L. Fabris

Novel Readout Design Criteria for SiPM-Based Radiation Detectors

ISBN: 9788897413134

UNIVERSITÀ DEGLI STUDI DI BERGAMO

NOVEL READOUT DESIGN CRITERIA FOR SiPM-BASED RADIATION DETECTORS

Lorenzo Fabris
doctoral thesis

Tesi di Dottorato in Tecnologie per l'energia e l'ambiente

L. Fabris

Novel Readout Design Criteria for SiPM-Based Radiation Detectors

ISBN: XXX-XX-XXXXX-XX-X

UNIVERSITÀ DEGLI STUDI DI BERGAMO

NOVEL READOUT DESIGN CRITERIA FOR SiPM-BASED RADIATION DETECTORS

Lorenzo Fabris
doctoral thesis

Tesi di Dottorato in Tecnologie per l'energia e l'ambiente

UNIVERSITÀ DEGLI STUDI DI BERGAMO
Dipartimento di Ingegneria e Scienze Applicate

DOTTORATO DI RICERCA
IN
TECNOLOGIE PER L'ENERGIA E L'AMBIENTE

XXVIII ciclo

anno 2015

NOVEL READOUT DESIGN CRITERIA FOR SiPM-BASED RADIATION DETECTORS

Doctoral thesis:
Lorenzo Fabris

Supervisor:
Prof. Valerio Re

Dipartimento di Ingegneria e Scienze Applicate. Università degli studi di Bergamo
ISBN XXX-XX-XXXXX-XX-X 9788897413134
Open access copy available at: [illegible]

Abstract

The advancements in integrated silicon technology over the last decades have enabled significant progress in all fields of engineering and science, and radiation detection is no exception. Several large-scale experiments have been made possible by the development of new ways of making circuits ever smaller, with good resistance to damaging radiation. Detectors, especially those for high-energy physics experiments, have also seen benefits from the ability of processing large silicon substrates into viable particle trackers. On the other hand, devices such as photomultiplier tubes (PMT) have been used for three quarters of a century and are still the workhorse of radiation detection in many applications. Only in the last four to five years, a new device has become attractive in overcoming some undesired shortcomings of PMTs, thanks to these advancements in silicon integrated technology. The new devices, called silicon-photomultipliers, or SiPM in short, seem to have all of the features of PMTs without the same constraints: large bulk, need for high voltage bias, sensitivity to magnetic fields and lack of ruggedness.
For these reasons, SiPMs tend to be considered as direct replacements for PMTs. This has been demonstrated true for some applications where the amount of signal is very large, however, when the signal becomes smaller, such as in large-area experiments or when observing weak interactions, the analogy breaks down in such a way that SiPMs have to be treated much differently than PMTs.

It has become obvious that different paradigms have to be defined in all those cases where the classic analogy between the two devices is not appli-

cable, and a boundary has to be determined in order to identify regions of applicability of the different models.
This work is meant to explore for the first time what happens when SiPMs are utilized for applications where the analogy with PMTs breaks down, and to supply a theoretical basis to generalize the observations made.

It is not expected that any reader will have strong knowledge of radiation detection topics, thus the first part of the manuscript will focus on supplying the basic knowledge needed to understand the concept presented later on. The main concepts will be introduced, followed by an overview of detectors. SiPMs will then be introduced and their mechanisms examined. From this examination, their analogy to PMTs will become clear, and so will the application limits of the analogy. A new boundary for the classic analogy will be introduced, based on the amount of signal to be detected in a given application, and a new theoretical approach to SiPM-based instrumentation will be presented.
Finally, the newly developed theory will be applied to circuit design in two extreme cases: one well within the classic boundary, the other outside of the boundary. It will be shown that the theory is applicable in both cases and that, while in the extreme case the theory completely solves an otherwise unsolvable problem, it leads to important considerations also within the classic approach boundary, as far as circuit design is concerned.

Keywords: Light Readout, Silicon Photomultipliers, Radon Detection, Health Physics Instrumentation, Environmental Remediation, Nuclear Physics Instrumentation.
PACS: 28.41.Rc, 29.90.+r, 07.50.-e, 07.50.Ek, 07.88.+y

