UNIVERSITY OF BERGAMO

SCHOOL OF DOCTORAL STUDIES
Doctoral Degree in Engineering and Applied Sciences
XXXII Cycle
SSD: ING-INF/01 Electronics

DEVELOPMENT OF LOW LATENCY, NON-INTERCEPTIVE, PHASE-BASED BEAM DIAGNOSTIC SYSTEMS FOR A PROTON THERAPY LINEAR ACCELERATOR

Supervisor
Chiar.mo Prof. Valerio Re

Doctoral Thesis
Francesco GALIZZI
Student ID 1021659

Co-supervisors
Prof. Gianluca Traversi
Dr. Michele Caldara

ACADEMIC YEAR 2018–2019

Contents

List of Figures

List of Tables

Introduction

This thesis discusses the development of a low latency, non-interceptive, phase-based beam diagnostic system for the first proton therapy linear accelerator.

Proton therapy is a type of therapy used to treat cancerous tissue. It is part of the broader category of external beam radiation therapy, representing one of the most recent options among these. Cancer is the second leading cause of death globally and it is responsible for about 1 in 6 deaths. Nowadays about 18 million people are diagnosed with cancer worldwide annually. By 2040 this number is expected to reach almost 30 millions. As the number of patients grows, more effective treatments have to be developed to be able to reduce the impact of cancer on society; this means that people diagnosed with cancer should be given a longer life expectancy but also that proton therapy should become more widespread and more affordable.

Following the onset of cancer, many types of treatment can be carried out. Radiation therapy is one of the primary options, alongside surgery and chemotherapy. Hadron therapy (of which proton therapy is a sub-category) presents numerous advantages with respect to other types of radiotherapy and can be used in situations where radiotherapy cannot. Although presenting clinical advantages, hadron therapy is not widespread due to its higher costs.

LIGHT is a novel linear accelerator for proton therapy designed and built by ADAM SA, a CERN spin-off. The design of medical linear accelerators has been made possible only recently thanks to major advancements in the field of accelerating radio-frequency cavities, a central technology in linear accelerators. This breakthrough enables a more effective treatment, by allowing physicians to actively and quickly tweak more aspects of the therapy (the particles energy in particular), and it also makes proton therapy cheaper. Furthermore, current hadron therapy accelerators are often very bulky and need considerable radiation shielding, so they need to be placed in ad-hoc buildings; LIGHT instead, thanks to its compactness and its low level of beam losses, can be installed in pre-existing hospitals so that proton therapy can be progressively adopted by medical centres which do not currently provide it.

This thesis is composed of five chapters. Chapter 1 gives an overview about cancer, current treatments and technologies enabling hadron therapy, with a focus on LIGHT and the role of this thesis in its development. Chapter 2 describes the design of the two diagnostic systems developed during this thesis, the first enabling online energy measurements and the second being a tool for accelerator operators to recover operability more quickly. Chapter 3 reports the signal processing exploited by those developed systems, which performs phase and energy detection for a bunched proton beam. Chapter 4 describes how such signal processing has been implemented to achieve low-latency results and, by future enhancements, how it can enable a fully real-time implementation which will allow to instantaneously suspend the treatment when an erroneous situation is detected. Finally, chapter 5 reports the achieved results based on on-beam measurements carried out during the accelerator commissioning which is currently still ongoing.

CHAPTER 1

Cancer, Proton Therapy & The LIGHT System

1.1 Cancer

What Is Cancer? Cancer is a generic term for a large group of diseases that can affect any part of the body. It is the second leading cause of death globally and it is responsible for about 1 in 6 deaths [1].

One defining feature of cancer is the rapid creation of abnormal cells that grow beyond their usual boundaries, and which can then invade adjoining parts of the body and spread to other organs. As cancerous cells grow and multiply, they form a mass of cancerous tissue, called a tumour, that invades and destroys normal adjacent tissues. The term tumour refers to an abnormal growth or mass. Tumours can be cancerous or noncancerous. Cancerous cells from the primary (initial) site can spread throughout the body; the latter process is referred to as metastasizing. Metastases are a major cause of death from cancer. [2]

What Causes Cancer? The majority of cancers (90 to 95 %) are due to genetic mutations from environmental and lifestyle factors. The remaining are due to inherited genetics [2]. In this context environmental means any cause that is not inherited genetically. Common environmental factors that contribute to cancer death include tobacco, alcohol, diet and obesity, infections, radiation, stress, immunodepression, lack of physical activity and pollution.

How Is It Diagnosed? Most cancers are initially recognized either because of the appearance of signs or symptoms or through screening. Neither of these leads to a definitive diagnosis, which requires the examination of a tissue sample by a pathologist. People with suspected cancer are investigated with medical tests, which commonly include blood tests, X-rays, (contrast) CT scans and endoscopy. [3]

The tissue diagnosis from the biopsy indicates the type of cell that is proliferating, its genetic abnormalities and other features. Together, this information is useful to evaluate the prognosis and to choose the best treatment. Cytogenetics and immunohistochemistry are other types of tissue tests. These tests provide information about molecular changes (such as mutations) and may thus also indicate the prognosis and best treatment.

How Can It Be Treated? More than 30 % of cancers could be prevented by avoiding risk factors including: tobacco, obesity, poor diet, physical inactivity, alcohol, sexually transmitted infections and air pollution. Many of these factors are controllable lifestyle choices, thus cancer is generally preventable. But not all environmental causes are controllable, such as naturally occurring background radiation and hereditary genetic disorders, which are not preventable via personal behaviour. [1]

Following the onset of cancer many types of treatment can be carried out. The types of treatment will depend on the type of cancer and on its advancement status. Some people with cancer will have only one treatment but most people have a combination of treatments, such as surgery with chemotherapy and/or radiation therapy. Other types of treatment include immunotherapy, targeted therapy, or hormone therapy.

The primary treatment options include surgery, chemotherapy and radiation therapy.

Surgery for Cancer Many people with cancer are treated with surgery [4]. Surgery works best for solid tumours that are contained in one area. It is a local treatment, meaning that it treats only the part of the body with the cancer, but, depending on how the surgery is performed, it can be more or less invasive. Surgery usually removes some, but not all, of a cancer tumour, because removing an entire tumour might damage an organ or the body. Removing part of a tumour can help other treatments work better.

The main drawbacks of surgery are pain, the possibility of infections and bleeding. This effects can be minimised by a proper hospitalisation.

Chemotherapy Chemotherapy works by stopping or slowing the growth of cancer cells, which grow and divide quickly. It does so by using drugs that target fast-growing cells. Chemotherapy can be used to shrink a tumour before surgery or to destroy cancer cells that may remain after the surgery; moreover, it can be used to kill cancer cells that have returned after treatment or spread to other parts of the body. [5]

As chemotherapy has the effect of also killing healthy cells (especially the fast-growing cells that share this trait with cancerous cells) it may cause mouth sores, nausea, hair loss and, most commonly, fatigue.

Radiation Therapy Radiation therapy (or radiotherapy) uses high doses of radiation to kill cancer cells and shrink tumours. At high doses, radiation therapy kills cancer cells or slows their growth by damaging their DNA. Cancer cells whose DNA is damaged beyond repair (both strands in the DNA's double helix are severed) stop dividing or die and are then removed by the body.

Radiation therapy does not kill cancer cells right away. It takes days or weeks of treatment before DNA is damaged enough for cancer cells to die. Then, cancer cells keep dying for weeks or months after radiation therapy ends. Being a localised treatment, its actual side effects depend on the part of the body that is treated. Side effects which do not depend on the treated body part are the same as for chemotherapy.

There are two main types of radiation therapy, external beam and internal [6, 7]. The type of radiation therapy used to treat a tumour depends on many factors, related not only to the tumour itself but also to other factors concerning the patient, such as age and their general medical condition.

As this thesis is concerned with a system for external radiotherapy, this topic will be expanded upon in the next section.

1.2 Radiotherapy

In external beam radiation therapy, beams of particles are directed toward the patient's cancerous tissue to kill its cells and thus it is most useful when the cancer is localized to one area of the body. The employed particles need to carry enough energy as to be ionising: this allows them to damage the cells they pass through so to cause their death and in this way also prevent them from further reproducing. This is necessary to eradicate the tumour, because whereas normal cells divide about 50 times and then die spontaneously, cancer cells can go on dividing indefinitely.

The damage to the cells is either direct or indirect ionisation of the atoms which make up the DNA chain. Indirect ionisation happens as a result of the ionisation of water, forming free radicals, notably hydroxyl radicals, which then damage the DNA.

There are mainly two factors to consider about treating a tumour with radiotherapy: the type of the tumour and the type of particles to use. This is because different cancers respond to radiation therapy in different ways, each particle has a different way of damaging the cells and not all types of particles are suited for treating every type of cancer.

The following paragraphs will illustrate benefits and downsides of the different types of particles that are nowadays used for radiotherapy.

Photons In photon therapy, most of the radiation effect is through free radicals. Cells have mechanisms for repairing single-strand DNA damage and double-stranded DNA damage. However, double-stranded DNA breaks are much more difficult to repair, so it's much more likely that cells will die. Cancer cells have a diminished ability to repair sub-lethal damage. Single-strand DNA damage is then passed on through cell division; damage to the cancer cells' DNA accumulates, eventually causing them to die or reproduce more slowly.

One of the major limitations of photon radiation therapy is that the cells of solid tumours become deficient in oxygen. Solid tumours can outgrow their blood supply, causing a low-oxygen state known as hypoxia. Oxygen is a potent radiosensitiser, increasing the effectiveness of a given dose of radiation by forming DNA-damaging free radicals. Tumour cells in a hypoxic environment may be as much as 2 to 3 times more resistant to radiation damage than those in a normal oxygen environment. [8]

Another downside is that the energy deposition of a photon beam cannot be controlled. In the human body it decreases more or less linearly, as depicted in fig. 1.1, with small variations which depends on the exact tissue it passes through. This implies that much of the radiation is released on healthy tissue, which is correlated with the onset of secondary tumour, especially in paediatric patients, and thus it's highly undesirable. To mitigate this a combination of beams coming from different directions is often used; however, this technique can be exploited also with other types of particles. Nevertheless, photon beams cannot be used to cure tumours that are close to vital organs due to the way they penetrate the human body.

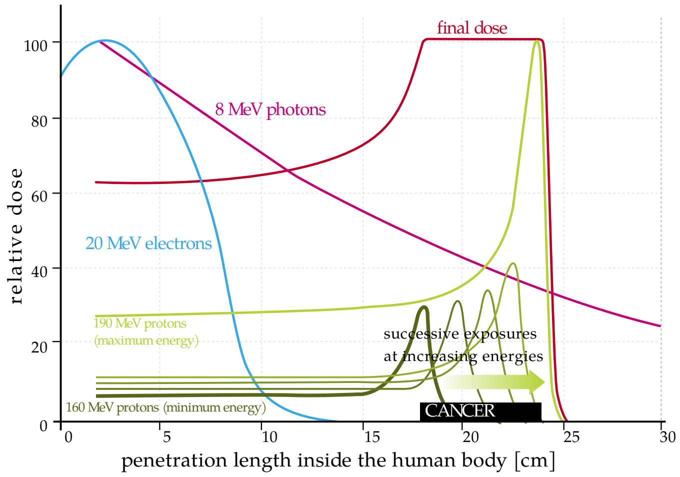


Figure 1.1 – In depth distribution of the absorbed dose, in the case of 8 MeV photons, 20 MeV electrons, 190 MeV protons [9]. Electrons and photons mainly affect the first layers of tissue, while protons release most of the energy to a precise depth, variable with the beam energy. The green line shows the distribution of the dose received by the patient in the case of treatment of a tumour with a thickness of 6 cm and located between 18 and 24 cm depth, irradiated with proton beams with controlled different energies (this curve is called Spread Out Bragg Peak). The dose is concentrated along the region, with limited damage to the surrounding tissues.

Electrons Electron beams are useful for treating superficial tumours; the reason why can be clearly seen in fig. 1.1. In fact, the maximum deposition of energy by electrons occurs near the surface of the irradiated volume. The dose then decreases rapidly with depth, sparing underlying tissue. Electron beams usually have nominal energies in the range from 4 to 20 MeV; depending on the energy this translates to a treatment range of approximately 1 to 5 cm in the body. For this reason they are used mostly for treating skin cancer. [10]

Neutrons The dose deposition of high energy neutrons is much like that of photons. The therapeutic difference is the stronger biological effect (cell killing per amount of deposited dose) of neutrons. For some kind of cancers and at certain locations in the body, this can be an advantage. [11]

Heavy Ions Charged particles such as protons and carbon ions can cause direct damage to cancer cell DNA through high LET (linear energy transfer, which is the amount of energy transferred to the cells per unit of travelled distance) and their effect is independent of tumour oxygen supply because these particles act mostly via direct energy transfer (unlike photons) usually causing double-stranded DNA breaks. Due to their relatively large mass, protons and other charged particles have little lateral side scatter in the tissue, which means that the beam stays focused on the tumour shape, and delivers small dose side-effects to surrounding tissue. [12, 13]

Another big advantage of heavy ions is that they can precisely target the tumour depth thanks to the Bragg peak effect (see fig. 1.1). The energy they transfer to the irradiated tissue is low when they enter the body, it increases slowly in the first part of the travelled path, it has a sharp peak (the Bragg peak) at a precise depth, and then it immediately resets to zero. Furthermore, the depth of the Bragg peak can be controlled by tuning the entrance energy of the particles in the body, so in a typical treatment with heavy ions the energy and direction of the beam are modulated as to deliver the prescribed dose of radiation to the tumour while minimising the dose delivered to the healthy tissue. This makes it possible to reduce damage to the tissue between the radiation source and the tumour and sets a finite range for tissue damage after the tumour has been reached. This is very important in cases where the close proximity of other organs makes any stray ionisation very damaging (e.g. for head and neck cancers).

1.3 Hadron Therapy

Nowadays hadron therapy (radiotherapy using heavy ions) can be carried out with two different particles: hydrogen ions (or protons) and carbon ions. Their behaviour is the same with respect to their main advantages over other types of particles, but they are not quite identical. Their main differences regard penumbras, fragmentation tails and LET.

Penumbra Given a target region which should be treated with a certain dose of radiation, the penumbra region consist of the area surrounding the target for which the dose is not zero. There are various reasons why this phenomenon exists, both physical and geometrical [14].

Penumbra can be narrower or wider depending on various aspects of the beam delivery system, but it's always narrower for carbon ions and wider for protons, with a dependency on the energy of the particles (slight for carbon ions, significant for protons).

Fragmentation Tail Carbon ions interact on a nuclear level with the tissue they pass through, fragmenting themselves into smaller particles, most of which are protons. Some of these fragments travel non-negligible distances beyond the range of the carbon ion beam and deposit their energy in the "tail" of the target, partly nullifying the main advantages of heavy ions, which is the Bragg peak. Fragmentation tails need to be taken into account when designing the treatment plan, making this matter more difficult for carbon ions than for protons.

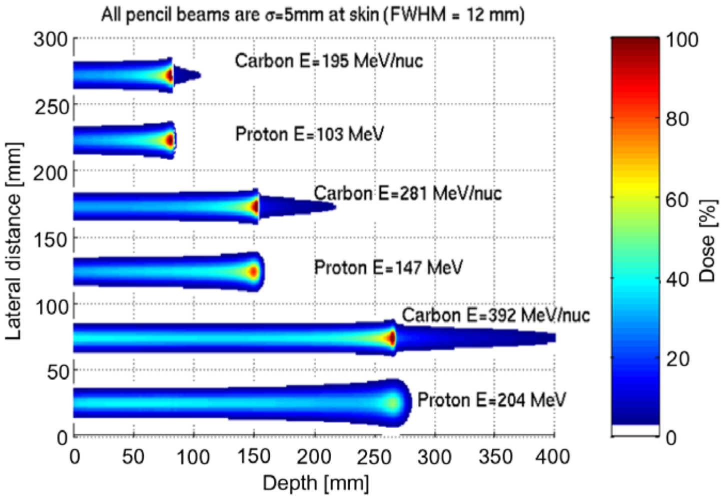


Figure 1.2 – Examples of fragmentation tails for proton and carbon ion beams [15]. The effect of penumbra and its dependency on energy is also perceivable. Compared to fragmentation tails for carbon ions, protons penumbras have much less influence on the dose distribution and they are also more easily described in physical terms.

LET Linear energy transfer (LET) is the amount of energy that an ionizing particle transfers to the material it traverses per unit distance. Thus it is a measure of how much radiation is delivered to a medium by a particle beam. In biological targets, this relates to the relative biological effectiveness (RBE), which gives an indication about how much the same dose delivered with different types of particles is effective in causing a certain biological effect (cells death for the case of cancer treatment).

LET is dependent on both the particle type and the traversed material. Clinical proton beams are low LET radiations; their biological effectiveness per unit of energy absorbed is very close to that of high energy photons. Variations of proton beams RBE are considered too small for accurate determination and application in the clinical environment, so the standard de facto is to use a fixed value for the RBE of protons.

Clinical carbon ion beams are high LET radiations; this means they are more effective in treating tumours, especially radioresistant ones. Unfortunately, carbon ions RBE is not as uniform as for protons, varying greatly because of tissue and depth.

Proton-based and carbon ion-based treatments differ also in another fundamental aspect, which is their dose fractionation. This specifies how the prescribed dose is delivered to the tissue in time. Clinical advantages are the basis for the adoption of this technique, which is applied differently for protons and for carbon ions.

All these differences result in protons and carbon ions being not freely interchangeable as cancer treatments. Equipment for both the types of particles has been developed to allow doctors the possibility of accomplishing a more suitable treatment for each patient.

1.4 Accelerator Technologies for Hadron Therapy

A hadron therapy treatment facility consists of a particle accelerator complex and at least one treatment room, although most nowadays facilities host several treatment rooms fed from a single accelerator. This is because of the bulkiness and cost of typical accelerators used for hadron therapy, which accounts for most of the space taken up by the whole facility.

Cyclotrons A cyclotron is a type of particle accelerator in which charged particles accelerate outwards from the centre along a spiral path. The particles are held to a spiral trajectory by a static magnetic field and accelerated by a rapidly varying electric field.

Cyclotrons were the most powerful particle accelerator technology until the 1950s when they were superseded by the synchrotrons. Nevertheless cyclotrons are still used to produce particle beams in physics, nuclear medicine and also hadron therapy, although carbon ions acceleration by cyclotrons is not feasible due to the high magnetic rigidity of carbon ion beams. Being a well established technology, cyclotrons allow the easiest setup for protons acceleration and there are several commercial solutions available [16]. They are a single, relatively compact piece of equipment which needs little setup and provide for a fixed energy beam.

This simplicity has various downsides. To achieve energy modulation the proton beam has to be degraded before being delivered to the patient. This is performed usually with a pair of absorbing wedges that reduce the energy without introducing substantial scattering and straggling (the former expresses a lack of uniformity in the particles direction and the latter expresses a lack of uniformity in the particles timing); a portion of this degraded beam is then selected with collimators and can be as little as 0.1 % of the initial beam intensity in some configurations. To make up for this, the extracted current from the cyclotron must be correspondingly higher (up to 1 μA, which is about $6 \cdot 10^{12}$ protons per second). This has the consequence that some of the compactness has to be given up to provide the accelerator room with proper radiation shielding, while the way energy modulation is achieved does not allow fast modulation, which would enable some treatment features that will be described later on.

Synchrotrons A synchrotron is a particular type of circular particle accelerator, descended from the cyclotron, in which the guiding magnetic field is time-dependent, being synchronized to a particle beam of increasing kinetic energy such that the bend radius (and hence overall path) of the particles remains constant. Due to its working principles, only one bunch of particles circulates in each acceleration cycle, so this single pulse is usually more intense than for cyclotrons. After acceleration, the energy is held constant for a certain period of time, the so-called "flat top", during which particles can be slowly extracted and directed to the treatment room. They may be progressively extracted over several seconds such that they may also be scanned at the patient.

As the energy may be varied with each acceleration cycle, there is no need for an energy degrader, which greatly reduces the average current requirement. To modulate the dose depth, different pulses are accelerated to different flat tops so that the overall dose conforms to the treatment planning. Unfortunately this does not allow to relax the requirements about the radiation shielding as, due to their more complicated geometrical arrangements, synchrotrons introduce considerable beam losses at the extraction point.

The biggest advantage of synchrotrons is that they can accelerate particles with larger magnetic rigidities than is achievable with a cyclotron, so at present they are the only accelerators used for carbon ion therapy. A single synchrotron can be designed to accelerate both protons and carbon ions, so the majority of treatment facilities based on synchrotrons offer both particles.

Linacs A linear accelerator accelerates particles by presenting them with a series of oscillating electric fields along a linear beamline. The accelerating gradient of an accelerator represents the increase of energy it can produce in a particle over a certain distance. In the last years, linacs were disadvantageous because the accelerating gradient possible was limited to a few MeV/m, and thus reaching high energies required a long accelerator. In recent years, significant developments allowed to achieve the high gradients needed to make linac-based therapy systems size- and cost-competitive with other technologies.

The current approach adopted for linac development achieves proton acceleration at treatment energies employing a fully linac structure while being cost-competitive with current cyclotron and synchrotron based solutions [17, 18]. The advantage of a linac lies in the pulsed nature of its beam, which has a high repetition rate (in the order of thousands of Hertz) allowing pulse-by-pulse energy variation, which removes the need for energy selection and which allows for a finer control of dose delivery. This active and fast control of the energy has the main advantage of enabling the treatment of tumours in moving organs, as the energy modulation rate allows to track the organ movements.

Currently there are no operating treatment facilities based on linacs. LIGHT represents the first attempt at employing a linear accelerator for a medical purpose.

1.5 LIGHT

LIGHT is a novel linear accelerator for proton therapy designed and built by ADAM SA. ADAM is a CERN spin-off company, founded on December 2007 to promote scientific know-how and innovations in medical technology. LIGHT can accelerate protons up to the maximum energy required for medical treatments, allowing to reach the deepest seated tumours at about 32 cm.

The LIGHT accelerator has a number of features, important for a proton therapy machine, that are specific to linacs. In particular:

- A pulsed beam structure with a repetition rate of up to 200 Hz and with electronic pulse-by-pulse tunable intensity and energy without the need of degraders and absorbers.
- An extremely small beam emittance (of the order of 0.25 mm mrad normalized), which makes it possible to use small aperture magnets allowing an overall longitudinal compactness.
- Almost no beam losses expected from simulations [19], implying a reduced size and cost for shielding and easier maintenance.

As previously stated, the above mentioned features make the LIGHT accelerator very well suited for active dose delivery with spot scanning technique, and adapted to the treatment of moving organs by volumetric rescanning with tumour tracking [20].

1.5.1 Structure & Layout

LIGHT is mainly composed of two parts: the linac, whose task is indeed the acceleration of the beam particles, and the high energy transfer line(s), which transports the accelerated beam to the treatment rooms. The acceleration line, in turn, consists of four sections: a proton injector (or source), a Radio Frequency Quadrupole (RFQ), a Side Coupled Drift Tube Linac (SCDTL) section and a Cell Coupled Linac (CCL) section. Each section is adapted to the increasing speed of the protons which are accelerated in about 25 m as indicated in table 1.1. In between those sections there are low and medium beam transport lines, the so called LEBT and MEBT.

Table 1.1 – Overall layout of LIGHT.

Section	Number of Modules	Length [m]	Energy Range [MeV]
RFQ	4	2.0	0.04–5.0
SCDTL	4	6.2	5.0–37.5
CCL	15	15.5	37.5–230

The proton source is able to accelerate the beam to 40 keV and to chop it in pulses of few microseconds at a frequency up to 200 Hz. After the source, the high frequency RFQ is used as a first accelerator, making the beam reach an energy of 5 MeV. The beam is then injected into 4 modules of SCDTL with an exit energy of 37.5 MeV. CCL structures are then able to accelerate the beam up to the final energy of 230 MeV. The accelerator is represented in fig. 1.3.

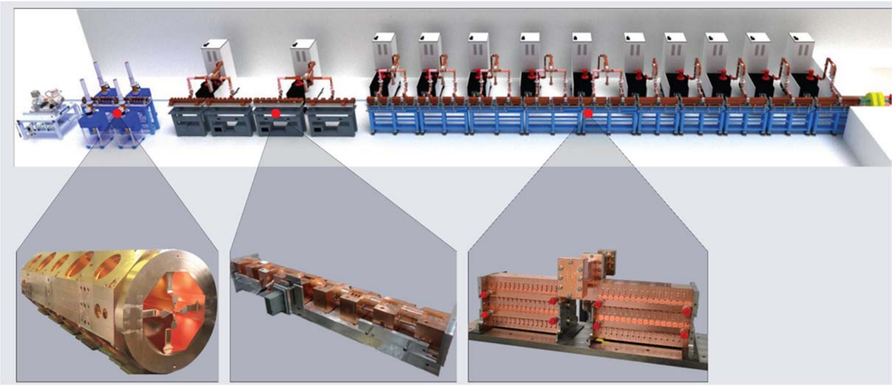


Figure 1.3 – LIGHT's acceleration line with a zoom on (from left to right): the RFQ; one of the SCDTL modules; one of the CCL modules. Reproduced from [21].

Having minimal beam losses, LIGHT could potentially be installed in a pre-existing medical facility (this is not the case with cyclotrons and synchrotrons because of their bulkiness and the radiation shielding they need). Thus a great effort has been put in making the acceleration line very compact in both length and width; in fact, this part of the machine has a fixed layout, so its compactness eases the placement of the accelerator. The transfer line(s) layout instead is very flexible, and indeed will be different for each installation. Furthermore, because of the modularity of the acceleration line it is possible to use a setup with lesss cavities to further improve the overall compactness of the machine. Of course this will mean that the deepest seated tumours cannot be treated, but there may be specialised medical centres interested in doing that (e.g. for treating tumours of the eye).

1.5.2 Beam Characteristics

As beam diagnostic systems are the subject of this thesis, it is essential to know the characteristics of LIGHT's beam. In particular, diagnostic systems will be needed from the output of the RFQ and potentially up to the transfer lines.

