructions with definite geometry and content and are homogeneous substances, which makes their analysis relatively easy.

The analysis is based on how the energy of the incident particle is lost in the medium. Main interactions by which the energy is lost are

- inelastic collisions with atomic electrons and,
- elastic scattering from the nuclei.

In an inelastic collision energy is transferred to the atom (or molecule) by exciting or ionizing it. Excitation is called "soft collision", where the configuration of the atom is changed such that the electrons go to a higher energy state, or the molecule goes to a higher energy vibrational state. Ionization is called "hard collision", where the transferred energy is higher than the binding energy of the electron and it is freed. Sometimes the freed electron has enough energy to cause secondary ionizations. They are called δ-rays or knock-on electrons.

Although the energy transfer at a single so-called "collision" [2] event is very small compared to the kinetic energy of the particle, but the matter is so dense that these events occure many times per unit path length. Therefore their cumulative effect makes the total loss a significant quantity.

[2]These are not collisions in the sense of the collision of two massive classical bodies, where they physically touch each other. They interact from a distance via photons.

The nature of inelastic collisions is statistical. They have some quantum mechanical occurrence probability. But because the number of events is high the statistical fluctuations becomes relatively small and hence the energy loss per unit length, $\frac{dE}{dx}$, is a meaningful quantity that can be used in the detection process.

Elastic scattering from the nuclei occurs less frequently and the transferred energy is small because the masses of the nuclei are large compared to the incident particle. A counter example may be incident α particles in a hydrogen target.

Other effects such as Cerenkov radiation, nuclear reactions, bremsstrahlung also occur but their contribution to the total energy loss is small.

Charged particles can be grouped into three categories according to the dominant reactions responsible for the total energy loss:

- light particles such as electron and positron,
- heavy particles ($m > m_e$) such as muon, pion, proton, α, light nuclei, "hadrons" in general,
- heavy nuclei.

Inelastic collision yields the most important contribution to the energy loss for heavy particles. (with $\sigma \simeq 10^8$ barn) Light particles are also being subject of these inelastic collisions but due to their small mass other effects such as emission of electromagnetic radiation in terms of bremsstrahlung, which are due to their acceleration in the electromagnetic field, becomes important. For heavy nuclei additional effects arise.

2.7.2.1. Ionization Loss. Figure 2.16, shows the dependence of stopping power ($\frac{dE}{dx}$ per density) to the momentum of the the incident muon passing through copper. As can be seen in the figure, the plot is divided into different domains where different effects are dominant and different approximations must be made.

First domain is where $\beta \leq 0.01$, which is in the order of the velocity of so called

orbiting outer atomic electrons in Bohr model. Lindhard found that stopping power is proportional to β in that region.

Second domain is where $0.01 \leq \beta \leq 0.05$. For that region there is not a rigorous theory. Andersen and Ziegler have found phenomenogical fitting formulae for protons [15].

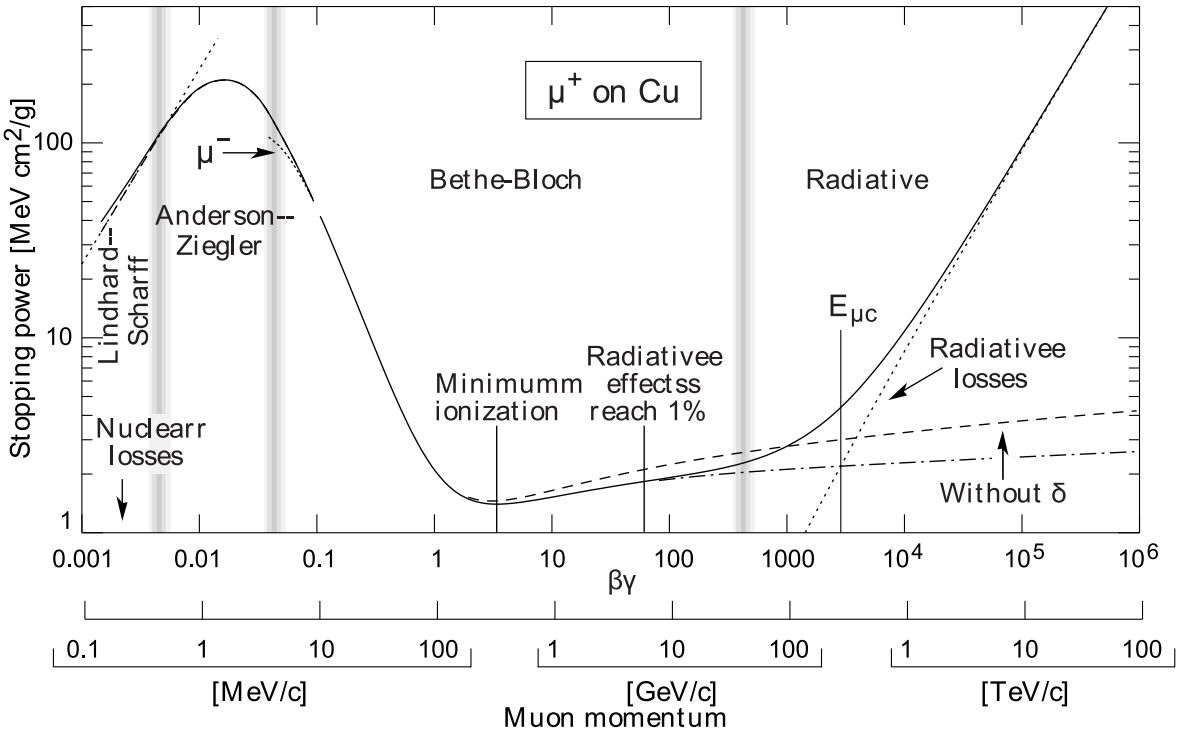


Figure 2.16: Stopping power [16]

HEP mostly deals with relativistic particles in non-radiative or radiative regions. Particles, except electron, with momentum between 10 MeV/c and 100 GeV/c which are in a moderately relativistic domain, lose energy according to the Bethe-Bloch equation (Equation 2.14), which is based on a first-order Born approximation. This formula was first obtained by Bohr in 1915 and developed further by Bethe by taking wave nature of electrons into account in 1930. The energy loss is mostly due to ionization and atomic excitation, and can be called electronic energy loss [16].

$$-\frac{dE}{dx} = Kz^2\frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\frac{2m_ec^2\beta^2\gamma^2T_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2}\right] \tag{2.14}$$

where $K/A = 4\pi N_A r_e^2 m_e c^2/A$ is a constant, I is the mean excitation energy, $T_{\max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1+2\gamma m_e/M+(m_e/M)^2}$ is the maximum kinetic energy which can be given to a free electron in a single collision, δ is the density effect correction to ionization energy loss due to relativistic length contraction.

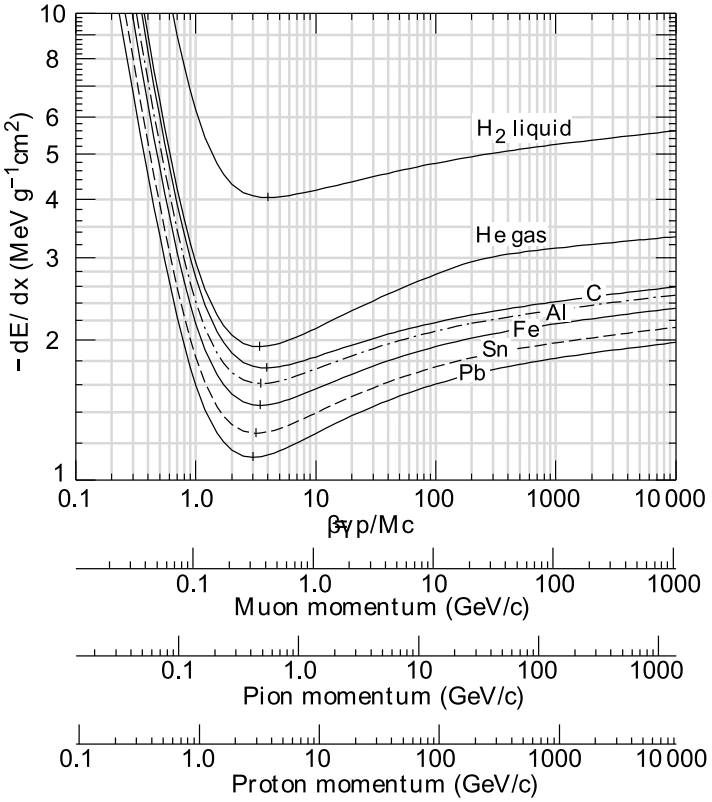


Figure 2.17: Mean energy loss rate in liquid (bubble chamber) hydrogen, gaseous helium, carbon, aluminum, iron, tin, and lead [16].

Although there is a dependency on particle mass, M, which is included in the equation by $T_{\max}$ and which becomes important only for highest energy regions, in HEP $\frac{dE}{dx}$ mainly depends on β. Particles with same velocities have similar energy loss rates in different materials, with a slight decrease as Z goes higher. This can be seen in Figure 2.17 and Figure 2.18.

According to these dependencies, to get $\frac{dE}{dx}$ from Bethe-Bloch formula, one has to determine the mean excitation value, I. Because it is very hard to derive it from first principles, there are estimates made by experimental stopping power measurements for protons, deutrons, α particles etc.

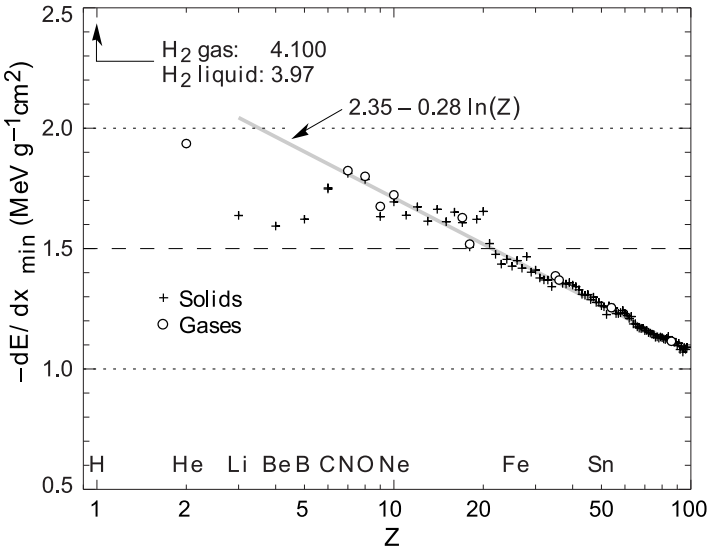


Figure 2.18: Stopping power at minimum ionization for the chemical elements. The straight line is fitted for $Z > 6$.

Range: $\frac{dE}{dx}$ formula can be integrated to find and approximate range, R, of the incident particle assuming its only energy loss is due to ionization and atomic excitation.

$$\text{Range}(T_i) = \int_0^{T_i} \left(\frac{dE}{dx}\right)^{-1} dE \tag{2.15}$$

This approach ignores the multiple Coulomb scatterings, which makes the path of the particle a zigzag. The path as a straight line from the hit point to the last position will be shorter than the actual zigzag path travelled so far. But because the effects of multiple scattering is small for heavy particles and this integral gives a valid result for the range.

This approximation is called "total (or partial) continuous slowing-down approximation". Because $\frac{dE}{dx}$ depends only on β, R is a function of it. R/M vs. $\beta\gamma = p/Mc$ is plotted in Figure 2.19.

Both stopping power and range formulas are expressed energy loss per length per density. For example, energy loss per distance, $\frac{dE}{dx}$ is given in units of MeV g^{-1} cm^2. After multiplying this value with the density of the material, ρ (in units g cm^3), one gets correct quantity in units MeV/cm.

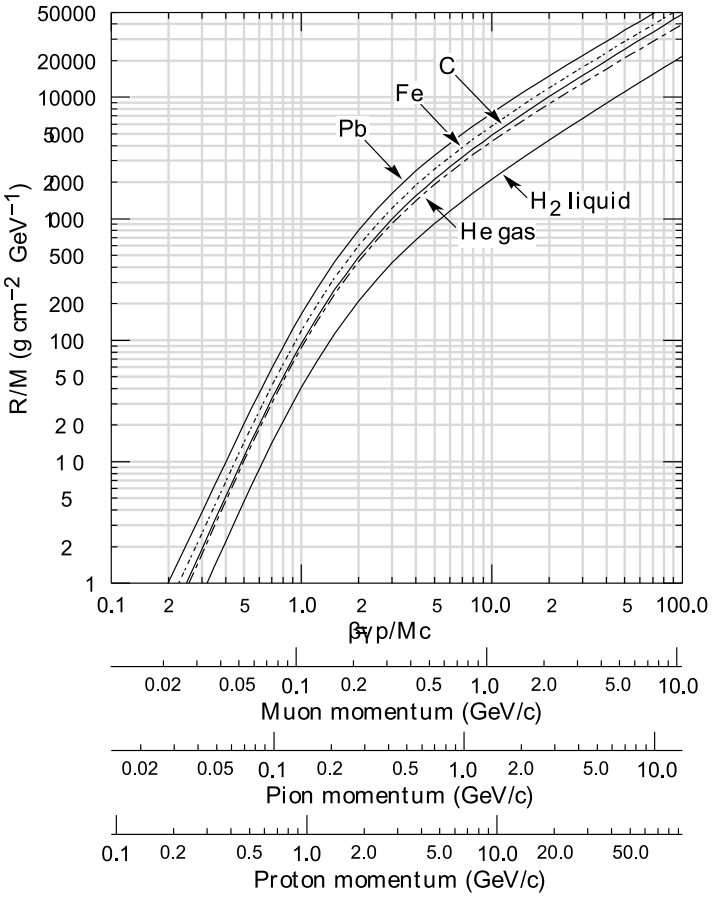


Figure 2.19: Range of heavy charged particles in liquid (bubble chamber) hydrogen, helium gas, carbon, iron, and lead [16].

In the highly relativistic domain, as the energy increases there is a rise in stopping power, which is due to the length contraction effects. The ionizing particle is exposed to a larger electric field perpendicular to its direction, hence corresponds and increase in $\frac{dE}{dx}$ by $\ln \beta\gamma$. Other important reason of the increase in energy loss rate is the Cherenkov radiation.

The correction term δ includes this relativistic rise and the phenomenological resistance against it due to the polarization of the medium. This is called the density effect, because the polarization effect in denser media is stronger. It become important when the fields become so strong such that charges are arranged such that the effect of the transverse electric field is canceled. Therefore at very high energies [16]:

$$\frac{\delta}{2} \rightarrow \ln(\hbar\omega_p/I) + \ln(\beta\gamma) - 1/2 \tag{2.16}$$

where $\hbar\omega_p = \sqrt{4\pi N_e r_e^3} m_e c^2/\alpha$ is called the plasma energy. Figure 2.16 shows the plots for both included and excluded δ in the Bethe-Bloch region.

In the radiative region where the energy is in the order of 100 GeV radiative effects dominates the energy loss and ionizations becomes less important.

Fluctuations: $\frac{dE}{dx}$ is a statistical quantity and hence is a distribution around a mean value. If the detector is thick enough $\frac{dE}{dx}$ measurements give a nearly gaussian distribution. But if the detector is a thin layer of material, where total energy loss is less than 1 MeV, the number of total collisions is not high and the distribution will has a long Landau tail and becomes asymmetrical. This is caused by the small number of collisions in which a large energy is transferred.

When these distributions are taken into account, the most probable energy loss is [16]:

$$E_p = \xi \left[\ln \frac{2mc^2\beta^2\gamma^2}{I} + \ln \frac{\xi}{I} + j - \beta^2 + \delta(\beta\gamma) \right] \tag{2.17}$$

where $\xi = KZx/2A\beta^2$ MeV for a thickness x in g cm^{-2} and $j = 0.2$. In the high energy domain E_p becomes

$$E_p \xrightarrow[\beta\gamma \geq 100]{} \xi \left[\ln \frac{2mc^2\xi}{(\hbar\omega_p)^2} + j \right] \tag{2.18}$$

Compounds: From the perspective of energy loss, a compound which is different elements mixed together, is equivalent to pure element layers with right proportions. The exact positions of the atoms are not important. What is important is the fact that they are included homogeneously with some proportion. In a compound

$$\frac{dE}{dx} = \sum_j w_j \left. \frac{dE}{dx} \right|_j \tag{2.19}$$

where $\frac{dE}{dx}\big|_j$ and w_j are the stopping power and proportion of j^{th} element respectively.

Similarly Z/A becomes $\sum w_j Z_j/A_j$ etc. But using this way, the I will be found is going to be less than original value because in a compound electrons are more tightly bound than in pure elements.

Multiple Scattering Through Small Angles: A charged particle going through a medium scatters many times with some small angles. These deflections are mostly due to Coloumb scattering, hence it is also called "multiple Coulomb scattering". For hadrons, strong interactions with the nuclei may also contribute to multiple scatterings.

The Coulomb scattering distribution is explained by the theory of Molière. The cumulative deflection angle θ_{plane}, seen in Figure 2.20, is a gaussian approximation for the central 98% of the projected angular distribution with a width given by [16]:

$$\theta^{\text{rms}} = \frac{13.6\text{MeV}}{\beta cp} z\sqrt{x/X_0}\,[1 + 0.038\ln(x/X_0)] \tag{2.20}$$

where z is the charge number of the incident particle, and x/X_0 is the thickness of scattering medium in radiation lengths.

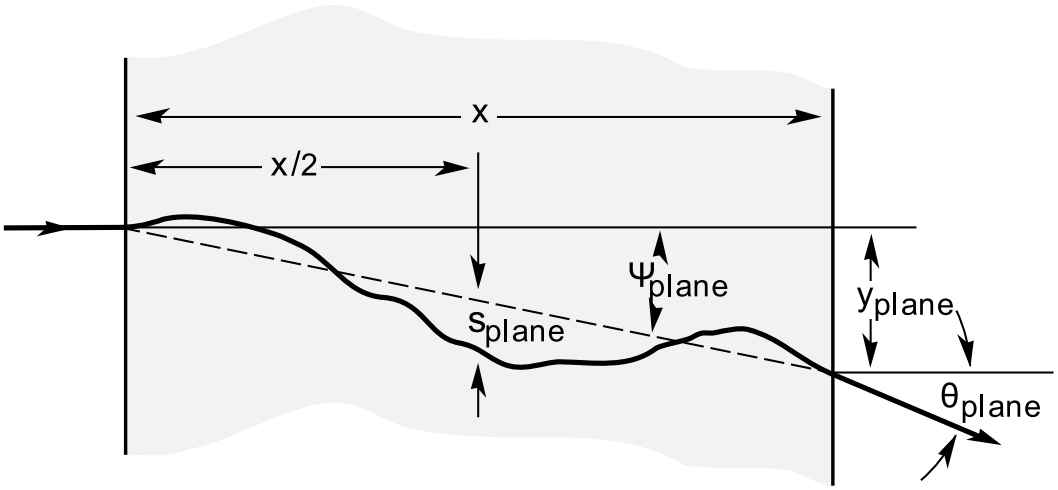


Figure 2.20: Quantities used to describe multiple Coulomb scattering [16].

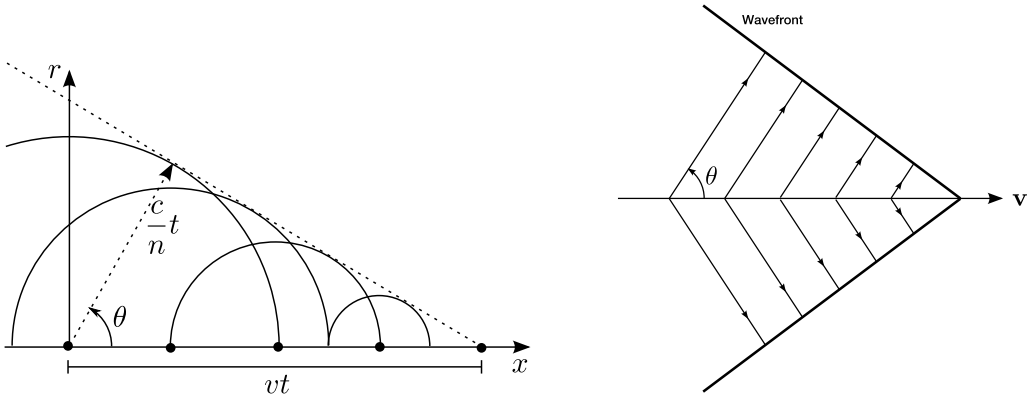
This equation describes scattering from a single material. Scattering from a compound is not simply equal to the sum of the θ_j's calculated using the formula for

every element. But some phenomenological approaches are developed by Lynch and Dahl.

From the viewpoint of an experimenter who wants to determine some properties of the incident particles, the processes occurring inside the matter can be divided into two categories: Destructive and non-destructive. "If the interaction with a detecting medium transfers but little energy to that medium, the measurement is called non-destructive" [17].

2.7.2.2. Cerenkov Radiation. The speed of a wave depends on the properties of the medium in which it propagates. For light, the speed in a medium is c/n, where n is index of refraction of the material.

A particle can not exceed the universal maximum speed, the speed of light in vacuum, c. But in a medium its speed v may higher than the local phase velocity of light in that medium, $v > c/n$. If that is the case, the radiation of the excited atoms becomes coherent. That radiation is emitted with a fixed angle with respect to the trajectory of the particle, as seen in Figure 2.21b. It is first observed by Cherenkov in 1935. It is the source of the blue glow in the cores of nuclear reactors.



(a) Cerenkov cone construction using Huygens' principle

(b) Cerenkov radiation

Figure 2.21: Cerenkov

The angle of emission, which can be deduced from the Figure 2.21a where the

phenomenon is explained using Huygens' principle, is given by:

$$\cos(\theta) = \frac{1}{\beta n} \tag{2.21}$$

Because the radiation angle depends on the velocity of the particle, it can be used to measure the velocity of a charged particle.

The spectrum of the Cerenkov radiation is continuous. It has some parts in the visible domain. The energy given by Cerenkov radiation is small compared to total ionization loss. For example for a particle with $\beta \approx 1$ in water energy loss due to the visible Cerenkov radiations is $\left(\frac{dE}{dx}\right)_{\mathrm{C}} \approx 40\mathrm{keV/m}$, which is 1/5000 of the total energy loss [18]. Hence it is a non-destructive process.

It is used in particle identification, to distinguish light particles, such as leptons, from heavier particles, such as hadrons. Simplest version of this PID mechanism is the threshold counter, which shows whether the velocity of a charged particle is lower of higher than a threshold value, $v_{\mathrm{th}} = c/n$, just by looking at the presence of Cerenkov light in the medium. If the momentum of the particle also is known, then one can have threshold for mass and separate lighter ones from heavier ones.

2.7.2.3. Transition Radiation. When a charged particle with a relativistic velocity crosses the boundary between two media with different index of refractions, it emits photons. This is called the transition radiation due to the instant acceleration caused by the transition in the index of refraction.

The number of photons emitted is related to γ. The radiation is coherent and the angle at which the photons are emitted is $\theta = 1/\gamma$ with respect to the particles direction. The energy of the photons depend on plasma frequency, which is

$$\omega_p = \sqrt{\frac{4\pi\alpha^2 c^2 (\hbar c)^2 N_A Z \rho}{mc^2 A(10^{-3}\mathrm{kg})}} \tag{2.22}$$

When $\gamma > 100$ ($\beta > 0.99995$) the radiation is in the x-ray domain. Hence transition radiation is used for velocity measurements of very high energetic particles [18].

Maximum energy of a transition photon is $E = \gamma\hbar\omega_p$. Photon emission probability is α. Therefore average total energy emitted per transition is

$$E_{\text{tot}} \simeq \frac{\gamma\hbar\omega_p\alpha}{3} \tag{2.23}$$

For a material with $\rho = 10^3$ kg/m^3, $\hbar\omega_p \simeq 20$ eV and hence $E_{\text{tot}} \simeq 5$ eV. That is very small compared to ionization energy loss, hence this process is non-destructive.

2.7.2.4. Brehmsstrahlung. Bremstrahlung is the German word for "braking radiation". Accelerating charged particles radiate and this causes a decrease in their kinetic energy. The name comes from the deaccelerating effect of the radiation.

Electron (e^-) and positron (e^+) are the only particles for which radiation is an important energy loss factor under GeV range because they are lightest charged particles. The cross-section is proportional to the inverse square of the mass.

$$\sigma \propto \left(\frac{e^2}{mc^2}\right)^2$$

If the masses of the muon, $m_\mu c^2 = 106$ MeV, and electron, $m_e c^2 = 0.5$ MeV are compared, it can be seen that $\sigma_\mu/\sigma_e \approx 1/40000$. Hence muons radiates much less than electrons.

Ionization loss is proportional to charge (Z) and the logarithm of energy. $\frac{dE}{dx}\big|_{\text{ion}}$ $\propto Z\ln(E)$. Radiation loss is proportional to energy and Z^2. $\frac{dE}{dx}\big|_{\text{brems.}} \propto Z^2E$. See Figure 2.22

Total ionization loss is realized by summation of many little energy losses to the individual molecules, which makes it a quasi-continuous process. Whereas the total

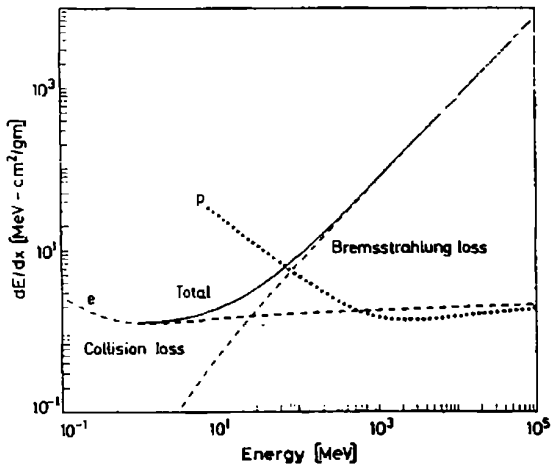


Figure 2.22: Radiation loss vs. collision loss for electrons in copper [15]. ($\frac{dE}{dx}$ for proton is also shown)

radiation loss may be done by one or two photons only. Hence the fluctuations are much higher for monoenergetic electron beams than beams with heavier particles.

2.7.3. Passage of Photons Through Matter

Photons do not have electrical charge. This make their behavior inside matter different than charged particles. For example, it is impossible for them to make inelastic collisions with the atomic electrons.

The main types of interactions for x-rays and γ-rays are

- Photoelectric effect,
- Compton scattering,
- Pair production.

Photons penetrate much further inside the matter, because the cross-section of these three processes are much smaller than inelastic collision. Individual photons do not loose energy on their ways, only the intensity of a photon beam decreases, because when one of these interactions occur, that photon is either absorbed (hence disappeared) or scattered and removed from the beam. Therefore a photon which

succeeds to pass through did not interacted at all and kept its initial energy.