Acknowledgments

The idea behind this work came to me thanks to my involvement in the nEXO collaboration. I can certainly say I would know much less about SiPM-based instrumentation if it were not for Dr. Mike Heffner of Lawrence Livermore National Laboratory and Prof. Giorgio Gratta from Stanford University who invited me to join the collaboration as responsible manager for the electronic instrumentation. Giorgio also graciously offered to proof read the nEXO instrumentation chapter to make sure I am not misrepresenting the collaboration's goals and the experiment's physics, so, special thanks to him. Of course I owe the entire collaboration as well for the many discussions had on the topic, especially to Prof. Liang Yang with the Illinois University, my co-manager, and Mr. Gerard Visser from Indiana University for designing and testing with me some of the circuit solutions. I am in debt to the Brookhaven National Laboratory folks (Drs. Veljko Radeka, Gianluigi De Geronimo and Shaorui Li) for the concerted initial effort that ultimately lead to most of this work.
To Benedetta Nodari at the University of Bergamo, goes my deepest gratitude for designing the radon readout board and most of the circuit beyond the front-end section, as well as doing many of the system measurements reported. She has been instrumental in the success of that system.
Such a system would not have been viable without the detector: thanks go to my colleague Dr. Zane Bell with the Oak Ridge National Laboratory for selflessly providing his ZnS(Ag) screens and his insights to the project. Big thanks also to the rest of my colleagues at ORNL for supporting me during the duration of the PhD program. Along the same lines, I must point out that the whole endeavor would not have been possible if it were not for the

ORNL Laboratory Director's fund that provided me with the funding necessary for research sabbatical leave during most of the doctorate program. The radon detection system development is supported and motivated by ST-Microelectronics through the University of Bergamo.
My thanks to Prof. Valerio Re for his support of this work, and to his colleagues for their hospitality, both in the office and in the lab.

Finally, I want to thank my wife and kids for their moral support and understanding during the academic and experimental phases of the work and for forgiving the frequent, related travel and absences. Same goes for my parents who, for some reason, felt it was necessary for them to provide for the academic fees, despite the fact that I do have an income on my own.

Disclaimer: Parts of this research were sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U. S. Department of Energy.

Contents

Chapter 1

Radiation and Matter

1.1 Introduction

In general, radiation is defined as the transport of energy by moving subatomic particles or by waves that occurs in a medium or in vacuum. Such motion implies energy transport: for waves it is easy to visualize the energy transported in relation to the wave amplitude; for particles, in terms of momentum. The simplest way to characterize radiation is to think about it as propagating waves of a certain frequency, and categorize it based on the frequency itself. The Planck-Einstein equation however:

$$E = h\nu$$

where h is the Planck constant, shows an interesting relationship between frequency and an energy E. This energy is not the actual energy related to the amplitude of the waves, as indicated above; it is rather a more fundamental type of energy. If we think about the de Broglie wave-particle duality, we realize immediately that this is the energy corresponding to the elementary particle (*photon*) that describes the wave in the quantum mechanics formalism. Therefore, we can characterize radiation via either its frequency or the energy associated to its radiation quanta. It should

also be noted that frequency and wavelength (λ) are related to the speed of light, c, by the equation:

$$\lambda\nu = c$$

Sometimes the use of wavelength in lieu of frequency or photon energy is preferred. The wavelength is expressed in meters and the frequency in Hz, but for the energy is preferable to use a non-SI unit named electronvolt (eV), which corresponds to the amount of energy gained by one electronic charge accelerated by an electric potential of 1 V and equals about 1.6×10^{-19}J.

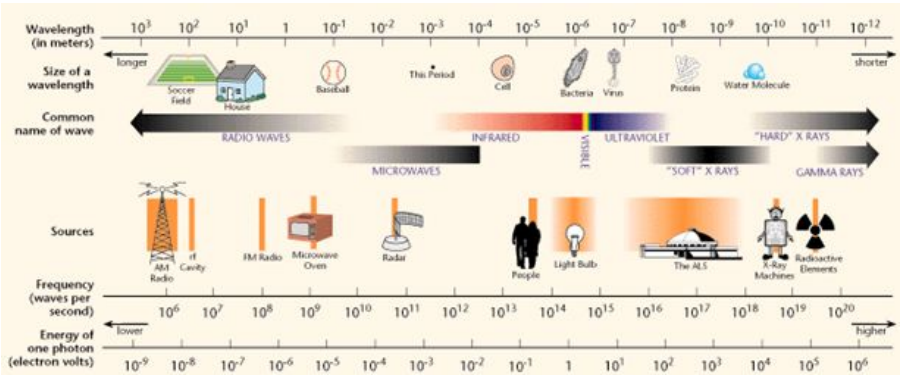


Figure 1.1: Radiation spectrum for most commonly encountered types if radiation.

In theory, the ways of characterizing radiation are completely interchangeable, but for practical uses one is preferred over the other. For example, when radiation is viewed as wave propagation, such as radio waves, the wavelength is the preferred description method, as it directly illustrates the relationship between the radio wave and its propagation in the atmosphere or in wave guides of certain sizes; however, when designing filters, amplifiers or receivers, the use of frequency is dominant. In optics, it is often convenient to visualize radiation as quanta of energy (photons) but it is equally important to visualize radiation as propagating waves. On

the other hand, in high-energy physics, where is more convenient to visualize the radiation as elementary particles interacting via forms of scattering with other particles, energy is preferred.