The beam is composed of pulses, which are generated at a maximum repetition rate of 200 Hz and have a duration of up to few microseconds. Each pulse is essentially a train of proton bunches. Bunches inside a pulse are evenly spaced; they repeat at the RFQ resonating frequency, which for LIGHT's RFQ is approximately 750 MHz.

By simulating the beam dynamics one can understand what shape bunches can assume. The code used for these simulations is called TRAVEL [22]; it is a multi-particle beam tracking code developed at CERN, which was used as a reference during the Linac4 commissioning and it was benchmarked there with the measurements. The first remark is that the most influent factors the bunch shape depends on are the observation point along the accelerator and the energy at which they are accelerated (that is equivalent to say which accelerating modules are switched on). Figure 1.4 shows some examples of three consecutive bunch shapes for different energies immediately downstream from the end of the acceleration line: it can be seen that also for a single position the shape of bunches can be very different depending on the energy. This will have to be taken into account when designing the diagnostic systems. Fortunately, within a single pulse all bunches have very similar characteristics, such as energy and intensity, and thus also the same shape.

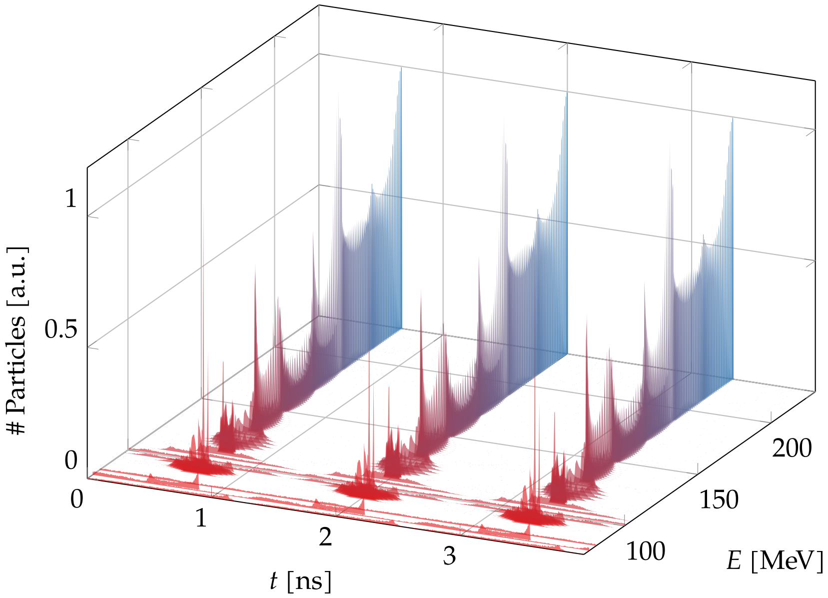


Figure 1.4 – Examples of bunch shapes for different energies at few tens of centimetres after the end of the acceleration line.

From a pulse to the next one, the treatment plan can request that both intensity and energy of the beam vary within their whole range. This is relevant as the diagnostic systems need to have a bandwidth wide enough to allow the measurement of two consecutive pulses with energies at the range boundaries. Table 1.2 summarises some of the key parameters of LIGHT's beam. The smallest energy useful for treatment is 60 MeV, but values as low as 5 MeV need to be measured during the accelerator commissioning. The maximum value for the bunch length is actually the inverse of the bunch repetition period, but bunches could be even longer and overlap one another.

1.5.3 Diagnostic Systems

Beam diagnostics is the ensemble of all the sub-systems in an accelerator which have the purpose of measuring some property of the beam. Diagnostic systems are essential constituents of any accelerator, as they allow to perceive what properties a beam has and how it behaves in the accelerator.

A great variety of physical phenomena are exploited for diagnostics, de-

Table 1.2 – Main parameters for LIGHT's beam.

Parameter	Min	Typical	Max	Unit
Beam energy	5		230	MeV
Pulse intensity	10^7		10^{10}	protons/pulse
RF frequency		750		MHz
Pulse width	0.5	1	4	μs
Pulse repetition rate			200	Hz
Bunch length	5	50	1333	ps

pending on the parameter of interest, the particles considered and, usually, some requirements about the conditions under which the measurement has to be carried out. A rough distinction between two different modes of operation may help to summarise the impact on the beam instrumentation:

- Diagnostic systems for accelerator commissioning:
  - Used to enable adjustments of the beam transport through different accelerator sections.
  - Required for the characterization of the beam behind each accelerator section.
  - Simple and robust devices with high sensitivity, allowing to operate with single or few bunches of low intensity.
  - Low or modest demands on accuracy.
  - Beam-destructive methods are possible.
- Diagnostic systems for standard operation:
  - Used for precise beam characterization in order to control and improve the accelerator operation.
  - Required for the diagnosis of unwanted errors and to trigger interlocks.
  - Devices are typically based on more or less sophisticated schemes.
  - High demands on accuracy.
  - Non-destructiveness is mandatory.

Therefore, there exists today a vast choice of different types of diagnostic devices, each in different variants.

The full LIGHT accelerator will include beam diagnostics between some modules and in the beam transfer line between the accelerator and the treatment room. However, these will not allow the beam to be fully characterised at all locations. During the commissioning stages, though, a more in-depth measurement is required.

The initial aim of this thesis was the development of a beam energy measurement system. The research and work that led to its realisation entailed a growth of expertise in the domain of signal processing, in particular in the field of spectral analysis, such that it was then possible to take advantage of this knowledge for the development of another diagnostic system which monitors the loss of synchronisation between the accelerating modules.

Particle accelerator development processes are very different when the accelerator is used to treat patients and not for scientific research. In the former case, the particle accelerator (or at least a part of it) must become part of a certified medical device, with all that entails. Regarding diagnostic systems, some of them may be employed also with medical purposes, while others may be confined to be technical tools used as feedback to the control system regarding the operational status of the accelerator.

Medical diagnostic systems need to provide the following guarantees:

Real-timeness The result of the measurement has to be provided so that the accelerator control system will be able to act on the machine as to prevent the beam from further reaching the patient. Obviously this guarantee needs to be respected deterministically, but its implication on the diagnostic system can be different based on the nature of the particles beam. This requirement is ment to ensure that the dose delivered to the patient which is not conforming to the treatment plan does not exceed a certain threshold. If the threshold is exceeded, the beam anomaly has to be detected and handled.

Non-destructiveness Non-destructive diagnostic systems exploit phenomena which do not alter the properties of the particles that are being measured. Medical systems are required to operate while the treatment is ongoing, so they need to be non-destructive, otherwise the measurements would not represent the properties of the beam when it reaches the patient.

Accuracy The measurement result has to be provided with a certain accuracy, which is usually dictated by medical regulations.

1.5.4 Beam Energy Measurement Requirements

The ideal energy measurement system should be "medical-ready" and able to work in every condition that LIGHT might put it in. Giving the high variability of the beam parameters and its single pass nature, this is quite a demanding request. However, non-destructive energy measurement systems are already being employed in various experimental facilities around the world, so the selected approach was to analyse those solutions and use them as a starting point, then trying to adapt that strategy to LIGHT features.

The following are they key requirements for the energy measurement system:

- Non-destructiveness.

- Real-timeness. The energy has to be measured pulse-by-pulse providing the result within 1–2 ms, such that the next pulse can be stopped before reaching the patient.
- The accuracy of the measurement should be better than 0.17 % and its resolution about 0.03 %. This is deduced from medical regulations about the precision the accelerator must have in the depth selection: 1 mm depth variation is equivalent to, in the worst case, an energy variation of 0.17 %, so a resolution 5 times better has been chosen to have a reasonable margin.

Given the non-destructiveness requirement, the only possibility is to measure the energy of the particles by exploiting the interaction of the EM field carried by the beam with a detector placed around it. The Time-of-Flight is one of the most used techniques. This actually measures the velocity of the particles, which can be directly translated into energy:

$$
\begin{aligned}
E &= A \cdot E_0 \cdot \left(\frac{1}{\sqrt{1-\beta^2}} - 1 \right) && (1.1) \\
&= A \cdot E_0 \cdot (\gamma - 1), && (1.2)
\end{aligned}
$$

where E is the energy to be computed, A is the atomic mass number of the particles, E_0 is the particle rest mass, β is their relativistic velocity and γ is their Lorentz factor. Protons in the 5–230 MeV energy range have a β from 0.1 to 0.6: this velocity can be measured by fixing a distance over which the particle travel time is measured. As times on such a small scale are best measured indirectly through the phase of electronic signals, what this thesis actually deals with is beam phase measurements, the characteristics of which will be discussed in detail in the next chapter.

CHAPTER 2

Phase-Based Measurement Systems

This chapter discusses the design process for diagnostic systems based on phase measurements. The first section outlines some design principles derived from the measurement requirements. The second section describes the embodiment of the detectors employed, their transfer function and the factors that have an influence on it. The third section discusses the motivations behind a general phase measurement system and its possible integration in the control system workflow. The last section lists the additional requirements for an energy measurement system based on phase measurements.

2.1 Design Principles

In accelerators, the longitudinal axis is the one the particles travel along. As such, kinetic energy is a longitudinal property of the beam, because it is proportional to the particles velocity whose longitudinal component is very predominant in relation to the transversal ones.

As sketched in the previous chapter, the aim is to measure the phase of the beam in two points along the beam pipe: the phase shift between the two signals will give a clue about the travel time of the beam between the two positions, thus giving an information regarding its energy. The aim is to have a single energy measurement per pulse. Given the bunched structure of the beam, the actual aim is to measure the phase of the centre of mass of the bunches. The only way to achieve this in a non destructive way is by means of electromagnetic pickups.

Electromagnetic pickups can be used to detect the timing of bunches, whether it is in relation to the phase of the electromagnetic field in a certain accelerating cavity or in relation to the phase of the same bunches at a different point along the longitudinal axis. To enable the former case, not only beam pickups have to be used, but also RF probes which are inserted directly in the accelerating cavities. These probes will give a reference phase against which the phase shift of signals from different pickups along the accelerator can be computed. These phase shifts will be measured when an optimal setup is achieved for a certain section of the accelerator so that, when a failure occurs, the operator can quickly recover the machine by tuning its parameter so to find back those phase shifts.

2.2 Physical Embodiment

Pickups can be capacitive or inductive. The former are generally preferred as inductive pickups are very sensitive to rapidly changing magnetic fields, which are always present in the environment of RF accelerators.

Phase probes are capacitive pickups most commonly used for, as suggested by their name, phase measurements. A phase probe consists of a metallic ring inserted in the beam pipe, connected to an electric feed-through connector accessible on the air-side. Its output signal represents the charge induced by the electric field of the beam particles on the metallic electrode, which is insulated from the external vacuum chamber. The output signal of a phase probe is usually very small, so an amplifier is always put near its output.

Figure 2.1 shows the scheme of a phase probe and its equivalent circuit, while fig. 2.2 shows its embodiment and table 2.1 reports their dimensions. In LIGHT, two different mechanical designs have been used: the phase measurement requires some pickups to be put in between accelerating modules, thus the design should aim for minimal obstruction; the energy measurement requires the largest resolution possible, so the design should aim for maximal beam sensitivity. As these two requirements are in trade-off, the decision was to pursue two different designs to avoid jeopardising the development of both systems.

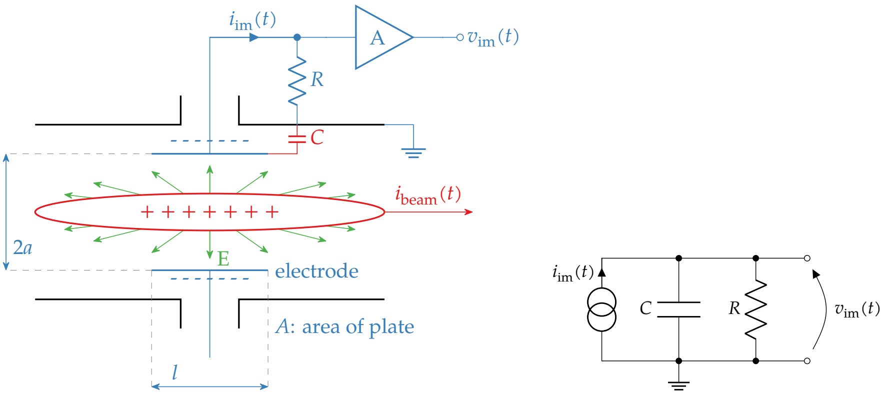


Figure 2.1 – Scheme of a phase probe electrode in the beam pipe (left) and its equivalent circuit (right). Reproduced from [1, p. 190, fig. 3].

It has to be noted that also a third type of capacitive pickups is used in LIGHT: these are the BPMs (Beam Position Monitors), which are very similar to the phase probes used for the energy measurement, the only difference being that the electrode of a BPM is divided in four sections so as to be sensitive to beam displacements in the transversal plane. The recombination of the four output signals of a BPM can thus be used in the same manner as a single phase probe output. This fact can be exploited to avoid putting too many

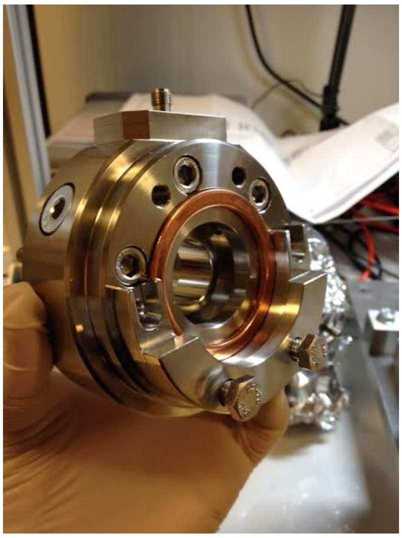


Figure 2.2 – Physical embodiment of LIGHT's phase probe. Here the model that will be used in the transfer lines is shown (PPR); the one that will be used between the accelerator modules (mPPR) is identical, except it is smaller.

Table 2.1 – Dimensions of the electrode for each type of pickup used.

Pickup	Elec. length [mm]	Elec. aperture [mm]	Flange-to-flange [mm]
PPR	15	20	50
BPM	15	20	55
mPPR	10	7	30

additional probes in between the accelerating modules, which would have a negative impact on beam features such as the emittance.

The transfer function of a phase probe can be found in [1, p. 190]:

$$\frac{V_{\mathrm{im}}(s)}{I_{\mathrm{beam}}(s)} = \frac{A}{2\pi a} \cdot \frac{1}{\beta c} \cdot \frac{1}{C} \cdot \frac{sRC}{1+sRC} \tag{2.1}$$

where

A is the area of the electrode,

a is the electrode aperture (radius),

R is the input resistance of the amplifier, and

C is the sum of: the capacitance between the electrode and the beam pipe; the capacitance of cables used to connect the probe to the amplifier; the input capacitance of the amplifier.

Three terms can be clearly distinguished:

- A term dependent on the probe geometrical features; as LIGHT's probes are cylindrical, the expression can be further simplified:

  $$Z_G = \frac{A}{2\pi a} = l \qquad [\mathrm{m}] \tag{2.2}$$

  where l is the longitudinal extent (length) of the probe electrode. The magnitude of this term is limited by the minimal longitudinal space a phase probe has to occupy; this is not only to limit the overall bulkiness of the machine but also, for the mini phase probes, to enable their use in inter-tank spaces.

- A term dependent on the physical properties of the beam:

  $$Z_P = (\beta c)^{-1}. \qquad [\mathrm{s/m}] \tag{2.3}$$

  This means that, for a fixed beam intensity, the probe will output a weaker signal for a more energetic beam.

- A term dependent on the electronic features of both probe and beam:

  $$Z_E(j\omega) = \frac{1}{C} \cdot \frac{j\omega RC}{1 + j\omega RC}. \qquad [\Omega/\mathrm{s}] \tag{2.4}$$

As the beam pulse is composed of evenly spaced bunches, its spectrum will be comb-like, with a spacing between its peaks given by the bunches repetition rate. This repetition rate will be called f_{RFQ} in this thesis as it corresponds to the working frequency of the RFQ. Thus, the filtering effect of $Z_E(j\omega)$ on the first harmonic of the pulse spectrum can be computed as following:

$$|Z_E(j2\pi f)|_{f=f_{\mathrm{RFQ}}} = \frac{1}{C} \cdot \frac{2\pi f_{\mathrm{RFQ}} RC}{\sqrt{1 + (2\pi f_{\mathrm{RFQ}} RC)^2}}. \tag{2.5}$$

Figure 2.3 depicts $Z_E(j\omega)$ distinguishing two cases for R: low (50 Ω) and high (1 MΩ) impedance.

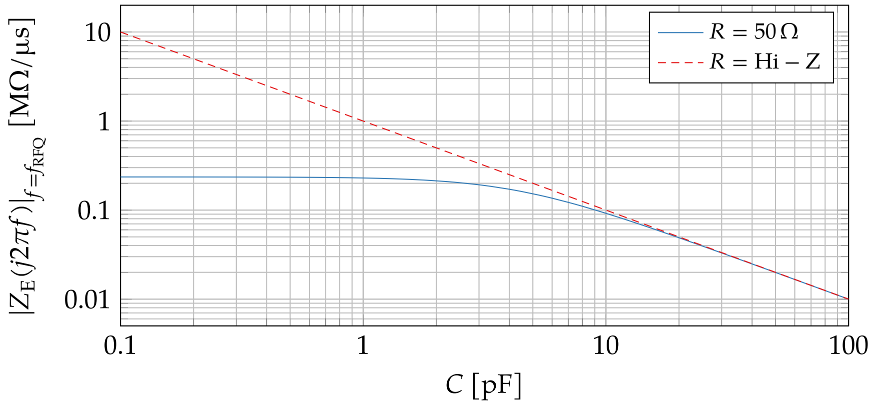


Figure 2.3 – Capacitive-dependent term of a phase probe transfer function.

The phase probes have been built to match 50 Ω so impedance mismatches can be avoided only by using a low input impedance amplifier. The best results are achieved by reducing C but acting on it is difficult. A way to minimise it would be by screwing the amplifier directly on top of the phase probe SMA connector (therefore avoiding to use a cable) and by employing an amplifier with a low input capacitance.

Figure 2.4 represents the phase probe transfer function for the maximum and minimum value of β. It can be seen that Z is a bit higher for the energy measurement phase probes (PPR) than for the phase measurement phase probes (mPPR) due to their slightly longer electrode (15 mm vs. 10 mm) but the particle velocity has a greater impact anyway.

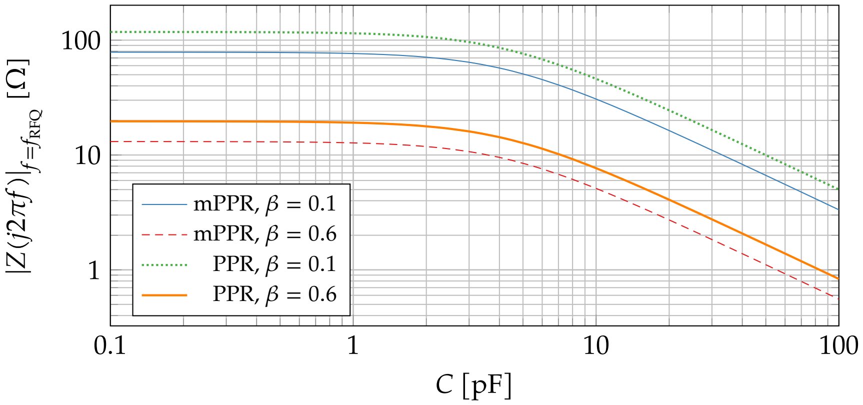


Figure 2.4 – Phase probes transfer impedance at β limits.

Note that, for the given measurement task, only the first harmonic of the pulse signal is relevant, for two reasons:

- The shape of bunches can vary along the beam pipe. So the phase of harmonics beyond the first does not depend only on the bunch centre of mass, but also on other features which have an influence on the bunch shape, like the exact settings of the accelerating cavities before the probe under analysis.
- The second harmonic is at 1.5 GHz: acquiring signals at such high frequencies requires particular precautions which would lead to a more complicated and costly acquisition system, the added cost not only being due to the more sophisticated acquisition chain but also to the extra computational power needed for signal processing.

2.3 Phase Measurement System (PMS)

Particles are accelerated by the sinusoidal electric fields present inside the acceleration cavities. The beam can reach the transfer lines only if each module is correctly synchronised with all the other modules. This condition is guaranteed by the LLRF (Low Level Radio-Frequency) control system which controls the phase of each cavity power supply.

Despite the presence of the LLRF, RF phases might drift; this can be due to different reasons: temperature and humidity variations on cables and radio-frequency networks in specific parts of the LLRF system, which cannot be compensated; unwanted drift in the Modulator Klystron Systems (MKS); second order effects. As an example, a variation in the environmental temperature can lead to an expansion of the concrete under the accelerator, with an expansion factor in the order of $10\,\mu\mathrm{m}/(\mathrm{m\,K})$, leading to a phase drift of $0.1\,^{\circ}/\mathrm{K}$. Additionally, the RF phase might be lost in case of a shutdown or failure of the RF subsystem controlling a particular module. Whenever one of these condition happens, the beam bunch phase changes as well because the energy "kick" given by the dysfunctional module will be incorrect, thus changing the beam timing from that point on. Phase probes can act as feedback mechanism for the LLRF, allowing the automatic compensation of phase drifts which is currently done manually, typically about once per hour. Furthermore, downtimes can be tremendously reduced by making their recovery also automatic.

The phase of the RFQ is used as a reference because it is the first accelerating module and it also has the task of chopping the beam into bunches. An RF probe is used to sense the electric field inside the RFQ. The signal coming from it is much more powerful than the ones coming from beam probes but it is otherwise very similar to them and thus, after an attenuation stage, can be processed the same way as PPR, mPPR and BPM signals.

The requirements for this system are more or less the same as for the energy measurement system, which are listed in the previous chapter (real-timeness, non-destructiveness and high accuracy). The amount and position of probes needed for this system to be useful has been extrapolated from beam dynamics studies. This is because, as previously said, the particles would not reach the end of the linac in case the RF phases are lost, so having a single probe after the last cavity would be of no use. Five mPPR have been foreseen, one of which will be placed in between SCDTL modules 2 and 3 while the other four will be interposed between CCL modules. Moreover, the whole accelerator diagnostic includes three exploitable BPM positioned: directly after the RFQ; in between SCDTLs and CCLs; at the end of the linac.

There is no need for the PMS to perform measurements all the time. PMS measurements can be orchestrated by the accelerator control system so to happen:

- Periodically: to compensate for drifts, the most useful schedule is to perform measurements starting from the most upstream beam probe to the furthest from the RFQ, measuring each time the phase shift with the RFQ probe.

- On request: when a module failure has been detected and manual intervention is required to bring it back to working conditions, the control of the PMS can be given to an operator, which would need to carry out manual checks.

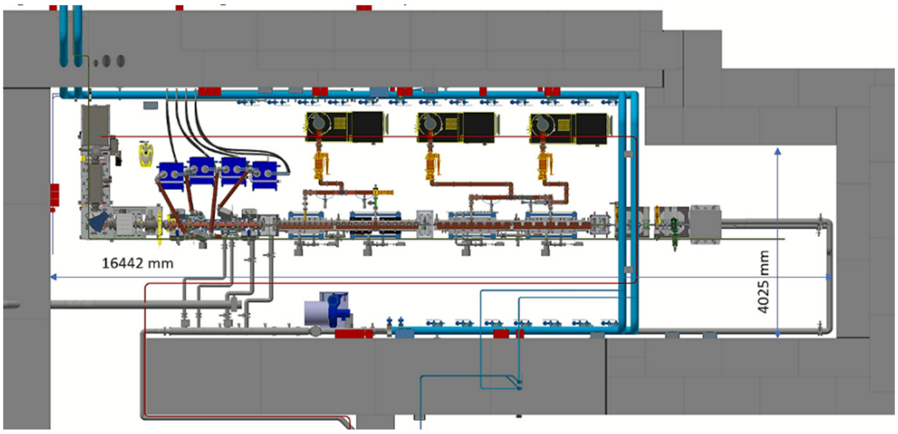


Figure 2.5 – LIGHT's in-development prototype for energies up to 37.5 MeV. Fixed BPMs are installed at the RFQ exit and after the last SCDTL module. More BPMs are present in the diagnostic test bench placed at the end of the acceleration line, along with phase probes and other diagnostic systems. Reproduced from [2].

2.4 Energy Measurement System (EMS)

LIGHT's energy measurement system builds upon the phase measurement system. The main additional requirement that has to be taken into consideration is that it has to be possible to trace back an unambiguous timing information from the acquired signal phases. For this to hold over the whole energy range that LIGHT can produce, the most common practice is to implement a Time-of-Flight system made of three probes, as depicted in fig. 2.6: the phase shift between the two closest probes is used to get a rough estimate of the particle β; this information allows to compute the number of bunches (N) between the two probes which are further apart (1 and 3); the phase shift between probes 1 and 3 can then be used to have a finer evaluation of β.

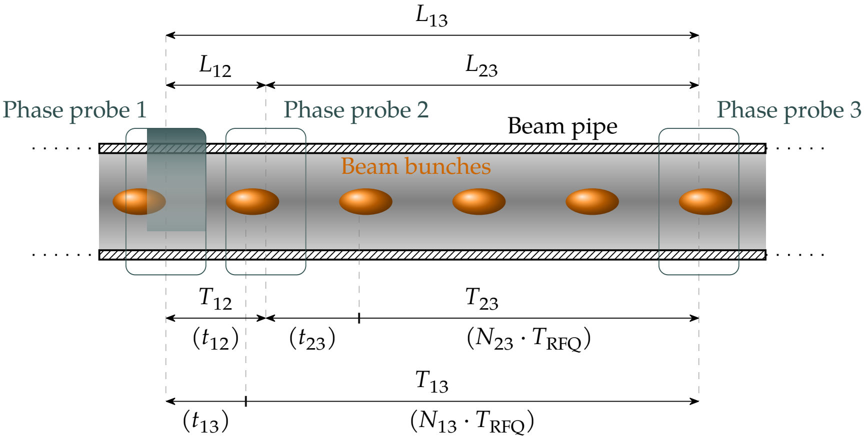


Figure 2.6 – Three probes layout for a Time-of-Flight system. See text and eqs. (2.8) to (2.14) for information about the depicted quantities.