The attenuation in the intensity of a photon beam can be shown by:

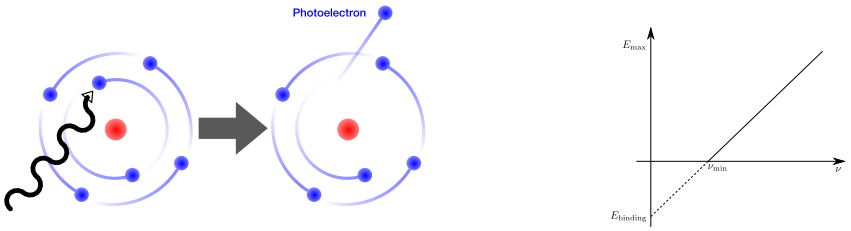
$$I(x) = I_0 e^{-x/\mu} \tag{2.24}$$

where I_0 is the initial intensity of the incident beam, x is the thickness of the absorber, and μ is the radiation length. The radiation length is the distance over which the electron energy is reduced by a factor $1/e$ due to radiation loss. μ, can be found using the total cross section with atom density, N and it is the inverse of the mean free path of the photon.

$$1/\mu = N\sigma = (N_A\rho/A)(\sigma_{\text{photo}} + Z\sigma_{\text{compton}} + \sigma_{\text{pair}}) \tag{2.25}$$

2.7.3.1. Photoelectric Effect. Photoelectric effect is the name of the physical process where an atom absorbs a photon and immediately afterwards ejects one of its electrons. The energy of the ejected electron is

$$E_{\text{max}} = h\nu - E_{\text{binding}} \tag{2.26}$$



(a) Photoelectric effect [19] (b) E_{max} vs. ν

Figure 2.23: Photoelectric effect

This process occur only for electrons bound to an atom, because a free electron can not absorb a photon and conserve momentum, but the nucleus can recoil while electron ejection.

The relation between cross section and atom number in MeV range is: $\sigma \propto Z^4$ or Z^5. Therefore higher-Z materials are more favourable for photoelectric absorption.

2.7.3.2. Compton Scattering. Compton scattering is the simplest electromagnetic interaction, which is scattering of a photon on a free electron. (If the photon energy is high enough such that $E \gg E_{\text{binding}}$, then bound electrons can be assumed to be free.)

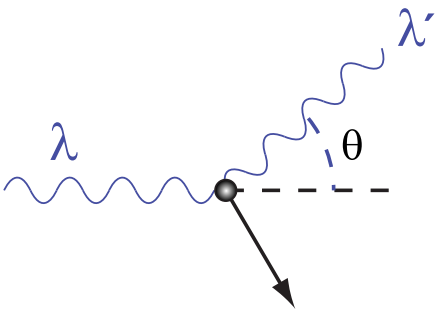


Figure 2.24: Compton scattering [20]

From energy and momentum conservation relations for θ and ϕ can be found [15]:

$$\begin{aligned} \cos(\theta) &= 1 - \frac{2}{(1+\gamma)^2 \tan^2\phi + 1} \\ \cot(\phi) &= (1+\gamma)\tan\frac{\theta}{2} \\ \text{where} \quad \gamma &= \frac{h\nu}{m_e c^2} \end{aligned} \tag{2.27}$$

The cross section of Compton scattering is one of the first calculations made using QED by Klein-Nishina [15].

$$\frac{d\sigma}{d\Omega} = \frac{r_e^2}{2}\frac{1}{\left[1+\gamma(1-\cos\theta)\right]^2}\left(1 + cos^2\theta + \frac{\gamma^2(1-cos\theta)^2}{1+\gamma(1-\cos\theta)}\right) \tag{2.28}$$

Its integral over Ω gives the total cross section, σ_c, which is related to the total probability per electron.

Total cross section can be thought as the sum of two cross sections $\sigma_c = \sigma_a + \sigma_s$, where σ_a is the absorption and σ_s is the scattering cross section. σ_s and σ_a are related

to the average energy remained in the scattering photon and transferred to the recoil electron respectively. Recoiled electron is stopped by the material, hence its energy is the energy absorbed by the material [15].

$$\frac{d\sigma_s}{d\Omega} = \frac{h\nu'}{h\nu}\frac{d\sigma}{d\Omega}$$

For energies lower than $m_e c^2$ the Klein-Nisina formula approaches to classical scattering formula, called "Thomson scattering". $\sigma = \frac{8\pi}{3} r_e^2$. Sometimes all electrons contribute to the scattering together in a coherent manner. This phenomenon is called "coherent scattering" or "Rayleigh scattering". In both processes no energy is transferred to the medium by excitation or ionization. Only the direction of the photon is changed.

At higher energies their cross section is so small that these effects can be neglected.

2.7.3.3. Pair Production. If the energy of a photon is higher than the total masses of an electron and positron, $E_\gamma > m_{e^-}c^2 + m_{e^+}c^2 = 1.002$ MeV, there is a probability that an electron-positron pair is created. To preserve momentum, this can occur only around a third body, which is usually a nucleus.

The combined effect of pair production and bremsstrahlung of high energetic photons may give result to an electron-positron shower. A photon in matter turns into an electron-positron pair which then emits high energetic photons by bremsstrahlung, which turns into another e^-, e^+ pair, and so on, until last pairs with low energies are stopped by atomic collisions not by bremsstrahlung.

After number of n radiation lengths the total number of particles (γ, e^-, e^+) will be $N \simeq 2^n$ and each one have an energy on the average $E_n \simeq E_0/2^n$. The maximum value for n can be found using the assumption that the electrons stops abruptly when their energy is below a critical energy E_c. $n_{\max} = \ln\left(\frac{E_0}{E_c}\right) / \ln 2$ and $N_{\max} \simeq E_0/E_c$

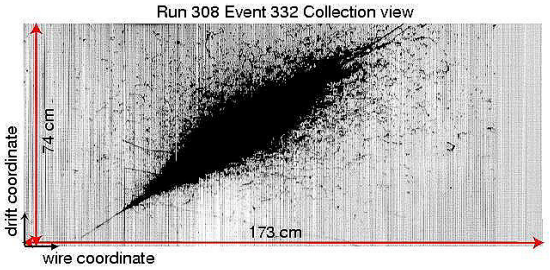


Figure 2.25: Electromagnetic shower observed in the ICARUS LAr drift chamber during the technical run with cosmic rays at Pavia, summer 2001 [21].

But of course the electrons do not stop instantly and go some radiation lengths further, and because there are many bremsstrahlungs involved fluctuations from the average is high. dE/dx can be fit well with a gamma distribution

$$\frac{dE}{dx} = E_0 b \frac{(bn)^{a-1} e^{-bn}}{\Gamma(a)} \tag{2.29}$$

where a and b are parameters which depend on the material.

While the cascade goes through the matter its width also increases due to finite opening angle between electron and positron while pair production which directs particles away, multiple scattering of electrons and the emission of bremsstrahlung photons away from the axis. Knowledge of the longitudinal and transverse profiles of the showers, called "jets", is important for the design of electromagnetic calorimeters.

2.7.4. Detection Methods and Some Important Aspects of Detectors

The basic principle of all detectors is to transform some or all energy of the radiation into a form which is more accessible to human perception and information processing technologies. The ionizing material, the type of the converted energy and its observation method also defines the type of the detector. For example,

- Gaseous detectors collects the freed electrons due to ionization and creates an electrical current signal.

- Scintillator detectors transfers the radiation energy to excited states of molecules which later turns into emission of light while decaying into the lower energy states.
- In photographic emulsions, chemical reactions are induced by ionization which creates an image of the trajectory.

Because of the advances in electrical technologies in terms of precision, signal manipulation, information processing and storage, detectors, in the end, transform their information about the radiation into electrical impulses which are further processed by electrical means.

2.7.4.1. Sensitivity. A detector can not detect every kind of radiation. Detectors are specialized for a specific type of radiation and they are most responsive to that radiation if it is in a specific energy range. Sensitivity of a detector is its capacity to convert the radiation energy into a usable, meaningful signal. Hence detectors are designed to be sensitive to certain types of radiation in a certain energy range.

Sensitivity depends on

- The cross section of the main interaction between detector material and radiation,
- The size and mass of the detector,
- Inherent noise and external noise sources.

As the amount of the sensitive volume in the detector (the mass of the detector) and the cross-section, the sensitivity also increases (if the detector is able to convert the ionization energy efficiently.) The inherent noise is the fluctuations in the voltage independent of the presence of any radiation. Hence the ionization energy must be higher than the average noise level.

2.7.4.2. Detector Response. The detector response is the relation of the electrical output signal and the energy loss of the incident particle which converted to ionization energy. The output signal is generally in the form of pulses. Therefore the integral of

the pulse over time, or if the pulse shape does not change for successive pulses, the pulse height contains information about the energy loss. Altough for many detectors the response is approximately linear over a certain energy range, the relation is not always linear, the output may saturate or depend on the radiation etc. If it depends in the radiation source, that information can be used to identify the type of the particle.

2.7.4.3. Energy Resolution. The energy resolution of a detector is its ability to distinguish two different energy values. It can be measured by looking at the energy spectrum of a monoenergetic radiation. Theoretically the spectrum of a radiaton with single energy component is a dirac delta distribution, but detectors give a gaussian distribution around a mean value (if they absorb all of the energy). Resolution is a function of energy and defined by $\Delta E/E$ where ΔE is the sigma of the gaussian distribution. Resolution = $\Delta E/E$.

2.7.4.4. Response Function. The response function is the pulse height distribution of the detector output due to a given type of monoenergetic radiation. In real life this distribution may not even gaussian, especially for neutral radiation. For example, if some of the electrons leak out of the detector, then not all of their kinetic energy is absorbed, hence a low energy tail occurs in the spectrum. Similarly, photons due to compton scattering escaping from the detector leads to the same result etc. Figure

The pulse height spectrum of a measurement is the convolution of the incident radiation spectrum and the response function.

$$H(E) = \int S(E')R(E, E')dE' \tag{2.30}$$

where $H(E)$ is the pulse height spectrum, $S(E)$ is the radiation spectrum and $R(E, E')$ is the response function at energy E'.

2.7.4.5. Response Time and Dead Time. The response time is the time interval from when the incident radiation has arrived to when the output signal is formed. It is

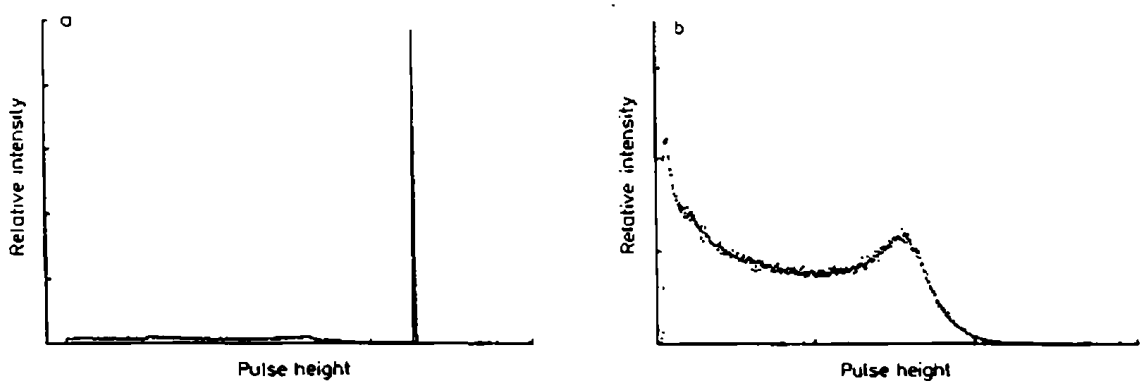


Figure 2.26: The response functions of tho different detetors for 661 keV gamma rays. (a) shows a germanium detector with a large photoelectric cross section relative to Compton scattering at this energy. (b) shows the response of an organic scintillator detector [15].

desired to have a rise time curve which is as close to vertical as possible, in order to be able to mark the event time with high precision. The decay time of the pulse is related to the dead time of the detector. If a new particle hits the detector before previous output signal decays they may pile-up.

Dead time is common to all signal processing components such as DAQ systems and detectors. It is the time it takes after the component gets the input until it is available to take a new one, in which it processes an event.

A detector may be insensitive during the dead time. Then, if a new input comes before the dead time, it can not be processed and lost. A detector may also be sensitive. If so, the new input may pile up with the previous one and both events may be wasted.

In a probabilistic process, it is impossible to eliminate all losses due to dead time. For example, if the event occurence probability is a uniform distribution, then the time interval between successive events is a Possion distribution, which is non zero for all time values. Hence there is always the probability that a new particle comes before dead time. Thus, if there is dead time in the system, then data loss is inevitable.

2.7.4.6. Detector Efficiency. The absolute (or total) efficiency is the ratio of the events registered by the detector to the number of particles emitted by the source.

$$\varepsilon_{\text{tot}} = \frac{\text{events registered}}{\text{events emitted by source}} \tag{2.31}$$

which is a function of detector geometry, placement of the source, and interaction probability.

There is also the intrinsic efficiency, which is the ratio of the events registered by the number of actual hits to the detector.

$$\varepsilon_{\text{tot}} = \frac{\text{events registered}}{\text{events impinging on detector}} \tag{2.32}$$

2.7.5. Main Detector Components

2.7.5.1. Ionization Detectors. Ionization detectors are working by collecting the freed electrons due to ionization of molecules of the medium and creating an electrical pulse signal from them. The space between opposite charged electrodes is filled with ionizable gas (or liquid) and an electric field is applied to the medium. When charged particle passes through it breaks electron bounds of the molecules and make them free. Because there is an applied electric field, the ion-electron pairs are not recombined under their own electrical attraction, electrons are collected at the positive charged electrode and create some current.

The amount of the applied voltage has different effects. When the voltage is increased more electrons are collected before they are recombined hence the current increases. From some point the freed electrons are accelerated so much that they can ionize the medium further, which creates a cascade. The number of secondary ionizations is proportional to the initial number of freed electrons, hence this system is called proportional chamber. If the voltage increased further, a situation where the gas as a whole discharges and the system saturates. There, the output amplitude is

always the same. This system is called Geiger-Müller

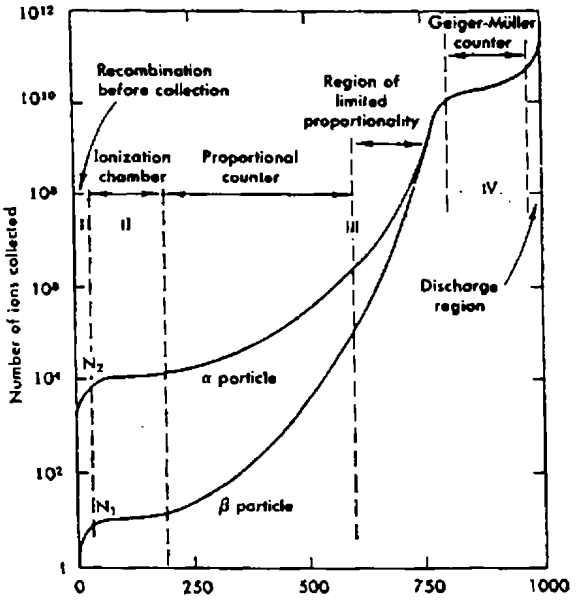


Figure 2.27: Number of ions collected versus applied voltage in a single wire gas chamber [15]

Some of the historical examples of ionization detectors are ionization chamber, proportional counter and Geiger-Müller counter. Today multi-wire proportional chambers, drift chambers and time projection chambers uses this mechanism.

The multi-wire proportional chamber and drift chamber consist of anode wires on a grid with a grid size in the order of milli meters. The ions are collected by the nearest wire. Every wire is connected to a different input of the DAQ system. By looking at from which wire the signals are coming and the incoming time that pulses, the trajectory of the particle is deduced. The spatial resolution is determined by the denseness of the wires.

2.7.5.2. Scintillator Counters. For some materials, a portion of the incoming particle's dissipated energy by the molecular excitation is reemitted as visible or ultra-violet light. The light flashes produced by individual ionizing particles are called "scintillations" and detectors using these fact are called scintillator counters.

Scintillator is a kind of material which has the property of luminescence, which

means emitting light at relatively low temperatures. (For example, phospors has the property of luminescence.) When hit by an incident charged particle, the particle transfers some of its energy to the molecules of the scintillator, and excite them. After the decay time the molecules return to their ground state by emitting photons. If the transition from higher state to lower state occurs in nano seconds, this kind of luminescence process is called fluorescence. If the excited state is metastable and returning to ground state takes more time, the process is called delayed fluorescence (microseconds) or phosphorescence (hours).

The distinctive property of scintillating materials is that they are not continuous radiation sources but emit light in discrete bursts as flashes. This property was first observed by William Crookes in 1903. In the beginning the scintillations are observed visually in dark rooms. Later, they are converted to voltage pulses using photo-multiplier tubes.

The decay time probability of one molecule is exponential distribution with a mean of τ. The number of the photons emitted in time is related to the number of excited molecules in an event and the decay times of different types of molecules. Generally N is modelled with linear superposition of two exponential decays [15].

$$N(t) = A \exp\left(-\frac{t}{\tau_{\text{fast}}}\right) + B \exp\left(-\frac{t}{\tau_{\text{slow}}}\right) \tag{2.33}$$

Scintillators are used widely because of their useful properties such as

- Energy Sensitivity: Most scintillators response linearly to the energy of the incident particles. PMTs are also linear devices. Hence the amplitude of the electrical output of the PMT is proportional to the number of emitted photons from the scintillator which is also proportional to the dissipated energy.
- Fast time response: Their response and recovery times are generally shorter than other detectors, which makes them to give more precise timing information and have faster event rates. Thus, they are used for measuring event time differences.

- Pulse shape discrimination: For certain type of scintillators, the shape of the output pulse signal depends on the type of incident particle, because different fluorescence mechanisms are excited by particles of different ionizing powers. Deciding the type of the particle by looking to the pulse shape is called Pulse Shape Discrimination (PSD). (See Figure 2.28)

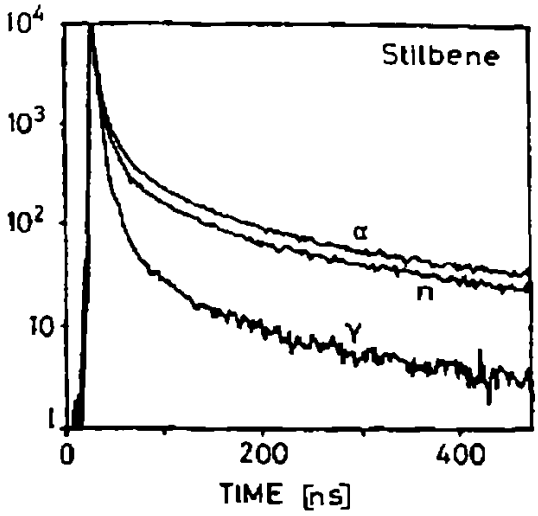


Figure 2.28: Pulse shape of stilbene light for α particles, neutrons and γ rays [15].

There are organic, plastic, crystal and gaseous types of scintillators.

2.7.5.3. Photomultiplier Tube. Scintillators are not useful by themselves because the amount of the light emitted is hard to see and light is not a good energy form to process information. Therefore they are used together with PMTs. The geometry of the scintillator is adjusted such that emitted photons are reflected mostly to the direction of a waveguide, which collects and directs the photons to the PMT and which is non sensitive to the ionizing radiation. The scintillator is also covered with a material for not letting the photons escape. Figure 2.29 show a schematic of scintillator PMT system.

A PMT is an extremely sensitive photon detector of which mechanism is based on photoelectric effect. It converts light to electric current, some of them react even to a single photon.

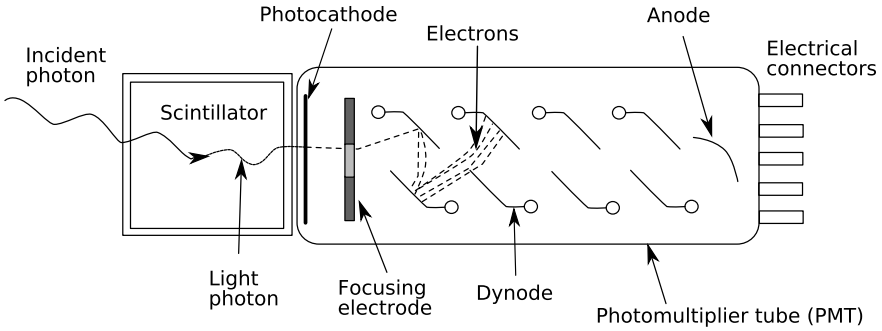


Figure 2.29: Schematic of a coupled scintillator PMT system with internals of PMT [22]

A PMT is a vacuum tube. On one side, there is a photocathode, which is a semiconductor with a very low work function. When a photon with enough energy hits the photocathode, an electron, called photoelectron, is emitted. The probability of producing a photoelectron after a hit of a photon is called "quantum efficiency", which depends on the wavelength of the photon. A reasonable quantum efficiency value for a good PMT at 400 nm wavelength is 0.25.

The photoelectron is accelerated with a voltage difference. Accelerated electron hits another electrode called "dynote". There it frees other electrons and they are accelerated together towards the next dynode. A typical PMT includes 10 dynode stages. After the last dynode the number of electrons exceed 10^6, which is an easily detectable amount with an external circuit, and an electrical signal is generated.

PMTs are isolated from external magnetic fields by coatings made of an allow called mumetal which has high magnetic permeability, because the electrons has to be focused on the dynodes via internal fields.

2.7.5.4. Calorimeters. Calorimeters are used to measure the total energy of a particle. Entering particles initiate a particle shower and the energy of that particles is deposited and measured by the calorimeter. There are electromagnetic and hadronic calorimeters.

The electromagnetic calorimeters are for particles with electromagnetic interactions such as electrons, positrons and photons. Their size is large enough to hold all of the electromagnetic cascade, which consist of successive brehmstrahlung photons and pair production electron/positrons, inside themselves. Typically the length to absorb all of the secondary particles corresponds to 15-25 radiation lengths.

The hadronic calorimeters are used to detect particles such as protons, charged pions, charged kaons, neutrons, etc., which interact with strong force. Their difference is that the particle shower is initiated by strong interaction.

From another perspective the calorimeters are divided into two: sampling and continuous calorimeters. The sampling calorimeter is filled with two different materials. One is chosen to produce and evolve the particle shower in a limited volume and other one is specialized in measuring the deposited energy. These materials are placed alternatingly along the particle trajectory.

They may be an array of plastic scintillator layers which are a few mm thick, and lead layers which are 1 mm thick. These calorimeters are called sampling calorimeters because a fraction of the all ionization is detected (sampled) by material placed at regular intervals.

The shape of the shower may give information about the type of the particle.

2.7.5.5. Other Types. In the "cloud chamber", invented by C.T.R.Wilson, a gas is supersaturated with vapour and the vapour condenses and create small droplets around the ions generated around the trajectory of the incident particle. It is used to determine the particle trajectory, and by applying a magnetic field, the electrical sign.

In "photographic emulsion" a photo-chemical reaction which makes the grains darker occurs along the path of the particle. It is used to measure the cross-sectional area location of the incoming beam.

2.7.6. Spectrometers

Spectrometers are combinations of different types of detectors, specialized detectors placed one after another, which are used to detect all possible decay products of an interaction.

Not all of the particles can be detected directly. Most of them has a very short lifetime, τ, which prohibit them to have a meaningful interaction with the detection medium. They decay to one of the stable particles in multiple stages. Decay process of charged pion, π^+ is given as an example:

$$\pi^+ \rightarrow \mu^+ + \nu_\mu, \quad \mu^+ \rightarrow e^+ + \nu_e \overline{\nu}_\mu \tag{2.34}$$

Table 2.2: Particles which can be detected with detectors [18]

Particle	c τ (m)	Detection method
γ	stable	Electromagnetic calorimeter
e^-, e^+	stable	Tracking and electromagnetic calorimeter
μ^-, μ^+	6.59×10^2	Tracking with penetration of many interaction lengths
π^-, π^+	7.80	Tracking and hadronic calorimeter
K^-, K^+	3.71	Tracking and hadronic calorimeter
K_L^0	15.5	Hadronic calorimeter
p	stable	Tracking and hadronic calorimeter
n	$2.69 \times 10^1$1 m	Hadronic calorimeter
ν	stable	Momentum conservation

Table 2.2 gives the list of absolutely stable particles and particles which has enough lifetime to be able to travel several meters, their path length in their mean lifetime if their velocity is speed of light and their detection method. Spectrometers are used to distinguish them by their means of interactions. The track they create along the spectrometer is like a signature.

A spectrometer may include a part for charged particle tracking in a magnetic field to measure the momentum, an electromagnetic calorimeter to measure photon and electron energy, a hadronic calorimeter to measure hadron energy, a muon chamber, etc. Figure 2.30 show a schematic of an ideal spectrometer which includes all of the mentioned parts, and the types of interactions when a high energetic particle from the previous table passes through it.