As mentioned previously, the quantum of energy transported by radiation can vary by several orders of magnitude. Figure 1.1 shows an approximate classification of radiation as a function of its descriptive parameters [1]. This range is known as *spectrum* of the radiation, more often called electromagnetic spectrum, because in this particular case, the associated particle is a photon; the radiation associated with particles with finite rest mass (atomic particles such as electrons, protons, neutrons etc.) is not indicated on this graph. This particle radiation will be discussed separately later on.

In the context of radiation detection, the preferred method of characterization for radiation is through its energy.

1.2 Ionizing Radiation

One important aspect of radiation is its ability of interacting with matter. Depending on the energy content, radiation may be absorbed by the medium it travels, it may not interact with it at all, or may alter the ionization state of atoms, or even transmute them. Radio waves and microwaves will generally exchange energy with a medium mostly via heat or molecular or atomic resonance and the atomic structure of the medium will remain unaltered (unless the heat triggers chemical reactions). At higher energy (higher end of visible light/ultraviolet and above), there is an increasing chance that electrons may be dislocated from their orbit, as the energy transported by the radiation quanta becomes comparable in magnitude to the binding energy of the electrons. In some case, the energy is such that atomic particles can be displaced from the atoms nuclei or even new particles/anti-particles may be observed. When this happens, the corresponding radiation is categorized as *ionizing radiation*, as opposed

to *non-ionizing radiation* for all other cases. The observation, characterization and quantification of ionizing radiation constitute the core of radiation detection.

1.2.1 Radiation Interactions with Matter

The best way to introduce the interaction of ionizing radiation with matter, is to focus on the specific particle: alpha or other heavy charged particles, electron, photon or neutron. Each of these particles has its own unique characteristics; therefore it is intuitive to assume that each particle will exhibit a unique way of interacting with matter. For example, alpha particles and other heavy charged particles are massive (at least 4 nucleons) and possess more than an elementary charge; electrons are relatively light and also have one elementary charge; photons have no charge and no mass. Finally, neutrons have relatively large mass and no charge. By examining the interactions of these particles (or the ones that are most commonly encountered), we can efficiently cover all range of interactions.

1.2.2 Alpha and Heavy Charged Particles

For this family of particles, the interaction occurs primarily via Coulomb force between their positively charged nuclei and the negatively charged orbitals of the media's atoms. As a particle passes a nearby atom, the electrons feel the attractive force of the heavy ion. The electrons absorb energy from the ion and, depending on the amount of this exchange they can be promoted to a different energetic level or completely expelled from the orbital shells. The heavy ion will leave behind either an excited or an ionized atom. As a result, the ion is decelerated. Since not all of the energy can be exchanged at once[1], the ion will undergo several more interactions until it is stopped. The average distance traveled by the ion is called the *stopping distance* of the medium or *range* of the ion in the material. The

[1]For an ion of mass m and energy E, the maximum energy that can be transferred to an electron of mass m_0 is $\frac{4Em_0}{m}$

trajectory of the ion is generally a straight line, as its large momentum does not allow for appreciable changes in trajectory.

An important parameter from these observations is the particle range, as it is an indication for how long a given particle beam travels in a given material. This is directly correlated, for example, to the location where most damage can occur, and it is not just a property of heavy charged particles, but it is rather a parameter common to most radiation interactions. One can experimentally determine the range for charged particles in different materials by comparing the number of particles n_{in} incident on a material (in this case called absorber) of given thickness t with the number of particles n_{out} surviving the passage through the absorber for different values of thickness. The situation is shown in Figure 1.2. An important application of this method is in determining shielding and detector geometries.

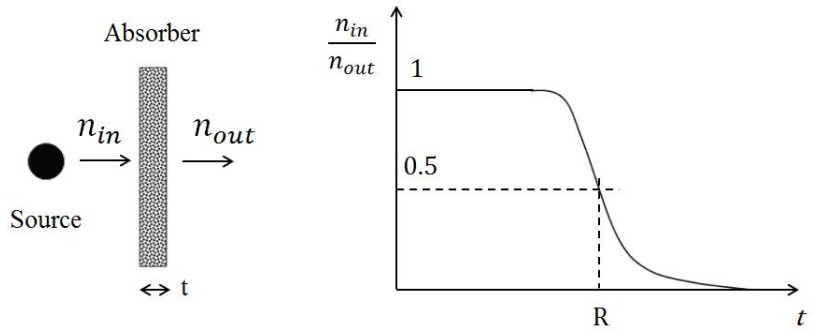


Figure 1.2: Simplified experimental determination of particle range in materials.