The acquisition of the signals from the three probes obviously has to be performed in synchronisation; this makes it possible to extrapolate their phases φ_1, φ_2 and φ_3, all referred to the same instant. In the following chain of formulas, Q_{xy} refers to the quantity Q as measured from probe x to probe y (e.g. L_{12} refers to the distance between probes 1 and 2, as depicted in fig. 2.6):

$$T_{\mathrm{RFQ}} = f_{\mathrm{RFQ}}^{-1}; \tag{2.6}$$

$$t_{xy} = \frac{\Delta\varphi_{xy}}{2\pi} \cdot T_{\mathrm{RFQ}}; \tag{2.7}$$

$$L_C = \min(L_{12}, L_{23}); \tag{2.8}$$

$$\Delta\varphi_C = \begin{cases} \Delta\varphi_{12} & L_{12} < L_{23} \\ \Delta\varphi_{23} & L_{12} > L_{23} \end{cases}; \tag{2.9}$$

$$N_{xy} = \left\lfloor \frac{L_{xy}}{L_C} \cdot \frac{\Delta\varphi_C}{2\pi} - \frac{\Delta\varphi_{xy}}{2\pi} \right\rceil; \tag{2.10}$$

$$T_{xy} = t_{xy} + N_{xy} \cdot T_{\mathrm{RFQ}}; \tag{2.11}$$

$$\beta = \frac{L_{13}}{T_{13}} \cdot \frac{1}{c}; \tag{2.12}$$

$$\gamma = \frac{1}{\sqrt{1-\beta^2}}; \tag{2.13}$$

$$E = A \cdot E_0 \cdot (\gamma - 1). \tag{2.14}$$

This is all of the computation that has to be performed to get the energy from the signals phase shifts. T_{xy} refers to the time actually taken by the beam to travel from probe x to probe y, while t_{xy} refers only to the part represented by $\Delta\varphi_{xy}$.

2.4.1 Phase Probes Placement

The placement of the phase probe is particularly critical for the effectiveness of the energy measurement.

General Rules As first rule, the probes should not be positioned between the accelerating modules, for the obvious reason that the energy of the particles would vary from a probe to the next. The only place where it makes sense to place the EMS is thus at the end of the linac:

- During commissioning a test bench is always present after the last installed acceleration cavity, and this test bench is the host of the EMS (along with other diagnostic systems).
- After commissioning, the beam exiting the linac enters the transfer lines which bring it either to the patient or to a beam dump; the EMS will be installed on these transfer lines.

No secondary EMS can be put before; this could be useful for low energies as in that situation the beam has to travel through all the turned off modules thus debunching along the way. This effect can be seen in fig. 1.4: roughly above 130 MeV each bunch is a sharp peak but under that value the beam sometimes debunches till the point of showing overlapping between the bunches. The effect of the debunching is hard to analyse for the generic case and thus will be evaluated through bunch simulations.

As second rule, the three phase probes have to be installed on a straight section of the acceleration line. This is because where the beam pipe bends, the beam trajectory is more sensitive to variations of the controlling magnetic field, so the effective distance travelled by the beam varies non-negligibly. It is expected that in straight sections the magnitude of the beam displacement would be small enough as to not have an influence on the energy measurement.

Unambiguous Energy Measurement The energy measurement has to be unambiguous and for that reason the three probes layout has been adopted. Now it has to be guaranteed that the measurement between the two closest probes is unambiguous over the whole energy range. This is true if

$$\Delta\varphi_{C,\max} - \Delta\varphi_{C,\min} \leq 2\pi - M, \tag{2.15}$$

where M is a mark up to be chosen (if the whole 2π range were used then the tiniest error in the phase measurement would result in an erroneous measurement). Note that this doesn't force the number of bunches between the two closest probes to be always smaller than 1. Elaborating on eq. (2.15), it is possible to find out what this means for L_C:

$$\begin{aligned}
\Delta\varphi_{C,\max} - \Delta\varphi_{C,\min} &\leq 2\pi - M; \\
\left(t_{C,\max} - t_{C,\min}\right) \cdot \frac{2\pi}{T_{\mathrm{RFQ}}} &\leq 2\pi - M; \tag{2.16} \\
\left(\frac{L_C}{\beta_{\min} c} - \frac{L_C}{\beta_{\max} c}\right) &\leq \frac{2\pi - M}{2\pi} \cdot T_{\mathrm{RFQ}}; \tag{2.17} \\
L_C &\leq \frac{2\pi - M}{2\pi} \cdot \frac{T_{\mathrm{RFQ}} \cdot c}{\left(\beta_{\min}^{-1} - \beta_{\max}^{-1}\right)}. \tag{2.18}
\end{aligned}$$

This conclusion can be tested for a couple of situations:

- The first commissioning stage of LIGHT produces an energy range from 5 to 37.5 MeV: this means β from 0.1 to 0.27, so L_C can be at most 63 mm.
- The final installation of the EMS will have to measure from 60 to 230 MeV (for lower energies the beam will not be able to reach the transfer lines): in this case L_C has to be smaller than 319 mm.

It has to be kept in mind that also the resulting $\Delta\varphi_C$ range is important, because the measured phase shifts will always be from 0 to 2π, so they have to be wrapped to the right range. This can be done by computing, from the chosen L_C, the midpoint of $\Delta\varphi_C$:

$$\Delta\varphi_{C,\mathrm{mid}} = \frac{2\pi}{T_{\mathrm{RFQ}}} \cdot \frac{L_C}{c} \cdot \left(\frac{\beta_{\max}^{-1} + \beta_{\min}^{-1}}{2}\right), \tag{2.19}$$

and from that the two extremes:

$$\Delta\varphi_{C,\text{min}} = \Delta\varphi_{C,\text{mid}} - \pi; \qquad \Delta\varphi_{C,\text{max}} = \Delta\varphi_{C,\text{mid}} + \pi. \tag{2.20}$$

As per the previous examples, it can be computed that

- for the commissioning $\Delta\varphi_{C,\text{min}} \approx 211°$ and $\Delta\varphi_{C,\text{max}} \approx 571°$, while
- for the final installation $\Delta\varphi_{C,\text{min}} \approx 473°$ and $\Delta\varphi_{C,\text{max}} \approx 833°$.

Resolution Constraints Further constraints about the probes placement, specifically their relative distance, can be derived considering the required resolution. Using eqs. (2.8) to (2.14) the propagation of the uncertainty from the measured quantities $(L, \Delta\varphi)$ to the energy can be computed. With respect to those equations it is useful to proceed backward:

$$\frac{\delta E}{E} = \gamma(\gamma+1)\cdot\frac{\delta\beta}{\beta}; \tag{2.21}$$

$$\frac{\delta\beta}{\beta} = \sqrt{\left(\frac{\delta L}{L_{13}}\right)^2 + \left(\frac{\delta t}{T_{13}}\right)^2} \tag{2.22}$$

$$= \sqrt{\left(\frac{\delta L}{L_{13}}\right)^2 + \left(\frac{1}{N_{13}}\cdot\frac{\delta\Delta\varphi}{2\pi}\right)^2 + \left(\frac{\delta T_{\text{RFQ}}}{T_{\text{RFQ}}}\right)^2}. \tag{2.23}$$

Equation (2.21) is represented in fig. 2.7 and has a quite simple interpretation: going from velocity to energy makes the relative error worse by a factor ranging from 2 to 3.

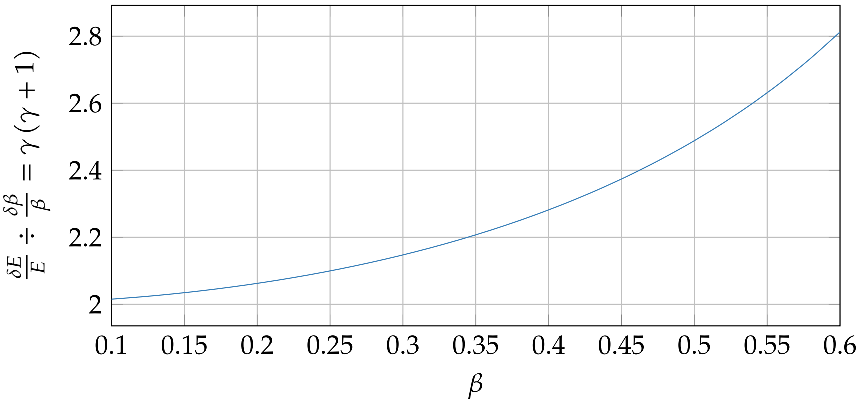


Figure 2.7 – Relative errors ratio from the particle velocity to its energy.

For β the question is more complicated because δL, $\delta\Delta\varphi$ and δT_{RFQ} are all independent. The expected stability for the RFQ frequency is better than 10^{-5} %, so it can be considered negligible with respect to the other two terms. Thus, carrying on from eq. (2.22) and considering $\delta T_{\text{RFQ}} = 0$:

$$\frac{\delta\beta}{\beta} = \sqrt{\left(\frac{\delta L}{L_{13}}\right)^2 + \left(\frac{\delta t}{T_{13}}\right)^2}$$

$$= \sqrt{\left(\frac{\delta L}{L_{13}}\right)^2 + \left(\frac{\delta t \cdot \beta c}{L_{13}}\right)^2} \tag{2.24}$$

$$= \frac{1}{L_{13}} \cdot \sqrt{(\delta L)^2 + (\delta t \cdot \beta c)^2} \tag{2.25}$$

$$= \frac{1}{L_{13}} \cdot \sqrt{(\delta L)^2 + \left(\frac{\delta \Delta\varphi}{2\pi f_{\mathrm{RFQ}}} \cdot \beta c\right)^2}. \tag{2.26}$$

Given that a limit for $\frac{\delta E}{E}$ (0.03 %) has already been computed, eq. (2.26) can be used to represent the limits on δL and $\delta\Delta\varphi$ that allows to comply with the limit on the energy. Figure 2.8 shows this through two plots: the first represents the worst situation for the energy but with different possibilities for L_{13}; the second represents the "worst" situation for the probes distance (going lower than 1 m would require too much precision both on distance and phase) but with different possible energies. Regarding the distance, a precision of about one tenth of millimetre is expected; this gives the possibility of having different requirements for $\delta\Delta\varphi$ and L_{13}. Clearly the 0.03 % limit has to be satisfied in all conditions, so also for the maximum energy, which is the worst case, but this is not necessarily always 230 MeV: given the modularity of LIGHT, some installations could be limited to lower energies, be it because of space limitations (for the case of already existing buildings willing to expand their treatment capabilities) or because some medical centres wants to specialise on a certain type of tumour.

Correct Bunch Detection Lastly, till now N_{13}, the integer number of bunches between probes 1 and 3, has always been considered to be computed exactly, but by looking at eq. (2.10) it can be seen that also N_{13} is computed from the phase shifts, so it has to be verified in which conditions the assumption is true. This can be done by applying the variance formula to eq. (2.10):

$$\delta {N_{13}}^2 = \left(\frac{\mathrm{d}N_{13}}{\mathrm{d}L_{13}}\right)^2 \delta L^2 + \left(\frac{\mathrm{d}N_{13}}{\mathrm{d}L_C}\right)^2 \delta L^2 + \left(\frac{\mathrm{d}N_{13}}{\mathrm{d}\Delta\varphi_{13}}\right)^2 \delta\Delta\varphi^2 + + \left(\frac{\mathrm{d}N_{13}}{\mathrm{d}\Delta\varphi_C}\right)^2 \delta\Delta\varphi^2 \tag{2.27}$$

$$= \left(\frac{\Delta\varphi_C}{2\pi L_C}\right)^2 \delta L^2 + \left(-\frac{\Delta\varphi_C L_{13}}{2\pi L_C^2}\right)^2 \delta L^2 + \left(\frac{L_{13}}{2\pi L_C}\right)^2 \delta\Delta\varphi^2 + + \left(-\frac{1}{2\pi}\right)^2 \delta\Delta\varphi^2 \tag{2.28}$$

$$= \left(\frac{\Delta\varphi_C}{2\pi L_C}\right)^2 \left[1 + \left(\frac{L_{13}}{L_C}\right)^2\right] \delta L^2 + \left[1 + \left(\frac{L_{13}}{L_C}\right)^2\right] \left(\frac{\delta\Delta\varphi}{2\pi}\right)^2 \tag{2.29}$$

$$\delta N_{13} = \sqrt{1 + \left(\frac{L_{13}}{L_C}\right)^2} \cdot \sqrt{\left(\frac{\delta L}{T_{\mathrm{RFQ}} \cdot \beta c}\right)^2 + \left(\frac{\delta\Delta\varphi}{2\pi}\right)^2} \tag{2.30}$$

It can be seen that this requirement goes in the opposite direction as compared to the previous two: eq. (2.18) gives a maximum limit for L_C; eq. (2.26) says

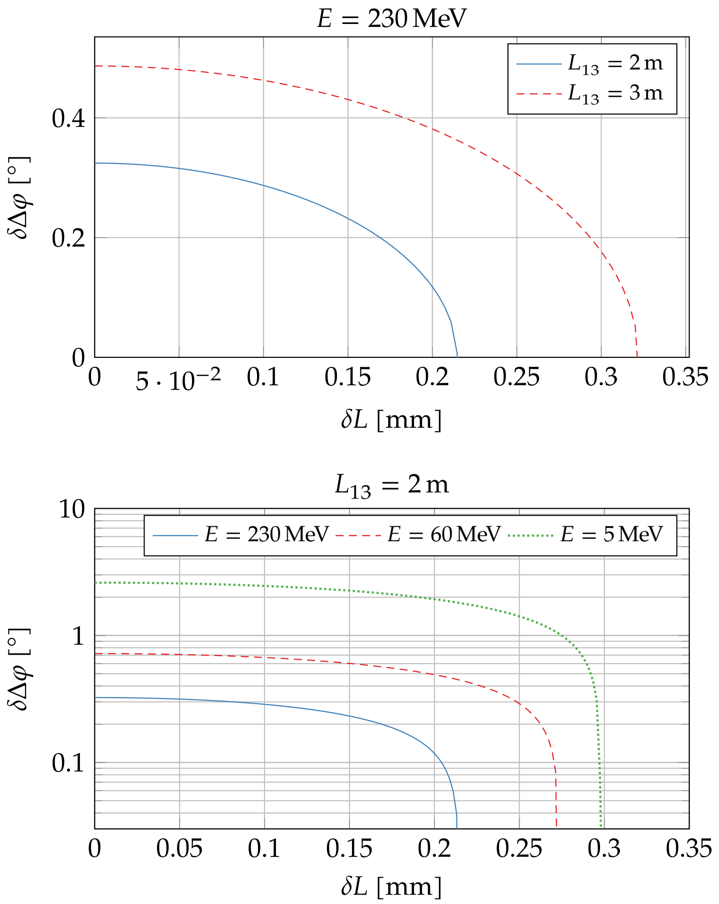


Figure 2.8 – Limits on δL and $\delta\Delta\varphi$ based on the required energy measurement resolution. Each line represents the maximum values that δL and $\delta\Delta\varphi$ can assume for $\frac{\delta E}{E} < 0.03\,\%$ to hold. For both the plots the blue line chart represent the same conditions ($L_{13} = 1\,\mathrm{m}$, $E = 230\,\mathrm{MeV}$).

that the greater L_{13}, the smaller the measurement error; eq. (2.30) instead wants to keep L_C and L_{13} close.

δN_{13} should be smaller than 0.5, so that when rounding its expression to the nearest integer number there is no error. It is thus possible to plot $L_{13} \div L_C$ for different δL and $\delta\Delta\varphi$ imposing $\delta N_{13} = 0.5$. The function for $L_{13} \div L_C$ is somewhat easier to grasp by defining beforehand two “help” functions:

$$\begin{aligned} \mathrm{hypot}(x, y) &= \sqrt{x^2 + y^2}; \\ \mathrm{cathetus}(h, c) &= \sqrt{h^2 - c^2}; \end{aligned}$$

$$\delta N_{13} = \mathrm{hypot}\left(1, \frac{L_{13}}{L_C}\right) \cdot \mathrm{hypot}\left(\frac{\delta L}{T_{\mathrm{RFQ}} \cdot \beta c}, \frac{\delta\Delta\varphi}{2\pi}\right) \tag{2.31}$$

$$\frac{L_{13}}{L_C} = \mathrm{cathetus}\left(\frac{\delta N_{13}}{\mathrm{hypot}\left(\frac{\delta L}{T_{\mathrm{RFQ}} \cdot \beta c}, \frac{\delta\Delta\varphi}{2\pi}\right)}, 1\right). \tag{2.32}$$

Figure 2.9 represents eq. (2.32) for the worst case, which this time occurs for the smallest energy.

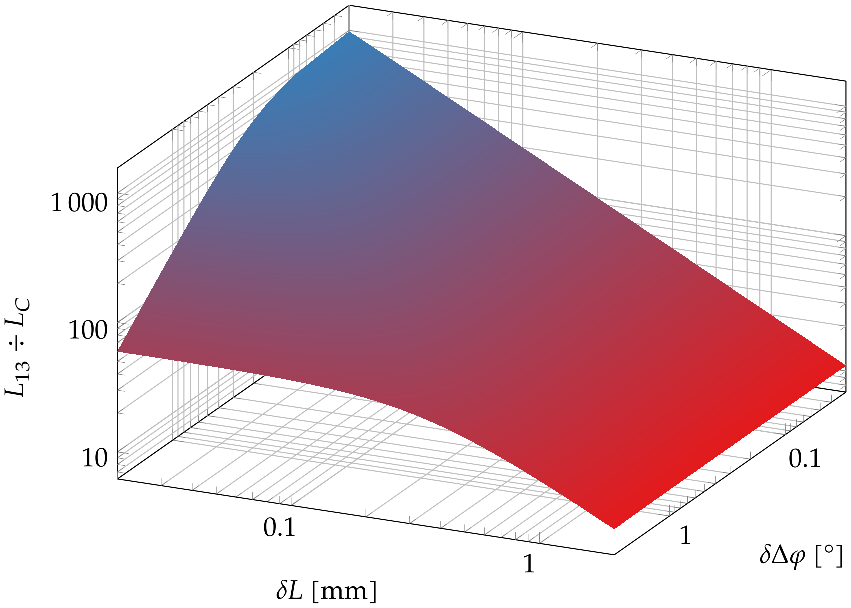


Figure 2.9 – Limit on the ratio of the furthest to the closest distance between probes to ensure a correct detection of the number of bunches.

It can be seen that δL is sensibly more influent on $L_{13} \div L_C$ than $\delta\Delta\varphi$. Given that the expected distance resolution is 0.1 mm, this means the distance ratio should be limited to about a hundred:

$$\frac{L_{13}}{L_C} < 100.$$

CHAPTER 3

Signal Processing & Phase Extraction

3.1 Signal Processing Driving Choices

A great effort has been spent, during this thesis, in carefully selecting the most appropriate way of extracting the phase information from the phase probes output signals. This investigation has been driven by the high variability of the characteristics of the signal to be processed (e.g. intensity and bunch shape) and by the highly demanding energy measurement resolution required for the system to be “medical-ready”. It has to be kept in mind that the latter property would be very desirable, but not strictly indispensable for the system to be useful in non-medical contexts. Another essential feature is the real-timeness of the measurement, which is to a degree in trade-off with a high resolution. This will be very important when choosing the kind of digital processing the signals will go through, because it limits all kind of processing to happen within 5 ms which is the minimum required beam pulse repetition period for the treatments.

3.1.1 Signal Characteristics

As already seen in fig. 1.4, bunches can have very different shapes, thus greatly influencing the pulse spectrum. Equation (2.1) expresses how to compute the image voltage at the output of a probe given the beam current, so only finding the value of the latter really gives an idea about what to expect from V_{im}.

It is expected that all the bunches within a pulse will be very similar to one another, so the probes output will be periodic signals and as such they will have comb-like spectra, with a spacing between the spectra lines of f_{RFQ}. Of these lines, only the first gives information about the centre of mass of the bunches and thus is the only one that can be exploited. Physical simulations regarding the beam dynamics have been performed to obtain an accurate estimation of the bunch shape. Figure 1.4 is the result of one of such simulations performed at 22.5 cm after the end of the linac, which is the foreseen position of the first EMS probe. Very similar results can be obtained for the foreseen position of the third EMS probe, which is 3.595 m after the first one.

Given the bunch particles, the bunch charge can be computed by multiplying that number by the elementary charge; the bunch current is then the derivative in time of the charge. From the beam current representation in time the Fourier transform can be computed to see how its power distributes over the frequencies and then sample it at the RFQ resonating frequency. The result of this operation starting from the simulated bunches is represented in fig. 3.1.

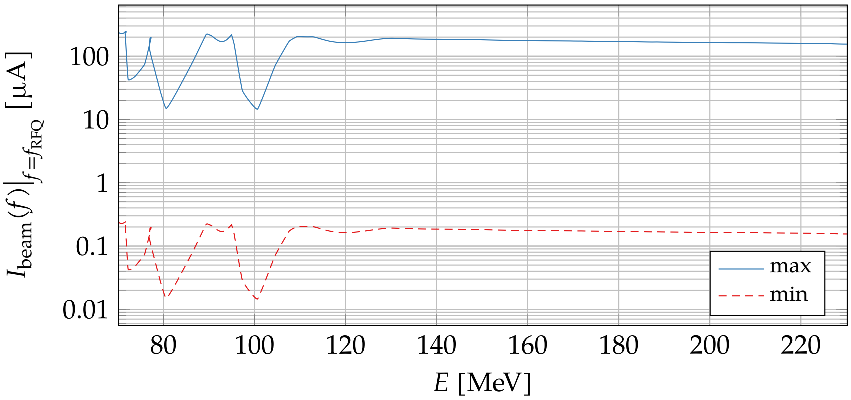


Figure 3.1 – Estimated beam current for the EMS first phase probe.

Figure 3.1 presents the best and worst cases, which happen respectively for the maximum and minimum foreseen beam intensity for patient treatments as reported in table 1.2. In some cases for low energies the beam debunches, meaning that the particle bunches de-focus in the longitudinal direction, leading to a drop in the foreseen beam current. This happens because the debunching is so strong that the signal has a non-zero mean value thus having weaker dynamics. Three representative cases are depicted in fig. 3.2, showing that it is in fact the extreme debunching the reason for the current drop at f_{RFQ} and not just a smaller peak height.

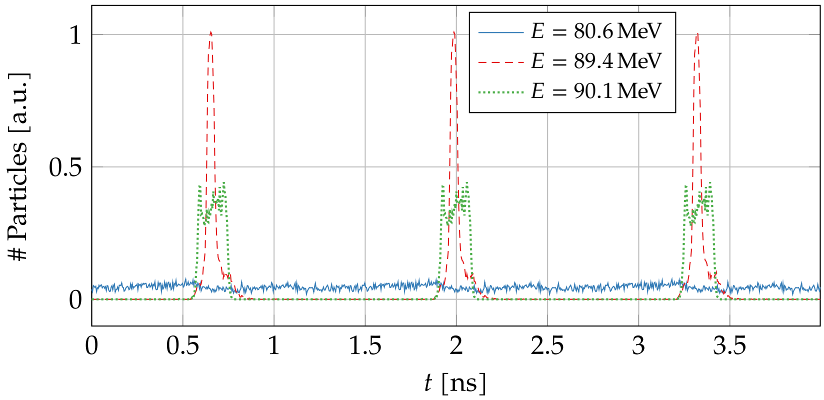


Figure 3.2 – Beam debunching effect on the bunch shape as predicted by beam simulations. These are actually examples, for three specific energies, of what is shown in fig. 1.4.

Putting together the information from eq. (2.1) and fig. 3.1 allows to estimate V_{im} bounds. The exact value of C to use in eq. (2.1) could not be measured; this is because the theoretical value of the probes electrode capacitance is smaller than 1 pF, so the input capacitance of the amplifier and the one of the cable will dominate, being in the oreder of some tens of pF. The exact model of cables that will be used is not known yet, but also their length (which influences the capacitance) will be decided based on how the analog components will fit mechanically with the rest of the equipment placed around the beam pipe. Figure 3.3 represents the expected bounds for V_{im} at the EMS first probe, assuming 100 pF as value for C.

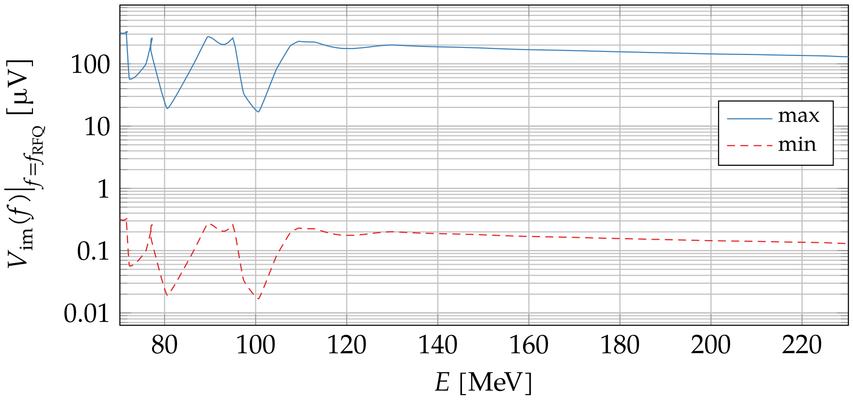


Figure 3.3 – Foreseen image voltage for the EMS first probe.

3.1.2 Energy Measurement Requirements

Section 2.4.1 introduces some equations which allow to establish a limit for the phase shift measurement precision and accuracy.

An attempt in achieving a better precision would be to try to maximise the signal-to-noise ratio while still in the analog domain; then, following the trend in Time-of-Flight based energy measurement systems, the signal will be digitised and the phase extraction will be done digitally.

Regarding the accuracy, the question is a bit more complicated. Analog electronic components usually introduce drifts and have environment dependent characteristics, such as impedance that vary with temperature or output signals power depending on the power supply. Moreover, analog components are never perfectly identical, and working at 750 MHz makes these imbalances disruptive: as an example, 1 mm of difference in the overall cables length of the acquisition chain of two probes would introduce 1.35° of offset in their phase shift. For this reasons, the energy measurement will have to integrate a calibration feature to allow periodic checks and corrections of the system conditions.