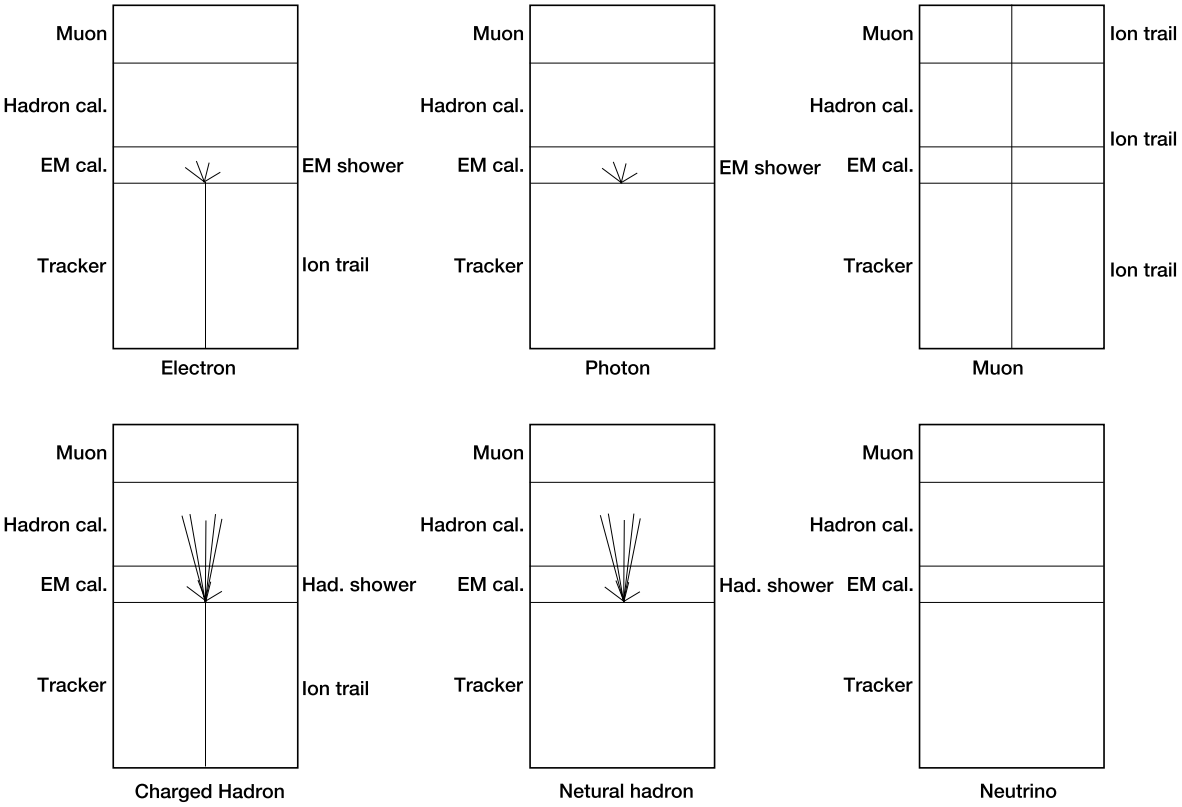


Figure 2.30: High-energy particle interactions. Adapted from [18]

One of the main jobs of the DAQ systems is to analyze the interaction patterns occurring in the spectrometer and to decide whether it is an important event for the research problem at hand. For example, in some experiments identification of rare events of primary particles, say an electron, is needed when there is a constant "noise" due to background hadrons, such as pions. Then the trigger in the DAQ system should signal only when the pattern of the electron, as depicted in the picture book, has taken place, but do not react to the pion pattern.

The spectrometers which are mainly cylindrical and cover all 4π solid angle, are

called hermetic detectors. ATLAS and CMS are this kind of huge detectors.

2.8. Data Acquisition Systems

DAQ is the part of the experiment where the raw data due to the events gathered by detectors (which are mostly electrical signals in pulse form) are filtered, collected, categorized and stored to make them ready for further processing and analysis.

Data acquisition systems are the essential parts of experimental physics and applied science. Although some branches of science are quite advanced, its rate of growth is held back by the DAQ systems, due to their several drawbacks such as their inability to handle high event rates. Experimental physics that depends on charged particle and photon detectors, photonics, science that depends on ultra-fast lasers, experiments that requires real-time and feedback systems, single-shot ultra short beam bunch-shapers, and even chemistry are among those which can take advantage of a better DAQ system.

To design a DAQ system for HEP experiments requires a collaboration of physics, electronics, computer science and networking, some of which are explained in this thesis to some extend.

A DAQ system usually starts with a trigger which is responsible for filtering uninteresting events such as noise or background radiation. Triggers should select interesting events as early, efficient, fast as possible by using all of the available information. The may have to do it in stages. Some information is available on the signal level (LV0) Some information needs signal processing and is available only later (e.g. energy sums) (LV1, LV2, . . .)

By using the signals coming from triggers and the detectors together, the DAQ system does following jobs.

- Readout: To gather the data produced by the detectors.
- Triggering: To feed higher level triggers of which decision systems are finer than

the first triggering stage.

- Event building: To form complete events using the fractions of the same event coming from different sources.
- Data logging: To store the event data.
- Run control and monitoring: To provide control, configuration and monitoring facilities.

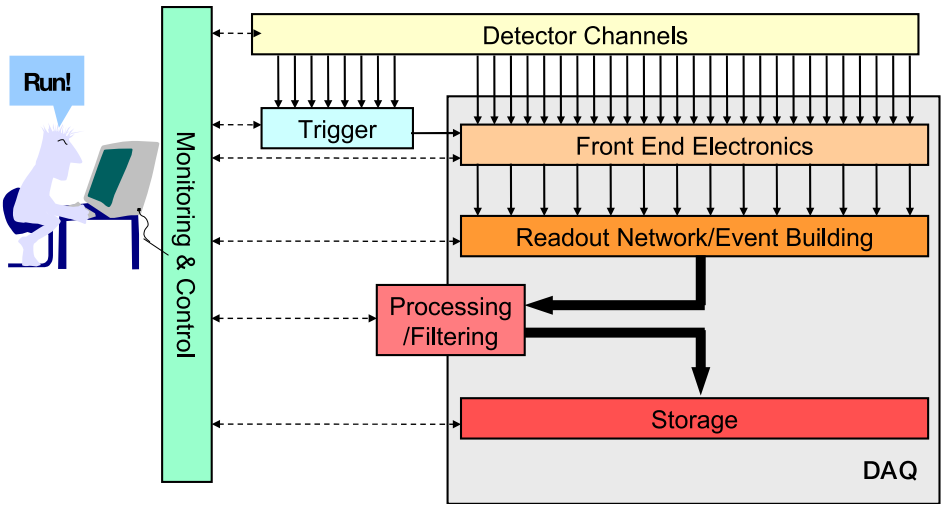


Figure 2.31: A generic DAQ system

Figure 2.31 shows a generic DAQ system for a moderate size experiment. Detector channels are connected both to triggers and front-end electronics. Front-ends digitize the detector outputs according to the warning signals from triggers. Then the digitized data is sent to readout networks for event building. Readout networks joint raw data and add some additional information as headers and build event frames. After some filtering and processing done on these event frames they are stored in computer memories. All of these stages are constantly monitored and controlled by another layer which is connected to every stage.

2.8.1. An Example of a Basic DAQ System with Trigger

Figure 2.32 shows the schematics of the DAQ system of β-decay measurement, which is a basic stochastic process.

If an electron is ejected and interacts with the sensor, the sensor generates an

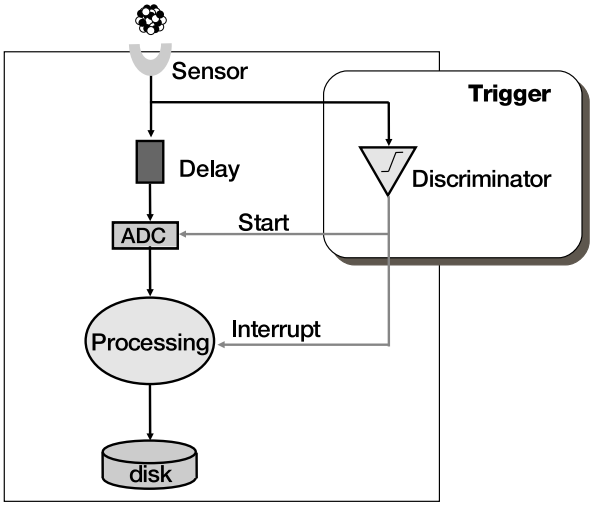


Figure 2.32: Basic DAQ system for recording stochastic events such as β decay [23]

.

electrical pulse, which goes to a delay line and to the trigger.

The trigger uses a discriminator to notice the presence of a pulse and sends "start recording" signal to the ADC, and "stop what you are doing and start processing the new pulse" signal, which is called "interrupt", to the processing unit.

Because the processing done in the trigger also takes some time, say T_{tr}, the signal is delayed before connecting to the ADC. The delay amount is adjusted to be equal to the process time of the trigger. $T_{\mathrm{del}} = T_{\mathrm{tr}}$ In this way, the pulse arrives to the ADC at the same time when the start signals arrives.

β decay is a stochastic process. The interarrival time between successive electrons is an exponential distribution with a mean value $\tau_\beta = 1$ ms. (Average event rate is $f_\beta = 1$ KHz.) Say, the DAQ systems processing time is the same, $\tau_{\mathrm{DAQ}} = 1$ ms.

In this situation, there is the probability that a new electron comes earlier than τ_β, hence before the DAQ finishes to process the previous pulse. If this occurs, according the design of the system, the signal which is being processed while the new one comes will be lost, by the interrupt from the trigger.

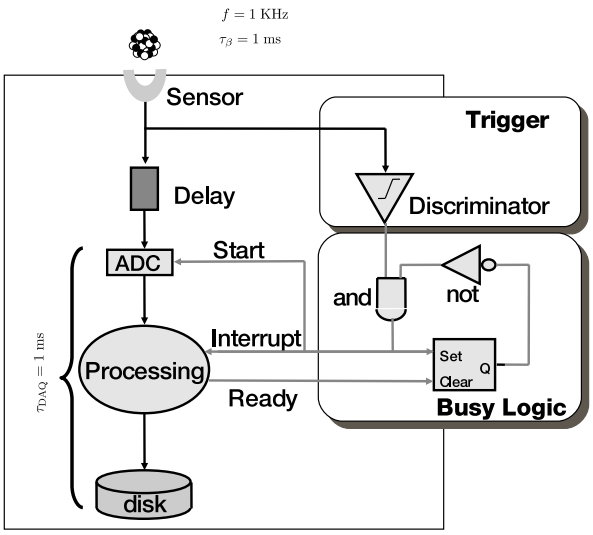


Figure 2.33: Basic DAQ system with busy logic

In this configuration there is a busy logic, where there is a register which holds the busy flag. When the new pulse comes and if system is non-busy, both inputs of the AND gate is high and "start" is sent to ADC and processing unit, and busy flag is set to high. The busy flag is lowered by processing unit after it finishes its processing. This way, if a new pulse comes while processing, the system is not interrupted and the new pulse is lost rather than the previous one.

The ratio of the number of registered events to the number of actual events is important. $\epsilon = \frac{N_{\text{reg}}}{N_{\text{tot}}}$ Number of total events can be estimated from the recorded events if this ratio is known. To find it, first, the average count rate of the detector-DAQ system, ν, should be known.

If the dead time and average event rate is given, which are τ and f successively, then in a time interval T,

- the total number of events is $N_{\text{tot}} = fT$,
- the number registered events is $N_{\text{reg}} = \nu T = n_r$,
- the sum of all dead times is $\tau_{\text{tot}} = n_r\tau$ (for every recorded event the system remained idle for τ ms),
- the number of lost events is $N_{\text{lost}} = \tau_{\text{tot}} f = n_r \tau f$ (this number of events are

generated while the system is dead),

The number of generated events must be equal to the sum of registered and non-registered (lost) events.

$$\begin{aligned} N_{\text{tot}} &= N_{\text{reg}} + N_{\text{lost}} \\ fT &= n_r + n_r \tau f \\ \Rightarrow f &= \nu(1 + \nu f) \\ \Rightarrow \nu &= \frac{f}{1 + \tau f} \end{aligned} \tag{2.35}$$

ϵ is also equal to the ratio of ν to f.

$$\epsilon = \frac{N_{\text{reg}}}{N_{\text{tot}}} = \frac{\nu}{f} = \frac{1}{1 + \tau f} < 1 \tag{2.36}$$

which is always less than 1. Hence some of the events are always lost in this setup. For the values we chose ($f = 1$ KHz, $\tau = 1$ ms) $\epsilon = 0.5$. Which means half of the events are lost.

To deal with these losses an approach called "de-randomization" is used. A "First In First Out" buffer (FIFO) is added on the signal chain between ADC and the processing unit. See Figure 2.34.

FIFO is a data structure which includes N number of elements. (N is called the depth of the FIFO.) When the object in the last cell is taken all of the elements are shifted from input to output. If the objects are added from the input without anything is taken from the output they fill the FIFO memory.

In this system: If the memory is full then no new object can be taken and FIFO sends a "full" signal to the busy logic. If the FIFO is non-empty (if it includes at least one recorded pulse) it sends "data ready" signal to the processing unit and the

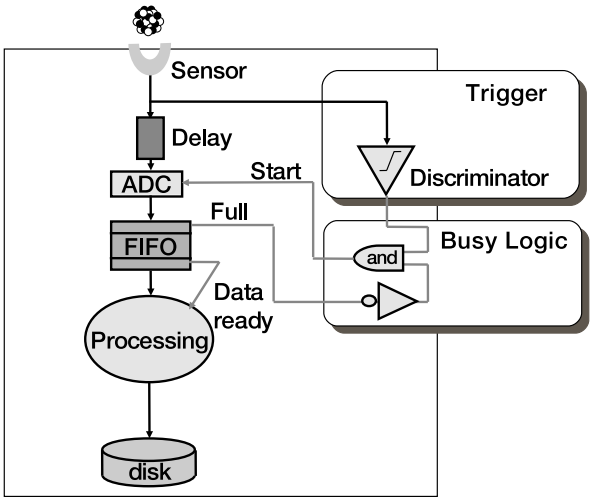


Figure 2.34: Basic DAQ with a derandomization mechanism

processor takes the oldest pulse data from the FIFO. If the FIFO is empty then nothing happens and the system waits for new pulses.

It can be seen that if the mean interarrival time is higher than the processing time, then eventually the FIFO will be filled and the system will start to lose events. But if the opposite is true, the FIFO eliminates the event losses by changing the stochastic behavior of the source to a periodic from the perspective of the processing unit.

The use of high energies, the need of increased time and spatial resolutions and higher event rates for discovering rare events, demands new technologies for information processing and data acquisition systems. In recent HEP experiments DAQs and successive analysis have to deal often with very challenging data rates. Hence they require efficient data flow and efficient algorithms

2.9. Literature Survey

The first functional and conceptual kind of the system designed in this work is introduced by Massaschusetts Institute of Technology (MIT) in 1999 [24, 25, 26] for a medium scale experiment [27] with approximately 2000 channels, used in Compton electron beam polarizer. And it is still in use.

MIT/Bates Compton polarimeter is used to measure the polarization of an polarized electron beam with 200 mA current and 800 MeV energy (which was world's highest current polarized electron beam), collected in a storage ring, at the center of Bates linear electron accelerator complex. With this setup, electron beam is hit with a 532 nm, 5 W monochromatic, continuous-wave laser. High energetic photons emitted due to Compton process are catch with a CsI detector, their energy is measured and their energy spectrum is obtained. See Figure 2.35. This process is repeated after inversion of the laser and electron polarization. At the end, from the asymmetry of the cross section which depends on the photon energy, the polarization of the electron beam is obtained.

But because the symmetries are under 1% level and the measurement must be finished within 15 minutes, the recorded event rate must be very high. To get 5% statistical resolution using this setup, a DAQ system which can handle an average pulse rate of 100 Hz is needed. The energy of the photons captured by the CsI detector must be measured sensitively minimum 10^5 times per second. A DAQ system which can handle this high trigger rate is not possible with traditional methods.

For this purpose, Akdoğan and his colleagues build a VMA and ADC based sampling system for the first time. Pulses sampled with 100 Msps samplerate were recorded into appropriate windows using a simple threshold trigger, and integration and pile-up rejection algorithms are applied on them. This is a dead-timeless system and provides the needed high resolution of photon energy easily. Altough the sampling is made into 12 bits, due to the integration of a 16 samples window, a dynamic range which is approximately 14-15 bits is obtained. The pulse-shape discrimination is also realized online for first time using a sampling-ADC. Because the time resolution is not needed for this application, no data about the time information is provided. All of the signal processing is done by the multi-purpose CPU on the control card of the VME. Because of that, it is limited in terms of signal processing abilities. However, it is an important work, because of its ability to make on-the-fly pulse-shaped signal processing without transmitting the raw data, and showing that this principle can work actually. Our work is based on this work to some extend.

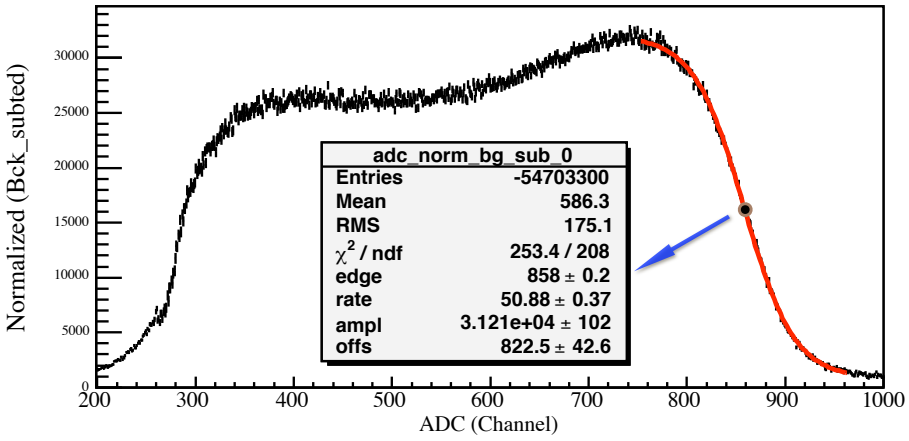


Figure 2.35: Photon energy spectrum (which is a Compton spectrum) obtained by the VME based system. It should be noted that the energy of each one of approximately 60 million photons is measured individually and all of measurements are done in a total time of 10 minutes.

In the work of Hien and Senzaki [28] from Japan in 2001 for nuclear spectroscopy, a FPGA is added to increase the data transmission speed and to decrease the dead-time. the FPGA is not used for signal processing, but it is a step to the right direction. The samplerate reached by this work is 40 Msps and the deadtime is at the 1 μs level for each pulse. Although it seems that it is behind the previous work, it is an important step of the FPGA usage in data transmission.

In 2002, Bolic and Drndarevic [29] developed a FPGA based photon spectroscopy system. Their digitizer was a 8-bit, 60 MHz modest unit. Using this system, pile-up rejection and pulse amplitude measurements are done on the FPGA. This work brings the "digital pulse processing" concept into the FPGA design world for the first time. Like MIT group, they reduced the number of analog circuits and hence the hardware complexity to minimum. The difference is that the processes are not made with a general purpose CPU but with a FPGA. This makes the channel number density scalable. As a matter of fact, it is possible to use one FPGA per channel.

In the first half of the 2000's with the expected fast developments in FPGA technology, the usage of the high density integrated circuits in the DAQ systems is

increased. In 2006, the German group of Streun [30] made a DAQ sub-system of a positron emission tomography (PET) system using non-stop sampling ADC. In a PET system, measuring the electron-positron annihilation time is as important as the energy resolution. Streun and his colleagues used an 12-bit 40 Msps system, of which sampling period is 25 ns. Nevertheless, by applying a linear interpolation to the samples belonging to the rise-time of the pulse with the FPGA, the pulse starting time is obtained with a 2 ns precision between successive sample times. This work has two important aspects: 1) it is a significant step for real-time signal processing using FPGA in physics DAQ systems, 2) it shows the need for such systems in medical research.

Roughly at the same time, Khomich et. al. [31] developed the FPGA usage in DAQ systems one step further. In that work, the German group build a FPGA based fast decision trigger system for one of the new detectors of CERN. Although it is not directly related to DAQ, it is significant because of its filtering of the incoming data, due to the limitations of the used DAQ systems. Here, not the pulse-based processing but the relations between channels is aimed.

Also in 2006, Nicolau [32] from Italy designed a FPGA based DAQ with a sampling ADC. The most important contribution of this work to the literature is that, it decreased the time resolution under 1 ns, down to 0.3 ns, with a 200 Msps 8-bit ADC. Unfortunately, here FPGA is used as a control unit, with the target of data compression. The system is again triggered with a simple threshold value, but there is no filtering according to the pulse shape.

Muller et. al., in an international collaboration, developed a DAQ system for a stable photodiode-based photon detector with high energy resolution (14-bit dynamic range) [33]. The ADC they used had 10-bit resolution and a limited buffer memory. They compromised from the triggering ratio and sampling rate (10 Msps) which can be handled. Here, signal processing can be done only offline. Because no processing is done on the system, the signal is filtered and conditioned before being digitized. This, however, in conditions of high noise or background signals, make the real signal and the background signal indistinguishable. Other than the resolution, the contribution

of this work to the literature is that it is a 32 channel compact system.

In 2007, the first literally dead-timeless DAQ system was made by an Italian group Giachero et. al. [34]. With this VME based system which uses a 18-bit sampling ADC quite sensitive measurements are realized. But in this system FPGA is not used for sample processing and it can only work with low sampling rates such as 5 KHz for high resolution.

After the previously mentioned work made by Streon et. al., a Japanese group, Shimazoe et. al., built again a DAQ system for a PET scanning system which is important for health services [35]. The multi-channelness of this type of systems is significant for spatial resolution of the tissue. For this reason, Shimazoe et. al. designed an application specific integrated circuit (ASIC). Eventhough this work is conceptually different from other works, it is an important example of how necessary the digital pulse processing and this kind of scalable, multi-purpose systems are. Because they had to develop high cost ASIC, when this kind of systems do not exist. Another importance of this work is that, although ASIC is used, some part of the signal processing is done on FPGA to gain some flexibility.

The work done by Guo [36] et. al. in 2005 should be mentioned. This work gives an idea of what are the future aims of DAQ with FPGA. The main goal is to have a fast and programmable data processing system. In this work, a FPGA with 10 GHz input is introduced, which is orders of magnitude faster than previously mentioned works. This shows that all of the pulse-based signals used in experimental physics can be handled with this kind of systems.

Finally, in the light of developments, the ADC and FPGA [37] which give quiet flexible application fields and which we are using in our project should be mentioned. With the high-tech chips and digitizers this cards are using, we are reaching our goals.

The usage of digitizers and real-time pulse processing is accelerated in the last 5-6 years due to digital revolution in technology. But it must be accepted that this is a

new field and it is needed to be standardized. Every system in the literature has some inadequate parts and is designed for one purpose. For that reason, because they are specialized and application-specific, it is impossible to use them in other projects.

Our goal is to have a multi-purpose system which can meet the DAQ needs of various kind of experiments. To be useful in experiments and measurement which are previously mentioned, it has to have high resolution in both timestamp and pulse height. Furthermore, the abilities that it process pulses in real-time and that the fit parameters are recorded as data, were not realized previously.

2.10. Our Work

In this section we will explain the aim and design of our work.

2.10.1. The Aim

The growth of technology has an exponential form, and the best example of this is the Moore's law, which can be witnessed in our daily life. A similar or faster growth is apparent in data storage, data transfer, analog-to-digital converter (ADC), and Field Programmable Gate Array (FPGA) technologies. We propose to design and build a prototype of a next generation digital data acquisition technology that makes use of the latest technologies mentioned above, and we propose to replace a nearly 40-year-old aged DAQ technology that is still being used today.

Our primary aim is to present a digital module with a significant computational power that has the same functionality as the collection of many special purpose modules used to record the properties of a pulse-based signal. Thus, this will reduce the failure rate of the hardware of many-channel systems by moving the complexity of the system from hardware to software.

Today, physics experiments require an amazing number of data-channels. This number reached as high as six digit numbers for some experiments, thus traditional

DAQ systems became highly complex in terms of hardware; the efficiency of these systems are low due to dead-time and failure rate of such a huge number of modules. As a result, we believe that the timing of this project is appropriate because of a need for such systems that require minimal deployment time and maximum flexibility due to the programmability of the module/system.

With only one pulse (See Figure 2.36) it is possible to get detailed information about the quantity to be measured. For example, pulse height gives information about (single shot) photon energy. Decay time, if the signal comes from a particle detector, gives information about the interaction with the detector and hence about the particle type. This kind of analysis methods which depends on the shape of the pulse are called "pulse shape discrimination" (PSD) or "particle identification" (PID).

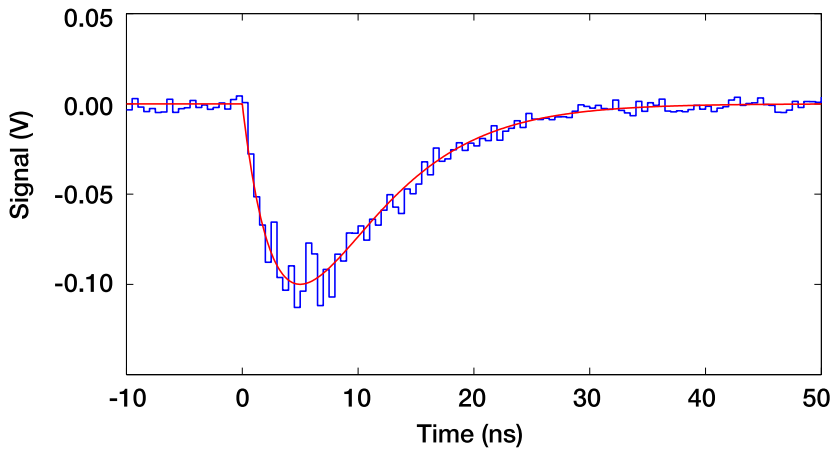


Figure 2.36: A typical pulse based signal. Digitized version of the analog signal and the fitted smooth function superimposed.

Trying this in real-time with traditional methods is cumbersome and inefficient. As an example, the one channel system still in use in Los Alamos is shown in Figure 2.37. Every element in the figure is a different module. This system only measures the pulse time and amplitude and it needs to be dead for some time after each pulse to reset itself.

Another area of use of of the decay time is the life time measurements, which is important in chemistry. Our prototype has sampling period of 2/3 ns, and with some special treatment, such as interpolation, its time resolution decreases below the

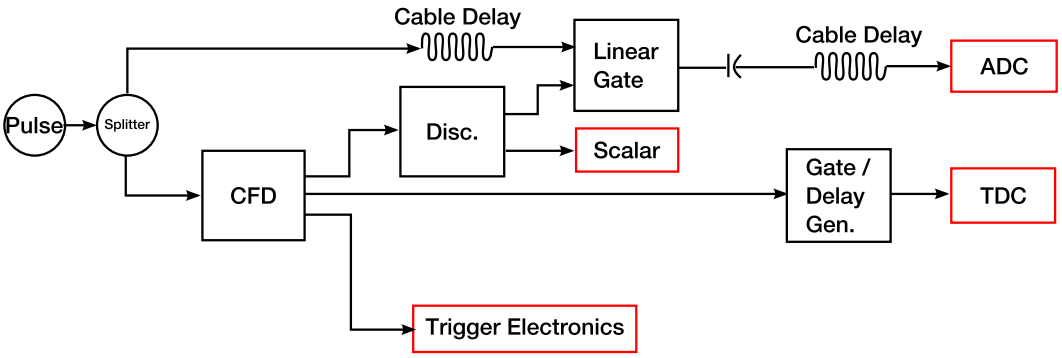


Figure 2.37: A traditional pulse recorder

sub-nanosecond region. Its vertical input resolution is 8 bits per sample. If the life time is on the order of micro seconds, than there will be 1000 sample measurements per one decay time constant, τ, and for a measurement which is 4 τ long, 4000 samples are collected by which a 17 to 20-bit dynamic range can be acquired. As far as is known, a measurement device which is at this level and works in real-time do not exist yet.

2.10.2. The Design

First, the signal in question, usually a pulse, will be digitized/sampled with an ADC at an appropriate sampling frequency depending on the expected features and width of the signal. Then it will be fit to a multi-parameter model by the FPGA unit on the module. The controller unit will continuously receive the fit parameters from the proposed modules as a stream of data. Thus, everything related to a pulse -including but not limited to- the pulse-time and height will be gathered.