The second noteworthy parameter, related to the range, is the energy absorption or stopping power, that is, the energy lost by the ions per unit length along their path, $\frac{dE}{dx}$. The classic expression describing the stopping power is the Bethe formula [2]:

$$-\frac{dE}{dx} = \frac{4\pi e^2}{m_e}\left(\frac{ez}{v}\right)^2 NZ\left[\ln\left(\frac{2m_e v^2}{I}\right) - \ln\left(1-\frac{v^2}{c^2}\right) - \frac{v^2}{c^2}\right]$$

In this expression, m_e is the rest mass of the electron, v, z, and E are respectively the velocity, charge and energy of the ion, N and Z the number density and the atomic number of the absorber, e the electronic charge and c is the speed of light in vacuum. I is an experimental parameter representing the average excitation potential of the absorber. The first thing to notice is that, for very high energy (velocity $\approx$ speed of light), above hundreds MeV, the formula yields, in normalized form, 2 MeV g/cm^2, independent of the specific particle type. These are the so-called Minimum Ionizing Particles (MIP), always present in cosmic-rays. When the energy is lower, the terms in square brackets are slowly varying functions of velocity and the behavior can be inferred by analyzing the multiplicative factor alone: for non-relativistic particles, the stopping power varies as $1/v^2$ (or roughly as $1/E$). For same-energy particles, the charge is the other param-

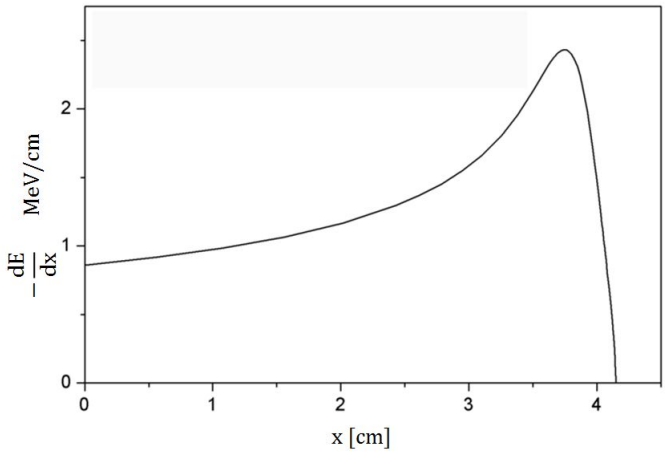


Figure 1.3: Typical Bragg curve for 5.49 MeV alpha particles in air. The Bragg peak occurs at about 3.8 cm.

eter that can change. In this case we see that particles with higher charge exhibit larger energy losses.
An interesting phenomenon, not directly visible in the Bethe formula, is found for highly charged particles: along most of the path within the absorber Bethe's formula applies (losses proportional to $1/v^2$), however, as the particles lose energy they tend to pick up electrons and sharply slow down, therefore the energy loss curve falls sharply. By plotting the stopping power vs. track length, one can, in these cases, identify a sharp peak indicative of this behavior. (These types of curves are known as Bragg curves).
A good example of a Bragg curve for 5.49 MeV alpha particles in air is shown in Figure 1.3. The peak, known as Bragg peak, is dependent on the absorber and, obviously, the particle characteristics. The most important observation about the Bragg peak is that at the peak there is the highest energy transfer between the particles and the absorber. This allows precise delivery of energy within a given material. This behavior is exploited in nuclear medicine for cancer radiation therapy.

1.2.3 Electrons (Beta-rays)

The main difference between electrons and heavy charged particles is their charge (lower and of opposite sign for electrons) and their mass. It is obvious, based on this, to expect a different behavior when electrons interact. Electrons will not lose energy as easy as more charged particles, because the Coulomb interactions will be rarer, but when they do interact, they can lose a large fraction of their energy, and their path can be altered. Given their smaller mass, there is also another process through which electrons can lose charge through radiative effect, known as *bremsstrahlung*. This is a known effect from classic particle theory [3]. The expression for energy loss via bremsstrahlung is the following:

$$\left(\frac{dE}{dx}\right)_b = \frac{NZ\left(Z+1\right)Ee^4}{136m_e^2c^4}\left(4\ln\left(\frac{2E}{m_ec^2}\right)-\frac{4}{3}\right)$$

From the formula, it is evident why the process is negligible for heavier particles (m^{-2} dependence).

An absorption curve obtained experimentally, shows important differences

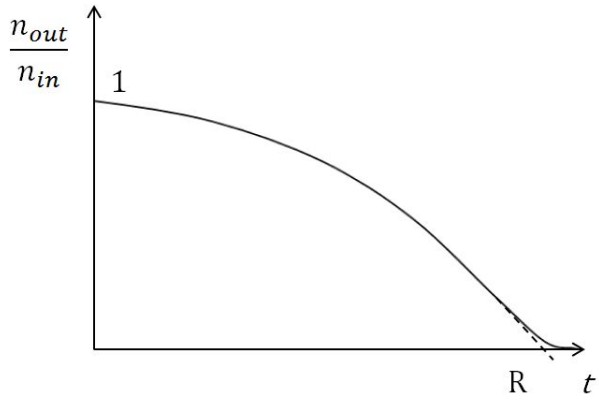


Figure 1.4: Typical electron absorption curve and range extrapolation.