3.2 Analog Acquisition Chain

The block diagram of the energy measurement system is shown in fig. 3.4. This is a general scheme which includes all the possibly desirable hardware features for the EMS and will be described and refined in the following sections.

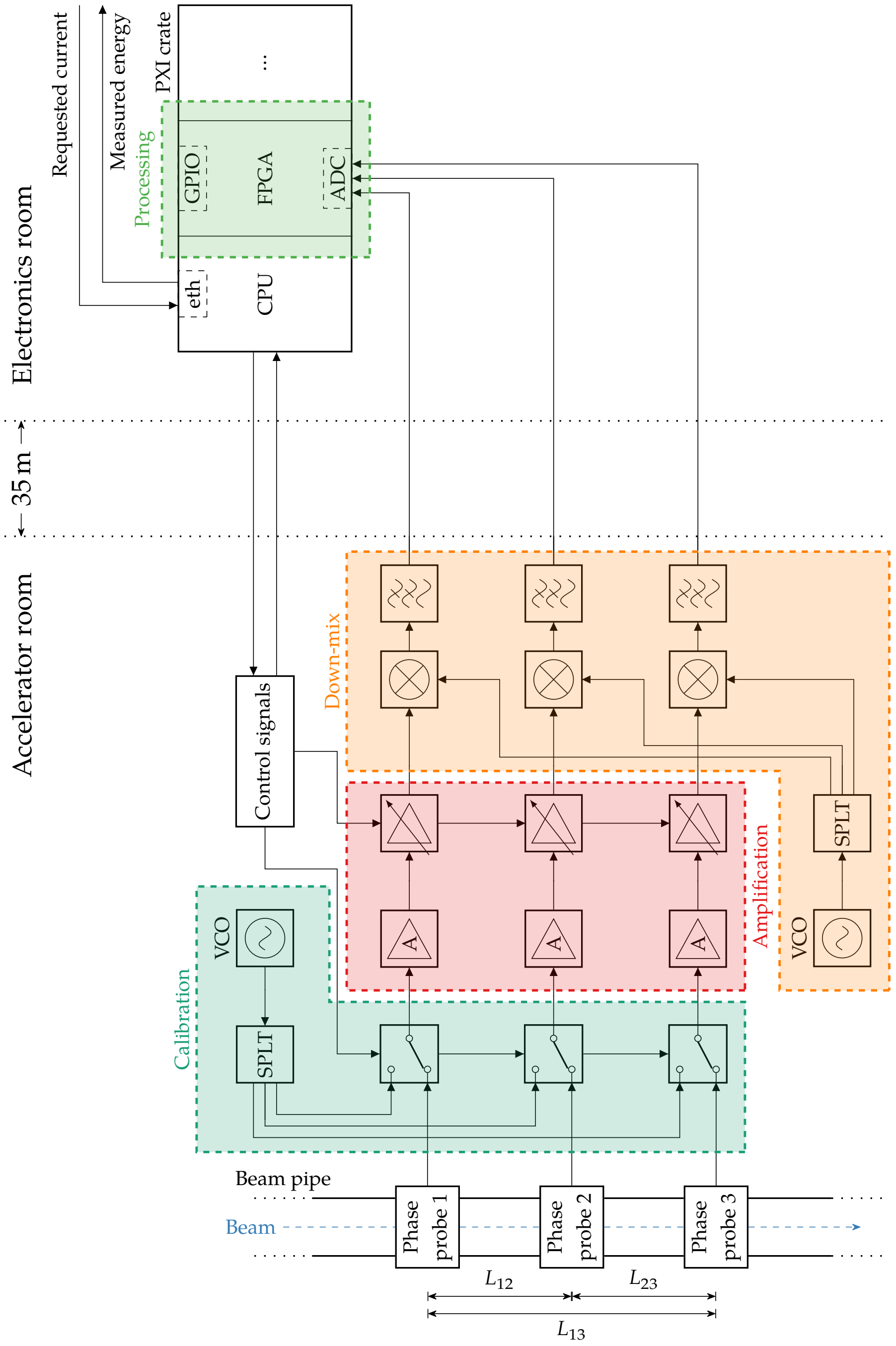


Figure 3.4 – Acquisition system hardware diagram for the EMS.

Calibration As each phase probe has its analog acquisition chain, a system calibration allows to cancel out the differences in the phase offsets introduced by each analog component. This subsystem consists of a sinusoidal oscillator, which emulates the relevant part of the beam signal, a three-way signal splitter and three relays, which allow to selectively inject into the analog chains either the oscillator signal or the probes outputs. The oscillator is voltage controlled but for our purpose the control voltage is fixed such that the signal frequency matches f_{RFQ}. The calibration is performed in two steps: first, the oscillator output has to be selected as input of the following processing stages, so that in this conditions the measured phase shifts are the offsets to be cancelled out ($\Delta\varphi_{xy,\mathrm{off}}$); second, the phase probes are selected as inputs and phase shift measurements are performed again ($\Delta\varphi_{xy,\mathrm{meas}}$). The correct phase shifts to be used for the energy computation will be

$$\Delta\varphi_{xy} = \Delta\varphi_{xy,\mathrm{meas}} - \Delta\varphi_{xy,\mathrm{off}}. \tag{3.1}$$

The frequency at which the calibration has to be performed is to be established based on experimental results; however, the required calibration rate is expected to be less than once per hour.

A further notice on the relays is that their inputs are 50 Ω-terminated while not connected to the outputs. This avoids charge accumulation on the probes electrodes in case they are crossed by the beam when the relays are set on the oscillator.

For the PMS, the accuracy is not so relevant, so the calibration feature is not needed. However, for simplicity of implementation, the PMS will be designed to be as close as possible to the EMS, so the available acquisition channels will be four for each system (further explanations on this matter will be given in chapter 4). For the PMS, eight detectors have been foreseen, so the calibration subsystem is replaced by a multiplexing layer which will allow to select between which detectors the phase shift measurement has to be performed.

Amplification This stage is necessary given the small signal amplitude foreseen in fig. 3.3. Figure 3.4 presents the first approach, which is a chain of two amplifiers with the second having a variable gain; this is to tackle the three orders of magnitude range for the beam intensity. However, this solution presents a major inconvenience: the phase offset introduced by the variable gain amplifiers changes with the gain and the change is different for each amplifier. Thus the calibration would have to be performed for each possible gain and the results stored and looked up for each measurement.

The current approach uses a single, fixed gain, low noise amplifier with 33 dB of gain (ZRL-1150LN+), available from Mini-Circuits. The main constraint in the selection of the amplifier model has been the high frequency at which it has to operate; the selected amplifier features a good combination of two relevant parameters, which are gain and noise figure. The noise figure is defined as the difference of the input signal-to-noise ratio and the output one:

$$\mathrm{NF} = \mathrm{SNR}_{in}\Big|_{\mathrm{dB}} - \mathrm{SNR}_{out}\Big|_{\mathrm{dB}}\,, \tag{3.2}$$

so what it represents is the worsening of the SNR as the signal goes through the amplifier. The selected amplifier has a noise figure of 1 dB; measuring its output with a spectrum analyser gives about −130 dBm of power over the relevant frequency range.

The input stage of the amplifier also influences the equivalent circuit of the phase probes, as reported in fig. 2.1. Figure 3.5 represents the impedance of the assemble of phase probe, SMA cable and amplifier. The real part of Z unsurprisingly starts at 50 Ω but then gets unstable above 40–50 MHz. Similarly, the imaginary part is equivalent to a 500–600 pF capacitor up to 60 MHz but from then on it jumps from capacitive to inductive behaviour unpredictably. This measurement does not seem reliable above about 50 MHz, so the exact behaviour and the equivalent capacitance of the phase probes will have to be estimated indirectly. One way to achieve this would be to measure a beam with known energy, intensity and bunch shape; then the amplitude of the acquired signals would depend only on the capacitance.

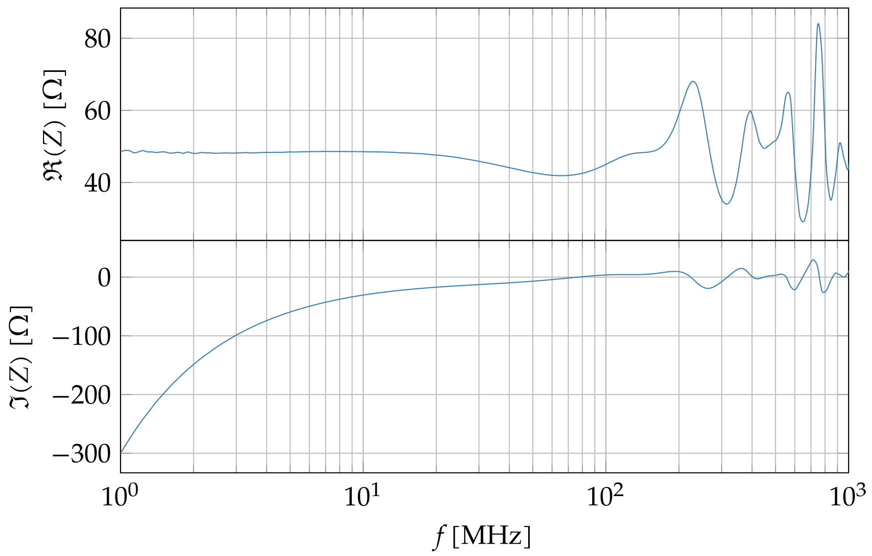


Figure 3.5 – Phase probe impedance as measured with a network analyser.

Down-mix This stage performs a frequency shift of the probes signals. It is desirable to do so to ease the digitisation of the signals but also their transmission outside the accelerator room, which takes place over several metres of cables. Transmitting signals with frequency components at 750 MHz can be problematic for two reasons, one being the higher attenuation of the signal power and the other being the higher sensitivity to the cable conditions: a variation in the signal transmission timing translates to the phase as

$$\mathrm{d}\varphi = 2\pi f\,\mathrm{d}t\,, \tag{3.3}$$

so this may influence the rate at which the system calibration has to be performed.

Frequency shifting is accomplished using a frequency mixer; this component performs the equivalent of a multiplication between its two input signals. Here an example is given using pure sinusoidal signals so to be as general as possible:

$$\begin{aligned}
v_{\mathrm{IF}}(t) &= v_{\mathrm{RF}}(t) \cdot v_{\mathrm{LO}}(t) && (3.4)\\
&= A_{\mathrm{RF}} \sin(\omega_{\mathrm{RF}}\, t + \varphi_{\mathrm{RF}}) \cdot A_{\mathrm{LO}} \sin(\omega_{\mathrm{LO}}\, t + \varphi_{\mathrm{LO}}) && (3.5)\\
&= \frac{A_{\mathrm{RF}} A_{\mathrm{LO}}}{2} \cdot [\cos(\omega_{\mathrm{RF}}\, t + \varphi_{\mathrm{RF}} - \omega_{\mathrm{LO}}\, t - \varphi_{\mathrm{LO}}) + \\
&\qquad - \cos(\omega_{\mathrm{RF}}\, t + \varphi_{\mathrm{RF}} + \omega_{\mathrm{LO}}\, t + \varphi_{\mathrm{LO}})] && (3.6)\\
&= \frac{A_{\mathrm{RF}} A_{\mathrm{LO}}}{2} \cdot [\cos((\omega_{\mathrm{RF}} - \omega_{\mathrm{LO}})\, t + \varphi_{\mathrm{RF}} - \varphi_{\mathrm{LO}}) + \\
&\qquad - \cos((\omega_{\mathrm{RF}} + \omega_{\mathrm{LO}})\, t + \varphi_{\mathrm{RF}} + \varphi_{\mathrm{LO}})]\,. && (3.7)
\end{aligned}$$

RF refers to the first input signal, which is usually the signal of interest; LO refers to the second input signal, which usually is a local oscillator; IF refers to the mixed output. Note that to get only the down-mixed signal, low-pass filters have been placed after the mixers with a cut-off requency such that their outputs are only

$$v_{\mathrm{IF}}(t) = \frac{A_{\mathrm{RF}} A_{\mathrm{LO}}}{2} \cdot [\cos((\omega_{\mathrm{RF}} - \omega_{\mathrm{LO}})\, t + \varphi_{\mathrm{RF}} - \varphi_{\mathrm{LO}})]. \qquad (3.8)$$

It is easy to see that the down-mix operation preserves the phase shift between the RF signals, given that the same LO signal is used for every mixer.

However, this operation also has downsides: the signal frequency will have to be estimated, since now it depends on the oscillator which is not as stable as the RFQ; the output signal phase stability is worsen because both mixer and oscillator are not ideal components. Moreover, a real mixer introduces losses when mixing the signals; an experimental example is shown in fig. 3.6. For this reasons, also in this case the decision about whether using or not the down-mix stage will be made based on experimental data given the complexity of tackling such a problem analytically.

Processing This stage performs the digitisation and processing of the signal to extract the phase shift and the energy. Given the real-timeness requirement, it is inevitable that this needs to be performed on some kind of embedded platform which can offer the possibility of communicating using a real-time interface.

The selected digitiser is ADQ14AC-4C-GPIO from Teledyne SP Devices. There are a number of factors which led to this choice:

- The 1 GHz sampling frequency allows to have a fair number of samples also for the shortest beam pulses.
- The 14 bits (ENOB: 9.6 bits) ADCs give a good resolution also for lower beam intensities; currently, signals weaker than 1 mV have been ac-

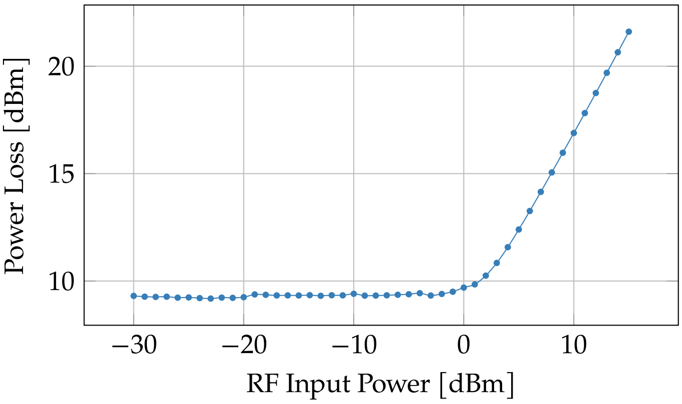


Figure 3.6 – Frequency mixer power loss from RF to IF. Real world mixers accept inputs in a certain power range above which they saturate. Also if the inputs have a proper power, for this model the power loss is quite high (9 to 10 dB, as can be seen in the plot).

quired with fairly good results with respect to the phase shifts measurements.

- The analog bandwidth goes from 80 Hz up to 900 MHz, potentially allowing both the down-mix and the non-down-mix cases.
- The GPIO lines enable the possibility to implement real-time data exchange in a very flexible way.
- The card hosts an FPGA which can be accessed through an FDK (Firmware Development Kit) to implement custom signal processing. The custom firmware can interoperate both with the host PC and the GPIO lines.
- Various form factors are available, which allows to choose the most suitable one based on the rest of the equipment. For example, the PXI form factor has been used throughout this thesis and will be used for the PMS, while for the final EMS implementation a stand-alone card (e.g. in the USB form factor) would be preferable. The different interfaces are implemented in such a way that no changes have to be made to the inter-operating software when changing the card form factor.

Usually digitisers come with an even number of acquisition channels, so the four channels model has been chosen.

The implementation of the FPGA custom firmware and the software developed to communicate with the digitiser will be analysed in depth in the next chapter.

It is maybe useful to clarify at this point how the direct digitisation of signals at 750 MHz is possible using a 1 GHz sampling frequency. The common interpretation of the Nyquist–Shannon sampling theorem leads to the choice of a sampling frequency which is greater than twice the maximum frequency present in the signal to be sampled, but actually the theorem only requires a sampling frequency greater than twice the *bandwidth* of the sampled signal. Considering only the first harmonic, the bandwidth of the signal is very narrow, so in principle a sampling frequency in the order of tens of megahertz would suffice but, given the short duration of the beam pulse, it would give very few samples. Obviously, aliasing the higher harmonics can be avoided by using low-/band-pass filters at the digitiser inputs.

In any case, performing the discrete Fourier transform (DFT) on the digitised signals gives valid information only in the frequency range $(-0.5f_s, 0.5f_s)$; the properties of the 750 MHz signal can be retrieved anyway thanks to the following properties of the sampled signal, of the sampling operation and of the Fourier transform:

- The Fourier transform of a sine wave is

  $$\mathcal{F}\left[A_g \cdot \sin\left(\omega_g\right)\right] = \frac{A_g}{2} \cdot \left[\delta\left(f - f_g\right) + \delta\left(f + f_g\right)\right], \tag{3.9}$$

  where $\delta(x)$ is the Dirac delta function.

- A sampling operation in time corresponds to a repetition operation in frequency, i.e. given a signal $g(t)$ whose Fourier transform is $G(f)$ then the Fourier transform of the sampled signal is

  $$\mathcal{F}\left[g(t) \cdot \mathrm{comb}_{T_s}(t)\right] \propto \sum_{n=-\infty}^{+\infty} G(f - nf_s), \tag{3.10}$$

  where

  $$\mathrm{comb}_X(x) = \sum_{n=-\infty}^{+\infty} \delta(x - nX) \tag{3.11}$$

  is the function modelling the sampling operation.

- The modulus of the Fourier transform of a real-valued signal is an even function while the argument is an odd function:

  $$|G(f)| = |G(-f)|, \tag{3.12}$$

  $$\arg G(f) = -\arg G(-f). \tag{3.13}$$

From these rules it can be deduced that in the frequency range $(0, 0.5f_s)$ the only frequency component is the alias $0.5 \cdot A \cdot \delta\left(f + f_g - f_s\right)$ of the negative component $0.5 \cdot A \cdot \delta\left(f + f_g\right)$ of the probe signal: the amplitude of this alias is the same as the one of the 750 MHz harmonic while its phase is the opposite of the one to be retrieved (as per eqs. (3.12) and (3.13)).

3.3 IQ Demodulation for Phase Extraction

IQ demodulation is the most widely used technique for extracting the phase information of a digitised signal. The formulation is quite simple: given a sinusoidal signal $g(t) = A_g \cdot \sin\left(\omega_g t + \varphi_g\right)$, this signal can also be written as

$$g(t) = A_g \cdot \cos\varphi_g \cdot \sin\left(\omega_g t\right) + A_g \cdot \sin\varphi_g \cdot \cos\left(\omega_g t\right); \tag{3.14}$$

this is easily derived from trigonometric equivalences. By defining

$$I = A_g \cdot \cos\varphi_g, \tag{3.15}$$

$$Q = A_g \cdot \sin\varphi_g, \tag{3.16}$$

then

$$g(t) = I \cdot \sin\big(\omega_g\, t\big) + Q \cdot \cos\big(\omega_g\, t\big). \tag{3.17}$$

Given the definition of I and Q it is easy to see that

$$A_g = \sqrt{I^2 + Q^2}; \tag{3.18}$$

$$\varphi_g = \operatorname{atan}(Q, I). \tag{3.19}$$

It is instead less obvious that I and Q can be computed from the digital samples of $g(t)$ using the following formulas:

$$I \approx \frac{2}{N} \cdot \sum_{n=0}^{N-1} g(n\, T_s) \cdot \sin\big(\omega_g\, n\, T_s\big); \tag{3.20}$$

$$Q \approx \frac{2}{N} \cdot \sum_{n=0}^{N-1} g(n\, T_s) \cdot \cos\big(\omega_g\, n\, T_s\big). \tag{3.21}$$

In fact, the summed terms of eq. (3.20) can be expanded as follows:

$$g(t) \cdot \sin\big(\omega_g\, t\big) = \big[I \cdot \sin\big(\omega_g\, t\big) + Q \cdot \cos\big(\omega_g\, t\big)\big] \cdot \sin\big(\omega_g\, t\big) \tag{3.22}$$

$$= I \cdot \sin^2\big(\omega_g\, t\big) + Q \cdot \cos\big(\omega_g\, t\big) \cdot \sin\big(\omega_g\, t\big) \tag{3.23}$$

$$= \frac{I}{2} \cdot \big[1 - \cos\big(2\,\omega_g\, t\big)\big] + \frac{Q}{2} \cdot \sin\big(2\,\omega_g\, t\big); \tag{3.24}$$

now the full right-hand side of eq. (3.20) can be analysed:

$$\frac{2}{N} \cdot \sum_{n=0}^{N-1} g(n\, T_s) \cdot \sin\big(\omega_g\, n\, T_s\big) = I - I \cdot \left[\frac{1}{N} \cdot \sum_{n=0}^{N-1} \cos\left(2\pi n\, \frac{2f_g}{f_s}\right)\right] + \\ + Q \cdot \left[\frac{1}{N} \cdot \sum_{n=0}^{N-1} \sin\left(2\pi n\, \frac{2f_g}{f_s}\right)\right]. \tag{3.25}$$

The bracketed terms of eq. (3.25) are equivalent to computing the mean of, respectively, the cosine and the sine of evenly-spaced points. As such, they can be considered negligible if one of the following conditions is verified:

1. The sampling frequency is an exact multiple of the signal frequency:

   $$\frac{2f_g}{f_s} = \frac{k}{N}, \qquad \frac{f_s}{f_g} = \frac{2N}{k} \in \mathbb{N}^* \tag{3.26}$$

   In this case the two terms are exactly zero as the arguments of the two trigonometric functions are evenly spaced over an interval which is an exact multiple of the circumference, the lower bound being 0 and the upper one being $2\pi k \cdot \frac{N-1}{N}$. This is the most common technique as usually f_s is chosen to be $4f_g$ so that there are four samples per period distant 90° from each other allowing to perform the IQ demodulation online and also removing possible voltage offsets from the signal [1]. However, this technique presents the downside of fixing the sampling frequency to the mechanical parameters of the RFQ, and thus usually requires to design a custom acquisition system with the required f_s.

2. Also if the frequencies ratio is not an exact integer, the same effect can be reached asymptotically if N is big enough. This second condition is the one relied upon for this system, as the other condition would require at least a sampling frequency of 3 GHz to be achieved.

If aither of the two conditions is valid then eqs. (3.20) and (3.21) are no longer approximations but equivalences.

All of the discussion has been carried out using the formulas for I, but the same can be said also for Q.

3.3.1 DFT-Related Interpretation

IQ demodulation can be shown to be equivalent to a single DFT point computed at the exact signal frequency. In fact, the discrete Fourier Transform of $g(t)$ can be computed as

$$\begin{aligned}
G(f) &= T_s \cdot \sum_{n=0}^{N-1} g(nT_s) \cdot e^{-j2\pi f n T_s} && (3.27)\\
&= T_s \cdot \sum_{n=0}^{N-1} g(nT_s) \cdot [\cos(-2\pi f n T_s) + j\,\sin(-2\pi f n T_s)] && (3.28)\\
&= T_s \cdot \sum_{n=0}^{N-1} g(nT_s) \cdot [\cos(2\pi f n T_s) - j\,\sin(2\pi f n T_s)] && (3.29)\\
&= T_s \cdot \sum_{n=0}^{N-1} g(nT_s) \cdot \cos(2\pi f n T_s) + \\
&\quad - j \cdot T_s \cdot \sum_{n=0}^{N-1} g(nT_s) \cdot \sin(2\pi f n T_s). && (3.30)
\end{aligned}$$

Then the phase of the sinusoidal component of g for a certain frequency f can be computed from its DFT as

$$\begin{aligned}
\arg[G(f)] &= \operatorname{atan}\left(\frac{\Im(G(f))}{\Re(G(f))}\right) && (3.31)\\
&= \operatorname{atan}\left(-\frac{\sum_{n=0}^{N-1} g(nT_s) \cdot \sin(2\pi f n T_s)}{\sum_{n=0}^{N-1} g(nT_s) \cdot \cos(2\pi f n T_s)}\right). && (3.32)
\end{aligned}$$

By defining

$$\begin{aligned}
I(f) &= \frac{2}{N} \cdot \sum_{n=0}^{N-1} g(nT_s) \cdot \sin(2\pi f n T_s), && (3.33)\\
Q(f) &= \frac{2}{N} \cdot \sum_{n=0}^{N-1} g(nT_s) \cdot \cos(2\pi f n T_s), && (3.34)
\end{aligned}$$

eq. (3.32) can be rewritten as

$$\arg\left[G(f)\right] = \operatorname{atan}\left(-\frac{I(f)}{Q(f)}\right). \tag{3.35}$$

Beside the fact that the DFT cannot be usually computed at arbitrary frequencies, IQ demodulation does just that with only a further step: indeed, IQ computes $\operatorname{atan}\left(\frac{Q(f)}{I(f)}\right)$ which can be proven to be equivalent to $\operatorname{atan}\left(-\frac{I(f)}{Q(f)}\right) + 90°$, so for phase *shift* measurements the two computations are equivalent.

This has the consequence that the knowledge about the DFT can be used with the IQ computation, to tackle possible computational problems or to enhance its performance. In fact, it will become useful to estimate how the IQ demodulation influences the noise of the signal on which it is performed.

3.3.2 Least Squares Interpretation

Another interesting interpretation relates the IQ demodulation to approaches more typically seen in the statistic field.

Phase detection can be formulated as a parameter estimation problem for a linear model. Considering $g(t)$ again, the typical formulation of the problem is

$$\underline{x} = \mathbf{H} \cdot \underline{\vartheta} + \underline{w} \tag{3.36}$$

where, in our case:

$\underline{x}$ is the vector of the acquired samples,

$\mathbf{H}$ is the matrix-form of the sampled sinusoidal signal, its n-th row being $\underline{h}_n^{\mathsf{T}} = \left[\cos\left(\omega_g\, n\, T_s\right) \quad \sin\left(\omega_g\, n\, T_s\right)\right]$,

$\underline{\vartheta} = \left[A_g \cos\varphi_g \quad A_g \sin\varphi_g\right]^{\mathsf{T}}$ is the vector of parameters to estimate, and

$\underline{w}$ is the noise model, assumed to be unknown.