These will be achieved in a dead-timeless manner within the provided specifications using a deep buffer memory between the ADC and FPGA. The system will be designed to be completely scalable, and the trigger defined by the relations between the channels will be realized by a cluster of CPUs as an offline system. As a result, the following traditional modules/units will be replaced by this single smart-module: 1) Discriminator, 2) Linear Fan In Fan Out, 3) Scalar, 4) Integrating ADC, 5) TDC, 6) Circuits needed to establish a trigger, 7) A huge amount of active and passive delay units for timing, 8) PID circuits.

Four important aspects of this project are

- Multi-purposeness: Every data channel gives ultra-sensitive timestamp, height, decay time and similar properties of incoming pulses defined by the model.
- Dead-timelessness: In traditional systems there is always a dead-time in every system doing any function stated above. Therefore it is impossible to prevent to lose some part of the data. In our system, sampled pulses are kept in deep buffer memory, and are processed with a dead-timeless manner for a chosen average maximum pulse frequency.
- Triggerlessness: Traditionally, in large systems, a logic gate is built according to upon which aspect of pulse is looked, for all channels. This hardware determines when the signal contains information about that aspect, and data is processed only in that time interval. Therefore, an error or deficiency in that hardware will result in loss of data or in the renewal of the whole experiment.

  With this system, in every channel if a pulse meets its triggering criteria (such as channel threshold) digitized pulse parameters are obtained and saved in a wide data storage system. Hence the same data can be analyzed repeatedly with different triggering criteria, which is like going back in time and repeating the experiment. In this way, the risk of possible data loss due to triggering errors or inefficiencies in complex systems is reduced to zero.
- Scalability: Every channel works independently while recording. Hence, as long as synchronization is provided and there is enough room for data storage, the complexity of the system depends to the number of channels as $O(n)$. On the other hand, the backbone of the system has an $O(1)$ complexity.

Any kind of signal which belongs to the definition of "pulse" is analyzed, fitted to a model and the fit parameters are transmitted to the data bus in real-time. The fitting is done for roughly 30-50 samples, the χ^2 is calculated and minimized. Our aim is to handle 1 million pulses in per second. This corresponds to a high computational power which is roughly 0.1 Teraflops (10^{11} calculations per second). Because of that this approach is not realized before. But recent technologies make it possible.

The place of our work in a general framework which is responsible for pulse processing and data acquisition is shown in Figure 2.38 The job of every ADC + FPGA channel in here is to digitize the signal and analyze it in real-time on FPGA according to the purpose of the scientific research.

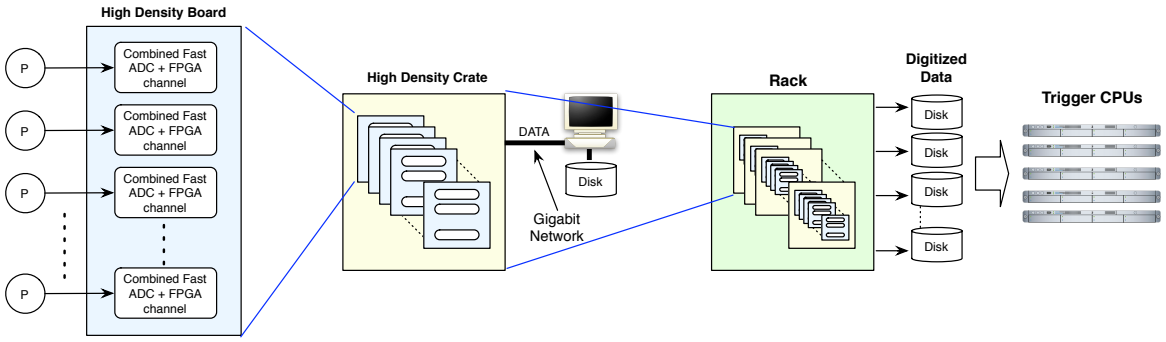


Figure 2.38: General structure of the aimed next generation DAQ system

2.10.3. The Timeline

During the first phase of the project, the required size of buffer memory and the logic-area of the FPGA was determined by a simulation code that simulates some common pulse shapes according to the horizontal (time) and vertical resolution of the digitizer for a given average pulse frequency to make the system dead-timeless.

The simulation code was implemented and optimized on an FPGA platform during the second phase. The signals which are produced by an arbitrary function generator, in a controllable manner, were supplied to the system to let it analyze them. Then, the results were compared with the expected values. This was the most critical part of the project and it will be the most important item in the list of success-criteria.

In the third and the last phase of the project, the system will be developed for a multi-channel (≥ 2) setup, and it will be tested through a simple physics experiment: The pulse parameters will be recorded to the disk, then the trigger will be established offline to demonstrate that physics analysis and calibration can be performed. Our design goal for the prototype is to achieve 0.1 ns time resolution for a 1 MHz average pulse rate.

3. MODELING OF EXPECTED SIGNAL

The electrical signal from PMT which is the final output of the whole system, is due to sum of different processes all of which are described by probabilistic distributions.

In our model the dimensions of the scintillator are, h, thickness in x direction, L, length in z direction), and w, width in y direction. The energy of the incident particle is E. The point where the incident particle hits the scintillator as a ratio to L is z. (Figure 3.1)

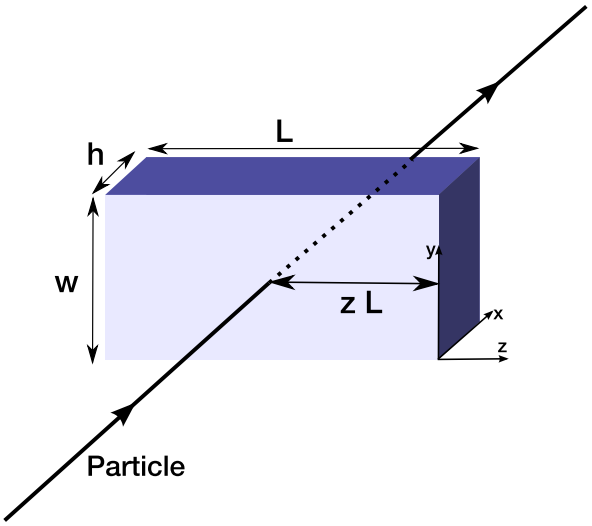


Figure 3.1: Dimensions of the scintillator, its orientation and the point where the incident particle hits it

The relativistic energy relation is

$$E = \sqrt{p^2c^2 + m^2c^4}$$

and the momentum is

$$p = \gamma m v$$

where

$$\gamma = \frac{1}{\sqrt{1 - v^2/c^2}}$$

. If these equations are solved for v then the relation between energy and velocity of the incident particle is found as

$$v = c\sqrt{1 - \left(\frac{mc^2}{E}\right)^2} \tag{3.1}$$

In our simulations we choose relativistic electrons as the incident particles, hence $m_e c^2 = 0.51$ MeV and $E = 40$ MeV. Therefore $v = 0.99992\ c \approx c$. If the particle hits the detector with a right angle then the time interval in which the particle passes through the detector is

$$t_{\text{pass}} \simeq \frac{h}{v} \tag{3.2}$$

The energy loss, E_{loss} in the kinetic energy of the particle due to the interaction with the scintillator is taken from ESTAR data [38]. The data consists of tables of empirical values of stopping "power" (energy per density per length), P, related to incoming particles kinetic energy for different materials. (Figure 3.2) The data is independent of the density because same material can have different densities for different temperatures and pressures. When multiplied with the density of the material the values give the E_{loss}/range, where

$$E_{\text{loss}} = \int_0^h \left(\frac{dE}{dx}\right) dx$$

Hence the range is

$$h = \frac{E}{P\rho}$$

We choose the properties from a real-life plastic scintillator, BC-408 by Saint-Gobain

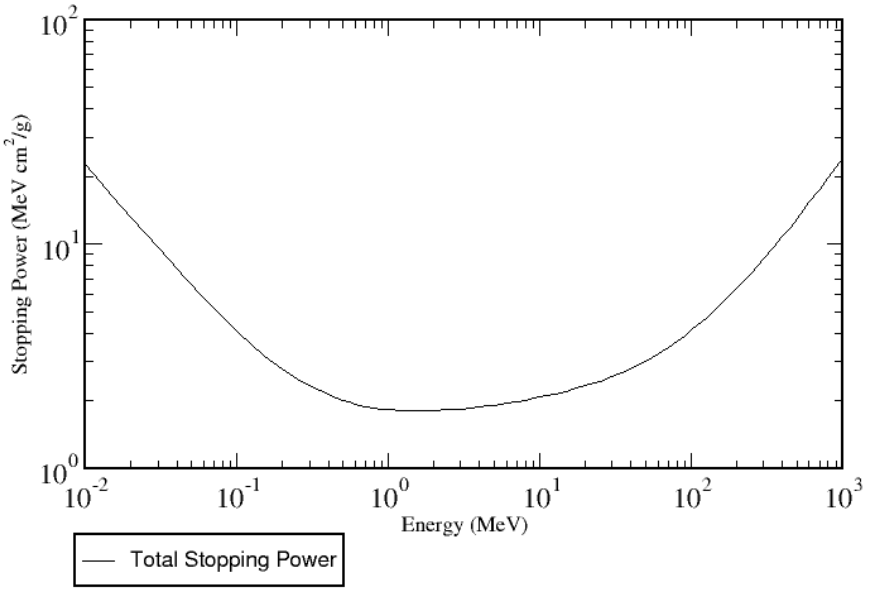


Figure 3.2: Stopping power, $\frac{dE}{dx}/\rho$, vs. energy of the incident particle, E for vinyltoluene based plastic scintillator [38].

Crystals, which has the $\rho = 1.032 \frac{\text{g}}{\text{cm}^3}$ and chose thickness as $h = 2.5$ cm. The $P = 2.796$ MeV cm^2/ g for vinyltoluene based plastic scintillator for $E = 40$ MeV from the ESTAR table. From these values the E_{loss} value is calculated. (After losing the energy

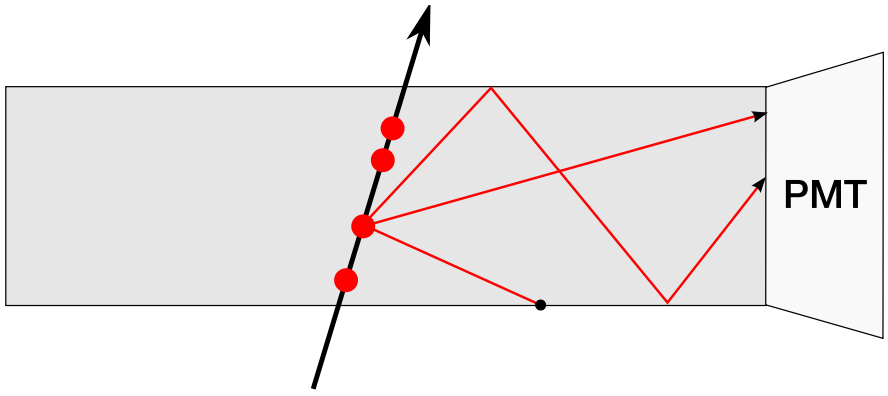


Figure 3.3: Interaction points in the scintillator and different photon trajectories

$E_{\text{loss}} = 7.21$ MeV, the new kinetic energy again corresponds to $v \approx c$, hence the t_{pass} calculation remains valid.) But not all of the E_{loss} is transformed to scintillation light. Some of the energy may turn into heat, or kind of radiation which is not in the visible range etc. The remaining energy, rE_{loss} is turned into photons to be detected by PMT.

The wavelength of the maximum emission, λ, of BC-408 is 425 nm [39]. (and the

scintillator is transparent to that wavelength) Using the relations $c = \lambda\nu$, $\omega = 2\pi\nu$, $E = \hbar\omega$ the energy of one photon is found as

$$\delta E = 2\pi\hbar c/\lambda \tag{3.3}$$

Hence the number of emitted photons, N, are

$$N = \frac{E_{\text{loss}} r}{\delta E} \tag{3.4}$$

The photon emission time is the sum of two time intervals, $t_{\text{excitation}}$ and t_{emission}.

t_{exc} is the time when a scintillator molecule is interacted with the incident particle and goes to an excited state. As seen in Figure 3.3 the red dots are some of the interaction points which result in emission. t_0 is the time when the particle hits the front side of the scintillator. Because the matter is homogeneous and the loss of velocity due to energy loss is negligible the possibility to have an interaction is the same everywhere on the trajectory. Hence the t_{exc} is chosen to form a uniform distribution between t_0 and $t_0 + t_{\text{pass}}$, $t_{\text{exc}} = \text{unidist}(t_0, t_0 + t_{\text{pass}})$ Figure 3.4 shows the histogram of the Monte Carlo simulation.

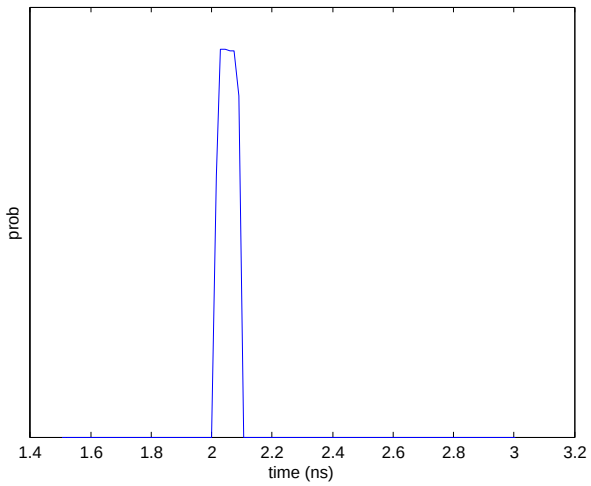
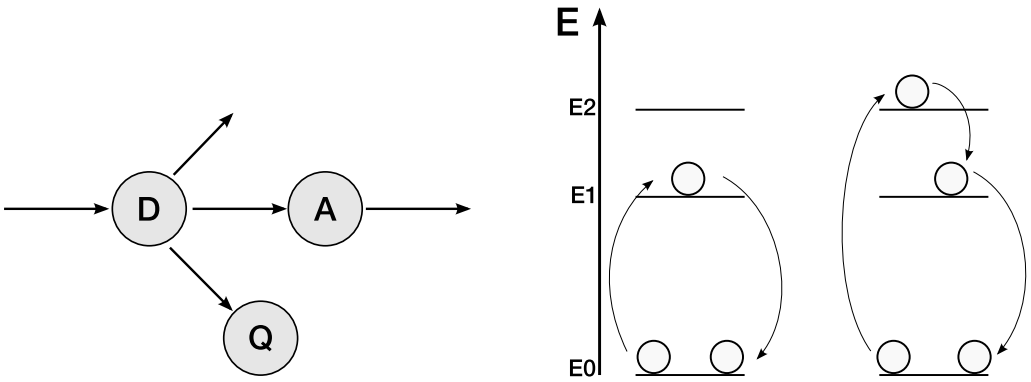


Figure 3.4: $t_{\text{excitation}}$ distribution

t_{emit} is the time when the excited molecule returns to its ground state by emitting a photon with the energy δE. This decaying process is a complex one and is related both to the spontaneous emission phenomenon and to the interaction of the molecule with its environment. The thermal interactions perturb the molecule in the excited state and makes it to go to a lower energy state.

As seen in Figure 3.5b, this process may not have only two steps such that at $t = t_0 + t_{\text{exc}}$ the molecule excites: $E_0 + \delta E \rightarrow E_1$ and at $t = t_0 + t_{\text{exc}} + t_{\text{emit}}$ it deexcites: $E_1 \rightarrow E_0 + \gamma_{\delta E}$. The molecule can first go to an higher energy state, E_2 by emitting a photon opaque to the scintillator and then after a while decays to E_1 and then decays to E_0 etc. This is depicted in Figure 3.5a. The particle transfers some portion of its kinetic energy to the "donor" state first. Then from that state, either a photon invisible to the PMT is emitted or energy is transferred to another state, called "quenching" state, which is unrelated to the detection process or transfered to the acceptor which emits the photons we are interested in.

To model the emission time t_{emit} we chose the probability distribution referenced in [40]:



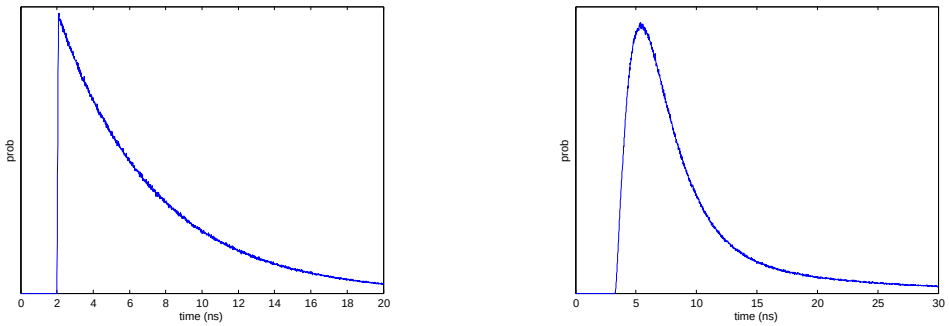
(a) Two step energy flow (Adapted from [41])

(b) Energy levels for simple and complex processes

Figure 3.5: Energy flow processes diagrams

$$p(t_{\text{emit}}) = \frac{1}{1+R}\frac{e^{-t_{\text{emit}}/\tau_2} - e^{-t_{\text{emit}}/\tau_1}}{\tau_2 - \tau_1} + R\frac{e^{-t_{\text{emit}}/\tau_3}}{\tau_3} \tag{3.5}$$

Here τ_1 is the rise time and τ_2 is the decay time and they are called fast decay constant. τ_3 is the slow decay constant. Together they model the complex interactions inside the scintillator responsible to the final emission. The Monte Carlo simulation of t_{emit} distribution is shown in Figure 3.6b.



(a) with one τ (b) with three τs

Figure 3.6: t_{emission} distributions

The photons are emitted isotropically, uniformly to all directions. t_{flight} is the duration the emission time and the time when the photon reaches to the PMT. To simulate t_{fl} we modeled the scintillator body as a 2D rectangle seen from the front. Photons are emitted with a uniform distribution of α between $\frac{\pi}{2}$ and $-\frac{\pi}{2}$. The horizon-

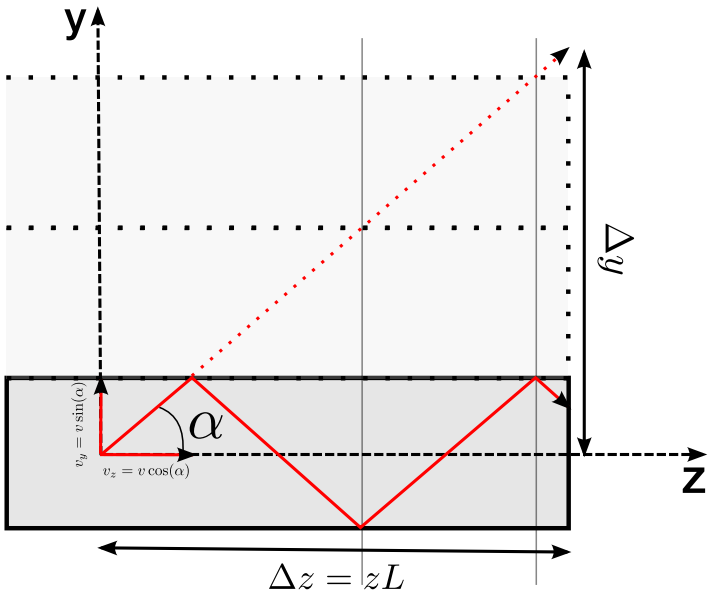


Figure 3.7: Calculation of the flight time

tal velocity component of the photon is $v_x = c\cos(\alpha)$. Vertical distance to the PMT is

zL.

$$t_{\text{fl}} = \frac{zL}{c\cos(\alpha)} \tag{3.6}$$

where $\alpha = \text{unidist}(-\frac{\pi}{2}, \frac{\pi}{2})$, a uniformly distributed random number between $-\pi/2$ and $\pi/2$.

According to the angle, some of the photons go directly to the PMT, some others hit the upper or lower sides of the scintillator and reflect. Photons are reflected with a 98% efficiency. To take account of the possibility that the photon reaches the PMT the number of reflections must be calculated. The vertical velocity component is $v_y = c\cos(\alpha)$. Then the total path gone on the vertical direction is $v_y t_{\text{fl}}$. Therefore the total number of reflections is (Figure 3.7)

$$N_{\text{ref}} = \lfloor \frac{v_y t_{\text{fl}}}{w} \rfloor$$

The Monte Carlo simulation of the photon flight distribution is shown in Figure 3.8.

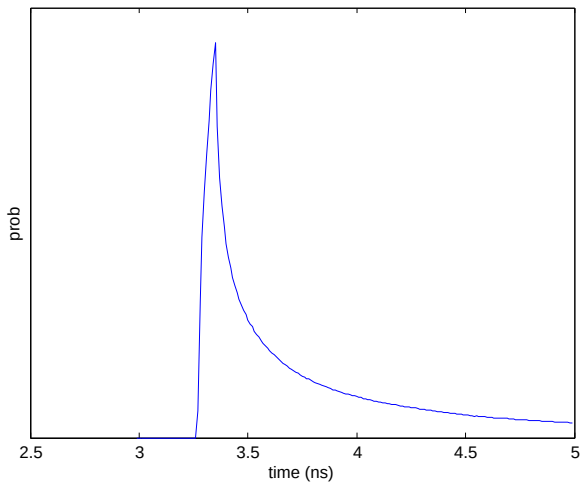


Figure 3.8: t_{flight} distribution

At every reflection there is a, $p_{\text{a}} = 2\%$, probability of absorption. To reach to the PMT a photon has to survive all the reflections. The probability to survive a reflection

is $p_{\mathrm{r}} = 1 - p_{\mathrm{a}} = 98\%$ and therefore the probability to survive N_{ref} reflections is

$$p_{\text{reach PMT}} = p_{\mathrm{r}}^{N_{\mathrm{ref}}} \tag{3.7}$$

According to our model the time when a photon reaches to the PMT depends on 4 factors: t_0, t_{exc}, t_{emit} and t_{fl}. t_0 is constant for all photons due to the same incident particle. The last three has their own distributions and the resulting event times are the sum of the time intervals taken from these three distributions. In Figure 3.9 some of the events produced with the Monte Carlo simulation are indicated on the time axis.

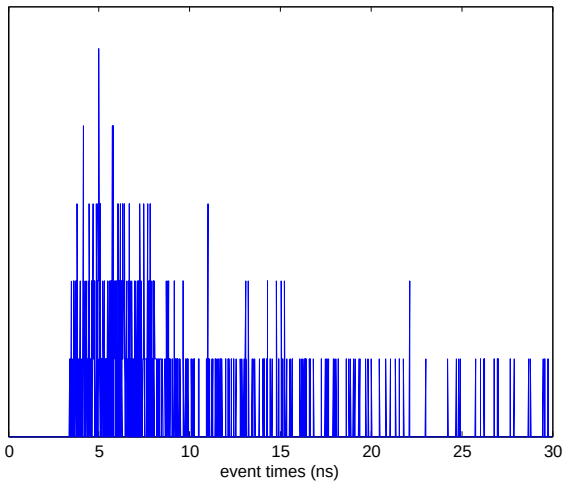


Figure 3.9: t_{all} distribution

When a photon goes into the PMT, using the photoelectric effect it generates an electrical signal. That signal is not a dirac-delta but rather a gaussian curve, $e^{\frac{-t^2}{\sqrt{2\pi\sigma}}}$, which can be modelled with a σ_{PMT} value. Therefore the final signal is the convolution of the time distribution histogram of the incoming photons to the PMT with a gaussian curve with $\sigma = \sigma_{\mathrm{PMT}}$. See Figure 3.10

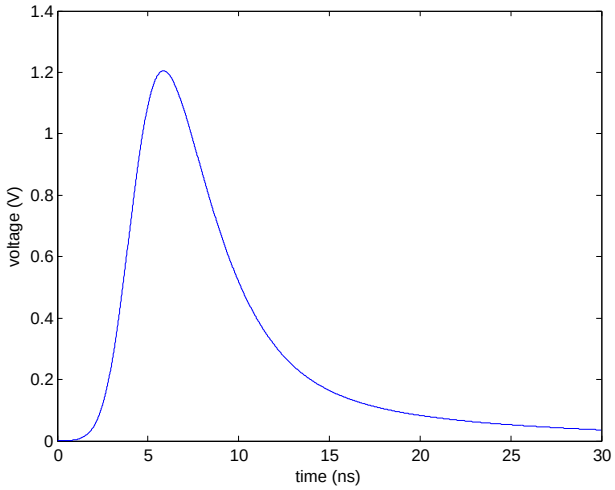


Figure 3.10: Output signal of the PMT

3.1. Fit Function

Figure 3.11 shows the electrical signal sent to the ADC to be analyzed further. (Noise due to the electrical circuits is added and the peak is downwards because electron has negative charge.) We want to capture its essential characteristics in a certain precision by using minimum number of parameters to reduce the computational costs. The output signal of the PMT is not an analytical function but a sum of different statistical processes.