compared to that of heavy charged particles: as the electrons enter the absorber, they are immediately scattered off-axis, therefore a typical curve for electrons does not show a well-defined region where most of the absorption takes place[2], instead, it shows an energy continuum. It is not obvious how to extrapolate the range for electrons; the most common method is to extrapolate the absorption curve to the intersection with the horizontal axis. An example is shown in Figure 1.4. In this case, the range is interpreted as the thickness of absorber needed to stop the electrons.
Overall, for the same energy, the electron range is greater than that of heavy charged particles in any material. For example, the range of 1 MeV alpha particles in air is about a centimeter, while 1 MeV electrons in air

[2]From simple reasoning on the Bethe equation one should also notice that electrons do not exhibit a Bragg peak, another justification for not having a sharp edge in the absorption curve.

travel for about 1.2 m. When electrons are emitted via decay of radioactive species, they are called beta rays.

1.2.4 Photons (X- and gamma-rays)

In contrast to heavy charged particles and electrons, photons have no mass and no charge. There are no Coulomb interactions possible for photons, therefore the way they interact with matter depends only on their energy, E_p, and the atomic characteristics of the media. In radiation detection, the photon's origin referred to when determining the name for the corresponding radiation. Photons emitted by atomic shells are called X-rays, while photons emitted by nuclei are called gamma-rays. X-rays have energies between a few 100 eV to several 100 keV; gamma-rays have energies from a few 100 keV to several MeV, with an overlap between high energy X-rays (also known as hard X-rays) and the low end of the gamma-ray range. The predominant interactions photons have with matter are of three types: *photoelectric*, *Compton* and *pair production*. In photoelectric absorption photons interact directly with the atoms in the media. When an interaction occurs, the photon is completely absorbed by the atom's orbiting electrons and its energy is in part acquired by the atom, and in part transferred to an electron (known as *photoelectron*) that is ejected by the atom. The energy balance for this process can be written as:

$$E_p = E_{bind} + E_{electron}$$

with E_{bind} and $E_{electron}$ respectively the photoelectron binding energy in the atom and the energy it appears with when freed from the original atom. The interaction leaves an ionized atom behind. Usually, this ion recombines with a free electron almost instantly, with another photon (usually an X-ray of energy equal to the electrons binding energy in the specific material) emitted in the process. An approximation for describing the probability of occurrence of this process is the following [4]:

$$P\,(\text{photoelectric}) \approx \text{constant} \cdot \frac{Z^n}{E_p^{3.5}}$$

In this expression, $4 < n < 5$. It is immediately noticeable that the photoelectric effect tends to be dominant for relatively low energies and for heavy materials. This probability is commonly expressed in terms of photon cross-section of the material in barns or as mass attenuation coefficient μ/ρ, in cm^2/g, being μ the linear attenuation coefficient (cm^{-1}) and ρ the density (g/cm^3) . Figure 1.5 shows experimental results for the cross-section in different materials [5]. Some interesting features, not explainable by us-

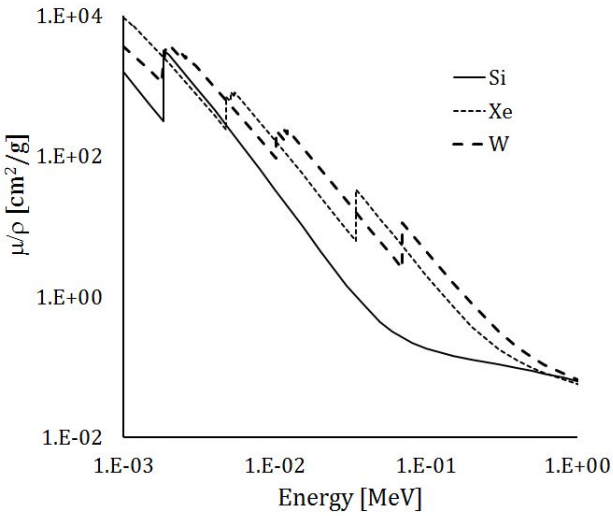


Figure 1.5: X- and gamma-ray photoelectric cross-section for different materials (Silicon, Xenon and Tungsten).

ing the approximated formula, are visible in Figure 1.5. The many edges seen are energy regions where the probability of absorption of the photon sharply increases: these regions correspond with the excitation energies of the atoms'orbital shells. The inner shells tend to be tightly bound, therefore, their edges are placed at higher energies.