The ordinary least squares estimation of $\underline{\vartheta}$ is then

$$\underline{\hat{\vartheta}} = \left(\mathbf{H}^{\mathsf{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathsf{T}}\underline{x}. \tag{3.37}$$

The expression for $\underline{\hat{\vartheta}}$ can be greatly simplified. Considering the parenthesised factor ($c_n = \cos\left(\omega_g\, n\, T_s\right)$ and $s_n = \sin\left(\omega_g\, n\, T_s\right)$):

$$\mathbf{H}^{\mathsf{T}}\mathbf{H} = \begin{bmatrix} c_0 & c_1 & \cdots & c_{N-1} \\ s_0 & s_1 & \cdots & s_{N-1} \end{bmatrix} \cdot \begin{bmatrix} c_0 & s_0 \\ c_1 & s_1 \\ \vdots & \vdots \\ c_{N-1} & s_{N-1} \end{bmatrix} \tag{3.38}$$

$$= \begin{bmatrix} \sum_{n=0}^{N-1} c_n^2 & \sum_{n=0}^{N-1} c_n\, s_n \\ \sum_{n=0}^{N-1} c_n\, s_n & \sum_{n=0}^{N-1} s_n^2 \end{bmatrix} \tag{3.39}$$

$$= \frac{1}{2}\begin{bmatrix} \sum_{n=0}^{N-1} (c_{2n} + 1) & \sum_{n=0}^{N-1} s_{2n} \\ \sum_{n=0}^{N-1} s_{2n} & \sum_{n=0}^{N-1} (1 - c_{2n}) \end{bmatrix} \tag{3.40}$$

$$= \frac{1}{2}\begin{bmatrix} N + \sum_{n=0}^{N-1} c_{2n} & \sum_{n=0}^{N-1} s_{2n} \\ \sum_{n=0}^{N-1} s_{2n} & N - \sum_{n=0}^{N-1} c_{2n} \end{bmatrix}. \tag{3.41}$$

The same reasoning used to simplify eq. (3.25) can be applied considering $\sum_{n=0}^{N-1} c_{2n} \ll N$ and $\sum_{n=0}^{N-1} s_{2n} \ll N$, so

$$\mathbf{H}^{\mathsf{T}}\mathbf{H} \approx \begin{bmatrix} \frac{N}{2} & 0 \\ 0 & \frac{N}{2} \end{bmatrix} \tag{3.42}$$

and thus

$$(\mathbf{H}^{\mathsf{T}}\mathbf{H})^{-1} \approx \begin{bmatrix} \frac{2}{N} & 0 \\ 0 & \frac{2}{N} \end{bmatrix}. \tag{3.43}$$

The conclusion is that

$$\underline{\hat{\vartheta}} = \begin{bmatrix} \frac{2}{N} & 0 \\ 0 & \frac{2}{N} \end{bmatrix} \cdot \begin{bmatrix} c_0 & c_1 & \dots & c_{N-1} \\ s_0 & s_1 & \dots & s_{N-1} \end{bmatrix} \cdot \underline{x} \tag{3.44}$$

$$\approx \begin{bmatrix} \frac{2}{N} \cdot \sum_{n=0}^{N-1} x_n \cdot c_n \\ \frac{2}{N} \cdot \sum_{n=0}^{N-1} x_n \cdot s_n \end{bmatrix}, \tag{3.45}$$

to be compared with eqs. (3.20) and (3.21).

This interpretation allows to derive some statistical properties of the IQ demodulation. In fact, when the noise is uncorrelated from the signal, homoscedastic and serially uncorrelated, then OLS provides a minimum-variance, unbiased estimation.

3.3.3 Effects of Using a Wrong Frequency

All of the discussion so far relies on the fact that the exact signal frequency is known. In our system this would not be true if the down-mix was performed. For this reason it is useful to have an idea about the effects of applying the IQ demodulation with a slightly off frequency.

The issue can be tackled starting from eq. (3.22) by supposing that a pulsation $\hat{\omega}_g = \omega_g + \omega_\varepsilon$ was used:

$$g(t) \cdot \sin\big(\hat{\omega}_g\, t\big) = \big[I \cdot \sin\big(\omega_g\, t\big) + Q \cdot \cos\big(\omega_g\, t\big)\big] \cdot \sin\big(\hat{\omega}_g\, t\big) \tag{3.46}$$

$$\begin{aligned} &= I \cdot \sin\big(\omega_g\, t\big) \cdot \sin\big((\omega_g + \omega_\varepsilon)\, t\big) + \\ &\quad + Q \cdot \cos\big(\omega_g\, t\big) \cdot \sin\big((\omega_g + \omega_\varepsilon)\, t\big) \end{aligned} \tag{3.47}$$

$$\begin{aligned} &= I \cdot \left[\frac{1}{2}\cos(\omega_\varepsilon\, t) - \frac{1}{2}\cos\big((2\,\omega_g + \omega_\varepsilon)\, t\big)\right] \\ &\quad + Q \cdot \left[\frac{1}{2}\sin(\omega_\varepsilon\, t) + \frac{1}{2}\sin\big((2\,\omega_g + \omega_\varepsilon)\, t\big)\right]. \end{aligned} \tag{3.48}$$

Making the same assumptions as before about a proper selection of f_s or N, the previous equation can be simplified to

$$g(t) \cdot \sin\big(\hat{\omega}_g\, t\big) = I \cdot \frac{1}{2}\cos(\omega_\varepsilon\, t) + Q \cdot \frac{1}{2}\sin(\omega_\varepsilon\, t), \tag{3.49}$$

so

$$\hat{I} = \frac{2}{N} \cdot \sum_{n=0}^{N-1} g(n\,T_s) \cdot \sin\!\left(\hat{\omega}_g\, n\, T_s\right) \tag{3.50}$$

$$= I \cdot \frac{1}{N} \cdot \sum_{n=0}^{N-1} \cos\!\left(2\pi n \frac{f_\varepsilon}{f_s}\right) + Q \cdot \frac{1}{N} \cdot \sum_{n=0}^{N-1} \sin\!\left(2\pi n \frac{f_\varepsilon}{f_s}\right). \tag{3.51}$$

Similarly, for Q:

$$\hat{Q} = \frac{2}{N} \cdot \sum_{n=0}^{N-1} g(n\,T_s) \cdot \cos\!\left(\hat{\omega}_g\, n\, T_s\right) \tag{3.52}$$

$$= Q \cdot \frac{1}{N} \cdot \sum_{n=0}^{N-1} \cos\!\left(2\pi n \frac{f_\varepsilon}{f_s}\right) - I \cdot \frac{1}{N} \cdot \sum_{n=0}^{N-1} \sin\!\left(2\pi n \frac{f_\varepsilon}{f_s}\right). \tag{3.53}$$

Now, the following two quantities (where $r_f = \frac{f_\varepsilon}{f_s}$) have to be analysed:

$$k_s = \frac{1}{N} \cdot \sum_{n=0}^{N-1} \sin\!\left(2\pi\, n\, r_f\right); \tag{3.54}$$

$$k_c = \frac{1}{N} \cdot \sum_{n=0}^{N-1} \cos\!\left(2\pi\, n\, r_f\right). \tag{3.55}$$

These two formulas can be rewritten without the use of summations:

$$k_s = \frac{\sin\!\left(\pi\, r_f N\right)}{\sin\!\left(\pi\, r_f\right) \cdot N} \cdot \sin\!\left(\pi\, r_f\,(N-1)\right), \tag{3.56}$$

$$k_c = \frac{\sin\!\left(\pi\, r_f N\right)}{\sin\!\left(\pi\, r_f\right) \cdot N} \cdot \cos\!\left(\pi\, r_f\,(N-1)\right). \tag{3.57}$$

The error in frequency can be expected to be very small, so the small-angle approximation can be applied to the sine at the denominator:

$$k_s = \frac{\sin\!\left(\pi\, r_f N\right)}{\pi\, r_f N} \cdot \sin\!\left(\pi\, r_f\,(N-1)\right), \tag{3.58}$$

$$k_c = \frac{\sin\!\left(\pi\, r_f N\right)}{\pi\, r_f N} \cdot \cos\!\left(\pi\, r_f\,(N-1)\right). \tag{3.59}$$

As last step, it can be assumed that $N \gg 1$ and thus $N - 1 \approx N$:

$$k_s = \frac{\sin^2\left(\pi\, r_f N\right)}{\pi\, r_f N} = \frac{1 - \cos\!\left(2\pi\, r_f N\right)}{2\pi\, r_f N}, \tag{3.60}$$

$$k_c = \frac{\sin\!\left(\pi\, r_f N\right) \cdot \cos\!\left(\pi\, r_f N\right)}{\pi\, r_f N} = \frac{\sin\!\left(2\pi\, r_f N\right)}{2\pi\, r_f N}. \tag{3.61}$$

The independent variables are N and r_f. Figure 3.7 shows k_s and $1 - k_c$ (so that the aim is for both quantities to be as close as possible to 0). If only a search for the DFT peak were used as frequency estimation, the maximum error would be half the width of a frequency bin. This situation is represented by the red lines on the plots: it means that $r_f = (2\,N)^{-1}$ and thus

$$k_s = \frac{1 - \cos(\pi)}{\pi} = \frac{2}{\pi}, \tag{3.62}$$

$$k_c = \frac{\sin(\pi)}{\pi} = 0, \tag{3.63}$$

which would prevent a proper estimation of φ_g. This motivates the search for a better method for estimating the fundamental frequency of the possibly down-mixed signals coming from the phase probes.

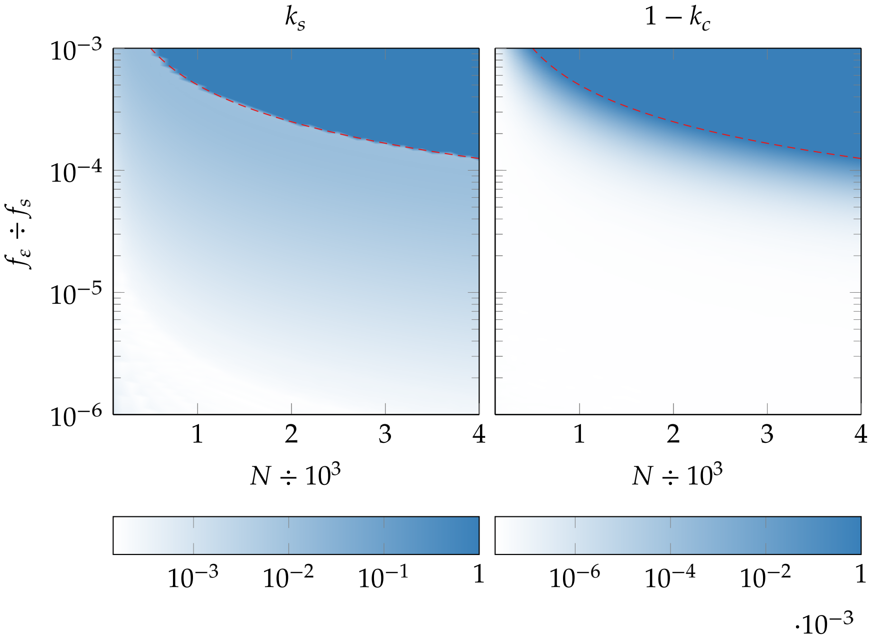


Figure 3.7 – Errors estimation for when a wrong frequency is used in the IQ demodulation.

3.4 Frequency Detection

Given the design of the system, the frequency detection is best performed digitally as a preliminary step of the IQ demodulation. As it is performed digitally, it will be almost unavoidably based on the DFT, with some refinement on top of its result. For the case of a simple DFT, the frequency is estimated as

$$\hat{f}_g = \frac{f_s}{N} \cdot \arg\max_k |\mathrm{DFT}_k\left[g(t)\right]| \tag{3.64}$$

where N is the DFT length and $\mathrm{DFT}_k\left[g(t)\right]$ denotes the k-th sample of the DFT of $g(t)$. In this case, f_ε is uniformly distributed in the interval $\left(0, \frac{f_s}{2N}\right)$.

The next section discusses the literature associated with possible refinements on top of this computation.

3.4.1 Sub-Bin Resolution

Buneman Formula For a sampled signal, if the sampling time window is not an exact multiple of the signal period, then the signal frequency doesn't fall on an exact DFT point. The Buneman algorithm allows to recover the exact frequency (perfectly for the case of pure sine waves) [2]:

$$i = \arg\max_k |\mathrm{DFT}_k\left[g(t)\right]|; \tag{3.65}$$

$$y = \sin\left(\frac{\pi}{N}\right); \tag{3.66}$$

$$x = \cos\left(\frac{\pi}{N}\right) + \frac{|\mathrm{DFT}_i\left[g(t)\right]|}{|\mathrm{DFT}_{i+1}\left[g(t)\right]|}; \tag{3.67}$$

$$\hat{f}_g = \frac{f_s}{N} \cdot \left[i + \frac{N}{\pi} \cdot \operatorname{atan}(y, x)\right]. \tag{3.68}$$

DFT Peak Weighted Average A simpler approach would be to consider k points before and after the DFT peak and perform a weighted mean on them, considering the magnitude of the DFT point as the weight [3]:

$$i = \arg\max_k |\mathrm{DFT}_k\left[g(t)\right]|; \tag{3.69}$$

$$\hat{f}_g = \frac{f_s}{N} \cdot \frac{\sum\limits_{j=i-k}^{i+k} j \cdot \left|\mathrm{DFT}_j\left[g(t)\right]\right|}{\sum\limits_{j=i-k}^{i+k} \left|\mathrm{DFT}_j\left[g(t)\right]\right|}. \tag{3.70}$$

DFT Peak Parabolic Interpolation One step further than the previous method is to observe that the neighbourhood of the DFT peak is usually shaped like a parabola. Thus the frequency estimation can be reformulated as a problem of finding the local linear regression of order 2 around the DFT peak. [4]

The generic algebraic problem for polynomial interpolation is

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \cdot \begin{bmatrix} \vartheta_1 \\ \vartheta_2 \\ \vdots \\ \vartheta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}, \tag{3.71}$$

$$\underline{y} = \mathbf{X} \cdot \underline{\vartheta} + \underline{\varepsilon} \tag{3.72}$$

and its OLS solution is

$$\underline{\hat{\vartheta}} = \left(\mathbf{X}^{\mathsf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathsf{T}}\underline{y}. \tag{3.73}$$

For the case of a parabolic interpolation $m = 2$. To have a unique solution therefore three points are needed:

$$\begin{bmatrix} S_{i-1} \\ S_i \\ S_{i+1} \end{bmatrix} = \begin{bmatrix} 1 & f_{i-1} & f_{i-1}^2 \\ 1 & f_i & f_i^2 \\ 1 & f_{i+1} & f_{i+1}^2 \end{bmatrix} \cdot \begin{bmatrix} c \\ b \\ a \end{bmatrix} + \begin{bmatrix} \varepsilon_{i-1} \\ \varepsilon_i \\ \varepsilon_{i+1} \end{bmatrix} \tag{3.74}$$

where $S_k = |\mathrm{DFT}_k\left[g(t)\right]|, f_k = \frac{f_s}{N} \cdot k$ and, as always, $i = \arg\max_k |\mathrm{DFT}_k\left[g(t)\right]|$.

The OLS solution would be

$$\underline{\hat{\vartheta}} = \left(\mathbf{F}^{\mathsf{T}}\mathbf{F}\right)^{-1}\mathbf{F}^{\mathsf{T}}\underline{S}, \tag{3.75}$$

This can be greatly simplified by considering that **F** is squared, so the inverse can be used in place of the pseudo-inverse:

$$\mathbf{F}^{-1} = \left(\mathbf{F}^{\mathsf{T}}\,\mathbf{F}\right)^{-1}\mathbf{F}^{\mathsf{T}}; \tag{3.76}$$

furthermore, the abscissa of the parabola can be scaled as following

$$k = f_k \frac{N}{f_s} - i \tag{3.77}$$

to simplify **F** to

$$\mathbf{F} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}. \tag{3.78}$$

It is now trivial to compute

$$\begin{bmatrix} \hat{c} \\ \hat{b} \\ \hat{a} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 0 & 2 & 0 \\ -1 & 0 & 1 \\ 1 & -2 & 1 \end{bmatrix} \cdot \begin{bmatrix} S_{i-1} \\ S_i \\ S_{i+1} \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 2\,S_i \\ S_{i+1} - S_{i-1} \\ S_{i+1} + S_{i-1} - 2\,S_i \end{bmatrix}. \tag{3.79}$$

The signal frequency is the abscissa of the parabola's vertex, for which the inverse of the previous scaling has to be applied:

$$\begin{aligned} \hat{f_g} &= \frac{f_s}{N}\left(i - \frac{\hat{b}}{2\,\hat{a}}\right) && (3.80) \\ &= \frac{f_s}{N}\left(i + \frac{1}{2}\,\frac{S_{i-1} - S_{i+1}}{S_{i-1} + S_{i+1} - 2\,S_i}\right). && (3.81) \end{aligned}$$

Given their complexity, these formulas cannot be compared analytically; instead, a numerical comparison for a set of plausible cases will be done. Before that, however, another effect has also to be considered, which influences the frequency estimation in a non trivial way when combined with different estimation methods.

3.4.2 Spectral Windowing

Spectral windowing can be used to enhance the DFT performances depending on the feature that will be extrapolated from it. From the plethora of possible window functions some numerical simulations has been performed using the most common ones, for the previously described sub-bin resolution methods. Figures 3.8 and 3.9 summarise some of the trade-offs that arise when choosing the sub-bin resolution method and the window function.

Every sub-bin resolution method is affected by the used window function in a different way. In any case, the choice of the sub-bin method appears to be more influential than the window function.

The Buneman formula is the most accurate but only when no window is used and when the number of samples acquired is very close to the length of

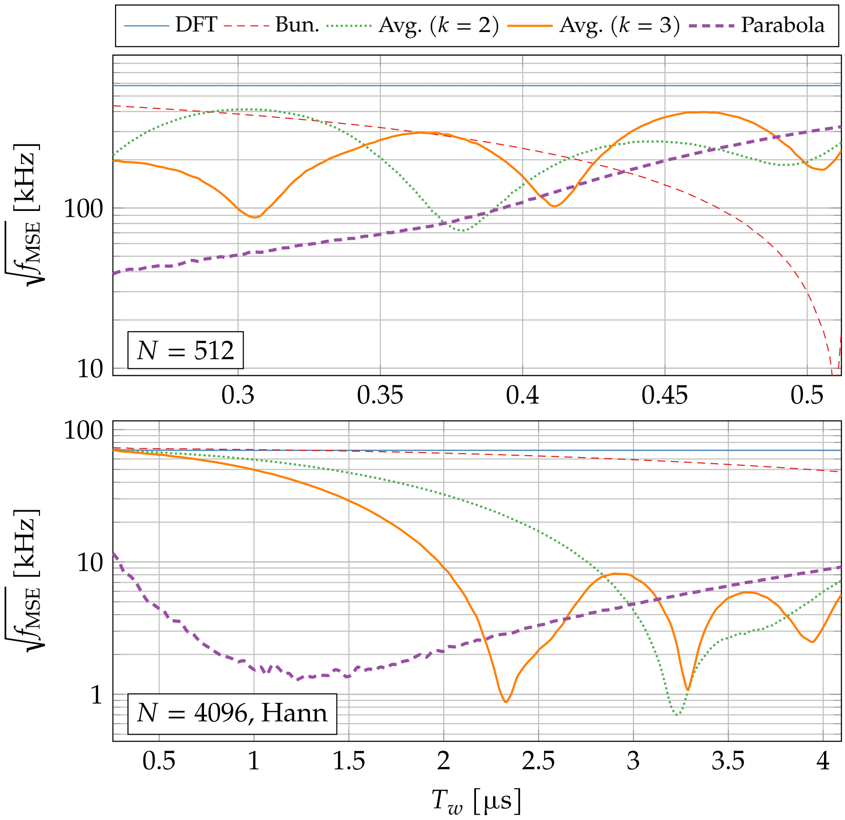


Figure 3.8 – Sub-bin frequency detection methods comparison for different T_w.

the performed DFT, so that very little zero-padding is performed on the signal prior to the DFT. It has to be kept in mind that the DFT will be performed by the FPGA, so it is to be expected that its implementation will pose some flexibility limits, the most common one being the inability to perform DFTs with a length which is not a power of two, so it will not be possible to limit the zero-padding for every beam pulse length. The DFT peak weighted average has good performance when the zero-padding does not exceed the 50 %, but it exhibits a non-linear behaviour, dependent on k. The DFT peak parabolic interpolation instead achieves good results in every conditions when coupled with the Hann window function.

Furthermore, table 3.1 compares the operations that have to be performed in order to apply the selected method to the DFT result. Clearly the Buneman method presents the disadvantage of needing an inverse trigonometric operation that the other two methods don't; the parabolic interpolation of the DFT peak is the cheapest one.

These considerations lead to the choice of the parabolic interpolation as sub-bin resolution frequency detection method, paired with the Hann window.

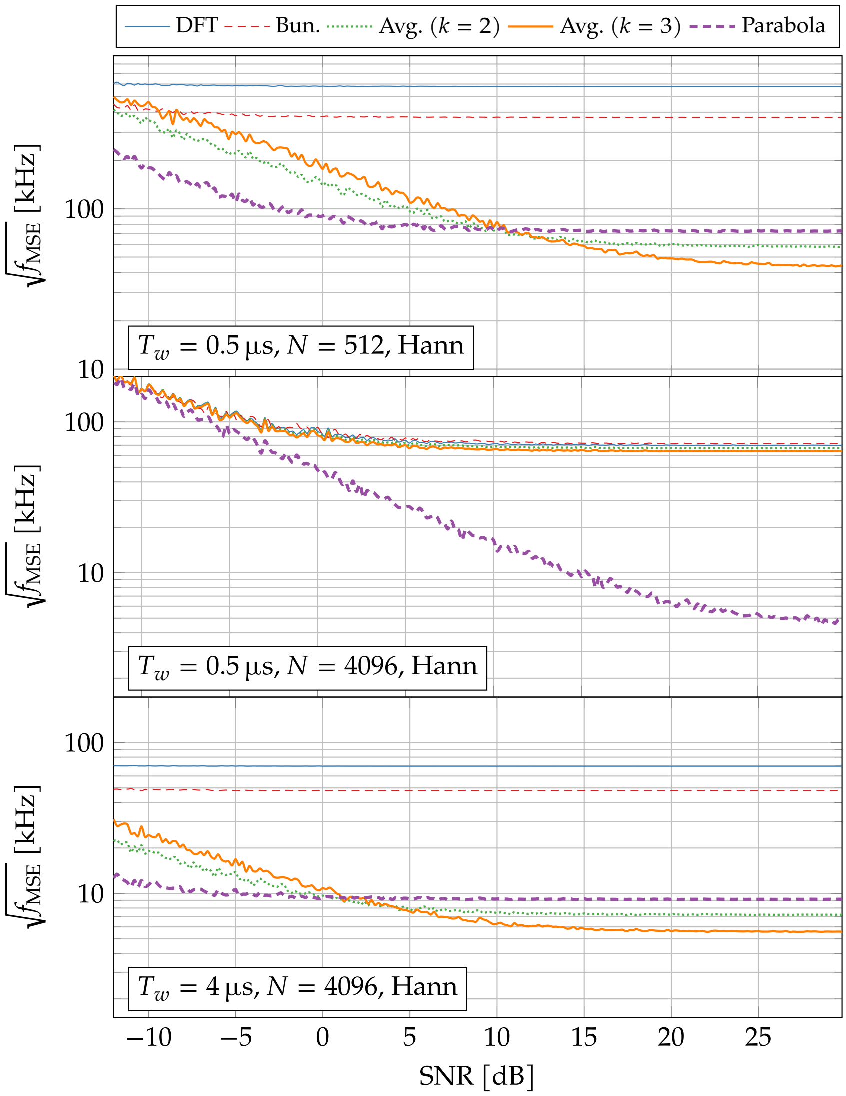


Figure 3.9 – Sub-bin frequency detection methods comparison for different SNR.

Table 3.1 – Sub-bin resolution methods resource comparison. In parentheses the additional operations needed when the DFT length is selected at runtime.

Method	Win.	Add./sub.	Mult.	Div.	Trig.	Inv.Trig.
Buneman		2	1	1	0 (2)	1
Average (k)	Hann	$4k$ (2)	$2k+2$	1 (1)	0 (1)	0
Parabola	Hann	4 (2)	1	1 (1)	0 (1)	0

CHAPTER 4

Processing Implementation & Systems Integration

This chapter discuss the details of the system implementation, from the prototype to the final integration with the LIGHT control system. As such, it cannot avoid describing also the selected digitiser, its features and the associated development kits.

4.1 ADQ14 Overview

Figure 4.1 – ADQ14 interfacing options. From left to right: MicroTCA, PCIe, PXIe, SuperSpeed USB and 10GbE (10 Gigabit Ethernet) [1].

Figure 4.2 shows the main components of the ADQ14 card and how it interacts with the surrounding world. The back of the card is home to the host PC interface, while the front panel hosts the IOs: analog acquisition channels, trigger, synchronisation input, a possible external clock reference and (optionally) a connector for general purpose DIOs. Except for the latter, all of the IOs directly interfaces the internal FPGA, which is also connected to the host PC interface and to an expansion board which manages the direction and status of the GPIOs.

While several, purpose-specific options are available, the digitiser is provided with a default firmware which handles the triggering, the sampling and the transferring of the data to the host. This firmware can be customised at two different points along the acquisition chain:

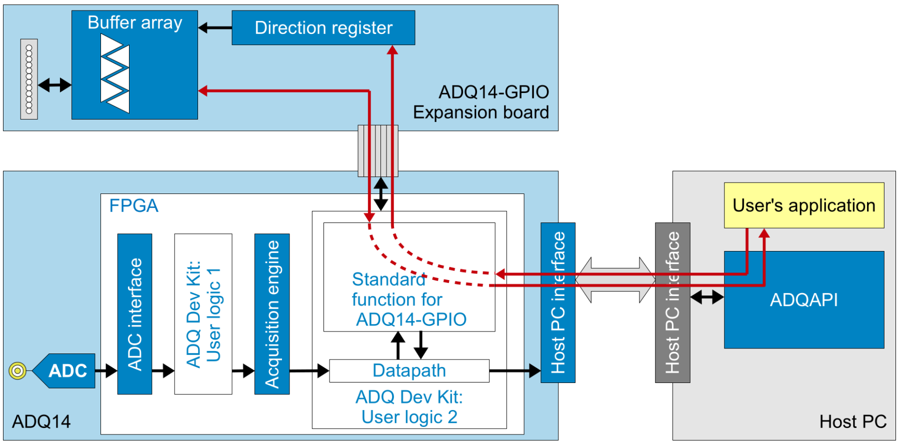


Figure 4.2 – ADQ14 structure, connections and peripherals.