We are trying to fit it to an analytical function in the form of a linear rise and exponential decay of which simplest form is te^{-t}. To start the pulse at t_0, $t \rightarrow (t - t_0)$. And f becomes $f = (t - t_0)e^{-(t-t_0)}$. To be able to adjust the decay time the parameter τ is added. (Bigger τ means longer decay) $(t - t_0) \rightarrow \frac{t-t_0}{\tau}$. $\Rightarrow f \rightarrow (t - t_0)e^{-\frac{t-t_0}{\tau}}$. To set the amplitude of the pulse, the parameter A is added $\Rightarrow f \rightarrow A(t - t_0)e^{-\frac{t-t_0}{\tau}}$. To make A the maximum value of the pulse, it is normalized by multiplying with $\frac{e}{\tau}$ $\Rightarrow f \rightarrow \frac{Ae}{\tau}(t - t_0)e^{-\frac{t-t_0}{\tau}}$ When there is no pulse, there are only small fluctuations around the offset value, f_0. To represent this situation f is defined piecewise:

$$f(t) = \begin{cases} f_0 + Ae\frac{t-t_0}{\tau}e^{-\frac{t-t_0}{\tau}} & t \geq t_0 \\ f_0 & t < t_0 \end{cases} \tag{3.8}$$

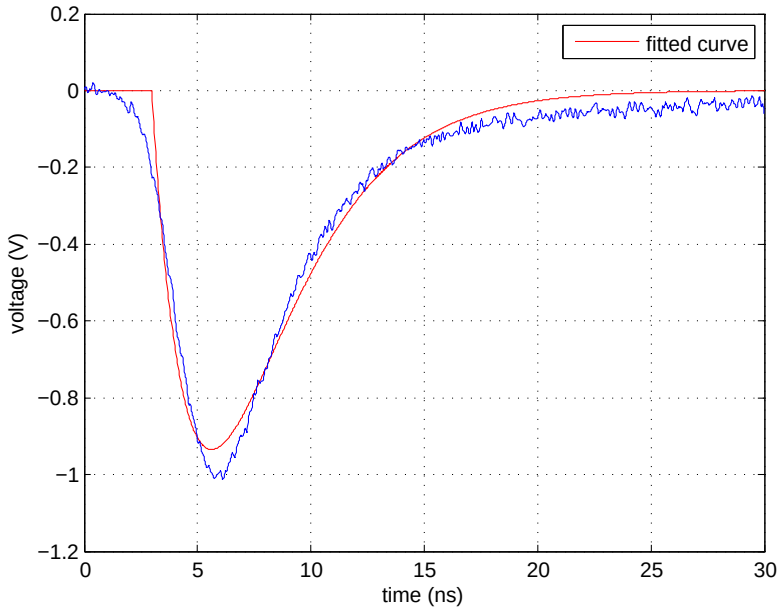


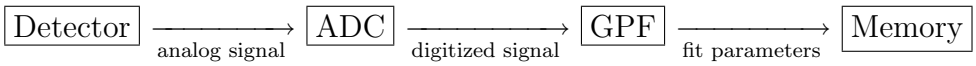
Figure 3.11: Output of the PMT (noise is added) and the fit function

This way we model every pulse using 4 parameters: f_0, t_0, A and τ. Offset is very easy to calculate and there is no need to fit it. Therefore we have three parameters to fit for minimizing the χ^2. For example, the pulse from the Monte Carlo simulation in Figure 3.11 is fitted best with the parameters: $A = -0.937$, $\tau = 2.646$, $t_0 = 2.976$, $f_0 = 0.0$.

4. THE ELECTRONIC EQUIPMENT

Because of the much higher event counts than bubble chamber photographs in today's DAQs of scintillation detectors used in HEP experiments, it is impossible to make the analysis by hand without the help of automated systems. And, it is very hard to do the analysis in real-time just by programming general purpose computers, because of the cost and hardware needs. Hence we choose to design integrated circuits with many parallel cores to get the benefits of specialized hardware and parallel processing.

At one end of the DAQ system we have the ADC and at the other hand we have the computer memory. It processes the digitized information coming from the detector and puts its results into a computer memory for recording and further analysis.



The analysis system, we call Gigahertz Pulse Fitter (GPF), is a "Field Programmable Gate Array" (FPGA). An FPGA is a digital computing resource, which is basically a bunch of logic gates and memory units, which can be connected arbitrarily by the wish of the designer to have a special purpose IC with intended functions. It is possible to change the structure of the FPGA at any given time, which makes it a very useful tool to design an IC.

The "programming" of an FPGA is called "digital design", a process which is different than traditional computer programming. Programming in software world means to arrange a sequence of instructions in machine language that the CPU architecture understands into a memory. The CPU then will read both the instructions and the data on which that instructions will to be applied from the memory and execute them. To produce this machine code, software programmers use higher level languages which are closer to the human understanding. They write the algorithms and describe the data structures using that programming language and then special software called compilers translate them to machine code.

Programming a FPGA is like to design that CPU with a fixed software. The switches between the logic units and gates are set up to build the intended design. Because the number of basic building blocks in today's chips are enormous, making the design on the level of the logic gates is virtually impossible. Hence the design process itself should be automated. There are "higher level programming languages", called hardware description language (HDL) such as Verilog and VHDL. The design is expressed in terms of them and special software synthesize the design into logic gates according to properties of the chosen FPGA. We used a combination of Verilog and VHDL during the course of this study.

4.1. Digital Design Fundamentals

The main difference of digital electronics from analog ones is that the signals have discrete values rather than values in a continuous range. Every discrete value is a combination of bits. A bit is a physical system which can only be in two different macro states labelled as LOW and HIGH states. They represent the numbers 0 and 1 in base 2 and are represented by two different voltage values, one is near zero (ground) and other is a chosen value which may be different in every system. These bits are made of circuit elements that have two stable values.

The advantage of the digital system is that putting a system into a state as exactly wanted is easier than as in analog system. Therefore a digital system is more immune to errors due to the noise and easier to transmit and store. If Nyquist-Shannon sampling theorem is kept considered, an analog signal can be converted to a digital one and then reconstructed without losses. [3]

Also in analog systems to increase the resolution of the signal (to have a greater range for the signal) one needs to make fundamental design changes to keep the linearity etc. But in digital systems adding more bits to the signal (increasing the bit depth) is enough for the same purpose.

[3] Actually the world itself is quantized, ie. the total charge in a system is an integer multiple of some unit charge, the charge of the electron. But because they are incredible small compared to macroscopic quantities, the concepts like continuity, integration, analog signal etc. are useful approximations.

In the digital domain the calculations can be converted to Boolean algebra and hence any kind of calculations can be made. A data processing algorithm can be implemented digitally in a certain precision much easier then an analog structure. Although there are analog computational techniques such as using op-amps to integrate or RLC (resistor-inductor-capacitance) circuits to apply a low- or high-pass filter to a signal. But using analog circuit elements there is no straightforward method to make arbitrary calculations. The abilities of using only analog circuit components is limited and not versatile as the possibilities of digital systems. Hence finding the best fit parameters to a function is a problem to be solved in digital domain.

In digital design the main building blocks are logic gates and flip-flops (latches, registers). The synchronisation is done by a common clock signal.

4.1.1. Data Representations in Binary Systems

4.1.1.1. Natural Numbers. The most straightforward data type to be represented by bits are the natural numbers. They are mapped to to their corresponding values in base 2, which may also be called in radix 2. (These natural numbers themselves can be used to represent another quantities.)

4.1.1.2. Signed Integers. To indicate the positiveness or negativeness of a number we can use + and − signs. But in the hardware everything must be expressed in terms of binary digits. For signed integers, integers with a positive or negative sign, there are different approaches. One of them is to use the first bit of a representation to indicate the sign. By using N bits one can represent the numbers from 0 to $2^{N-1} - 1$ both negative and positive. For example $+3 \rightarrow (+, 11) \rightarrow 011$ or $-2 \rightarrow (-, 10) \rightarrow 110$, where the 1 in the first bit represents − sign and 0 represents +. This convention is called signed-magnitude convention.

In this mapping there is a degeneracy in 0's representation. Both 1000000 (7-bit number) and 0000000 represents the integer 0: There are positive and negative zeros,

-0 and $+0$. To remove this degeneracy some other representations have been invented. Signed-complement system is one of them.

As an example of signed-complement system, in 2's-complement system a number is negated by taking its complement, whereas in signed-magnitude system it is negated by changing the sign bit. The 2's complement of a n-digit number x is the number y where $x + y = 2^n$. For example, -5 is the 2's-complement of 5, where $n = 3$. 5 is $(101)_2$. $y = (1000)_2 - (0101)_2 = (0011)_2$. In signed-complement system, the positive numbers starts with 0 and hence their complements start with 1 which also indicates the negativity.

Table 4.1: Signed integers and their representations in different systems

number	signed-magnitude	signed-complement
$+3$	011	011
$+2$	010	010
$+1$	001	001
$+0$	000	000
-0	100	—
-1	101	111
-2	110	110
-3	111	101
-4	—	100

The data itself, the bit values, are meaningless by themselves. Its meaning is given by the user which evaluates it according to a representation scheme/map chosen and accepted beforehand. The same set of bits can be read as an unsigned integer or a signed fixed number etc. Hence the memory by itself can not determine its content.

Another advantage of this representation is the implementation of addition and subtraction operations which takes care of the signs without any special handling.

Table 4.2: Same bits according to different representation schemes

bits	signed-comp. int.	unsigned int.
011	+3	3
010	+2	2
001	+1	1
000	0	0
111	−1	7
110	−2	6
101	−3	5
100	−4	4

Converting a quantity into an representation form is called encoding and getting the information from the encoded bit strings is called decoding.

4.1.1.3. Fixed Points. The integers can exact representations in the digital domain if enough number of bits are enough. But the same is not true for all of the real numbers, because some of, such as $\sqrt{2} = 1.41421\ldots$, has infinite number of digits in any base.

The numbers with a decimal point are represented by fixed points. They are interpreted as if there is a decimal point between some predetermined digits. If there are n bits in the integer part and m bits in the fractional part, such as $x = b_n \ldots b_2 b_1 b_0 . b_{-1} b_{-2} \ldots b_{-m}$, then

$$x = \sum_{i=0}^{n-1} b_i 2^i + \sum_{i=1}^{m} b_{-i} 2^{-i} \tag{4.1}$$

Their resolution is the smallest non-zero magnitude representable, which is 2^{-m}. A real number can only be represented by truncating at the nearest fixed point number in a given resolution.

The word length (number of bits of a number) and place of the decimal point is determined by educated guesses and knowledge about world length of the inputs of the

operations and the amount of the needed resolution.

For example, to calculate the offset value, the sum of samples in one block is needed. offset $= \sum_{i=1}^{8} y_i$. One sample is 8-bit. There are 8 samples in one block. The total sum is addition of 8 8-bit numbers. Addition of 2 n-bit number yields an $n+1$ bit number. If these 8 numbers are grouped into doublets, they correspond to 4 9-bit numbers, which is equivalent to 2 10-bit numbers. Hence the world length of the sum must be 11-bits.

$$\begin{aligned}\sum_{i=0}^{7} \left(b_7^{(i)} b_6^{(i)} b_5^{(i)} b_4^{(i)} b_3^{(i)} b_2^{(i)} b_1^{(i)} b_0^{(i)}\right)_2 &= (b_{10}^s b_9^s b_8^s b_7^s b_6^s b_5^s b_4^s b_3^s b_2^s b_1^s b_0^s)_2 \\ &\xrightarrow[\text{average}]{} \left(b_7^s b_6^s b_5^s b_4^s b_3^s b_2^s b_1^s b_0^s . b_{-1}^s b_{-2}^s b_{-3}^s\right)_2\end{aligned} \tag{4.2}$$

Then their average can be represented by a 11-bit fixed point which is interpreted such that the first 8 bits are the integer part, and last 3 bits are the fractional part.

4.1.2. Registers and Register Transfer

The bit representations should have a physical existence in a real medium to store and process individual bits. A device with two stable state is called a "binary cell", which has the ability to store a single bit. The valve is 1 if the cell is in one stable state, and 0 if it is in the other stable state.

A cell has an input which is used to set its content and has an output which is a physical quantity (current, voltage etc.) that distinguishes between two states.

An array of binary cells is called a "register". A register made of n binary cells (n-bits) have 2^n different combinations of the set. Therefore a register with n bits can be used to distinguish between 2^n different states, in other words, can store a discrete quantity of information that contains n-bits.

A digital system is basically its registers and its components responsible from the data processing operations that are applied between them. The transfer of a binary information from a register to another one is called "register transfer" operation. A transfer can keep the info untouched or make some manipulations on the bits. Hence register transfer is the translation of the term "information processing" to the digital design domain.

A digital processing system must have components to hold the data to be processed and components to manipulate the bits. Registers are used to hold the data and digital logic circuits are used to change binary variables.

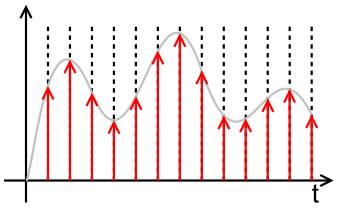
GPF is the processor which manipulates the data from output register of the ADC (a 64-bit register which stores 8 samples of incoming voltage signal) and transfer that binary information to the PCI interface by manipulating them such that PCI bus contain the best fit parameters and information about the incoming pulse.

4.2. Digitization

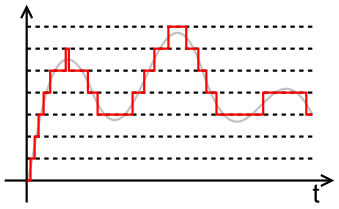
Digitizers are the devices which convert an analog signal to digital one. This is done by two operations: sampling (Figure 4.1a) and quantization (Figure 4.1b). Sampling is to measure the value of an analog quantity, a signal which is a continuous function of time, at discrete time instants. These measurements can not be stored in a digital domain with infinite precision. Hence quantizaton is needed. Quantization is a type of mapping of continuous ranges to integers.

For example, if the input from the ADC is between -0.5 and 0.5 volts and we are using 2 bits to represent them we get the Table 4.3).

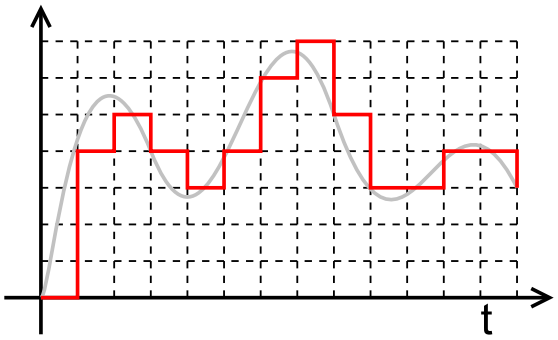
Important parameters of an ADC are sampling rate and quantization bit depth. Sampling rate, f_s is the number of measurements per second. The number of recorded samples in time T is $N = Tf_s$. Bit depth, N_{bits}, is the number of bits to store one sample. It gives the vertical resolution, dynamic range, of a digitized signal. A



(a) Sampling [42]



(b) Quantization [43]



(c) Digitization [44]

Figure 4.1: Digitization of an analog signal

Table 4.3: Representation of voltage values in bits by mapping the voltage ranges to integers

Voltage	Digitizer Bits	Integer
$[-0.50, -0.25]$	00	0
$[-0.25, +0.00]$	01	1
$[+0.00, +0.25]$	10	2
$[+0.25, +0.50]$	11	3

measurement can have $2^{N_{\text{bits}}}$ different values in digital domain. As an example, digitized audio recorded in audio CDs has $f_s = 44100$ Hz and $N_{\text{bits}} = 16$.

We used a flash ADC in our DAQ system, which is a 8-bit, 1.5 GHz device.

4.3. FPGA Design

FPGA is the acronym of Field Programmable Gate Array. The word field indicates that it can be programmed outside the factory. With an FPGA one can build

electronic circuits without using soldering iron or without producing a chip in the factory.

The traditional integrated circuits are called ASIC (Application-Specific Integrated Circuits). Comparing FPGA with ASIC may show their benefits [45].

- ASIC has lower cost per device in large volume production. It has a high initial cost for preparing the factory, but after the preparations are done, ASICs are cheap. FPGAs do not have a initial cost, but the cost of a unit is much higher.
- ASIC has higher development cost because they are produced at a semiconductor fabrication facility.
- ASIC has longer development cycle which take weeks or months, whereas a design and test cycle on a FPGA may take only minutes.
- FPGA may be reprogrammed in the "field" with firmware upgrades for adding new features and bug fixes. ASIC design can not be changed once it is produced.
- ASIC has better radiation protection.
- Analog designs can be embedded inside an ASIC.

To deal with pulses which are coming with random time intervals in real-time without a deadtime, each one has to be dealt very fast. Our design goal, which is to have an average pulse rate of 1 MHz, can not be achieved even with today's ultra fast general purpose CPUs. We showed that the extraordinary amount of the needed computation is reaching near Teraflops. And for that reason we decided to realize this application not on a general purpose CPU but special purpose FPGA.

Before starting to FPGA design we had written a computer program (a model) with the same function. (The details of the program is expressed in the next chapter.) Because FPGA is a programmable chip, designing a FPGA may be seen like translating a program from one language to another. But this is not right. Today, FPGAs of serious sizes are designed by coding in Verilog or VHDL. These languages, especially Verilog, has a similar syntax to C language. It is possible to write the Verilog code by converting the C code line by line. But, unfortunately, the resulting FPGA circuit will

become huge and very slow.

On the other hand, today, chip and FPGA design has more similarities to software projects than electronic circuit design. Because it is virtually impossible to do what a C program of a few thousands of lines does by connecting logic gates one by one.

Although FPGA design resembles more to coding, it does not have much common grounds with writing a software for modeling purposes. While programming for modeling, one does not pay much attention on the speed or memory consumption of the program. Only when the program will run for hour and/or will use all of the available computer memory, the programmer starts to look at optimizations rather than the functions itself.

The real difference of designing a hardware is not dealing with hardware but designing a real-time system. Not only FPGA/CPU/hardware designers but embedded system designers also deal with speed and size optimizations besides speed optimization. In fact, they spend most of their time to these optimizations. They try to implement the functions by obeying the speed and size limitations given within the specifications. For this purpose, they can be create routines with the same function which are much different than the routines in the modelling code, which are also runnable specs of the intended function. In brief, FPGA design is a type of parallel algorithm development which has constraints on computational complexity and speed.

Our sampling rate is 1.5 GHz. Eventhough the needed computation speed will be near Teraflops the gain in speed obtained by the FPGA based design is 2 order of magnitudes faster than a modern PC CPU. But if we would implement and run the same routines of modelling code into the FPGA, it would be even slower than the PC implementation. Because a FPGA runs roughly 50 times slower than a good PC processor. (Our FPGA runs in 200 MHz.) Than what is the advantage of a FPGA over a CPU? Function-specific hardware such as FPGAs, in contrast to CPUs, may run many processes in parallel and/or pipelined manner. Nevertheless, if it is written without taking parallelization into account, in other words, if only its function

is considered, then it has serialization in terms of data flow and its parallelization can not be possible. All in all, FPGA design is creating new routines which have the same functionality of the C code but which run much faster and take small space.

4.4. Actual Hardware and Requirement Considerations

Our DAQ architecture is made of cards which communicate with the PC over a PCI data bus. Every signal channel is connected to another card and every card has a signal input and an output to control its perimeter devices, which is called "I/O extension".

Each of these cards is a DAQ on which special purpose algorithms are run. The components of a DAQ card which are related to sampling are general purpose cards. Hence, especially in the prototyping phase, to design a whole DAQ card by ourselves would be a time waste. Therefore we decided to use DAQ cards from VMetro.[14] The unique feature of these cards is that they have FPGAs on board which can run application specific algorithms. This FPGA is XC5VSX95T which is one of the Virtex-5 FPGA series from Xilinx [37].

On this FPGA there are over 6 million logic gates and more than 8 million bits of memory which are implemented using 65 nm technology.

Before starting the detailed design, one should derive estimations about whether the circuit at hand can fit into the FPGA, by making rough calculations and design sketches. The general approach for this kind of estimations is as such: First the approximate size of the circuit is estimated as if the application is made for ASIC. Then, the number of gates is multiplied with 7 and compared with the number of gates available. The reason of the multiplication with 7 is that the circuits, and calculation units in the FPGA are general purpose and are not special for our application, which causes to loss of efficiency. It should be noted that the size of the circuit is also related to the speed it provides.

We found that if 1.5 GHz sampling rate and 1 MHz pulse analysis rate is aimed, the number of ASIC gates will be 50k. Which corresponds to 350k gates in FPGA and that can fit into our FPGA which has 6M gates. In FPGA designs which include so much routing (if there are many long wires), slowing down can be observed compared to ASIC design. Although this has a low probability, it can be overcame by cloning some units and let them run in parallel. This will increase the number of gates, which may pass the limit, we are certain that the system will handle a pulse rate of minimum 100 kHz.

One should also look at the memory requirements. Our design will use 20kb on-chip memory. The amount of the memory of the FPGA on the VMetro board is 8Mb which is much higher than the required amount. In the memory calculations, besides the amount of the memory, the number of memory blocks is also important. Because processes run in parallel, synchronized access to different memory locations is needed, which is quiet possible with 300 block RAMs on the XC5VSX95T.

4.5. Modules

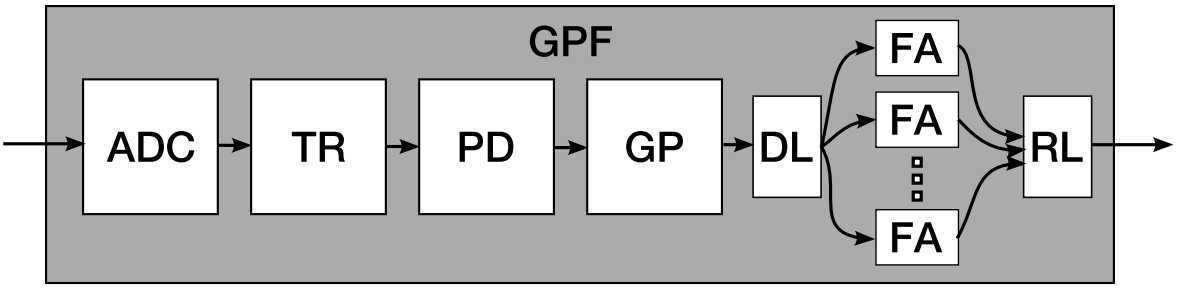


Figure 4.2: Diagram of the GPF system

The GPF system has a modular design made of modules with determined functions and defined ports between them. Hence every module is a block box. This modularity gives the advantage of the ability to change inner structure of one module without affecting the whole system in the design phase.

The system will work at 200 MHz (except the sampling rate of ADC.) Every module has an "valid" signal output which means that the module is working and its outputs are correct. For example, if the ADC can not provide a new block it gives 0

value as the valid signal and TR stops and does not take the same block again and makes its valid signal 0. Having 1 as the input valid signal is necessary but not sufficient condition to output 1 as valid signal. This way, the valid signal propagates along the signal path.

4.5.1. Analog to Digital Converter

The ADC samples the incoming analog signal with a samplerate, $R = 1.5$ GHz. It measures the voltage at every $\frac{2}{3}$ nano seconds. The bit width of one sample is 8. Hence the voltage values are converted to the sample values between 0 and 255. The data recorded by the ADC is $1.5 10^9/1024^3 = 1.4$ giga byte per seconds.

ADC collects a group of successive 8 samples, which we call a block of samples and send them to the TR as an 64 bit signal. Internally it has a FIFO which collects incoming samples and an interface with the GPF which makes a 64 bit block out of 8 8 bit values.

ADC is valid if it has prepared one new block.

4.5.2. Trigger (TR)

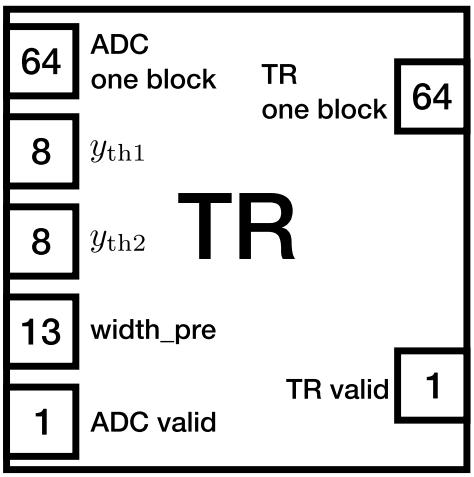


Figure 4.3: Input and output ports of TR

TR is the starting point of the analysis process. It looks whether there is an

important activity, in other words, whether there is a pulse. If a pulse is coming, TR raises its valid flag and informs other modules about the presence of the pulse.

Most of the time the signal is just the noise around the offset. If the event rate is 1 kHz and the mean pulse length is 50 samples, then the ratio of the number of samples belonging to a pulse to the samples belonging to noise is 1/300. TR gives invalid signal when there is no pulse hence the system remains idle most of the time.

To detect the pulse we first tried to use the standard deviation of one block. When there is no activity the mean value, $\mu = \frac{1}{N}\sum_{i=1}^{8} y_i$, of a block is approximately equal to the offset and the standart deviation, $\sigma^2 = \sum_{i=1}^{8}(y_i - \mu)^2$, is low. If there is a pulse the samples fall down quickly and σ of the blocks are higher. But this method did not give good results in real life tests and we tried another method.

As a finitie state machine TR has 3 states: WAITING, PULSE TOP, PULSE DECAY (See Figure 4.4) and two threshold values, y_{th1} and y_{th2}, which are set using the registers controlled by the kernel drivers. In the absence of a pulse it is in the WAITING state. When the signal crosses y_{th1} from above to below TR changes to PULSE TOP state. There it waits until the signal crosses y_{th2} from below to above and meanwhile records the time interval between to crossing times, Δt_{th}. After second crossing TR goes to PULSE DECAY state. In that state it waits until the pulse decays and returns again to the WAITING state. TR stays in the PULSE DECAY state for $\Delta t = c_{\mathrm{dec}}\Delta t_{\mathrm{th}}$, where c_{dec} is constant adjusted by its register. (See Figure 4.5)

If TR would raise its valid flag when the signal crosses y_{th1} and the pulse would be collected from that point, it would miss the beginning portion of the signal. The solution of these problem is to delay the signal by using a circular buffer. Incoming signal is always recorded to into a large circular buffer, where the read index follows the write index by a predetermined distance "prewidth" which corresponds to Δt_{pre} in the Figure 4.5. When the TR valid flag is raised the blocks from the read index are sent to PD, not the latest block which comes into TR yet. This mechanism adds only a few clock cycles delay to the whole system, which is insignificant.