As energy increases and photoelectric interaction probability decreases, a

new process, in which ballistic effects come into play, becomes dominant. In this process, known as *Compton scattering*, photons undergo collisions with electrons (assumed at rest) and scatter at a given angle θ after transferring some of their energy to the electron. The energetic electron also recoils away at an angle ϕ. Figure 1.6 illustrates this process. By applying

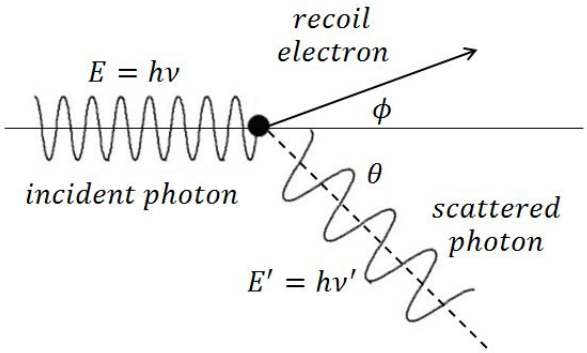


Figure 1.6: Sketch of a Compton scattering interaction.

energy and momentum conservation equations, it is easy to show that [6]:

$$h\nu' = \frac{h\nu}{1 + \dfrac{h\nu}{m_e c^2}(1 - \cos\theta)}$$

where $m_e c^2$ is the rest mass of the electron in electronvolts. What is more interesting, as it relates more to the material properties, is the probability of Compton interaction for an absorber. Intuitively, the more target electrons are available, the higher this probability must be, therefore the number of Compton interactions should show an increase with Z.
The rest mass of the electron is 511 keV; when the photon energy is more than twice this value, a third interaction process can take place. This process, called *pair-production*, is a nuclear reaction where the energetic photon

interacts with a nucleus. The photon's energy is completely absorbed to create an electron-positron pair: 1.022 MeV are used to create the pair and the energy above this value is transformed into kinetic energy of the pair. Eventually, the positron will annihilate and the electron will be captured. The probability for this interaction to occur follows, in first approximation, a Z^2 behavior. A simple picture showing the pair-production process is shown in Figure 1.7.

An interesting characteristic of this process is the fact that the electron

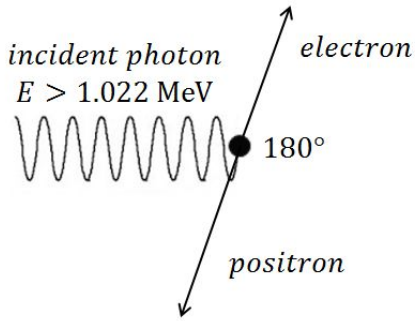


Figure 1.7: The pair-production process.

and positron leave the interaction region at opposite angles (for energy conservation reasons): this fact is greatly used in medical imaging for what is known as PET, or positron-emission tomography. Here, a patient is given a special chemical that is absorbed preferentially by tissue with particular characteristics (for example, tumoral masses). The chemical contains a pair-producing isotope, therefore after some time, the tumor becomes an electron-positron emitter. Detectors placed at 180 degrees axially around the patient detect all radiation present in the environment at any given time; however, only almost-simultaneous events observed in opposing detectors are recorded, as this is a clear signature of an electron-positron pair. The difference in time of arrival (usually a few ns) and the detector location allow reconstructing where the pair was created. Once sufficient statistics

are accumulated, an image of the tissue of interest can be made.
The important picture emerging from having considered the three main phenomena that occur in photon interactions with materials, is that the overall photon absorption characteristics are strongly dependent on the material type and on the photon energy, in a way that is more profound than what can be observed for other particle types. In particular, we can state that for a given Z, at low energies the interactions are all of the photoelectric type; there will be a range of energies where Compton interactions dominate, but it is clear that as the photon energy increases, pair-production settles in, to the point where photons will be more likely to interact via this process rather than Compton scattering. As an example, a complete picture for a classic radiation detection material (sodium iodide or NaI) is given in Figure 1.8. With the exception of the X-ray absorption edges, already discussed previously, cross-section curves for all materials look just like the one in Figure 1.8. The only differences, due to the different atomic numbers, are the relative weights of where in energy one interaction type prevails over the other.

Just as seen for heavy charged particles and electrons, it is useful for photon radiation as well to comment on attenuation and range. With reference to the experiment in Figure 1.2, we find that, for photons, the absorption law is rather simple and is represented by an exponential decay or, in a formula:

$$\frac{n_{out}}{n_{in}} = \exp\left(-\mu t\right)$$

where μ, called linear attenuation coefficient, is the probability per unit length that a photon is removed from the incident path. Clearly, μ will be the sum of the probabilities per unit length that a photon will undergo photoelectric, Compton or pair-production interactions. More commonly, a different parameter is preferred, called the mean free path. That is, the average path length a photon travels before undergoing an interaction in a

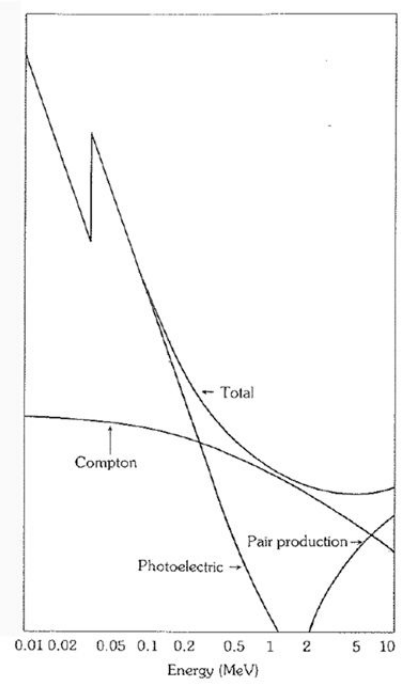


Figure 1.8: Total photon cross section for NaI [Harshaw catalog].