- The *User logic 1* block provides the ADCs samples, a validity bit for every sample based on the analog front-end settings and some information regarding the level trigger, but no information regarding where in time the sample is with respect to the trigger event (i.e. the current sample might be discarded later due to it not belonging to an acquisition record). It is mostly suited for streaming-like data processing, e.g. digital filtering, and for custom triggering requirements.

- The *User logic 2* block provides the output samples of the *User logic 1* block that made it through the more advanced acquisition features (e.g. sample skipping) with additional information regarding how the data has to be chunked as requested by the host. This allows to implement record-like processing, e.g. DFT and the like. Moreover, it allows to interface with the GPIOs, making it possible to control them directly from the hardware.

The onboard FPGA is produced by Xilinx [2], hence the FDK consists of a project in the native format of Xilinx's most recent EDA software suite: Vivado. Vivado is a comprehensive tool which allows to develop, analyse and synthesise HDL designs. The FDK allows to define what's inside *User logic 1* and *2* through the development of custom HDL (either Verilog or VHDL) blocks.

Finally, a software API is provided allowing to communicate and exchange data with the ADQ14. Specifically, the API allows to:

- Scan the host interfaces for connected digitisers, providing the information needed to be able to identify the different units.

- Query the device status for possible hardware faults, firmware errors and configuration parameters.

- Modify various acquisition settings, like the triggering mode and the way data is transferred from the ADQ14 to the host.

- Enable or bypass the custom implementation of the *User logic 1* and *2* blocks.
- Read and write registers inside the custom FW logic. This is very useful to implement host-controlled data exchange between the FPGA and the host which doesn't fit the flow of the ADC samples (e.g. additional acquisition settings or once-in-a-while parameters adjusting).
- Read and write the GPIOs. This obviously has to be coordinated with how the GPIOs are handled by the *User logic 2*.
- Effectively transfer the sampled data from the digitiser to the host in different ways depending on the use case.

The majority of the needed features is already there; what needs to be implemented on top is

- an interface for making the EMS and the PMS available to the accelerator control system,
- an FPGA implementation of the signal processing,
- a way for the host to set the values of non "hardcodable" parameters in the FPGA for the signal processing (e.g. the distances between the probes), and
- a protocol for transmitting the measurement results in real-time through the GPIOs.

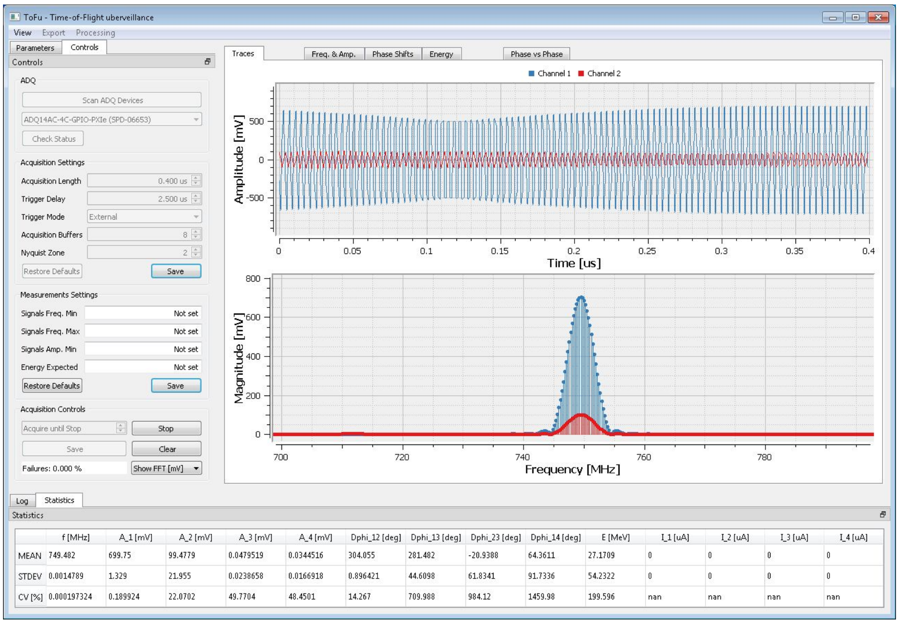
During this thesis all but the latter point have been implemented.

4.2 Standalone Controller Application

Prior to the accelerator control system integration, a standalone application has been developed for testing and tweaking the system: this way the EMS and the PMS can be operated independently from the control system, which is useful in the initial stage of development. It has been chosen to conceive firstly the overall architecture and only later focus on the signal processing so that the problems of implementing the data exchange and the processing can be addressed separately.

As SP Devices API are provided as a C/C++ library, and being the application quite critical in terms of performance, C++ has been chosen as implementation language. The application has to present a discrete amount of data to the user, possibly in a quasi real-time manner, and the most comprehensible way of doing so is through plots: therefore, the graphical aspect of the application is very important. To implement the GUI, the Qt framework has been chosen and coupled with the Qwt library: Qt is a widget toolkit for creating graphical user interfaces for cross-platform applications, while Qwt (Qt Widgets for Technical Applications) is a set of custom Qt widgets and utility classes useful for programs with a technical background. Being very straightforward, Qt allowed to speed up the GUI design (which is not really the gist of this work), and while Qt itself already provides widgets for charts representation, Qwt ones feel more feature complete and mature, so they have been preferred to the former ones.

4.2.1 Overview



ADQ14AC-4C-GPIO-PXIe (SPD-06653)

Acquisition Length: 0.400 us
Trigger Delay: 2.500 us
Trigger Mode: External
Acquisition Buffers: 8
Nyquist Zone: 2

Signals Freq. Min: Not set
Signals Freq. Max: Not set
Signals Amp. Min: Not set
Energy Expected: Not set

Failures: 0.000 %

	f [MHz]	A_1 [mV]	A_2 [mV]	A_3 [mV]	A_4 [mV]	Dphi_12 [deg]	Dphi_13 [deg]	Dphi_23 [deg]	Dphi_14 [deg]	E [MeV]	I_1 [uA]	I_2 [uA]	I_3 [uA]	I_4 [uA]
MEAN	749.482	699.75	99.4779	0.0479519	0.0344516	304.055	281.482	-20.9388	64.3611	27.1709	0	0	0	0
STDEV	0.0014789	1.329	21.955	0.0238658	0.0166918	0.896421	44.6098	61.8341	91.7336	54.2322	0	0	0	0
CV [%]	0.000197324	0.189924	22.0702	49.7704	48.4501	14.267	709.988	984.12	1459.98	199.596	nan	nan	nan	nan

Figure 4.3 – Screenshot of the standalone controller application.

Figure 4.3 shows the GUI of the developed application, which will be briefly described here below.

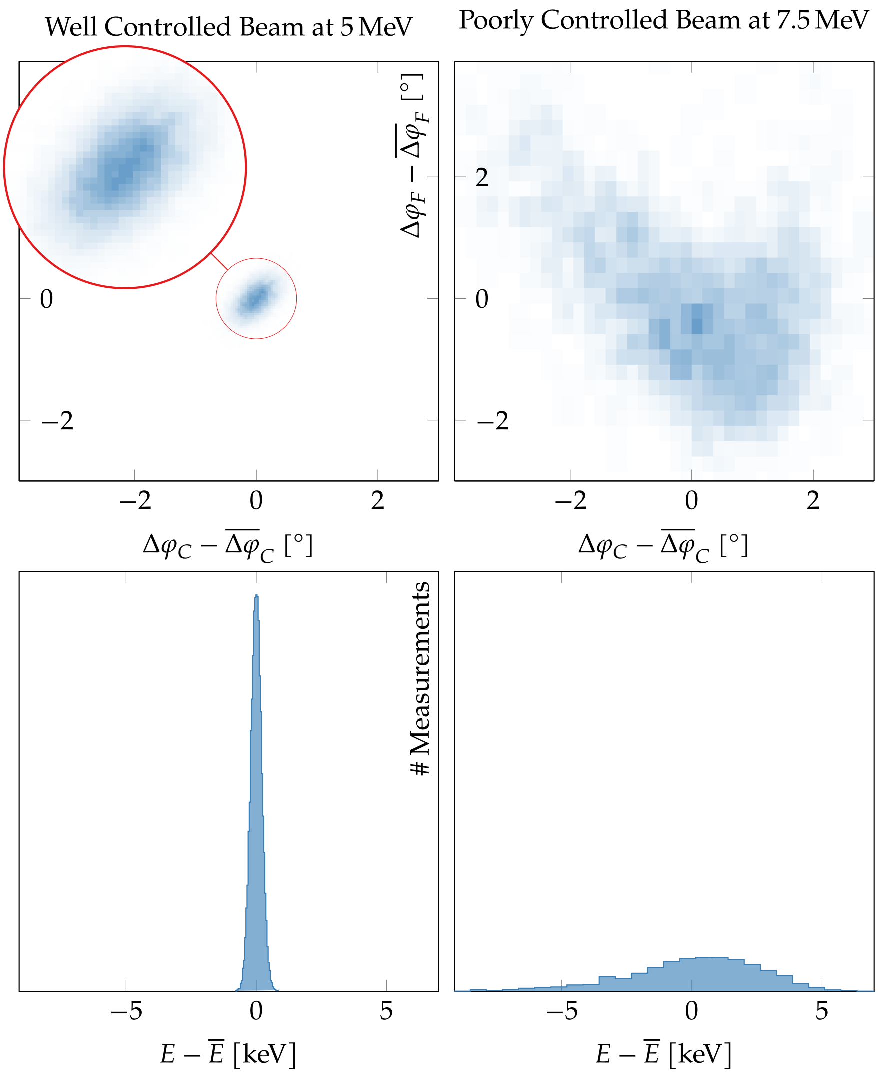
The main area hosts the plots, organized in tabs: the first tab shows the traces as they are acquired by the ADQ and, optionally, their DFT; the second tab shows the measured frequency and amplitudes for each beam pulse; the third tab shows the measured phase shift and the fourth tab shows the measured energy. The fifth tab shows the time elapsed from the trigger event to the availability of the energy information (in the FPGA); when the computation happens in the CPU, this is actually the time taken by the processing, as it is predominant. The last tab shows a plot of $\Delta\varphi_C$ vs. $\Delta\varphi_F$ (the former being the phase shift between the two probes which are closer to each other, the latter being the phase shift between the two probes which are further apart): this can be used to get a rough estimation of how properly the beam is controlled. In fact, both $\Delta\varphi_C$ and $\Delta\varphi_F$ are inversely proportional to the beam β, and thus to the energy:

$$\Delta\varphi_C \propto \frac{L_C}{\beta}; \qquad\qquad \Delta\varphi_F \propto \frac{L_F}{\beta}; \tag{4.1}$$

so if there were no measurements errors on the phases the following would be true:

$$\Delta\varphi_F = \frac{L_F}{L_C} \cdot \Delta\varphi_C - 2\pi k, \qquad k \in \mathbb{N}. \tag{4.2}$$

Two examples of this type of plot can be seen in fig. 4.4. On the left is represented a situation where the beam is well controlled; in fact, the energy variance is very small (less than 1 keV), but still in the top plot a slight correlation between $\Delta\varphi_C$ and $\Delta\varphi_F$ is visible, meaning that the phase resolution of the measurement is better than the energy variance. For the situation on the right instead the energy variance is quite higher, but the correlation between $\Delta\varphi_C$ and $\Delta\varphi_F$ seems to have two contributions, as it is positive for $\Delta\varphi_C - \overline{\Delta\varphi_C} > 0$ and negative when $\Delta\varphi_C - \overline{\Delta\varphi_C} < 0$, so all that variance does not entirely depend on variations of the energy (also the beam intensity varies more when the beam is not well controlled, amplifying the variability of the measured phase shifts).



Well Controlled Beam at 5 MeV

Poorly Controlled Beam at 7.5 MeV

Figure 4.4 – $\Delta\varphi_C$ vs. $\Delta\varphi_F$ measurements with corresponding energy histograms. Notice that the plots at the top have the same scale both in x and y. The same is true also for the x scale of the two energy plots.

On the left, the applications controls are organized in two tabs. The *Controls* tab allows to control the communication with the digitiser, from the ADQ's scanning to the start and stop of the acquisitions. The parameters tab allows to set all the values needed for the energy computation (probes distances, phase shift offsets from the calibration system and the energy range).

At the bottom there is a log window which lists various information about the operations performed by the application and reports possible failures. This window is tabbed with a pane showing some basic statistics for each measured quantity. This stats can be updated based on the area zoomed in the respective plot. Also, each plot can be directly saved as PDF and all the measurement results can be saved in a *csv* file.

4.2.2 Data Exchange

As previously hinted, the default firmware is quite flexible in terms of the ways it allows to transfer the acquired samples from the digitiser to the host. Of all the acquisition modes, the *triggered streaming* has been chosen as most suitable for the purpose. Figure 4.5 shows SP Devices' recommended way of employing this acquisition mode; this has been implemented as task of a dedicated thread with the aim of keeping the application responsive also under intense data exchange.

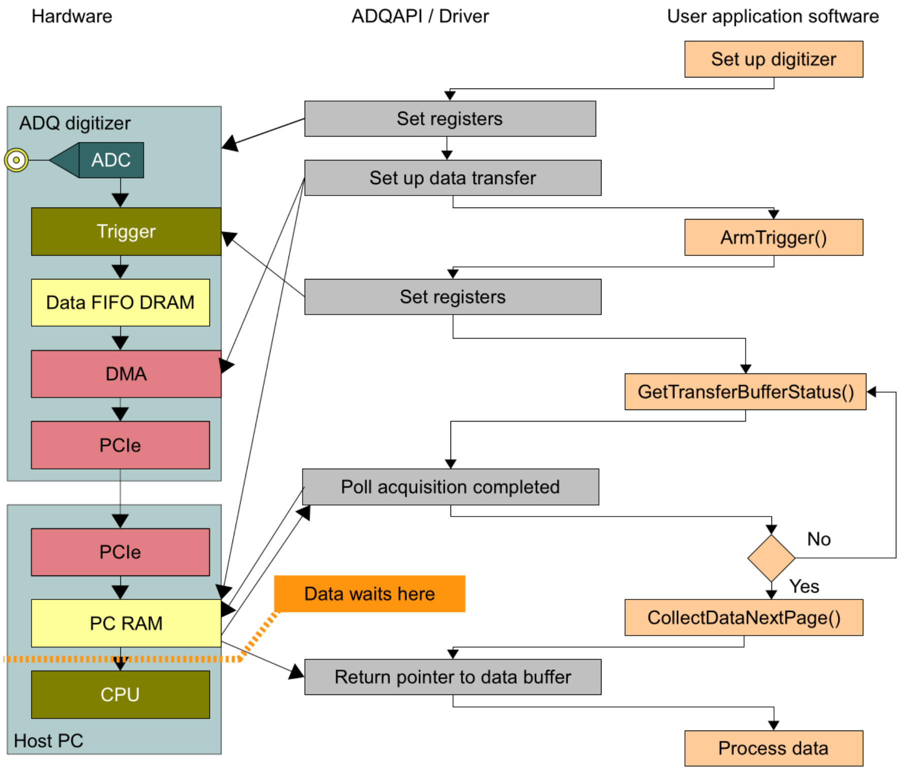


Figure 4.5 – SP Devices' recommended way of implementing the triggered streaming acquisition mode.

When a trigger event happens, a record of a user-set number of samples is acquired, for each user-enabled channel, and sent immediately to the host. The host needs to prepare beforehand some buffers so that the digitiser can fill them directly without requiring CPU time. The CPU needs to be able to empty the buffers at least as quickly as the digitiser fills them but only on average, as the FPGA RAM is used as intermediate buffer. (Although the board also hosts an additional, bigger RAM, it is not used in the *triggered streaming* mode to achieve an optimized data transfer.) The dedicated thread periodically polls the ADQ14 API for new incoming buffers, interprets their content and forwards the relevant information to the main thread, which handles the GUI and the user inputs.

4.2.3 Signal Processing CPU Implementation

For the first prototype of the system, the default ADQ14 firmware without any modifications has been used, so the signal processing has been implemented also on the host side. This will be very useful also later when testing the correctness of the HDL implementation as the latter will be much more difficult to carry out.

To maximise the performance of the application it has been decided to perform the processing in the thread dedicated to the communication with the ADQ; the other option would have been to do so in the GUI thread, requiring to pass all the samples to the GUI thread for every record, which is for each beam pulse. Doing the computation in the ADQ thread allows to add an option by which the user can enable or disable the visualisation of the acquired traces: when disables the samples don't have to be exchanged between the two threads and when enabled this exchange can be limited to a rate of about 60 Hz as these data are used only for visual feedback.

Another optimisation was to allocate all the memory needed for the computation when acquisitions are being started. In fact, the *triggered streaming* acquisition mode requires the number of samples in each record to be set before the acquisitions start; also, it cannot be changed while the acquisitions are being performed. This means that the amount of memory needed for the signal processing can be known in advance and the memory can be allocated once at the beginning of the acquisition. This allows to minimise the number of system calls in the ADQ thread, thus also making its performances more predictable.

As the signal processing to be performed is not too complicated it has been decided to implement it from scratch by using only the C++ Standard Library and none of the publicly available mathematical libraries. This kind of libraries usually aim to cover a wide range of use cases, sometimes sacrificing performance or control granularity, or yet they have a license which is unfriendly to closed-source code. For this reasons, an ad hoc solution seemed more appropriate.

Figure 4.6 shows the performance of the whole processing performed on three ADQ channels. Employing single precision floating-point numbers would not make any performance gain, so double precision will be used as data representation. Extended precision is represented just for reference. This plot shows that at a low pulse rate, the CPU should have no problem keeping up with the data exchange and processing; at 200 Hz instead it is better to stay below the 4000 samples per record. In any case, the performance achieved by

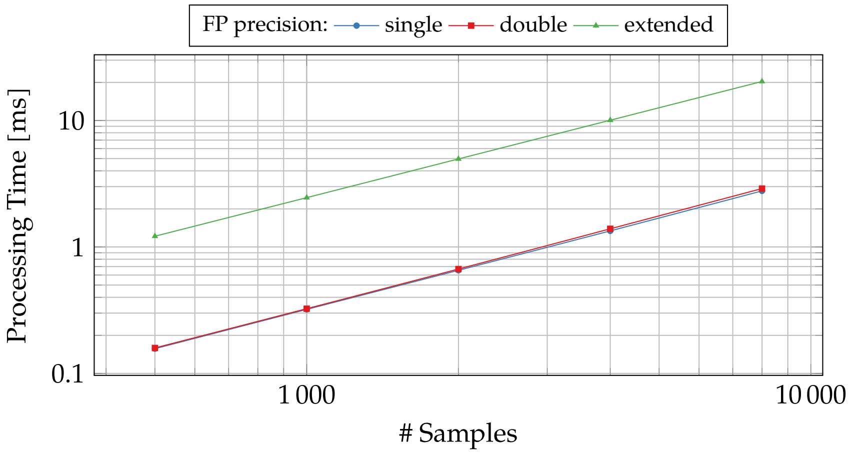


Figure 4.6 – Signal processing performance on the host PC for different floating-point precisions.

this implementation is good enough for testing purposes.

4.3 Signal Processing FPGA Implementation

The actual FPGA implementation of the digital processing that has to be performed on the outputs of the phase probes acquired through the digitizer is a bit more complex than what has been described in the previous chapter; this is mainly because of hardware limitations and non-idealities. The full processing can be outlined as follows:

1. For each channel, compute the mean value of the samples record and subtract it to every sample. This is necessary as the employed digitiser has a non predictable, non zero baseline for each channel which would interfere with the IQ computation.
2. For each channel, perform the spectral windowing.
3. For each channel, append zero valued samples so that the record length matches the required length of the DFT as can be performed by the FPGA.
4. For each channel, translate the complex representation of the DFT result from Cartesian coordinates to polar coordinates.
5. Compute the sum of all the channels' DFTs. This has a smoothing effect on the DFT and allows to achieve better results than, for example, extracting the frequency information from the DFT of the first channel alone.
6. Find the index of the DFT magnitude's peak and perform the parabolic interpolation in its neighbourhood.
7. For each channel, perform the IQ demodulation.

8. For each channel, translate I and Q into amplitude and phase.
9. Compute the energy from the phases of the first three channels and from user-provided parameters (e.g.: the probes distance).
10. For each channel corresponding to a phase probe, check that the amplitude is greater than a user-defined threshold: if this doesn't hold then act appropriately (by now this information is just sent back to the host; in the final implementation this could be used to trigger some more critical actions).
11. For each channel, create a data chunk to be sent to the host comprising the acquired samples, the detected frequency, the amplitude, the phase and the energy.

On top of the *triggered streaming* acquisition mode two more features have been implemented. The first (which is actually more of a workaround than a feature) consists in padding the data to send back with random data to effectively achieve the immediate transmission of the data to the host. In fact, the FW always sends to the host chunks of 512 bytes so, for example, if the record contains 500 samples ≡ 1 kB and the measurement results are 24 bytes then 8 chunks are sent to the host (2 for each channel), but if the samples selected are 600 then part of the data would be transmitted with the next trigger. By padding the data it is ensured that each trigger corresponds to an integer number of data chunks.

The second feature is the possibility of not sending the samples to the host. This obviously implies that the traces are not displayed on the relative plot, but allows to reduce the data rate, especially for longer pulses. When the transmission of the samples is disabled only one chunk per channel is sent, thus achieving a maximum data rate of 400 KiB at 200 Hz of trigger rate. This could be reduced to 100 KiB by "packing" all the channels together; it has been decided to not do so to keep the program logic simple. As the standalone application allows to select which computational unit should perform the computations, obviously when the host is chosen to do so this "optimisation" cannot be performed as the host needs to receive the samples to perform the computation.

To simplify the implementation process, the same firmware will be used for both the EMS and PMS.

4.3.1 Resources Usage

The development of the FPGA processing implementation has been driven more by a desire of ergonomicity than by the will to limit the resources usage, also because the latter requires a lot of experience to achieve.

HDL development has IPs as fundamental way of modularising the design. Fourteen of them have been developed to implement the requested features, some in VHDL and some using the Block Design feature of Vivado, which is useful when the IP has to deal with very simple logic but lots of interconnections:

inner_product This IP computes the inner product between two series of data. It is used for the IQ computation.

iq_channel This IP "packs" together two inner_product IPs to achieve both I and Q computation starting from the samples and from a sine/cosine digital synthesiser.

iq_to_ampphi This IP performs a conversion from Cartesian coordinates to polar coordinates, allowing to translate the I and Q information of an input channel into its amplitude and phase.

my_user_logic_2 This is the top level IP containing all the needed processing. It is designed as a *block design* so as to be easily modifiable.

parabolic_interpolator This IP performs the parabolic interpolation on the DFTs sum.

record_padding This IP contains a simple logic that pads the input data to adapt it to a required length and is used to achieve the instantaneous transmission of the data to the host.

samples_count This is a simple IP which counts the number of samples in the record.

samples_demean This IP subtracts the mean of the samples record to each sample, given a precomputed mean.

samples_mean This IP simply computes the mean of a record of samples.

samples_padder This IP is used to pad the input data to the required length as requested by the DFT IP.

samples_sum This IP computes the sum of all the samples in a record. It is used by the samples_mean block.

tof_energy_computation This IP performs the sequential operations needed to go from the phase shifts to the energy value.

tof_params_storage This IP is a wrapper around the SDK feature of being able to transmit data from the host to the ADQ and is used to transmit the various parameters needed for the energy computation.

user_logic_fsm This IP contains the finite state machine governing the various computation steps, from the samples queueing to the construction of the output record based on the user settings and the computed values.

This modularisation, despite seeming too fine grained, has proven to be flexible enough to easily allow the addition of (minor) features and small modifications.

4.3.2 Correctness Verification & Performance Assessment

Vivado allows early verification of the design functionalities through test benches. These are HDL blocks which have no IOs. Each test bench should instantiate the HDL block to be tested, drive its input signals as to emulate a certain operative condition and verify that its outputs satisfies the requirements.

Using test benches allowed to address implementation bugs early in the development process and to verify the correctness of the implementation, but it is not useful to build statistics about the design performance. This is because running the test bench can take a very long time: the whole processing takes about 100 µs when executed inside the actual FPGA, but running on a PC a simulation of that same operations takes about 10 minutes, so simulating the processing with more than a handful of example signals is not feasible.

A very insightful test which has been carried out is reported in fig. 4.7. This shows mean-difference plots for the detected frequency, amplitude and phase shifts and for the measured energy: for each data point in each plot, the x value is the mean of the CPU and the FPGA results while the y value represents their difference; the red, dashed line has the mean of the y values as ordinate and the grey lines are placed at $\mu \pm 1.96\,\sigma$ of the differences (if the differences distribution were normal, the interval between the grey lines would be equivalent to a 95 % confidence interval). For the frequency, we can see that the two computations basically agree and that the variability of the difference is the same as the variability of the value itself. The situation is similar for the amplitude except for the fact that the differences are not centred in 0 but in 0.05 mV, meaning that the FPGA always underestimates by more or less that amount; this is not really a concern as the amplitude is not a crucial parameter and the bias is quite small. The situation is a bit different for the phase shift because its value spans in a definite range and the whole range should be measurable with the same resolution, and in fact this is the case: the confidence interval for the differences roughly extends from −0.008° to 0.008°, which is one order of magnitude better than the desired measurement resolution. The plot for the energy case presents a very peculiar trend: the deviation of the results of the FPGA from the CPU ones is within a very defined limit, and always positive. This suggests that the difference may be due to a loss of resolution in some of the computations to get from the phase shifts to the energy and in particular from β to E, which involves squares, square roots and reciprocals: as they are performed using fixed-point values, the choice of the precision becomes critical. The precision of every computational step will have to be properly tuned as the results achieved are about one order of magnitude above the desired energy resolution and as such they do not satisfy the medical requirements.