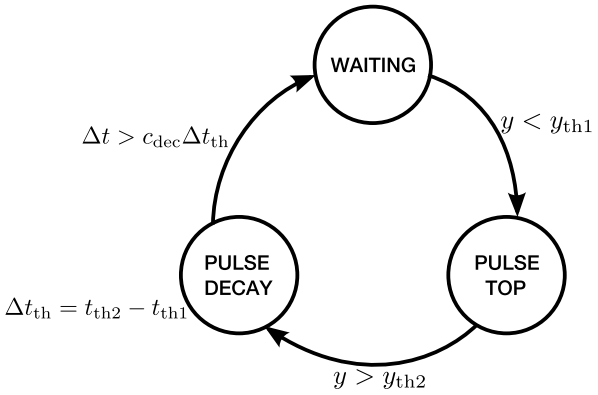


Figure 4.4: State machine diagram of TR

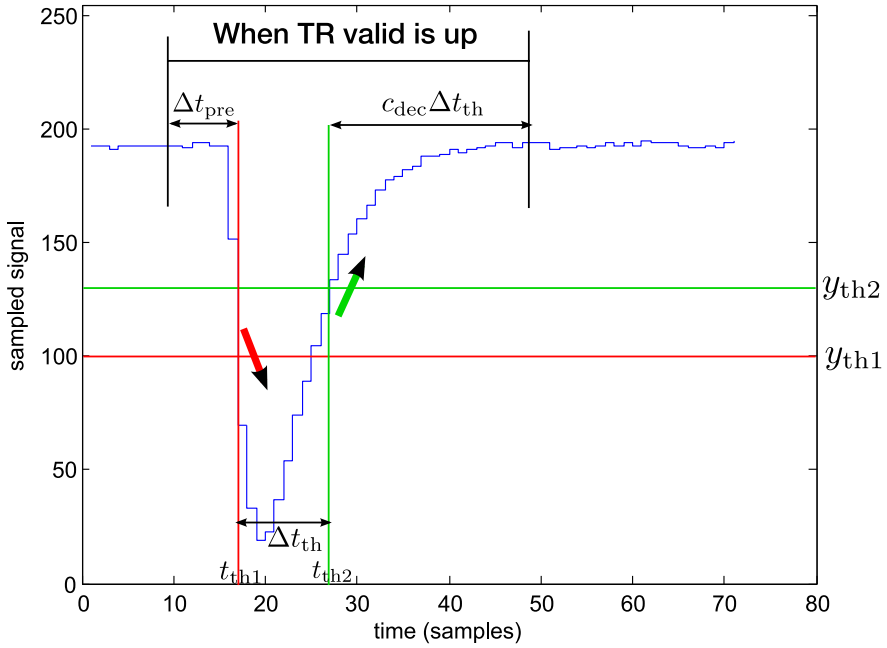


Figure 4.5: Threshold values and demonstration of $\Delta t_{\mathrm{prewidth}}$ and Δt_{th} mechanisms

4.5.3. Pulse Detector (PD)

The task of the PD is to extract some parameters while the blocks are flowing through it without the need to store the whole pulse. These parameters are:

- $n_{\mathrm{pulse,PD}}$, an estimation of the number of samples in the pulse. This is simply 8 times the number of blocks when the TR valid flag is up.
- $t_{0,\mathrm{PD}}$, an estimation of the start time of the pulse, which is called the "timestamp"

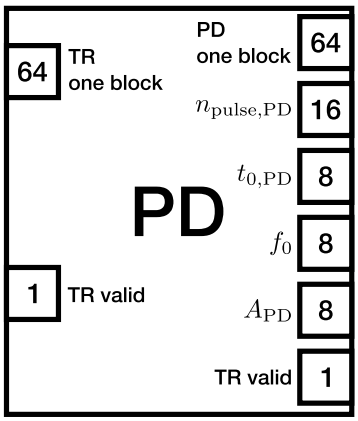


Figure 4.6: Input and output ports of PD

of the pulse. GPF has an internal counter which ticks with every new incoming block from ADC. Timestamp in the PD is just the number read on that counter, which is a rough approximation.

- f_0, offset value. f_0 is the mean value of the first block where the pulse is not started yet. $f_0 = \frac{1}{8}\sum_{i=0}^{8} y_i$.
- A_{PD}, an estimation of the pulse amplitude. This is the difference between the maximum and the minimum sample values in the whole set of blocks belonging to the pulse. The maximum (or minimum) value is calculated on-the-fly by comparing 9 number. 8 of them are the samples in the current block and the last one is the maximum (or minimum) value of the previous block.

Figure 4.7 shows these values on the pulse graph.

4.5.4. Guess Parameters (GP)

GP stores the pulse in a FIFO and makes last calculations of the guessed parameters. The important signals used in GP are:

- V_{half}, the voltage value which is half amplitude below the offset.

$$V_{\mathrm{half}} = f_0 - A_{\mathrm{PD}}/2$$

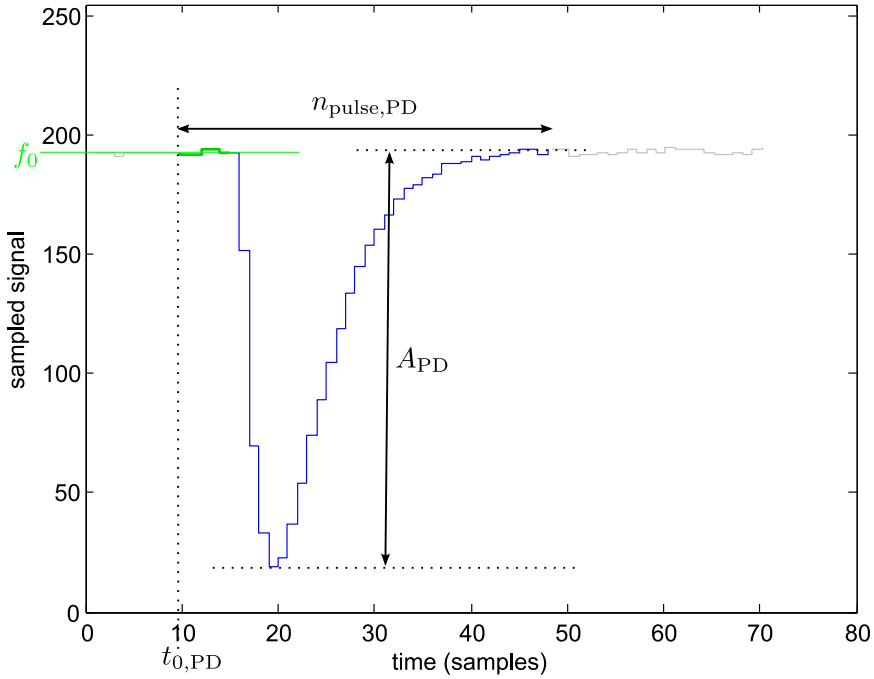


Figure 4.7: Calculations made inside PD

- t_1, the time when the pulse crosses V_{half} from above to below. $y(t_1) = V_{\text{half}}$
- t_2, the time when the pulse crosses V_{half} from below to above. $y(t_2) = V_{\text{half}}$ A linear interpolation is made in the calculations of t_1 and t_2. See the green lines in Figure 4.8. Let n_i the sample number where the inequality $y_{i+1} < V_{\text{half}}$ holds first.

$$\frac{y_{i+1} - y_i}{n_{i+1} - n_i = 1} = \frac{V_{\text{half}} - y_i}{t_1 - n_i}$$

- $t_{0,\text{GP}}$, starting time of the pulse with more precision.
- τ_{GP}, inverse tau, the inverse of the decay time of the pulse.

$t_{0,\text{GP}}$ and τ_{GP}^{-1} are calculated using t_1 and t_2. Let $\Delta t_{1,2} = t_2 - t_1$. Their guesses are found by the crude approximation that there is a simple relation between them and $\Delta t_{1,2}$ such that:

$$\tau_{\text{GP}} = c_2 \Delta t_{1,2} \quad \Rightarrow \tau_{\text{GP}}^{-1} = \frac{c_2^{-1}}{\Delta t_{1,2}}$$

and

$$t_{0,\mathrm{GP}} = t1 - c_1\tau_{\mathrm{GP}}$$

The values of c_1 and c_2 are chosen by plotting a pulse with reasonable arbitrary parameters and measuring them numerically. It is true that any choice of c's lead to better guesses for some pulses and worse guesses for other ones but these τ and t_0 values are just initial starting points for the fit algorithm.

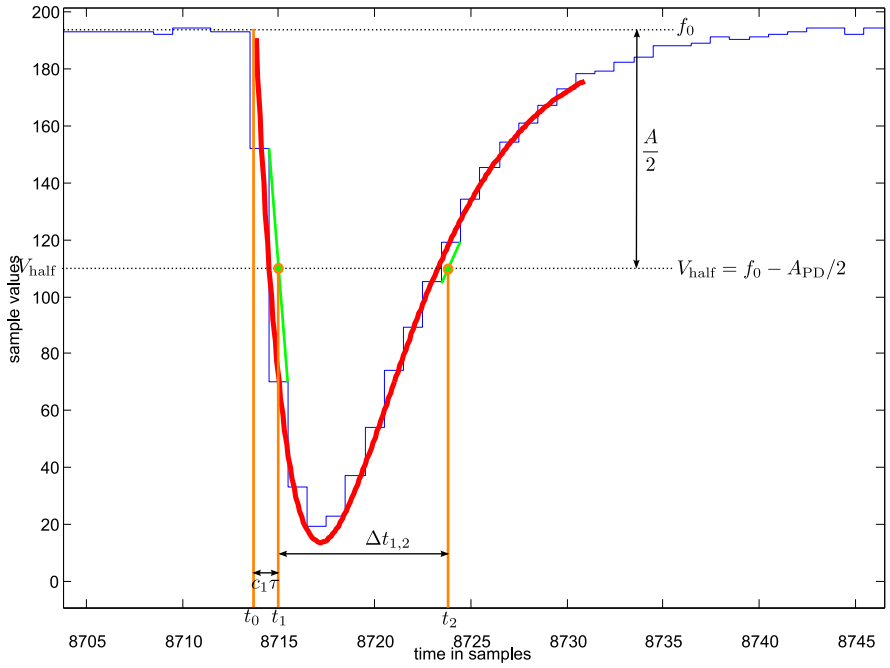


Figure 4.8: Calculations made inside G

In Figure 4.8 the red curve shows the fitted function to a simulation pulse (blue steps) using the guess parameters.

4.5.5. Dispatch Logic (DL)

After the GP we have the samples of the pulse in a memory and its guess parameters in registers. Up to this point every modules job takes only a few clock cycles and anything can be done in a pipelined manner. But the task of the FA take a much longer time, hundreds of clock ticks. Hence it is very probable that new pulses come

before FA finishes to find the fitting parameters. The solution to this problem is to have many fitting cores inside the GPF. We will fit 60 cores to the FPGA. Every core will have a "busy" flag, which will be raised when working. The job of DL is to send the pulse and its guessed parameters to a free FA.

4.5.6. Fitter Agent (FA)

RTL not implemented yet. FA takes the samples of a pulse and guessed fitting parameters from DL and find the parameters which give the minimum χ^2 using downhill simplex algorithm.

5. SOFTWARE

5.1. SystemC

We decided to make a software equivalent of our system. First of all, before starting to the electronic design one has to be sure that the intended algorithms are working correctly. Second, we needed input output data for testbenchs in Verilog simulations.

Algorithms could be tested in the Verilog environment, too. But creating and testing algorithms using traditional software programming is much easier than thinking in digital design domain. Hence we decided to use SystemC as the first step of the realization of algorithms and for preparing testing data.

SystemC is an IEEE standard (IEEE 1666 - 2005) created by Open SystemC Initiative and approved in 2005 by IEEE and in 2006 by ANSI. It is a C++ library which includes basic elements for digital design and simulation. "SystemC is an ANSI standard C++ class library for system and hardware design for use by designers and architects who need to address complex systems that are a hybrid between hardware and software" [46].

Object-oriented programming (OOP) languages are developed to fulfill the need of handling of large and complex projects by dividing the problem into basic building blocks called objects which consists of some data and some functions to be applied on that data.

C++ is on of the most common OOP languages today. Some of its abilities fits well for the system design purposes. In C++ new data types can be defined as classes and objects can be created from them. This makes it possible to define logical gates, ports, fixed-point signals, modules etc. as classes. Electronic components can be created as object instances of these classes and they are interconnected using special

methods of the objects used as ports. The operator overloading facilities of the C++ embeds the SystemC extentions into the language and makes its usage transparent. Inheritance and macros of the C++ is used for easy creation of modules and ports.

5.1.1. Important SystemC Classes

SystemC has classes for following tasks:

- The hierarchical decomposition of a system into modules
- The structural connectivity between those modules using ports and exports
- The scheduling and synchronization of concurrent processes using events and sensitivity
- The passing of simulated time
- The separation of computation (processes) from communication (channels)
- The independent refinement of computation and communication using interfaces
- Hardware-oriented data types for modeling digital logic and fixed-point arithmetic [46]

SystemC has n-bit wires on which some arithmetic and logical operations can be done. sc_bit, sc_logic are 1-bit data types with 4 possible values: 0, 1, x and z. x means unkown and z means high-impedance. Logic gates are implemented using overloaded operators &, |, $\sim$. Using C++'s template notation they can be defined as arbitrary-width signals: sc_bv$< n >$, sc_lv$< n >$.

To be able to use the native integer data types and the advantages of math library sc_int, sc_uint classes are used. Mathematical operations $+$, $-$, $/$, $\times$, % and shift operations $\ll$, $\gg$ can be used on them. Also calculations can be done in C++ and then cast to sc_int. Again the bit width is determined while the variable definition sc_int$< n >$.

For fixed point numbers there is the sc_fixed$< n, m >$ class. n is the total number of bits used for the signal and m is bits for the integer part. For both integers and fixed

points there are signed and unsigned versions. Also for fixed point there is a faster implementation which uses native double type for making its calculations internally, sc_fixed_fast< n, m >.

SystemC has sc_module class in place of Verilog modules. They are defined by SC_MODULE macro. The macro lets the programmer to define the input and output ports and the function which does the main task of the module. Ports are SystemC the classes sc_in, sc_in_clk, sc_out. The type of the port is determined using the template notation while the variable definition, sc_in<*type*>. For example: sc_in<sc_int< n >> defines an input port which is a n-bit integer. The inner circuitry is the local variables and they can be any of the SystemC classes or native C++ types. sc_in_clk is used for the special signal class sc_clock.

The constructor of the module class is created by another macro SC_CTOR which defines the main function and sensitivity list. Sensitivity list includes the signals which will trigger the module. Mainly it includes the clock and reset signals. If they are defined as input ports "sensitive $\ll$ reset; sensitive $\ll$ clock.pos();" commands make the module sensitive to changes in the reset signal and positive edge of the clock signal. When an event defined in the sensitivity list occurs the function chosen by SC_METHOD macro is called. That function determines the function of the inner parts of corresponding module. There are methods for explicitly define the functions to be run as parallel processes (threads), by using different macros such as SC_THREAD.

The modules are interconnected in another module or in the main function. There, modules are created as instances of the defined classes. The wires connecting them are created as instances of various signal classes and the connections are done by giving same signals as arguments to the different ports to be connected. The simulation starts with "sc_start();" command.

5.1.2. SystemC Executable

After the compilation, a SystemC kernel which can run the defined modules is created. When the executable is run, the kernel adds the modules into the memory and creates a list of events according to the sensitivity lists and runs the functions of the modules succesively according to the event times list.

Along with the SystemC code all facilities of the C++ can be used. Creation of the data to fed the input ports, visualisations of the outputs, comparisons for debugging, automization of determination bit widths can be done

5.1.3. The Simulation

We used SystemC for a full simulation of our system, from the ADC input to the determination of the fit parameters.

Simulation begins with the creation of the input signal. Pulses are created using our model function.

$$f(t) = \begin{cases} f_0 + Ae^{\frac{t-t_0}{\tau}}e^{-\frac{t-t_0}{\tau}} & t \geq t_0 \\ f_0 & t < t_0 \end{cases} \tag{5.1}$$

Random values for A, τ and t_0 are picked. A is chosen from a uniform distribution between predefined minimum and maximum values. τ is chosen from a normal distribution with predefined mean and standard deviation values. The time difference between successive t_0's are chosen from an exponential distribution with a predefined mean value. (All of the random functions are seeded by a given command-line argument to be able to have the same inputs and results when needed.) These pulse parameters are recorded for further comparisons. Offset, f_0, is a superposition of a DC offset, a gaussian noise and a sinusodial fluctutation. Both have predefined amplitudes. The frequency of the sinusodial is chosen to resemble the 50-60 Hz oscillations of the voltage from the outlet.

The input signal is created as an array of samples by discritization of the time domain and quantization of the voltage domain. First the t_0 of the next pulse is picked. The number of samples between the two t_0s is calculated. (If Δt is short enough, two successive pulses may overlap. This situation is called pile-up. (See Figure 5.1) Our system do not deal with that case.) That samples are created by sampling the model function (Equation 5.1) at discrete time values with a period $T = \frac{1}{1.5}$ nsec. That voltage values are converted to the sample values by quantizing, in other words by mapping the voltage interval to integers between 0 and 255.

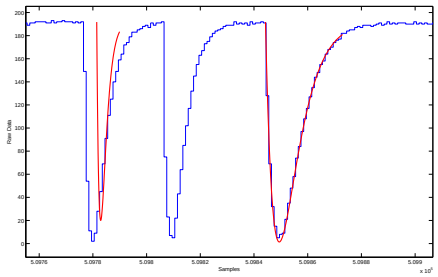


Figure 5.1: Pile-up

$$
\begin{aligned}
f_0[n] &= f_{0,\text{const}} + A_{f_0}\sin(2\pi\nu_{f_0}nT) \\
V_{\text{pulse}}[n] &= \begin{cases} 0 & , nT < t_0 \\ A_{\text{pulse}}(nT - t_0)e^{(nT-t_0)/\tau} & , nT > t_0 \end{cases} \\
\text{noise}[n] &= \text{uniformdist}(0,\sigma_1) + V_{\text{pulse}}[n]\text{uniformdist}(0,\sigma_2) \\
V[n] &= f_0[n] + V_{\text{pulse}}[n] + \text{noise}[n] \\
V_{\text{normalized}}[n] &= \frac{(V[n] - V_{\min})}{(V_{\max} - V_{\min})} \\
V_{\text{quantized}}[n] &= \lfloor V_{\text{norm}}[n]2^8 + 0.5 \rfloor
\end{aligned}
\tag{5.2}
$$

After generating the input samples in the pulse generator module, they are both recorded to a file and sent by ADC_input module with a clock rate 1.5 GHz. Another module ADC_inter picks 8 of them with a clock rate 200 MHz to simulate the real

world conditions and to match the empty signals of the ADC. TR module takes the sample blocks, delay them, and looks whether thresholds are crossed and sends signals accordingly. PD module calculates A, n_{pulse}, offset etc. GP finalizes the guess parameters. Fitter takes the pulse samples array and the guess parameters and apply them the downhill simplex algorithm, explained in the next section, to minimize the χ^2. Debug module records the calculated fit parameters after converting them to their unnormalized units, also gives information about the χ^2 values of the initial guesses and final fit parameters, and about the number of steps taken in the downhill simplex process.

A run of the executable crates random pulse signals added noise, converts that "analog" signal to digital, apply the algorithms of the modules, calculates the fit parameters and record both the produced/real parameters and the parameters found after fitting.

A script is applied to the simulation data after the execution. That script plots some statistics about the "wellness" of the system. It first display how many pulses are catched correctly, the total number of the pulses created, and the average number of steps taken by the downhill simplex algorithm in terms of average number of function calls. First two indicates whether the system is working, and last informs about its efficiency.

Figure 5.2 shows some statistics about an example run. The upper part is the superposition of the generated input signal (blue) and the best fitted functions (red). When there are enough pulses to get a statistically significant info, the amount is too high, hence the details can not be seen well. Figure 5.3 is a close-up of the input, fit function superposition.

Second row shows histograms of Δts (differences between t_0s) and fit parameters of both generated signals (producer) and found by the GPF (consumer).

Third row shows the histograms of the differences between values produced by

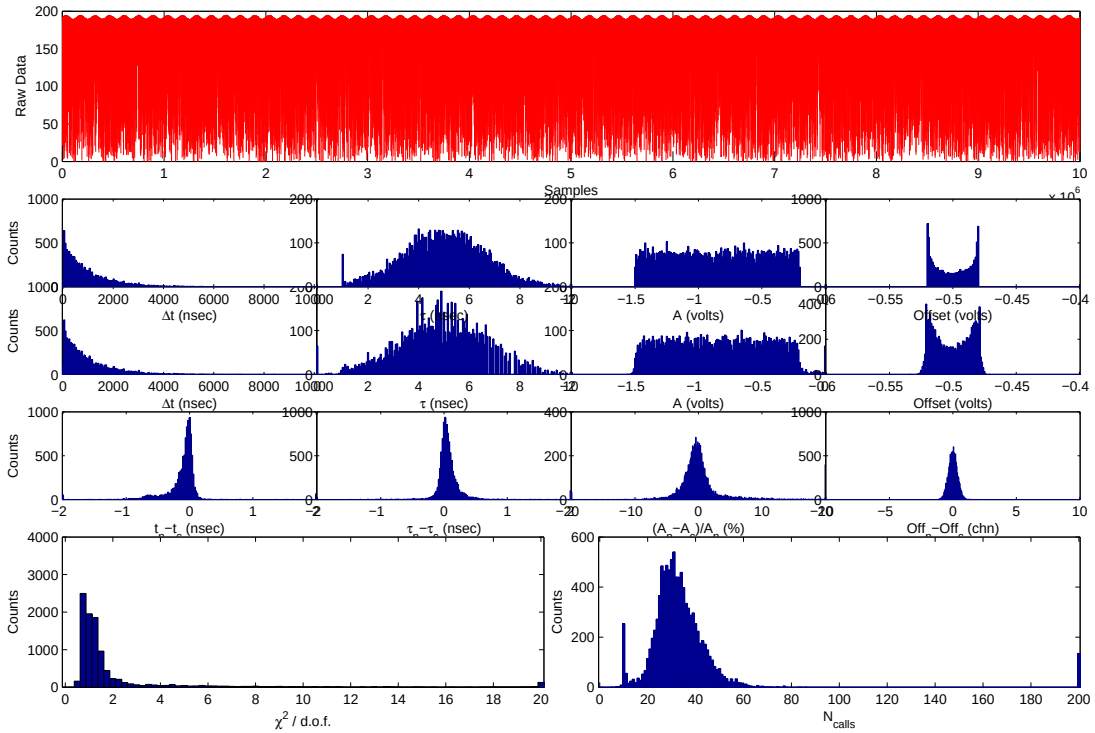


Figure 5.2: Results of a simulation

the producer, and the values found by the consumer and it yield the resolution for each parameter. The smallness of the width of the distributions indicates correct fitting.

The last row has histograms of χ^2 values and the number of function calls. A peak around 1 for χ^2 distribution indicates a good fit and a small mean value for N_{calls} show the efficiency of the fitting algorithm. Right most bars are due to the pile-up conditions or the insufficiencies of the algorithms.

5.1.4. Using SystemC and Verilog Together

Using a C model is easier for the verification of the FPGA design. Micro architecture and RTL design of each block is based on that model. In chip design the verification, in other words testing with simulation and debugging, is as important as the design itself and usually it takes more time than actual design. Verification and design are made in loop again and again until the design passes all of the simulation

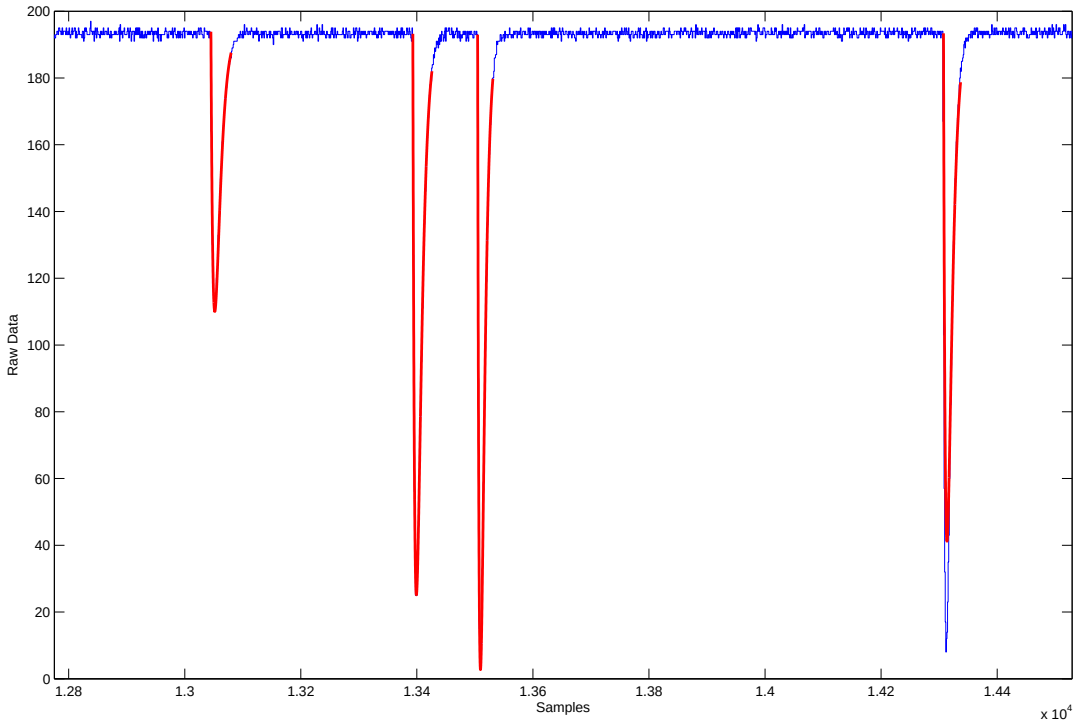


Figure 5.3: Close-up to a few pulse fits

tests.

We wrote the SystemC modules such that they show bit by bit the same input-output behavoir but written with higher level terms. For every module the binary values flowing through the input and output ports are recorded to files. Later they are compared by the outputs of the Verilog simulations. The Verilog code is adjusted until the outputs of Verilog and SystemC are the same. Hence the behavorial correctness of the system design is gained. But this do not ensure that the FPGA in the hand has enough resources to implement that design or that it satisfies the design constraints such that the clock frequency.

Although they behave the same, in other words, they send the exact bit streams to their outputs if they take the say input, the SystemC code do not perfectly match the Verilog design. Because making the design at the same level with two different languages is useless and pointless. We kept the SystemC design at a higher level. In

this way, when a design change came our minds, it was first tested in SystemC domain.

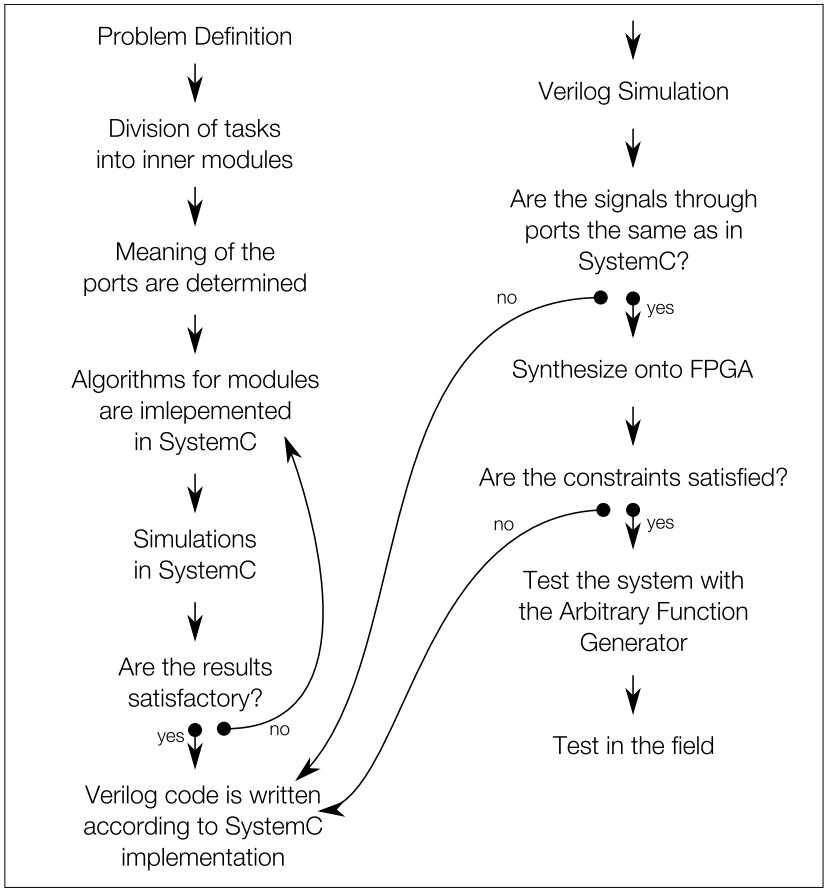


Figure 5.4: Design flow

5.2. Minimization of a Non-linear Function for Best Fit Parameters

"Take the derivative and equal to zero" is the famous method for finding extremum points of a function. In other words to find x^* in following equation.

$$\left.\frac{df}{dx}\right|_{x=x^*} = 0 \tag{5.3}$$

The solution x^*s are where the function $f(x)$ has an extremum and $f(x^*)$s are the corresponding values.