given material. The mean free path, λ, is defined by the following formula:

$$\lambda = \frac{\int_0^\infty t\mathrm{e}^{-\mu t}\mathrm{d}t}{\int_0^\infty \mathrm{e}^{-\mu t}\mathrm{d}t} = \frac{1}{\mu}.$$

1.2.5 Neutrons

The main characteristics of neutrons are their rest mass, comparable to that of a proton, and the fact that they carry no charge, therefore, like photons, they do not undergo Coulomb interactions. It is Coulomb forces that domi-

nate the interaction properties of massive particles, therefore it is legitimate to assume that neutrons tend to interact relatively weakly with materials. It is in fact observed that neutrons can go several centimeters through almost any material before interacting at all. When interactions occur, neutrons either change their energy and direction, or disappear completely, leaving new particles behind. Their interaction characteristics depend strongly on energy and not so much on material properties. It is therefore convenient to separate neutrons in two categories: fast and slow neutrons. The distinction is completely arbitrary and based on the cross-section drop of a particular material, Cadmium, which is one of the few offering appreciable cross-section to neutrons (other materials are Hydrogen, and isotopes of Lithium and Boron). Such threshold is 0.5 eV.

The main difference between slow and fast neutrons is in their way of in-

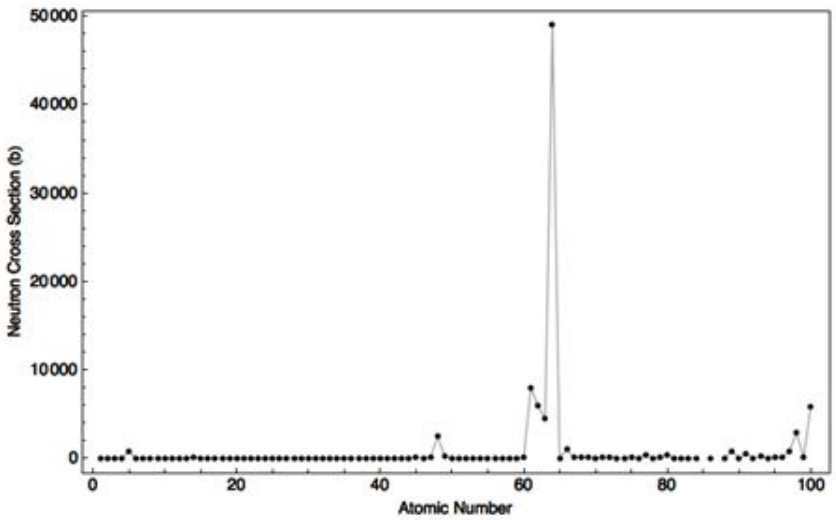


Figure 1.9: Neutron cross-section for natural elements. Note that some isotopes have much different cross-section than their naturally occurring counterparts.

teracting. Slow neutrons mostly interact via elastic scattering with nuclei and via nuclear reactions of capture. These have some importance because they can produce secondary radiation often easier to detect than neutrons themselves. Fast neutrons, on the other hand, interact mostly by scattering off nuclei or protons, and can transfer a considerable amount of energy per collision. When a fast neutron loses energy, it is said that the neutron is *moderated*, a common term found in nuclear power production. The neutron cross-section for materials is historically measured in barns (symbol b), a non-SI unit equal to 10^{-28} square meters. An example of neutron cross section for most materials, expressed in barns is given in Figure 1.9.

1.3 Radiation Sources

So far in our discussions, we assumed radiation was just present, disregarding where it actually came from or how it was generated. Although from now on we will mostly focus on specific sources of radiation, it is useful to give a brief overview of what sources of radiation are in existence.
There are two fundamental types of radiation sources: one type is naturally occurring radiation, the other is man-made radiation. Earth's crust contains several elements that are naturally radioactive and have been in existence since Earth formed (called *primordial nuclei*). The periodic table sorts all known elements based on their atomic number (equal to the number of protons in the nucleus), however, nothing is said about the number of neutrons as for chemistry purposes neutrons have no influence on behavior, except for making the atomic mass slightly different. The same chemical element can be present with different number of neutrons: these elements are called isotopes of the chemical element. It is the number of neutrons in the nucleus that determines the radioactive behavior of an element. Nature allows only a limited set of conditions where a given isotope can exist indefinitely (or be *stable*). When these conditions are not met, the isotope must cease to exist at a certain point in time. When this happens, the atom transforms into a different isotope or species that can be either stable or still unstable. The process continues until a stable configuration is reached.