Figure 4.8 shows instead a benchmark of the processing. As it is implemented in hardware, there is no source of fluctuation and thus the performance can be extracted directly from the Vivado simulations. It can be seen that, compared to the results in fig. 4.6, the FPGA processing time is less than one tenth of the CPU processing time; furthermore, the performance is more inde-

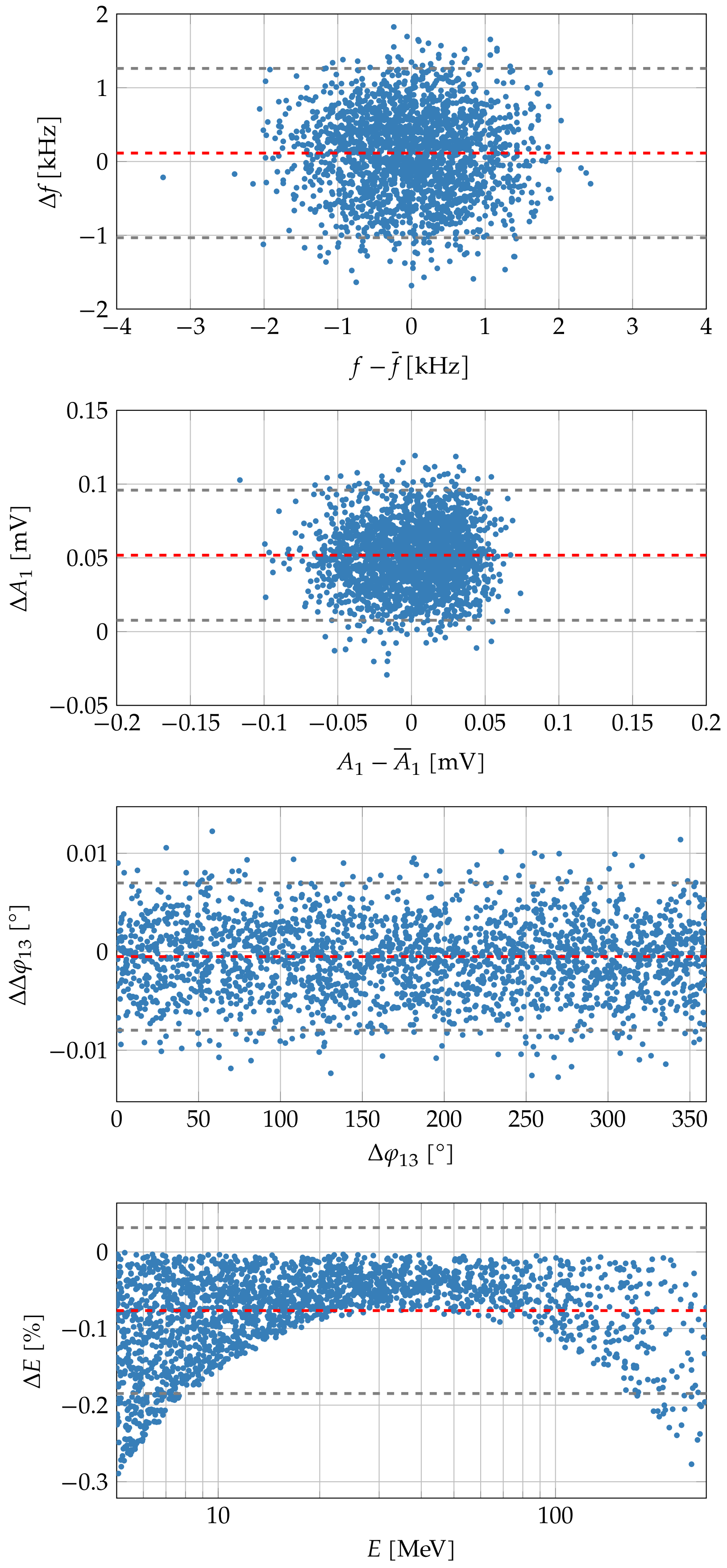


Figure 4.7 – Mean-difference plots comparing CPU and FPGA measurements as performed on the same signals.

pendent of the number of processed samples: this is because the processing time is dominated by the DFT, which expects an input of fixed length (4096 in this case) so shorter records are zero-padded. Also, it has to be noticed that the IQ demodulation is performed sequentially over the channels instead of in parallel; this has been done to reduce the used resources so that the design can fit in the FPGA.

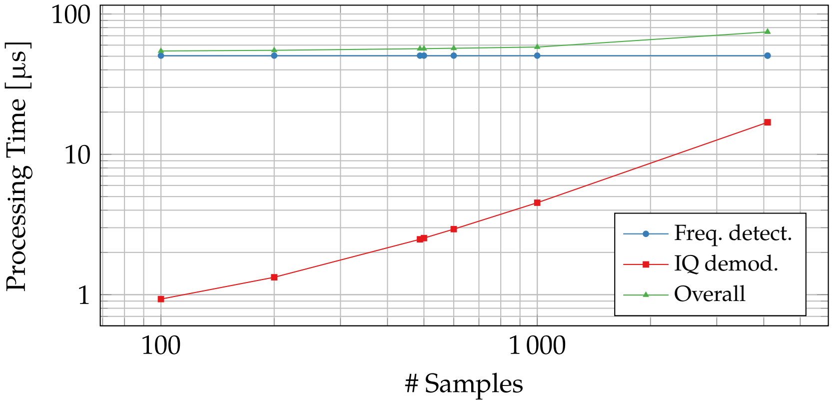


Figure 4.8 – Benchmark of the FPGA implementation of the signal processing.

4.4 Integration With LIGHT's Control System

The standalone application has been developed with just the energy and phase measurements in mind, so it is geared toward ease of use for that operations but it does not actually provide ways to extend its features or to integrate its working with other programs. For this purpose, a separated interface to the ADQ digitiser has been implemented.

LIGHT will be installed in different facilities. As such, the installations will differ in accelerator and beam transfer line layouts, number of treatment rooms (with an optional gantry), facility services, equipment suppliers and equipment versions. Thus, the control system needs to be extensible through configuration and modularisation. To achieve this, the control system relies on a multi-tier architecture with a clear separation between front-end devices and controllers [3]. From the point of view of the control system, the LIGHT prototype currently under commissioning is also the test ground for future installations, especially regarding the low-level control of hardware devices. Therefore the architecture of the current prototype control system, shown in fig. 4.9, can be taken as reference for the final one.

To minimize the time-to-market, the control sub-systems rely mostly on commercially available hardware and software; the diagnostics controller has been implemented in LabVIEW, so the EMS and PMS interface has also been developed in LabVIEW. This interface mainly consists of thin wrappers around the C functions of SP Devices SDK, with added helper functions to decode the data coming from the digitisers like converting the ADC codes to millivolts,

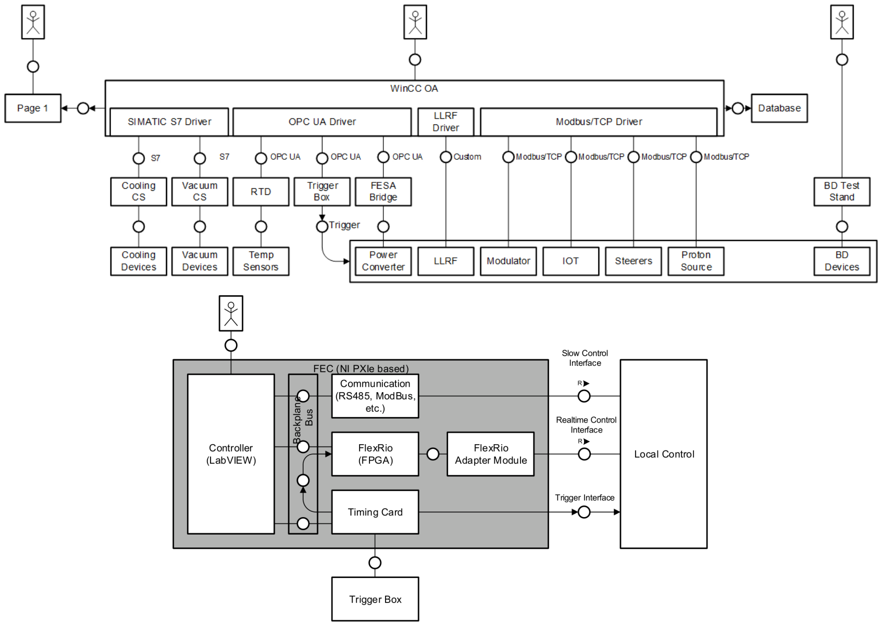


Figure 4.9 – The LIGHT prototype control system architecture with a close-up on the module controlling the beam diagnostics [4].

fixed-point to floating-point conversion, and so on. The implementation of this interface is very similar to the code of the dedicated thread of the standalone application, with the main difference being that the standalone application only has to manage one ADQ while the control system has to manage the entire machine and related diagnostic systems, so the data from the ADQ is polled much less frequently and for this reason the pre-allocated buffers have to be bigger by a couple of orders of magnitude: this is an approximation as the memory usage is not critical and so a significant mark-up can be taken while choosing the size of the pre-allocated buffers.

This interface has exactly the same behaviour as the standalone controller application, so its correctness does not have to be verified. Furthermore, its performance should be evaluated in the context of the whole control system, which is beyond the aim of this thesis.

CHAPTER 5

Systems Characterisation & On-Beam Measurements

To be able to test whether the limits shown in fig. 2.8 can be respected by the realised system, $\delta\Delta\varphi$ has to be estimated. This is not trivial as it cannot be estimated through direct measurements on the beam as it would not be possible to separate the measurement uncertainty from the inherent variability of the measured property. Given the embodiment of the system, one solution is to estimate a relation between the signal-to-noise ratio and the uncertainty on the estimation of the phase shift. This will be addressed in the following section. The section will also test the long-term stability of the system to establish a rate for the EMS calibration. Finally, the last section will illustrate some of the measurements that can be carried out through the EMS and how they are used to improve the accelerator performances.

5.1 Noise & Detection Limits

Figure 3.1 shows an estimation of the output current of the phase probes. This needs to be compared with the noise of the acquisition chain. The first active component on the path of probes signals is the fixed gain amplifier, so the noise it introduces is the most critical and it is the one against which the estimated signal power has to be compared.

The output noise of the amplifier has been measured by short-circuiting its input, connecting its output to the input of another amplifier and the output of this to a spectrum analyser: this was necessary as the power of the output of a single amplifier was below the noise baseline of the spectrum analyser. The result is that the output noise is about $-140\,\mathrm{dBm}/\sqrt{\mathrm{Hz}}$ over the whole spectrum, as represented in fig. 5.1.

To be able to compare it to the signal power it has to be kept in mind that the signal will later undergo the IQ demodulation. From section 3.3.1 it can be recalled that the IQ demodulation can be seen as the computation of a single point of DFT, so it can be deduced that the effect of the IQ computation on the signal is that of a filter. Its transfer function depends on two factors: the windowing function, which for simplicity has not been used (meaning that the

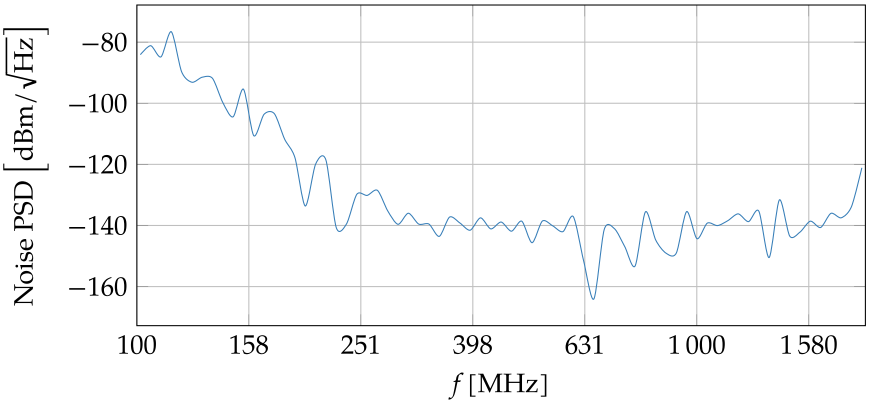


Figure 5.1 – Characterised amplifier output noise power spectral density.

rectangular window is used), and the number of samples. Figure 5.2 shows the magnitude of the IQ transfer function for different time windows. Some of the time windows have been chosen so as to show that a rippling effect arises when the time window is not exactly a multiple of the detected signal period ($((251.52\,\text{MHz})^{-1}$ in this case, which is equivalent to $(f_s - f_{\text{RFQ}})^{-1}$). The typical pulse length is 1 µs but a magnitude like the one for the 0.488 µs case can be used to be more general with respect to the time window (as it is the worse of the considered cases). Furthermore, the fact that the peak of the filter does not reach 0 dB for each time window is just an artefact of the numerical computation: in practice the magnitude will always be around 0 dB at the (detected) signal frequency, so the signal of interest passes unchanged through the IQ.

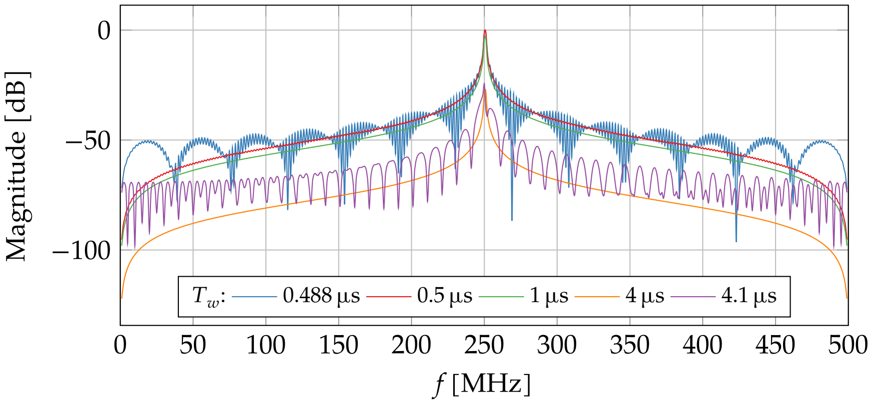


Figure 5.2 – Magnitude of the IQ demodulation transfer function.

The noise baseline has to be referred at the input of the amplifier, by subtracting the gain of the amplifier, and it has to be filtered according to the IQ filter magnitude. By integrating the result over the whole spectrum, the result is that the RMS power of the noise is $I_N \approx 1\,\text{nArms}$. Figure 3.1 shows the peak amplitude, so those values have to be down-scaled by a factor $\sqrt{2}$. The result is that the estimated SNR ranges from 32 to 92 dB for the energies at which debunching does not happen (from 110 MeV on) and from 9 to 69 dB when the beam debunches (below 110 MeV).

Numerical simulations can now be performed to assess the performance of the measurement system for these cases of SNR. These simulations consists in the generation of a sinusoidal signal to which white gaussian noise is added; the σ of this noise is chosen according to a given SNR; the computed signal is sampled at 1 Gsps for a time window of 0.5 µs, then the chosen signal processing is applied to the samples. Figure 5.3 shows the results in terms of f and $\Delta\varphi$ standard deviations, estimated over 10 000 measurements.

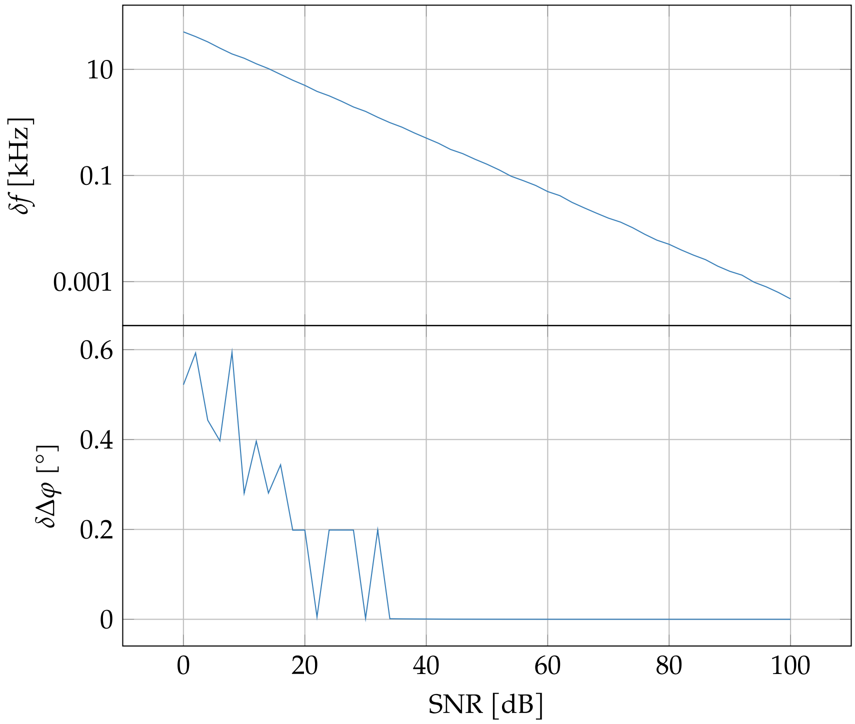


Figure 5.3 – Frequency and phase shift detection limits from numerical simulations.

The frequency trend is as expected: the estimation variability improves as the SNR grows. The phase shift trend instead raises some concerns about its reliability, mainly for two reasons: $\delta\Delta\varphi$ seems to assume definite values (0.6, 0.4, 0.2) with only few points in between and for high SNR it is always 0. Thus, the author's opinion is that another source of information should be used, that is measurements. The very same signal processing has been applied to signals from a voltage generator: in this case the noise is very small, so the output signal power has been measured instead of the SNR. The results are shown in fig. 5.4.

In both cases the frequency trend is similar so it can be used to correlate SNR values from simulations with the generator power from measurements. Comparing δf for the two cases allows to map the power range from −50 to −15 dBm to the SNR range from 0 to 35 dB. The $\delta\Delta\varphi$ values from the measurements can then be fitted for that SNR range so that it can be generalised to

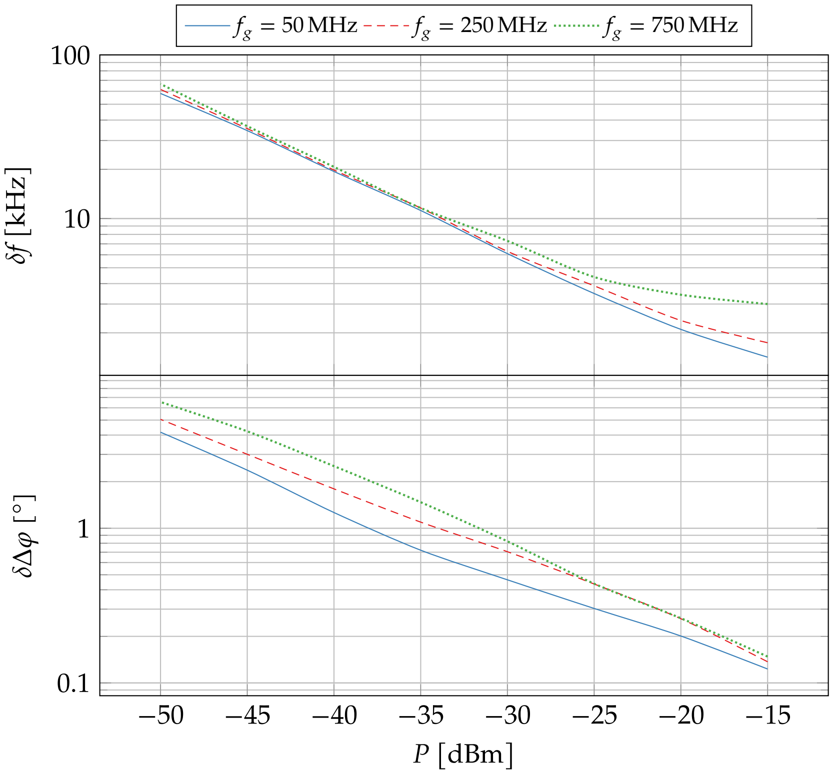


Figure 5.4 – Frequency and phase shift detection limits from measurements.

an arbitrary SNR:

$$\delta\Delta\varphi = 3.692 \cdot \exp(-0.0994 \cdot \text{SNR}). \tag{5.1}$$

This is represented in fig. 5.5.

Putting together the information from figs. 2.8, 3.1 and 5.5, the critical cases can be identified. These depend on the distance between the two furthest probes but values as small as 2 m should be allowed, so from fig. 2.8 $\delta\Delta\varphi$ should be at least 0.2°. Through fig. 5.5 this limit can be translated to having SNR $>$ 30 dB. Finally, fig. 3.1 and noise estimations highlights that the required energy resolution cannot be guaranteed when the beam debunches, as in those cases the signal-to-noise ratio drops below 35 dB.

As last notice, it has to be kept in mind that the real $\delta\Delta\varphi$ for the EMS is actually greater than what is represented in fig. 5.5 by a factor $\sqrt{2}$. In fact, every EMS measurement consists of two phase shift measurements: the calibration one and the pulse one.

5.2 Long-Term Drifts

To understand how often the system has to be calibrated it is useful to know if there are any drifts in the signals parameters. To investigate this, long-running measurements have been performed in three different conditions:

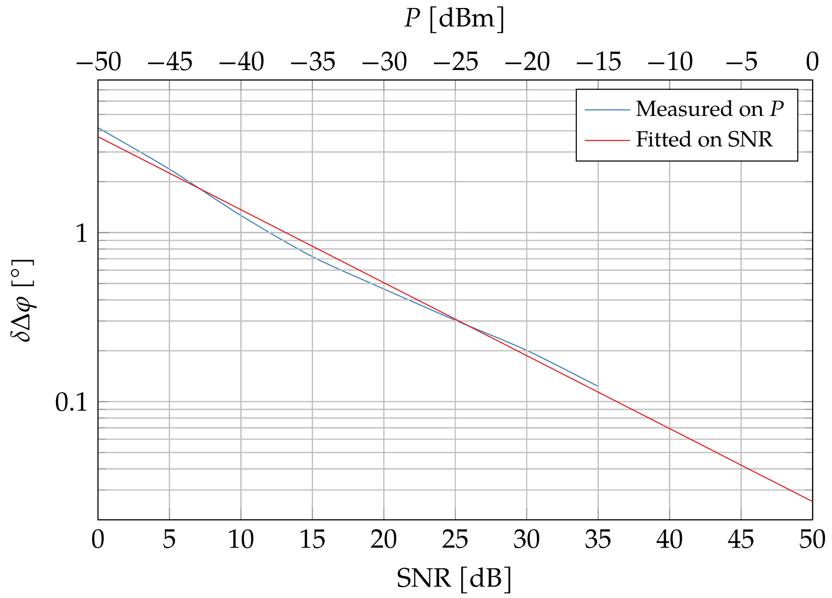


Figure 5.5 – Estimated $\delta\Delta\varphi$ from measurements.

1. using a voltage generator and the down-mixing stage;
2. with the generator again but without down-mixing;
3. with the VCO used for the calibration without down-mixing.

The fourth case, which is using both the VCO and the down-mixing, has not been measured as it is superfluous to measure both the sources of variability combined when they have been measured separately (although it might have been useful as a check for the other measurements). The hardware setup is identical to what is represented in fig. 3.4, except that the input signal comes from the mentioned source instead of the phase probes. The measurements are shown in fig. 5.6, while the statistics of interest are reported in table 5.1.

Table 5.1 – Summary statistics for long-run characterisation.

Quantity	Statistic	Case 1	Case 2	Case 3
$f - \bar{f}$ [kHz]	Minimum	−68.4	−2.70	−11.8
	1st Quartile	−24.8	−0.48	−2.0
	Median	2.1	0.00	0.0
	3rd Quartile	22.2	0.48	1.4
	Maximum	65.3	2.96	32.1
$\Delta\varphi - \overline{\Delta\varphi}$ [°]	Minimum	−0.523	−0.599	−0.125
	1st Quartile	−0.077	−0.096	−0.022
	Median	0.000	−0.001	0.000
	3rd Quartile	0.077	0.095	0.022
	Maximum	0.465	0.602	0.116

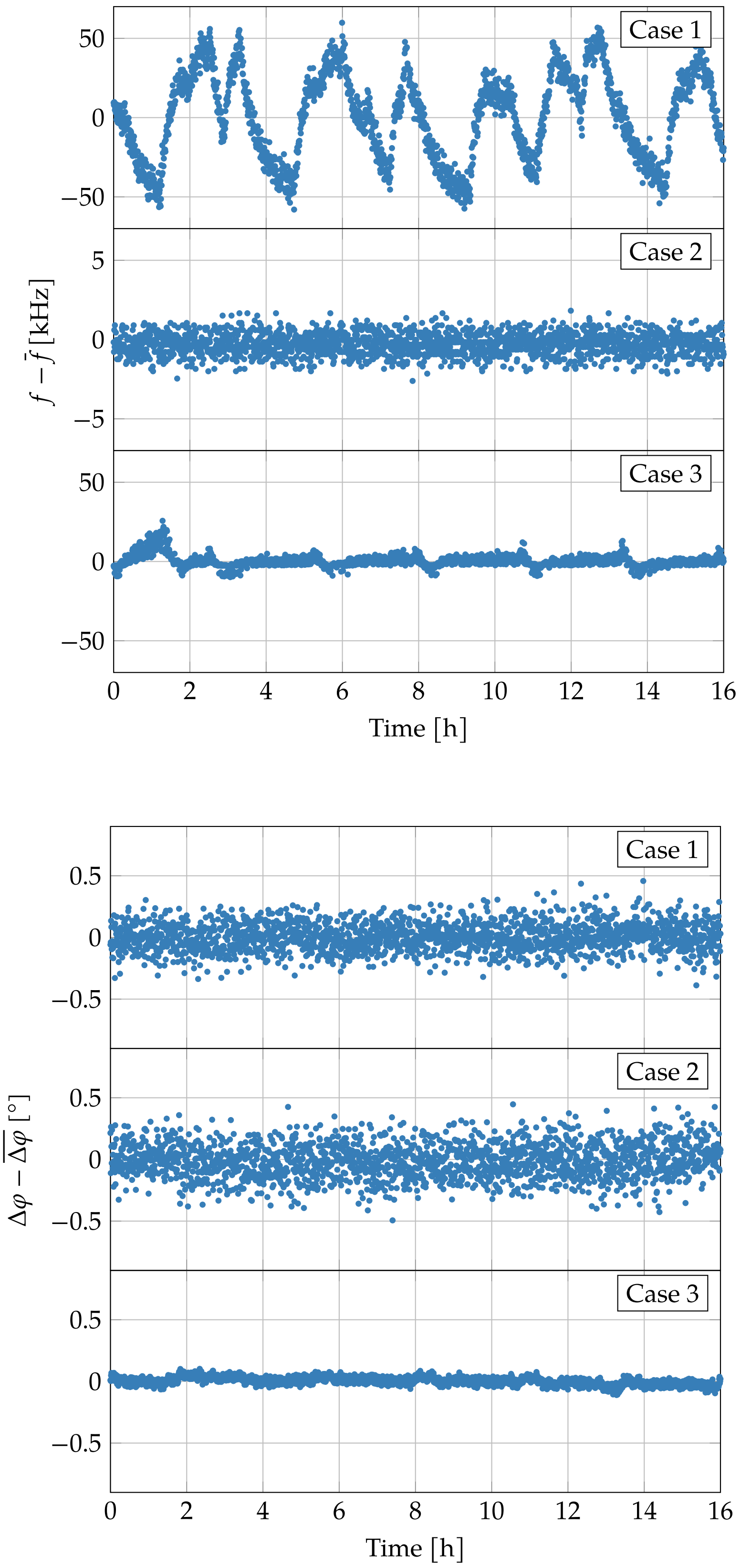


Figure 5.6 – Long-run measurements for investigating long-term drifts.