For functions with more than one arguments $f = f(x_1, x_2, \ldots, x_N)$ the equation

becomes a system of equations to be solved simultanously:

$$\left.\frac{\partial f}{\partial x_1}\right|_{x_1=x_1^*} = 0, \ldots, \left.\frac{\partial f}{\partial x_N}\right|_{x_N=x_N^*} = 0 \tag{5.4}$$

The solutions to the equations, the points $(x_1^*, x_2^*, \ldots, x_N^*)$, becomes the extremum points in the N-dimensional space.

An analytical solution to that equation system do not always exist, which means that one can not find explicit expressions in terms of elementary functions. They can be approximated using limits, series expensions or some numerical methods.

5.2.1. A Simple Root Finding Algorithm in 1D

To solve Equation 5.3 a simple numerical root finding algorithm can be used which includes some of the important concepts used in the N-dimensional case of downhill algorithm, such as iteration, step size etc.

First, for a given function f, $f'(x) = \frac{df}{dx}$ is calculated analytically. This will be the function we sample for different x values to find its roots. Then an initial value, x_0, and a step size, l, is chosen and the iterated variable x_i is let to be x_0. $(x_i = x_0)$

At every step

- the sign of $f_i' = f'(x_i)$, which is $s_i = \text{sgn}(f_i')$, is calculated and
- x_{i+1} is let to be $x_{i+1} = x_i + l$ and the sign of $f'(x_{i+1})$, which is $s_{i+1} = \text{sgn}(f'(x_{i+1}))$, is calculated.
- if the step size, l, is chosen reasonably
  - the differentness of the signs indicates a crossing of the x-axis. Hence there is a root between $[x_i, x_{i+1}]$.
  - the equality of the of the signs indicates that the x-axis is not crossed and there is no root between $[x_i, x_{i+1}]$.
  - (it is very improbable but if f_i' is zero, then we are lucky to step on a root.)

- x_i is let to equal to x_{i+1}, $x_i = x_{i+1}$
  - for the second case iteration is repeated with the same step size, l.
  - for the first case, whereas we crossed the x-axis, l is halved and inverted such that $l \leftarrow -\frac{l}{2}$ and the iteration is repeated.
- if l is shorter then a predefined value, $l < l_0$, the task is finished, which means that we are in the vicinity of a root nearer then l_0.

The point where the algorithm is finished is an extremum of the function f.

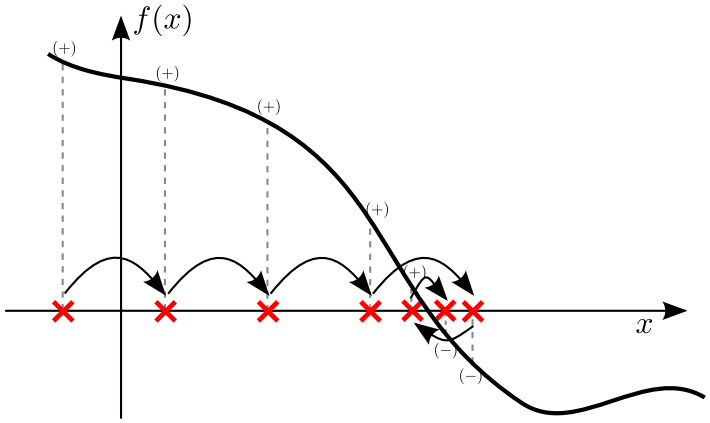


Figure 5.5: A simple root finding algorithm which samples the function and checks the signs

Every algorithm has strong parts and some weaknesses. This algorithm is easy to understand and implement, but could miss some roots. For example, when the step size is larger than successive roots, it can step on both of them and get the same sign.

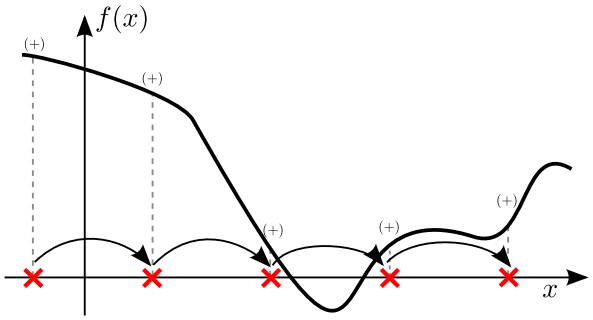


Figure 5.6: Deficiency of the root finding algorithm. Jumping over two roots

5.2.2. Meaning of the Best Fit for Our Model

As we said the FA module gets the sampled pulse signal as an array of 8-bit samples, initial guesses and tries to adjust the three parameters to find the parameters of the modeling function for the best fit.

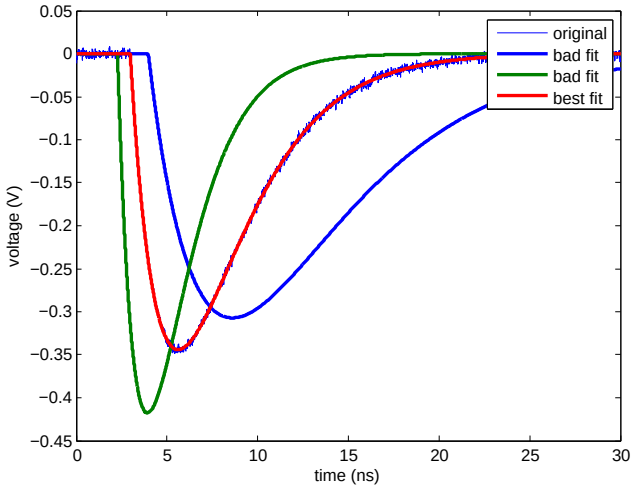


Figure 5.7: Fits with different fit triplets

The criteria for the goodness of a fit is chosen as the χ^2 function. The χ^2 is defined as such: The samples are $\{t_i, y_i\}$ pairs. But actually what we have is the y_i values. t_i is some timestamp, t_s, plus i times $T = 2/3$ nsec. χ^2 is the sum of the square of the differences between the fitting function and actual sample values at sampling times.

$$\chi^2(t_0, A, \tau) = \sum_{i \leftrightarrow t_0}^{N_{\text{samples}}} (f(t_0, A, \tau; t_i) - y_i)^2 \tag{5.5}$$

The meaning is that, at every sample point the fit funcition is different from the actual values. How different they are is measured by the difference between fit value, f_i, and the actual value, y_i. Because f_i may be larger or smaller than y_i the square of the difference is taken. (Actually in the hardware implementation the absolute value is used because it is a lot easier and faster to calculate the absolute value than making a squaring operation.) The total difference is defined as the sum of the squares.

χ^2 is a function of three variables: t_0, A and τ. It can be thought as a scalar field in a three dimensional space. On that space where the axes are t_0, A and τ, every point $P(t_0, A, \tau)$, correspondes to a χ^2 value. The initial guesses send us to a point near the minimal χ^2. From there to go to the mimimum is the task of the FA.

For non-pathalogical pulses there is only one extremum point, which is the mimimum value of the χ^2. Hence there is no need to worry about to be stuck in a local minima.

5.2.3. The Downhill Simplex Method

The "downhill simplex method", or "amoeba method", is proposed by John Nelder and Roger Mead in 1965 [47]. It is a heuristic nonlinear optimization technique for minimizing a function with many arguments. It needs only function evaluations, $F(\mathbf{P})$. The analytical derivatives of the function with respect to the parameters are not needed.

A simplex is a geometric object which is the generalization of the triangle to higer dimensions. For example the triangle is a 2-simplex and the tetrahedron is a 3-simplex. Downhill simplex uses a N-simplex (which has $N + 1$ vertecies) for a minimization problem in a N-dimensional space. As the iterations goes on, the simplex goes down the gradients and shrinks around the minima.

To construct the simplex one needs $N + 1$ points. The guessed point, $\mathbf{P_0}$, is one of them. The other N points may be chosen such that

$$\mathbf{P_i} = \mathbf{P_0} + \lambda_i \mathbf{e_i}$$

where $\mathbf{e_i}$ are basis vectors, and λ are constants to be chosen, according to characteristic length scale of the problem, to determine the inital size and shape of the simplex.

The idea is that, the function to be minimized is calculated at the vertecies of the

simplex. They can be called the test cases. At every iteration a new test case is chosen and replaced with an old one, according to the positions and values of old test cases. This can be visualized as the movement of one vertex of the simplex. At every step 1 vertex or N verticies changes its position. The algorithm is about how that change is made, in other words, how to move the simplex [48].

The vertex with highest value (worst point as a candidate for a minima) is reflected around the centroid of the remaining points of the simplex. If the new point is better than the best point of the previous simplex, which means that it is going to downwards, then the simplex is streched along that direction. If the new point is not better than the previous worst value, which means that simplex is stepped over a valley, then the simplex is shrinked. Figure 5.8 shows the possible movements can be made by the simplex along the iterations.

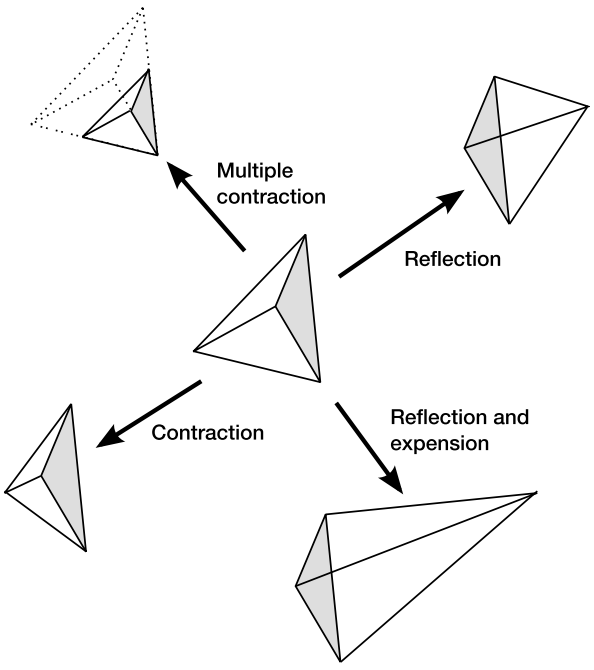


Figure 5.8: Movements of the simplex. Adapted from [48]

The iterations start after ordering the first $N+1$ points. The points correponding to the highest, next highest, and lowest values ($\mathbf{P}_{\mathrm{hi}}$, $\mathbf{P}_{\mathrm{nhi}}$, $\mathbf{P}_{\mathrm{lo}}$) of the function are determined. Then

- lowest point is reflected. $\mathbf{P}_{\text{lo}} \rightarrow \mathbf{P}_{\text{ref}}$
  - If the reflected point is better than the highest point $F(\mathbf{P}_{\text{ref}}) < F(\mathbf{P}_{\text{hi}})$, then $\mathbf{P}_{\text{ref}}$ is expanded by a factor of 2.
  - If the reflected point is worse than the second-highest point $F(\mathbf{P}_{\text{ref}}) > F(\mathbf{P}_{\text{nhi}})$, then $\mathbf{P}_{\text{ref}}$ contracted by a factor of 0.5. After contraction if the highest point do not changed than contract the whole simplex around the around the lowest point by a factor of 0.5.
- Check for the stopping conditions.

There are two stopping conditions: One is that if the fractional range from the $\mathbf{P}_{\text{hi}}$ to $\mathbf{P}_{\text{lo}}$ is small enough the algorithm is stopped. It is calculated as such

$$\left| \frac{F(\mathbf{P}_{\text{hi}}) - F(\mathbf{P}_{\text{hi}})}{F(\mathbf{P}_{\text{lo}}) + F(\mathbf{P}_{\text{lo}})} \right|$$

Second stopping condition is that the number of steps exceeds the predefined maximum allowed value. If it is exceeded it means that the simplex is lost in a flat region, or stuck somewhere else and could not succeed to shrink enough around a minima.

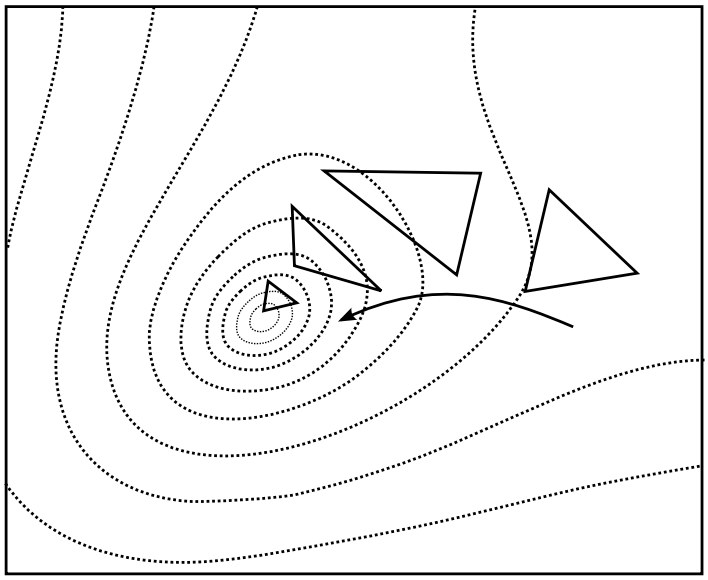


Figure 5.9: Simplex motion on 2D contour plot

The computational cost of the DSM depends on the number of steps to be taken to get an approximate value for the minima, and the computational cost of one function

evaluation. For our problem one evaluation of the χ^2 function takes a loop on the samples array. At every step the difference between actual sample and the fit function at that time is calculated. The evaluation of the fit function includes some multiplication operations and accessing to the look up tables. The divisions and exponentiations are made using LUTs. The absolute values of the differences are summed.

The number of steps to be taken is different for every pulse. Our estimation is that they will be between 20 and 60.

6. RESULTS

6.1. Measurement of Time Resolution

We designed a test suite for our hardware to measure the intrinsic time resolution of the system which can handle the input pulses in a deadtimeless manner. The test suite can also define its trigger condition on the boars, such that there is no external trigger electronics/logic.

The hardware consists of a two channel ADC, a FPGA and PCI bus to connect it to a computer, all of them on a single board which has PMC industry standard from factor. The task of the digitizer is to measure and send the incoming times of the pulses it detects. The specification we determined for this study was 100 ps.

For this purpose we made a special digital design and loaded it to the FPGA. This design has the important characteristics of our main DAQ system described in chapter 4 and is a basic working DAQ system in itself. [4]

There is a host software part of the system which consists of two parallel running threads which work simultaneously, one for event recording and one for event analyses.

We used an advanced digitizer board AD1500 manufactured by VMetro. (See Figure 6.1) It has PMC form factor. We fitted it into a PCI carrier card, and used it in a regular PC. It can also be fitted into industrial compact PCI carriers.

The ADC on the board has 8 bit vertical resolution. It must be derived with up to 1.5 Giga samples per seconds (Gsps) sampling clock, which corresponds to a sampling period of, $T = 1/1.5$ ns $= 0.667$ ps. The same clock signal can be distributed

[4]We presented this system as a laboratory exercise in "the International School Of Trigger and Data AcQuisition" (ISOTDAQ) 2010. It is a school organized by the TDAQ team of CERN to teach and guide M.S. and Ph.D. students who intend to work in experimental particle physics, specializing in trigger and data acquisition.

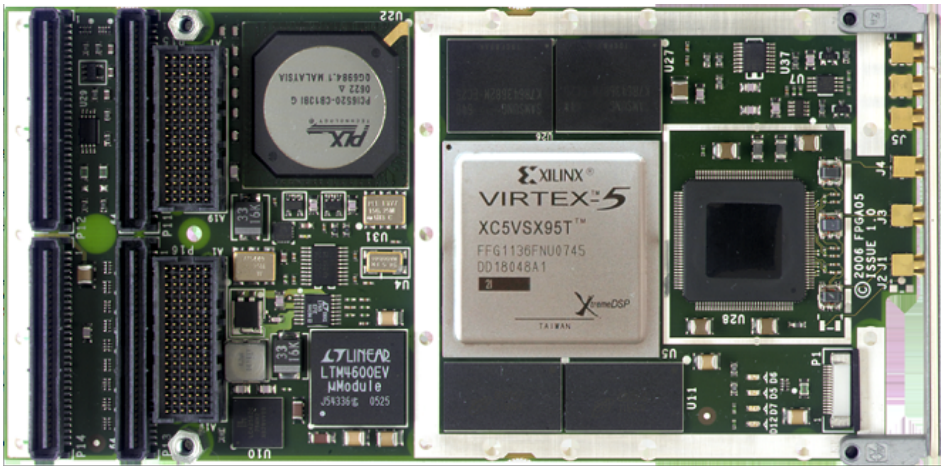


Figure 6.1: AD1500, digitizer board from VMetro

to multiple boards for synchronous operation. The clock signal we have is basically an oscillator that provides sinusoids.

Ideally, the period is constant and does not fluctuate over time. But actually it

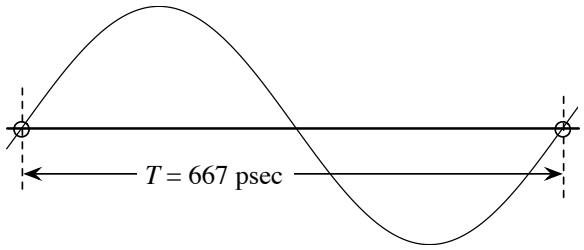


Figure 6.2: One period of ideal clock signal

differs from time to time. Sometimes the clock ticks earlier, and other times it becomes late for very short times. This fact is called "jitter". It happens in all digital systems

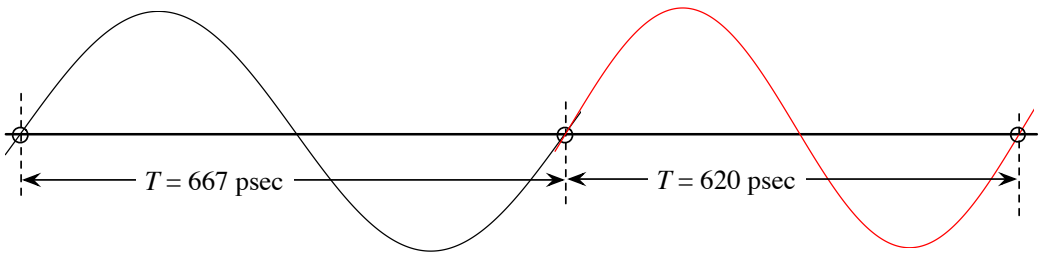


Figure 6.3: Jittered clock

synchronized with a clock signal. Therefore "sampling frequency" means actually "the average sampling frequency".

Figure 6.4 shows a pulse which is sampled with a clock with exaggerated "bad" jitter. If we want to determine when the voltage level of the signal crosses a threshold

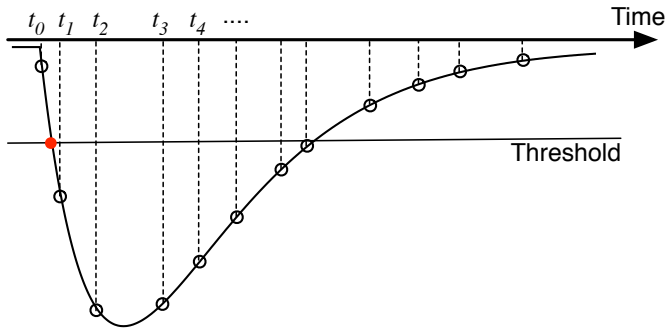


Figure 6.4: A pulse signal sampled with non-regular intervals due to jitter

value from high to low, first, we have to find t_0 and t_1, the time of two samplings in which that kind of transition happens. Then, we interpolate the time of interest. Interpolation according to a correct model of the signal enhances the time resolution, which equals to sampling period ($\Delta t = T$) if no interpolation is used. At this point the accuracy of sampling times become important.

Accuracy of such a time resolution depends on many other factors, such as the noise level in the analog part of the board, the quality of cables being used, and most importantly, the quality of the sampling clock used with the digitizer. Overall, this is called the intrinsic time resolution of the system and is considered to be a part of the specification of the system. It is the lower limit on timing resolution of any other measurement which is using this instrument.

6.1.1. Usage of Digitizer

The digitizer board is controlled by an FPGA. Its program is called firmware. We designed this firmware as part of this study. The firmware expects us to send:

- The threshold value: The ADC is an 8-bit digitizer, thus digitized values are in between 0 and 255, so the threshold value must be in this range. The digitizer will get a trigger, when the analog signal becomes smaller than this threshold value

for the first time. After that point, it starts to store the values of the samples in its own memory on the digitizer card.

- Width: The total number of samples it should store whenever a trigger condition is detected. This must be multiple of 8, because of the data width we use.
- Pre-width: The number of samples that need to be stored before the trigger point. It must also be multiple of 8. The minimum value can be 16. The maximum value is limited by the total number of samples explained above. For example, if width = 64 and pre-width = 16, then 16 samples are going to be before the trigger window, and 48 samples are going to be located at and after the trigger window, totaling 64 samples.

These parameters are set using a software which uses the device drivers we wrote as part of this study for controlling and communicating with the digitizer. The FPGA has input registers which can be accessed via computer memory. When that memory locations are changed, using the library we wrote, the parameters are updated by the driver, if the system is configured properly. Using these setup, the configurations can be made in a UNIX command shell environment and the device can be started and stopped.

After configuring the board with the above parameters, the board is ready to do its job. When the ARM command is sent to the board, it starts analyzing the signal, and sends packets of events each containing the samples of the pulse and the timestamp of trigger window. These packets are called event frames. Generally they may consist of three parts, the header, the footer and the data. The data is the digitized raw signal from the detector. The header and footer include some meta information about the data, such as the source (channel of ADC) and a timestamp etc.

Our event frame has the following syntax:

There is a 128 bit header. Its first 12 bits in the are left blank for future usage.

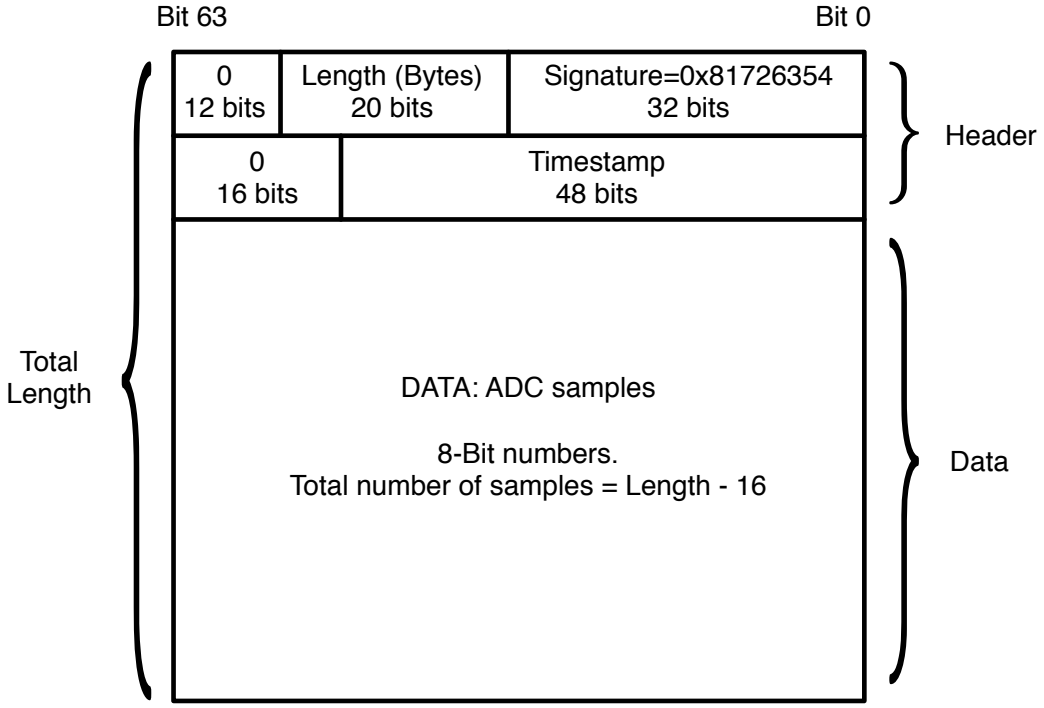


Figure 6.5: Event frame created by the digitizer for each pulse

Next 20 bits is the length of the event frame, in other words the number of samples to be recorded for that triggering plus the length of the header. This value is the same for each pulse in this test design.

The next 32 bits are reserved for the signature, which may used to label the event source. Each board connected to a different detector may be assigned a different signature value, and the DAQ system which analyses the event frames coming from different source can understand from which detector that frame has come. The value for the signature was chosen such that the occurrence of that pattern in the data is highly improbable. This non-similarity may also serve for correcting reading error. If a corruption occurs while reading event frames, and the read index placed at a random position, we start to search this signature within the data stream to determine the start of the next valid frame.

The next 16 bits are left blank. The timestamp is recorded to the next 48 bits after that. The ADC reads the signal in blocks of 8 samples and the timestamp is the

number of the block in which the threshold is crossed. At every block the timestamp value increases by one. To convert the timestamp value in the event frame to absolute time (in nano seconds) from the arming of the device: $t_{\text{abs}} = \text{timestamp} \times 8 \times 1/1.5$.

After the header comes the data part which includes the successive 8-bit sample values of the recorded pulse. The final version of the design will not include the raw data, but the fit parameters.

When the system is armed, the event frames are sent from the memory of the digitizer to a large ring buffer memory on the RAM of the PC using Direct Memory Access (DMA) feature of the PCI interface. These event frames may further analyzed via software on the computer.

6.1.2. Experiment

Our first experimental setup to determine the time resolution is given in Figure 6.6.