The best way to analyse the results is by considering the two cases with the generator first and the VCO case at a later moment. From the former ones it can be seen that the mixer represents a great source of variability for the frequency; however, this variability is limited and do not constitutes a drift in the long term. Furthermore, this seems to have no influence on the measured $\Delta\varphi$, which for both cases is very stable; the two phase shift measurements also have a very similar standard deviation. Because of this, the choice regarding whether to perform or not the down-mix has to be based on the attenuation introduced in the two cases. The evaluation of the attenuation introduced by the long cables at different frequencies will be performed once the exact setup (and thus, cable model) will be known; as an example, the model used during LIGHT's prototype commissioning introduces an attenuation of $2.4 \cdot 10^{-2}$ dB/m at 50 MHz and $9.5 \cdot 10^{-2}$ dB/m at 750 MHz, which over 30 m of length give 2.2 dB of difference in the introduced attenuation.

Unsurprisingly, a different situation is depicted when the VCO (used for the calibration functionality) is employed instead of a generator. It has to be noticed that the signal power is much greater in this case, because the generator was set to reproduce a situation similar to a beam acquisition while the VCO output power cannot be adjusted. Consequently, in accordance with the conclusions of the previous section, in this case the dispersion of the phase shift measurements is smaller (0.14° vs. 0.03°). The phase stability achieved by the calibration signal is comparable to what could be achieved using a signal generator, and it is almost one order of magnitude smaller than the desired phase accuracy. It can be concluded that the calibration feature has been, hardware-wise, adequately implemented.

5.3 On-Beam Measurements

The developed energy measurement system has been extensively used for the accelerator commissioning since its first working prototype. To fully characterise the beam properties, three different test benches have been designed by ADAM's beam diagnostics team: the first to measure the proton source output; the second used from the end of the RFQ until the last SCDTL module; the third to characterise the CCL modules [1–3]. As an example, the second test bench is depicted in fig. 5.7. The EMS is present in both the second and third test benches; from 5 to 37.5 MeV it is used alongside a spectrometer, while starting from 37.5 MeV there is not enough room in the bunker to host the spectrometer: a Multi-Layer Faraday Cup (MLFC) is employed instead. These two devices have very different characteristics and capabilities with respect to the system developed for this thesis, but they allow nonetheless to have another source of information regarding the beam energy. Furthermore, at the time the spectrometer was available, only the standalone controller application had been developed and the integration of the MLFC in the control system would happen only after the submission of this thesis, so it was not possible to compare the energy measurements from different detectors on a pulse-by-pulse basis, but only their statistics over a certain amount of pulses.

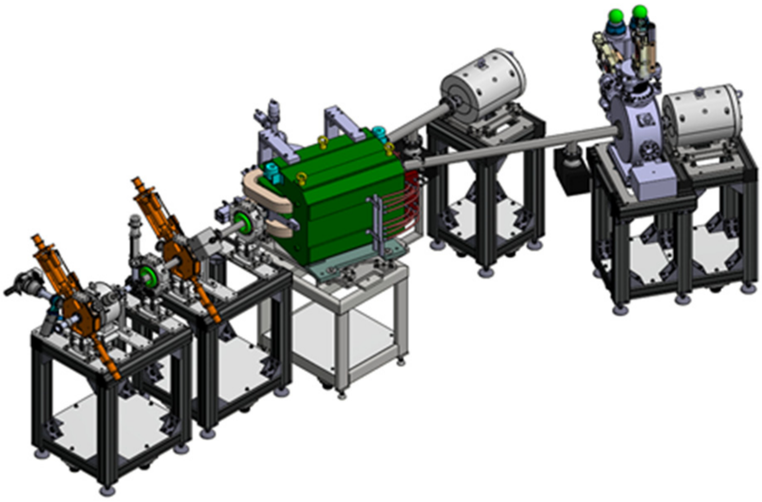


Figure 5.7 – Test bench designed for the characterisation of RFQ and SCDTLs. Beam enters on the left. Phase probes are placed before the dipole. Reproduced from [1].

Unfortunately, neither the spectrometer nor the MLFC can be used as absolute reference concerning the energy measurements. The spectrometer didn't allow for single shot measurement: the centre of mass of the beam in the spectrometer branch was computed by reconstructing the beam profile over several pulses. Moreover, no instrument was employed to measure the trajectory of the beam before entering the dipole, so the results of the energy measurements have a very poor accuracy. Regarding the multi-layer Faraday cup, the exact thickness and composition of its vacuum window were not known, so it was not possible to estimate accurately the energy loss given by the window. Because of this it has not been possible to verify the absence of offsets in the EMS measurements, which however seems extremely unlikely given that the system can be calibrated to cancel out possible phase offsets.

The EMS has been used from the RFQ output (5 MeV) all the way up to the maximum energy currently achieved by LIGHT, which is 52 MeV. The majority of these measurements were taken as part of a commissioning procedure whose main goal was not the stabilisation of the beam energy, and thus with an unstable beam. For example, *phase scans* have the aim to find a reference phase shift between the accelerating cavities such that the final beam energy is maximised; when scanning this phase shift, the beam is very stable only at the reference phase, while moving from that value produces very spread energy measurements. Furthermore, the current beam instabilities are greater than the resolution of the system, so it would be wrong to estimate the system accuracy from said measurements.

Nonetheless, it is useful to compare MLFC and ToF to understand what the latter can achieve compared to the former (apart from the obvious, which is the non-destructiveness of the measured beam pulse). Figure 5.9 compares sets of 2000 measurements for four different configurations of the accelerator.

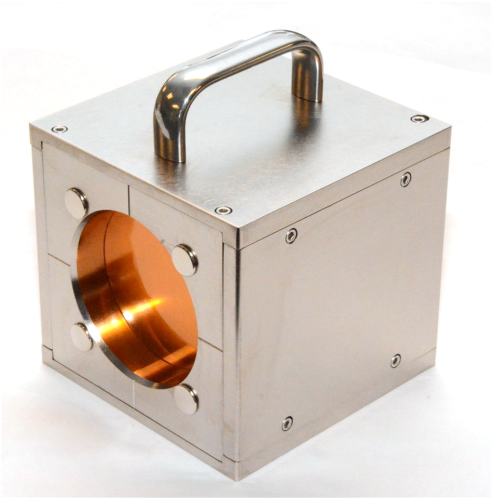


Figure 5.8 – The Multi-layer Faraday cup used in the third test bench of LIGHT.

For both the instruments, the measurements of weak beam pulses have been discarded.

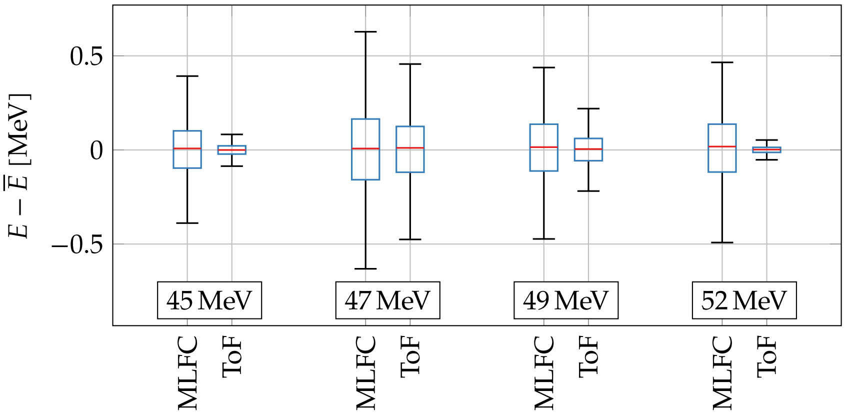


Figure 5.9 – Measurements comparing MLFC and ToF.

As the ToF measures the mean energy of a pulse, also for the MLFC the mean energy of each measurement has been computed. Figure 5.9 depicts box plots of such mean energies. Also, in light of the question regarding the vacuum window, the box plots have been centred around 0 MeV as comparing their means would have been pointless. These box plots allow to see that the ToF resolution makes it possible to distinguish the cases in which the beam is well controlled from the cases in which it is not, while the MLFC doesn't reach the required resolution.

Conclusions

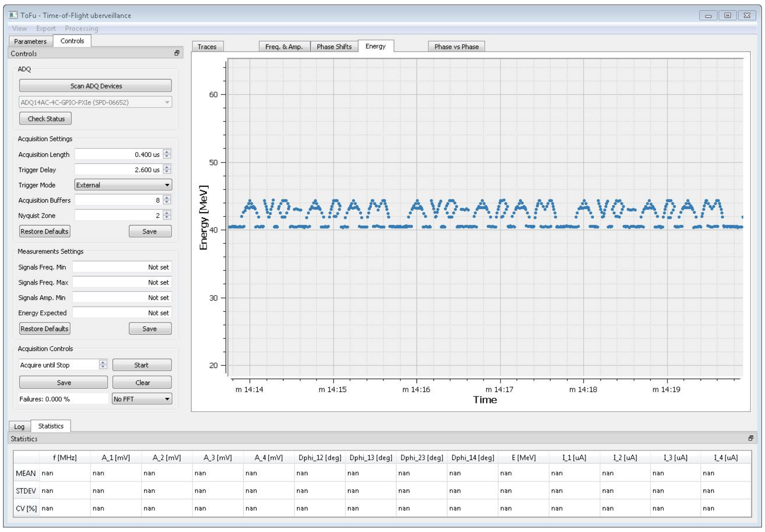
This thesis presented the development of low latency, non-interceptive, phase-based beam diagnostic systems for LIGHT, the first linear accelerator prototype for proton therapy.

The design of these systems proved to be considerably challenging, requiring both theoretical and in-depth technical knowledge in various subjects, from electronics to signal processing to FPGA and system programming.

This thesis has significantly contributed to the design and implementation of LIGHT's energy measurement system. Specifications have been produced about the conditions under which it is able to fulfil all the requirements for a medical conforming implementation, which have been found to be achievable in practice but, as per the current implementation, not for all the energies. Concerning the analog acquisition chain, various possibilities have been considered and finally narrowed down; to maximise the resolution of the measurement some aspects of the acquisition chain (most notably the down-mixing) could potentially be decided per-installation. Choosing a commercial digitiser allowed to limit the cost of the system. The signal processing has been carefully picked and tweaked to meet the medical requirements in terms of measurement resolution while at the same time achieving a processing time compatible with the 200 Hz measurement rate, enabling the treatment of moving organs by volumetric rescanning with tumour tracking.

The energy measurement system has also been instrumental for the accelerator commissioning. During all the commissioning stages it has been extensively used and it has been taken as the best source of information regarding the beam energy, which for this thesis is gratifying but also slightly inconvenient as a cross-check with a more mature measurement system would have been beneficial in boosting the confidence regarding its accuracy.

Even though spurious results can be produced by the energy computation algorithm, experience has shown that this results from the beam being poorly controlled upstream of the detectors. The fact that a greater instability of the beam is reflected in a greater variability in the energy measurement results allowed to employ it as a general diagnostic tool for detecting defects in the machine setup. Still, one of its major features is the fact that it can produce, for each pulse, both the energy and the amplitudes of the signals from which it has been computed, so that these amplitudes are a direct indication about the reliability of the energy measurement. Energy measurements with a standard deviation as small as 10 keV have been performed, so the resolution of the developed system is at least that small. Also, the system has been proven to perform measurements at a rate as high as 100 Hz, as shown in fig. 5.1, with tests planned for 200 Hz.



ADQ14AC-4C-GPIO-PXIe (SPD-06652)

Acquisition Length 0.400 us
Trigger Delay 2.600 us
Trigger Mode External
Acquisition Buffers 8
Nyquist Zone 2

Signals Freq. Min Not set
Signals Freq. Max Not set
Signals Amp. Min Not set
Energy Expected Not set

Acquire until Stop
Failures: 0.000 %
No FFT

	f [MHz]	A_1 [mV]	A_2 [mV]	A_3 [mV]	A_4 [mV]	Dphi_12 [deg]	Dphi_13 [deg]	Dphi_23 [deg]	Dphi_14 [deg]	E [MeV]	I_1 [uA]	I_2 [uA]	I_3 [uA]	I_4 [uA]
MEAN	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
STDEV	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
CV [%]	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan

Figure 5.1 – ToF standalone controller application showing the outcome of active energy modulation performed at 100 Hz.

Some of the presented aspects need either improvements or further developments. The estimation of the phase probes output signal power will have to be checked by comparing the measurements on the current accelerator setup with simulations, to verify that the estimations for the final setup are correct. This will be mandatory for a proper selection of the amplification stage of the acquisition chain. It was not possible to verify this during this thesis as the accelerator commissioning was still in an early stage.

The employment of the down-mixing stage will have to be decided based on the attenuation introduced by the cables used to carry the signals from the accelerator bunker to the back-end electronics rack. This decision will probably be taken on a per-installation basis, given that the length of these cables could be very different for different installations, such that some of them may justify the power loss introduced by the mixers to prevent a greater attenuation from the cables.

The FPGA firmware is the element of the system which needs more attention. The number of bunches between the two probes which are further apart is sometimes miscomputed; an extensive study will have to be carried out to discover which of the energy computation spurious results are due to inaccuracies in the instrument and which, instead, are due to instabilities in the beam or to improper settings of the accelerator.

Finally, upon the definition of a protocol for exchanging data through the GPIOs, the firmware will have to be augmented with the ability to receive the energy of the next beam pulse and to communicate back the measurement outcome to the accelerator control system.

Bibliography

References for Chapter 1

[1] World Health Organization. *Cancer*. URL: https://www.who.int/en/news-room/fact-sheets/detail/cancer.

[2] National Cancer Institute. *What Is Cancer?* URL: https://www.cancer.gov/about-cancer/understanding/what-is-cancer.

[3] National Cancer Institute. *Diagnosis and Staging*. URL: https://www.cancer.gov/about-cancer/diagnosis-staging.

[4] National Cancer Institute. *Treatment for Cancer*. URL: https://www.cancer.gov/about-cancer/treatment.

[5] National Cancer Institute. *Chemotherapy to Treat Cancer*. URL: https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types/radiation/internal-radiation-therapy-brachytherapy.html.

[6] National Cancer Institute. *External Beam Radiation Therapy*. URL: https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types/radiation/external-beam-radiation-therapy.html.

[7] National Cancer Institute. *Internal Radiation Therapy (Brachytherapy)*. URL: https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types/radiation/internal-radiation-therapy-brachytherapy.html.

[8] M. R. Horsman and J. Overgaard. "The impact of hypoxia and its modification of the outcome of radiotherapy". In: *J. Radiat. Res.* (Aug. 2016).

[9] U. Amaldi. "Acceleratori per la salute". In: *asimmetrie* 6 (2008), p. 21. URL: http://www.roma1.infn.it/rog/astone/didattica/asimmetrie6.pdf.

[10] EE Klein. "Electron-beam therapy: dosimetry, planning, and techniques". In: *Perez and Brady's principles and practice of radiation oncology, 5th edition. Philadelphia, PA: Lippincott Williams & Wilkins* (2008), pp. 190–217.

[11] Jason K. Rockhill and George E. Laramore. "Chapter 20 - Neutron Radiotherapy". In: *Clinical Radiation Oncology*. Fourth Edition. Elsevier, 2016. ISBN: 978-0-323-24098-7. DOI: https://doi.org/10.1016/B978-0-323-24098-7.00020-4.

[12] Radhe Mohan et al. "69 - Proton Therapy". In: *Leibel and Phillips Textbook of Radiation Oncology*. Third Edition. Philadelphia: W.B. Saunders, 2010, pp. 1493–1510. ISBN: 978-1-4160-5897-7. DOI: https://doi.org/10.1016/B978-1-4160-5897-7.00071-8.

[13] Joseph R. Castro et al. "70 - Carbon Ion Radiotherapy". In: *Leibel and Phillips Textbook of Radiation Oncology (Third Edition)*. Ed. by Richard T. Hoppe, Theodore Locke Phillips and Mack Roach. Third Edition. Philadelphia, 2010, pp. 1511–1522. ISBN: 978-1-4160-5897-7. DOI: https://doi.org/10.1016/B978-1-4160-5897-7.00072-X.

[14] Marco Durante and Jay S. Loeffler. "Charged Particles in Radiation Oncology". In: *Nature Reviews Clinical Oncology* 7 (2009). DOI: https://doi.org/10.1038/nrclinonc.2009.183.

[15] H. Suit et al. "Proton vs Carbon Ion Beams in the Definitive Radiation Treatment of Cancer Patients". In: *Radiotherapy and Oncology* 95.1 (2010), pp. 3–22. ISSN: 0167-8140. DOI: https://doi.org/10.1016/j.radonc.2010.01.015. URL: http://www.sciencedirect.com/science/article/pii/S0167814010000678.

[16] Ion Beam Applications (IBA). *World class proton therapy technology - IBA proton therapy*. URL: https://iba-worldwide.com/proton-therapy/proton-therapy-solutions.

[17] C. Ronsivalle et al. "The TOP-IMPLART Linac: Machine Status and Experimental Activity". In: *Proc. of International Particle Accelerator Conference (IPAC'17)* (Copenhagen, Denmark). International Particle Accelerator Conference 8. Geneva, Switzerland: JACoW, May 2017, pp. 4669–4672. ISBN: 978-3-95450-182-3. DOI: https://doi.org/10.18429/JACoW-IPAC2017-THPVA090. URL: http://jacow.org/ipac2017/papers/thpva090.pdf.

[18] P. Berra, U. Amaldi et al. "Study, Construction and Test of a 3 GHz Proton Linac-Booster (LIBO) for Cancer Therapy". In: *Proceedings of EPAC 2000* (Vienna, Austria). JACoW, 2000, pp. 2495–2497. URL: https://accelconf.web.cern.ch/accelconf/e00/PAPERS/WEP4A13.pdf.

[19] J. B. Farr, A. M. Kolano and A. Degiovanni. "Investigation on FLASH therapy using a high frequency linac for protons". In: (Manchester, England). PTCOG. June 2019.

[20] Y. Zhang et al. "An evaluation of rescanning technique for liver tumour treatments using a commercial PBS proton therapy system". In: *Radiotherapy and Oncology* 121.2 (2016), pp. 281–287. DOI: https://doi.org/10.1016/j.radonc.2016.09.011. URL: http://www.sciencedirect.com/science/article/pii/S016781401634333X.

[21] D. Ungaro, A. Degiovanni and P. Stabile. "LIGHT: A Linear Accelerator for Proton Therapy". In: *Proc. of North American Particle Accelerator Conference (NAPAC'16), Chicago, IL, USA, October 9-14, 2016* (Chicago, IL, USA). North American Particle Accelerator Conference 3. Geneva, Switzerland: JACoW, Jan. 2017, pp. 1282–1286. ISBN: 978-3-95450-180-9. DOI:

https://doi.org/10.18429/JACoW-NAPAC2016-FRB1IO02. URL: https://jacow.org/napac2016/papers/frb1io02.pdf.

[22] A. Perrin et al. *TRAVELv4.07: User Manual*. CERN Internal Notes. 2007.

[23] H. Owen, A. Lomax and S. Jolly. "Current and Future Accelerator Technologies for Charged Particle Therapy". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 809 (2016), pp. 96–104. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2015.08.038. URL: http://www.sciencedirect.com/science/article/pii/S0168900215009729.

References for Chapter 2

[1] P. Forck, P. Kowina and D. Liakin. "Beam Position Monitors". In: *CAS - CERN Accelerator School: Course on Beam Diagnostics*. 2009. URL: http://cds.cern.ch/record/1213277.

[2] G. De Michele et al. "Commissioning Status of the LIGHT Development Machine". In: *Proc. 29th Linear Accelerator Conference (LINAC'18), Beijing, China, 16-21 September 2018* (Beijing, China). Linear Accelerator Conference 29. Geneva, Switzerland: JACoW Publishing, Aug. 2018, pp. 352–354. ISBN: 978-3-95450-194-6. DOI: doi:10.18429/JACoW-LINAC2018-TUPO013. URL: http://jacow.org/linac2018/papers/tupo013.pdf.

[3] C. Jamet et al. "Phase and amplitude measurement for the SPIRAL2 Accelerator". In: *DIPAC 2009 - 9th European Workshop on Beam Diagnostics and Instrumentation for Particle Accelerators*. 2009, pp. 327–329.

References for Chapter 3

[1] C. Jamet et al. "Phase and amplitude measurement for the SPIRAL2 Accelerator". In: *DIPAC 2009 - 9th European Workshop on Beam Diagnostics and Instrumentation for Particle Accelerators*. 2009, pp. 327–329.

[2] National Instruments. *Buneman Frequency Estimator VI*. URL: https://zone.ni.com/reference/en-XX/help/371361R-01/lvanls/bun_frequency_estimator.

[3] Bernd Bischl, Uwe Ligges and Claus Weihs. "Frequency estimation by DFT interpolation: a comparison of methods". In: 2009.

[4] Paul Voglewede. "Parabola approximation for peak determination". In: *Global DSP Magazine* 3.5 (2004), pp. 13–17.

[5] Ronald Newbold Bracewell. *The Fourier Transform and Its Applications*. Vol. 31999. McGraw-Hill New York, 1986.

[6] Enrico Rubiola. *Phase Noise and Frequency Stability in Oscillators*. Cambridge University Press, Dec. 2008. ISBN: 9780521886772.

References for Chapter 4

[1] Teledyne SP Devices. *Home - Teledyne SP Devices*. URL: https://www.spdevices.com.

[2] Xilinx. *Kintex-7 FPGA Family*. URL: https://www.xilinx.com/products/silicon-devices/fpga/kintex-7.html.

[3] R. Moser and H. Pavetits. "The LIGHT Control and Interlock Systems". In: *Proc. of International Conference on Accelerator and Large Experimental Control Systems (ICALEPCS'17)* (Barcelona, Spain). International Conference on Accelerator and Large Experimental Control Systems 16. Geneva, Switzerland: JACoW, Jan. 2018, pp. 543–547. ISBN: 978-3-95450-193-9. DOI: https://doi.org/10.18429/JACoW-ICALEPCS2017-TUPHA064. URL: http://jacow.org/icalepcs2017/papers/tupha064.pdf.

[4] R. Moser et al. "Control and Interlock Systems for the LIGHT Prototype". In: *Proc. of International Conference on Accelerator and Large Experimental Control Systems (ICALEPCS'17)* (Barcelona, Spain). International Conference on Accelerator and Large Experimental Control Systems 16. Geneva, Switzerland: JACoW, Jan. 2018, pp. 1683–1687. ISBN: 978-3-95450-193-9. DOI: https://doi.org/10.18429/JACoW-ICALEPCS2017-THPHA130. URL: http://jacow.org/icalepcs2017/papers/thpha130.pdf.

References for Chapter 5

[1] A. Jeff et al. "A Diagnostic Test Bench for the LIGHT Accelerator". In: *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018* (Vancouver, BC, Canada). International Particle Accelerator Conference 9. Geneva, Switzerland: JACoW Publishing, June 2018, pp. 1808–1811. ISBN: 978-3-95450-184-7. DOI: doi:10.18429/JACoW-IPAC2018-WEPAF001. URL: http://jacow.org/ipac2018/papers/wepaf001.pdf.

[2] V. A. Dimov et al. "Beam Commissioning of the 750 MHz Proton RFQ for the LIGHT Prototype". In: *Proc. 9th International Particle Accelerator Conference (IPAC'18), Vancouver, BC, Canada, April 29-May 4, 2018* (Vancouver, BC, Canada). International Particle Accelerator Conference 9. Geneva, Switzerland: JACoW Publishing, June 2018, pp. 658–660. ISBN: 978-3-95450-184-7. DOI: doi:10.18429/JACoW-IPAC2018-TUPAF002. URL: http://jacow.org/ipac2018/papers/tupaf002.pdf.

[3] G. De Michele et al. "Commissioning Status of the LIGHT Development Machine". In: *Proc. 29th Linear Accelerator Conference (LINAC'18), Beijing, China, 16-21 September 2018* (Beijing, China). Linear Accelerator Conference 29. Geneva, Switzerland: JACoW Publishing, Aug. 2018, pp. 352–354. ISBN: 978-3-95450-194-6. DOI: doi:10.18429/JACoW-LINAC2018-TUPO013. URL: http://jacow.org/linac2018/papers/tupo013.pdf.