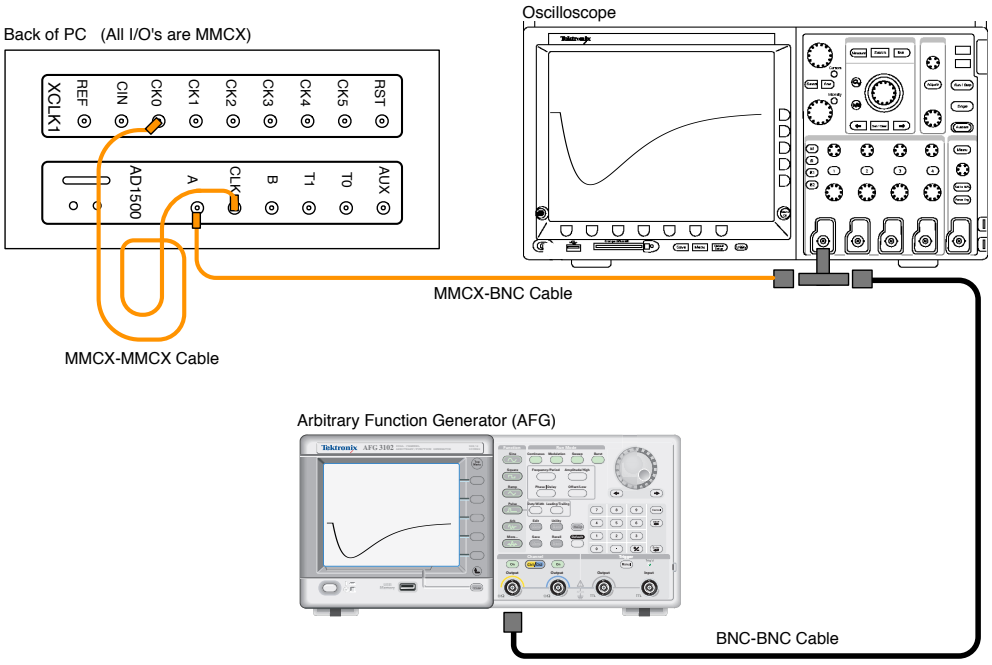


Figure 6.6: Experimental setup 1

"Back of PC" part shows two PCI cards. Upper one is the external clock and

lower one is our DAQ board AD1500. There is no other internal connections between clock and AD1500 other than the MMCX cable.

The clock has 6 identical outputs, to supply multiple acquisition boards. One of them (CK0) is connected to the clock input of the AD1500 (CLK).

The ADC of AD1500 has two input channel. The channel A is connected to the output of the Arbitrary Function Generator (AFG), AFG3252 from Tektronix. The connection goes through an oscilloscope which is used to control whether the system works properly, whether correct signals are generated.

AFG is controlled by a software which was written as part of this study. It communicates with the computer over on Ethernet connection and is adjusted with vxi11 protocol. Any parameter change can be done remotely from the host computer without touching the device by sending some commands to the IP address of the AFG through a TCP connection.

As its name suggests, the AFG is able to produce signals which are not simple waveforms such a sinusoid, sawtooth, square wave etc., but arbitrary waveforms. It has an user definable array in its internal memory which can hold up to 50k 16-bit points. According to a given frequency and operation mode that table is interpolated and output is produced.

AFG has two operation modes. In continuous mode the frequency (or period) is set and the shape defined in the table is oscillated continuously. In burst mode the length of one signal and the time interval between successive intervals are set. We used the burst mode to simulate incoming discrete pulses from a detector. We use an AFG instead of a real detector, because we want an absolute control over the pulses that will be analyzed on our board.

We used our software, which samples an analytical function and uploads that

array to the AFG, to generating the pulse

$$\frac{A}{e\tau}te^{-t/\tau} \tag{6.1}$$

where $A = -0.4$ V and $\tau = 6$ ns, with a event rate of 10 KHz. Hence we used AFG to simulate a scintillator type pulse which is generated at every 100 μs. The pulse displayed by the oscilloscope can be seen in the Figure 6.7. It is extremely predictable and reproducible.

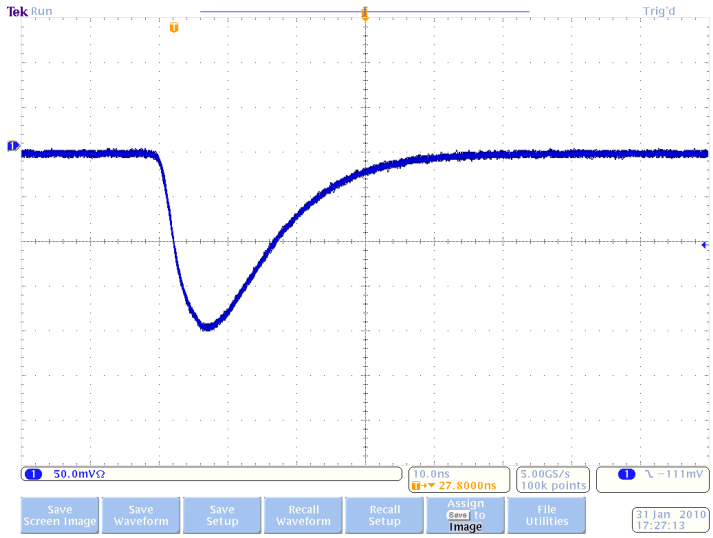


Figure 6.7: Pulse generated by AFG according to Equation 6.1 displayed by oscilloscope

Using this setup the time intervals between these repetitive pulses are measured. (See Figure 6.8) We measured these intervals tens of thousands times per second and created a histogram with ROOT library showing the distribution of T.

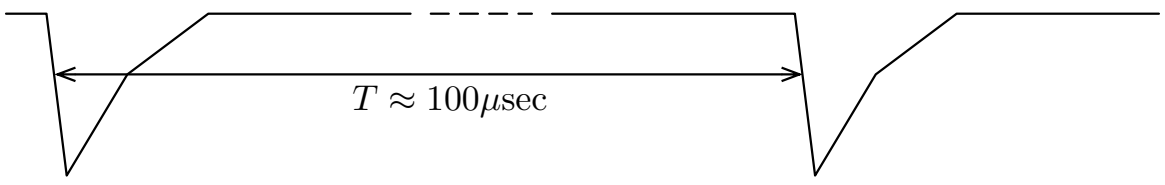


Figure 6.8: Successive pulses of which time difference is about 100 μs

For this special measurement we wrote a software which takes the threshold, width, pre-width values, and the histogram parameters as command-line arguments. The test application

- Sends the configuration parameters to the digitizer, and arm it to start the acquisition.
- A thread read the events from the PCI card and put into a local buffer.
- Another thread read the events one by one, and find the crossing point of the threshold time by interpolating a given number of samples around the interested point.
- After each time determination, its difference from the previous value is found and fed into a histogram, which can be seen on the screen in realtime.
- The samples of the signal, which are the raw data we get from the board, are plotted also on the screen, just like an oscilloscope, at every second for diagnostic purposes.

The t_is are calculated using four samples: two samples before and two samples after the threshold crossing. First, a linear interpolation is made on these four points, and the slope of the straight line, m, which goes through these points is found with linear regression method. Then by solving a simple equation, the time where this line crosses the threshold value, t_c is found. That value is relative to the time stamp, t_s, of the event frame. The absolute time of the event is $t = t_c + t_s$.

We started the program with threshold = 80, width = 256, pre-width = 16, number of histogram bins = 10000, lowest bin value = 99995, highest bin value = 100005. (Histogram unit is in ns.) After collecting about 50000 events, which is enough to determine the deviation of the distribution precisely, the histogram in the Figure 6.9 is get.

A Gaussian fit on the histogram give a mean-period, $\overline{T}$ = 99996.53 ns, and a sigma-period, $\sigma_T \simeq 0.025$ ns. The value for $\overline{T}$ is around 100 μs, which is as expected and as we programmed the AFG. The stability of that value, in other words the amount of fluctuations in T, the jitter, is given by the σ_T (the width of the distribution). If there is a jitter/error in the period, the time of the sampling is represented by $t_i \pm \sigma_t$.

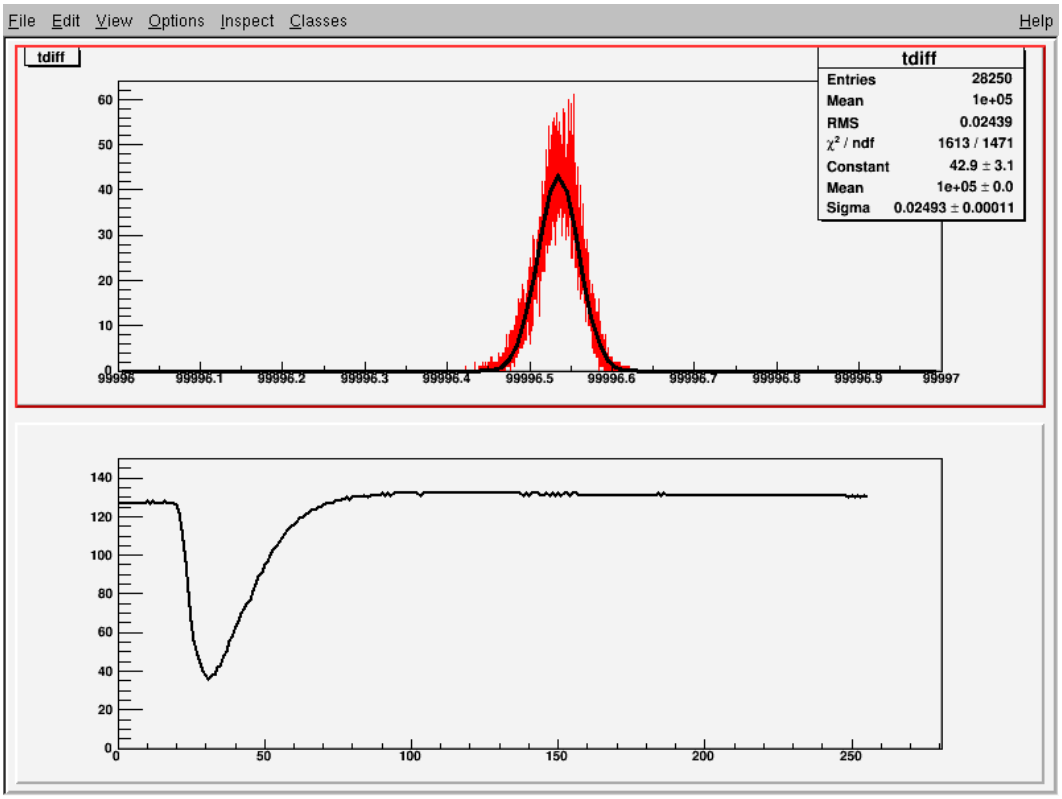


Figure 6.9: Upper plot is the histogram of interarrival times, T, the lower plot is a sample of raw pulse signal.

$$\begin{aligned} T_i &= t_i - t_{i-1} \\ \sigma_T &= \sqrt{2}\sigma_t \\ \sigma_t &\simeq 0.017\text{ns} = 17\text{ps} \end{aligned} \tag{6.2}$$

Hence we found that the intrinsic time resolution of this device is $\sigma_t = 17$ ps.

In this setup, the source of the jitter may be the external clock and/or the AFG. To be sure that the clock is mainly responsible from the jitter, we changed the setup a little bit. See Figure 6.10.

In the second setup a delayed version of the generated pulses is introduced, where the delay amount is adjustable using a delay box of which values, t_{delay}, are between 1 ns to 64 ns. ($t_{\text{delay}} \ll T$) The superposition of original and delayed signal is sent

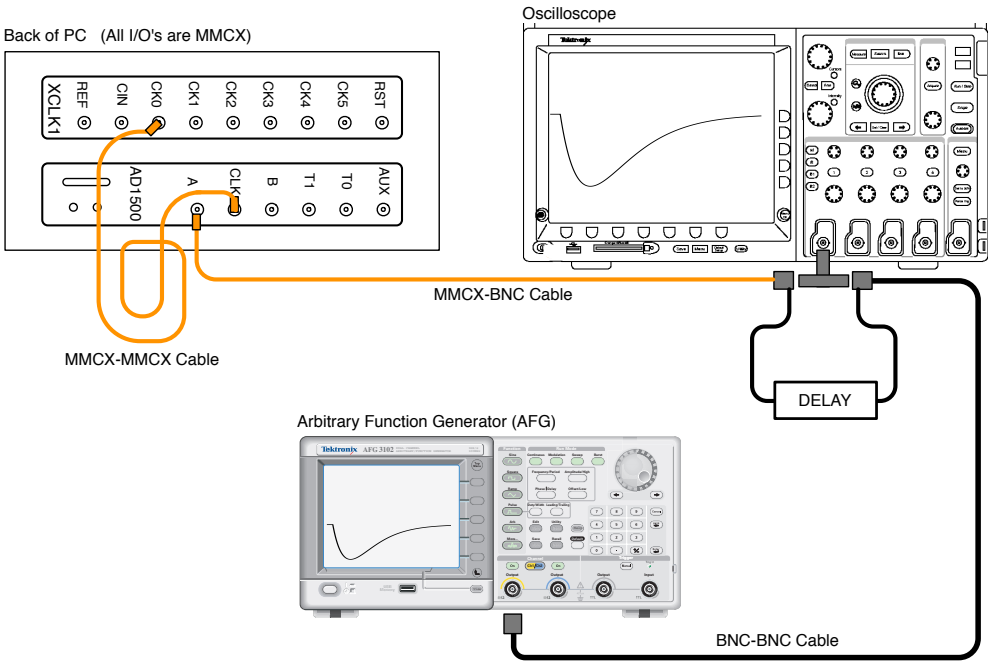


Figure 6.10: Experimental setup 2

to the ADC. A is set to -0.7. The result seen from the oscilloscope is as in Figure 6.11. (The third ripple is due to impedance mismatches at the connection point of the oscilloscope.)

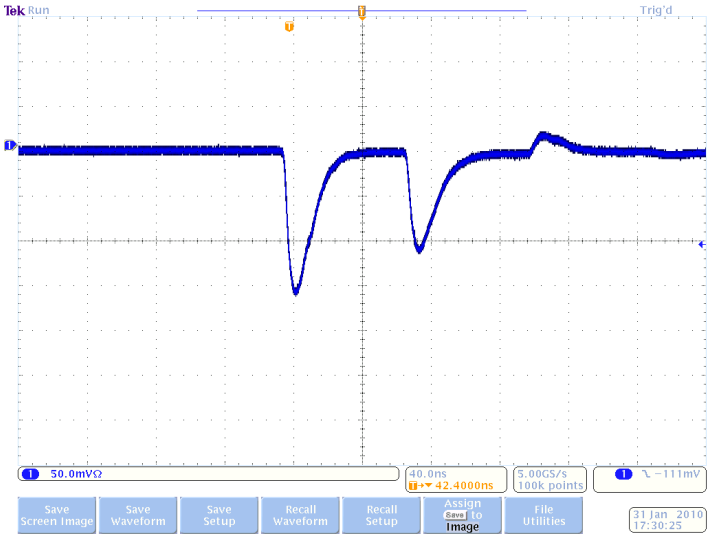


Figure 6.11: Superposition of the generated pulse and its delayed version

The amount of delay is set to 32 ns and the program is run with parameters, threshold = 110, width = 48, pre-width = 16.

This time the threshold value is increased to let it cross both pulses in their linear rise region and the width value is decreased so that an event frame includes only one pulse. If the width value would high, an event belonging to the big pulse will contain the beginning of the small pulse in its data, hence the second pulse could not trigger the system and create an event frame, and thus, again, only the time difference between original pulses is measured.

A delayed version of the signal is used to know the interarrival times exactly. In the first setup both t_i (timestamp of the first big pulse) and t_{i+1} (timestamp of the second big pulse), and hence their difference T_i, may be affected by the (hypothetical) jitter of the AFG. Thus we may not be sure that the found σ_t value belongs only to the jitter in the DAQ system.

In the second setup, although ${}_ti$ (timestamp of the first big pulse) may be affected by the jitter of the AFG, t_{i+1} (timestamp of the next small pulse) is not related to AFG in any way. Their difference depends only to the amount of the delay, which is the most simple component of all electrical devices, a cable, which has no jitter at all, because its length does not change in time.

The histogram parameters are determined such that only the interarrival times between successive big and small pulses is shown, of which mean values are equal to a constant time plus the delay amount. (There is another peak around the 100 μs value which is due to the interarrival times between successive small and big pulses.)

During data taking, the delay amount is changed from 32 ns, to 48, 56, 60, 62, 63, 63.5 ns respectively, staying in each delay amount for about 5 seconds such that same height for each peak is seen on the online histogram. The screenshot is shown in Figure 6.12.

A gaussian fit is performed on each peak one by one and the mean and sigma values related to every delay amount is calculated. Then the differences between delay amounts and calculated mean values are compared. It was seen that they are the same

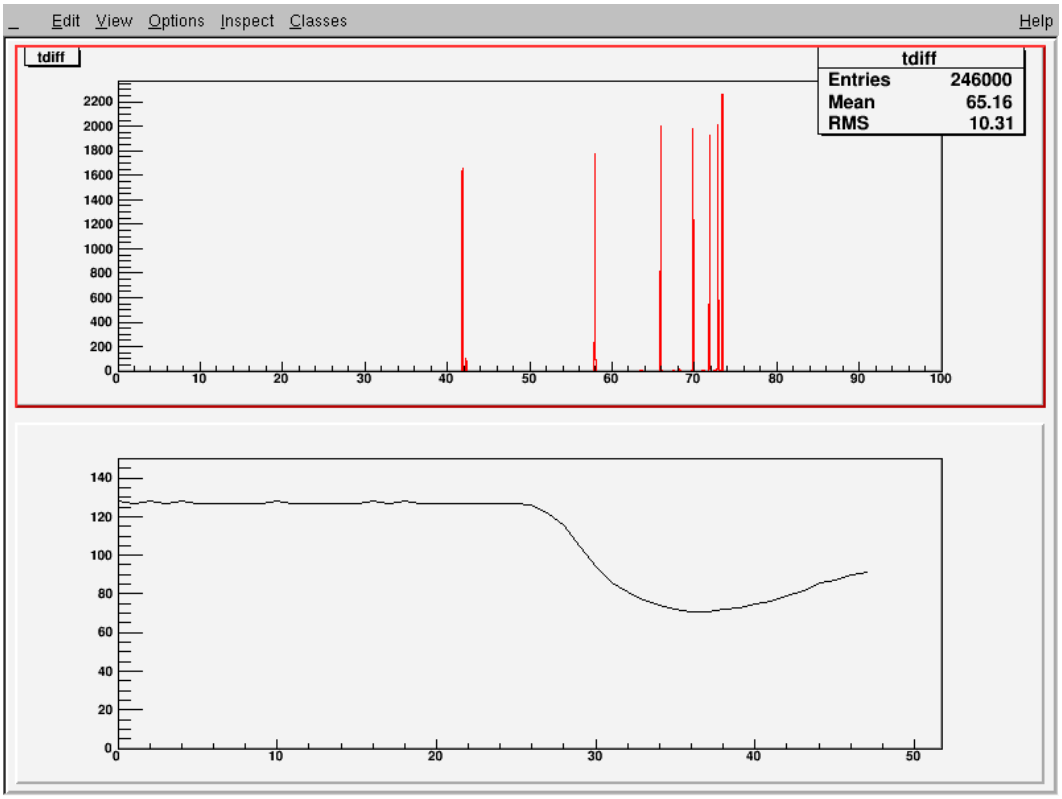


Figure 6.12: Histograms of the time interarrival between original and delayed pulses

in the corresponding sigma range, which shows that the system is consistent. And the σ value found is around 17 ps, which is the actual intrinsic time resolution of our DAQ system which is in the pico second region!

Thus we achieved a better timing resolution with this system than the original design parameters.

6.2. Future Work

We will include the fit routines into our firmware, so only the fit parameters will be passed to the host system, instead of raw data. This will reduce the computational load on the PC which may not be able to make complex calculations such as χ^2 minimization in realtime and make it possible to get the some result of the experiments on-the-fly, rather than having another DAQ layer for analyzing the raw data which work offline. It will be done after the implementation of fitting agent module is finished.

7. CONCLUSION

We designed and implemented a data acquisition system, which is deadtimeless, triggerless (with an autotriggering system) and it has a time resolution which can be considered one of the best using regular electronics.

These kind of systems will simplify the hardware of large scale experiments (like LHC experiment) greatly, and move the complexity of the system to the software side, which is a big advantage, because it can be upgraded without touching the actual hardware, thanks to the reprogrammability of FPGA, depending on the needs that may arise after the hardware setup is commissined.

REFERENCES

1. Plato. *Timaeus*. -300.

2. Griffiths D. *Introduction to Elementary Particles*. Wiley-VCH, 2008.

3. Nordmann A. Refraction on an Aperture. `http://en.wikipedia.org/wiki/File:Refraction_on_an_aperture_-_Huygens-Fresnel_principle.svg`, 2007.

4. Close F. *Particle Physics: A Very Short Introduction*. Oxford University Press, USA, 2004.

5. Rutherford E. The Stability of Atoms. *Proceedings of the Physical Society of London*, 33(1):389–394, December 1920.

6. Steere A. R. *A Timeline of Particle Accelerators*. Masters of science in physics, Michigan State University, 2005.

7. Cockcroft J.D. and Walton E. Experiments with High Velocity Positive Ions. *Proceedings of the Royal Society of London*, (136):619–630, 1932.

8. Van de Graaff R., Compton K., and Van Atta L. The Electrostatic Production of High Voltage for Nuclear Investigations. *Physical Review*, 43(3):149–157, February 1933.

9. Ising G. *Astronomi och Fysik*, (18):1–4, 1924.

10. Wideröe R. *Arch.für Elektrotechnik*, (21):387–406, 1928.

11. Lawrence E.O. and Livingston. The production of high speed light ions without the use of high voltages. *Phys. Rev.*, (40):19–35, 1932.

12. Chao A. et. al. 2001 Snowmass Accelerator R&D Report, 2001.

13. Panofsky W. The Evolution of Particle Accelerators and Colliders. *Spring*, 1997.

14. CERN. Accelerator Complex. `http://public.web.cern.ch/public/en/research/AccelComplex-en.html`, 2008.

15. Leo W. *Techniques for Nuclear and Particle Physics Experiments: A How-to Approach*. Springer, 1994.

16. Amsler C., Doser M., and Antonelli M. Review of Particle Physics. *Physics Letters B*, 667(1-5):1–6, September 2008.

17. Green D. *The Physics of Particle Detectors*. Cambridge University Press, 2005.

18. Rohlf J. W. *Modern Physics from a to Z*. Wiley, 1994.

19. Hünniger D. Photoelectric Effect Schematic. `http://commons.wikimedia.org/wiki/File:Photoelectric_Effect_Schematic-de.svg`, 2006.

20. Wikimedia. Compton Scattering. `http://en.wikipedia.org/wiki/File:Compton-scattering.svg`, 2007.

21. ETH. Particle shower. `http://www.ipp.phys.ethz.ch/aboutus/?file=poster`, 2006.

22. Horvath A. Photomultiplier Tube. `http://en.wikipedia.org/wiki/File:Photomultipliertube.svg`, 2006.

23. Vandelli W. Introduction to Data Acquisition, ISOTDAQ2010, 2010.

24. Akdogan T. Flash ADC Based DAQ System Design for the MIT/Bates Compton Polarimeter, 1999.

25. Franklin W., Akdogan T., and Marfuta P. A Compton polarimeter for the MIT/Bates South Hall Ring. *Progress in Particle and Nuclear Physics*, 44:61–

62, March 2000.

26. Akdogan T. Performance Report and Polarization Results of the MIT/Bates Compton Polarimeter, 2005.

27. Crawford C., Sindile A., and Akdogan T. Measurement of the Proton's Electric to Magnetic Form Factor Ratio. *Physical Review Letters*, 98(5):052301, January 2007.

28. Hien D. and Senzaki T. Development of a fast 12-bit ADC for a nuclear spectroscopy system1. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 457(1-2):356–360, January 2001.

29. Bolic M. Digital gamma-ray spectroscopy based on FPGA technology. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 482(3):761–766, April 2002.

30. Streun M. A PET system with free running ADCs. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 486(1-2):18–21, June 2002.

31. Khomich A. and Hinkelbein C. Using FPGA coprocessor for ATLAS level 2 trigger application. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 566(1):80–84, October 2006.

32. Nicolau C. An FPGA-based readout electronics for neutrino telescopes. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 567(2):552–555, November 2006.

33. Müller H., Pimenta R., and Yin Z. Configurable electronics with low noise and 14-bit dynamic range for photodiode-based photon detectors. *Nuclear Instruments*

and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 565(2):768–783, September 2006.

34. Giachero A., Guardincerri E., Musico P., Pallavicini M., and Ottonello P. Design and performances of a multichannel high resolution simultaneous sampling ADC card with on-board data elaboration capabilities. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 572(1):2, 2007.

35. Shimazoe K., Yeol Y., and Minamikawa Y. Development of 40 channel waveform sampling CMOS ASIC board for Positron Emission Tomography. *Nuclear Inst. and Methods in Physics Research, A*, 2006.

36. Guo J., You C., and Zhou K. A 10 GHz 4:1 MUX and 1:4 DEMUX implemented by a Gigahertz SiGe FPGA for fast ADC. *Integration, the VLSI Journal*, 38(3), 2005.

37. Xilinx Corporation. Virtex-5 Family Overview, 2009.

38. NIST. Stopping-power and Range Tables for Electrons. `http://physics.nist.gov/PhysRefData/Star/Text/ESTAR.html`, 2010.

39. Saint Gobain Crystals. Premium Plastic Scintillators (Data Sheet), 2010.

40. Nam J. A detailed Monte Carlo simulation for the Belle TOF system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 491(1-2):54–68, September 2002.

41. Buck C., Hartmann F.X., Motta D., and Schoenert S. Energy transfer and light yield properties of a new highly loaded indium(III) β-diketonate organic scintillator system. *Chemical Physics Letters*, 435(4-6):252–256, February 2007.

42. Adamek P. Sampled Signal. `http://en.wikipedia.org/wiki/File:Sampled.signal.svg`, 2006.

43. Adamek P. Quantized Signal. `http://en.wikipedia.org/wiki/File:Quantized.signal.svg`, 2006.

44. Adamek P. Digital Signal. `http://en.wikipedia.org/wiki/File:Digital.signal.svg`, 2006.

45. Sakulin H. Field Programmable Gate Arrays, ISOTDAQ2010, 2010.

46. Automation Design Committee Computer Standards IEEE. *IEEE Standard SystemC Language Reference Manual*. Number March. IEEE, 2006.

47. Nelder J. A. and Mead R. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 1965.

48. Press W., Flannery B., Teukolsky S., and Vetterling W. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.