Through these transmutations (or *decays*), energy is released, often in form of radiation. This is the primary reason for the presence of radiation in nature: the primordial nuclei are unstable elements and must eventually decay into stable ones. The most common naturally occurring isotopes are ^{14}C, ^{40}K, ^{232}Th, ^{238}U, ^{226}Rn, and others are present in very small amounts (isotopes are identified by their mass number followed by the name of the chemical element). The type of radiation isotopes emit is mostly beta, gamma and heavy charged particles (mostly alpha and fission fragments in some cases). Another source of natural radiation is represented by cosmic rays, the continuous flux of particles generated in supernovae by nuclear reactions and expelled at relativistic velocity. Cosmic rays contain mostly heavy nuclei and electrons, and their interaction with the upper atmosphere generates secondary particles that can reach the Earth's surface.

There are more isotopes observed in nature than those suggested above, such as ^{60}Co, ^{137}Cs, ^{239}Pu and ^{241}Am; but they should not be confused with naturally occurring radiation. These isotopes belong to the second category of sources: they are man-made. Large sources of man-made radiation, in fact, were the nuclear weapons tests performed during and after World War II. Those weapons produced considerable quantities of the radioactive isotopes listed above via neutron activation of stable elements.
Other man-made sources are the nuclear fuel cycle, medical, industrial and scientific applications. Examples of these are medical isotopes produced in special facilities and used mostly for medical imaging, sources used in X-ray fluorescence by the oil industry and particle accelerators used for scientific experiments.

The type of radiation generated by a specific source is characteristic of the physical processes occurring within the source. For example, gamma rays are generated by nuclear decay of radioactive isotopes such as ^{60}Co and ^{137}Cs; alpha particles and X-rays are emitted by nuclear reactions and electronic de-excitations in ^{214}Am. Particle accelerators produce beams of heavy ions, protons and neutrons for medical, industrial and scientific applications. A complete study of all radiation generators is outside of the

scope of this work; therefore we will only consider the reactions that are of interest for the following chapters.

1.3.1 Units, Exposure and Dose

For completeness it is important to discuss how radiation is measured (units), and what effects has on matter (exposure and dose). Since radiation is a flux of particles, the most intuitive way to measure it would be in number of particles per square area per unit time. This is true as long as the radiation is confined into a beam. Clearly, natural radiation does not behave as such, therefore, when considering radioactive materials is more useful to refer to their activity, that is, the total number of decays in a sphere surrounding the source per unit time. If there are N radioactive atoms, the rate of decay can be also defined as:

$$r = \frac{dN}{dt} = -\lambda N,$$

where λ is a decay constant, unique to the specific isotope. The rate of decay has dimensions of s^{-1}.
Going all the way back to the work of Pierre and Marie Curie, the unit used for the decays has been the *Curie* (Ci), defined as the activity of 1 gram of pure ^{226}Ra, or 3.7×10^{10} decays per second (in 4π). This unit is still widely used, although the proper SI unit would be the Becquerel (Bq), equal to 1 disintegration per second. It is obvious that 1 Ci = 3.7×10^{10} Bq, or 1 Bq = 2.703×10^{-11} Ci. Typical sources are practically measured in GBq, MBq, kBq, mCi, μCi, and nCi.

It should be noted that, while widely used, activity accounts for the number of decays within the atom, not the actual number of radiation quanta a decay produces. Therefore, in order to know the actual number of radiation quanta produced, one has to know the decay scheme of the specific isotope. For example, a 1 mCi ^{137}Cs source will produce roughly 3.7×10^7 photons per second; however, a 1 mCi ^{60}Co source will produce twice as many, because this isotope has two sequential states allowed in its decay

scheme. Occasionally this fact becomes a source of error in measurements.

Equally important as the units of measurement of radiation, is how radiation is absorbed by matter and what damage it can cause (especially when interacting with the human body) when absorbed. Changes in physical properties of materials are expected to be dependent on the energy absorbed per unit mass. Such energy is indicated as absorbed dose, D. Historically, the unit used has been the rad, with 1 rad = 100 ergs/gram in a given material. In SI units, the Gray (Gy) is preferred. Since 1 Gy = 1 Joule/kg, we have that 1 Gy = 100 rad. This energy can be correlated to the amount of charge produced by the absorbed dose by experimental observations. For example, in air, 33.8 Joules/kg of gamma-rays have been observed to produce a charge equal to 1 Coulomb/kg. This effect is indicated as total gamma-ray exposure and gives an idea of how much ionization a certain activity can generate in a given material. Historically, exposure has been measured in Roentgen (R), and relates only to the amount of ionization produced in STP air by gamma- or X-rays.

When addressing radiation effects in living organisms, the amount of damage induced by radiation can differ by orders of magnitude, depending on the radiation type. Heavy charged particles will likely damage more superficial areas than internal organs, and the overall effect is expected to be more dramatic. For this reason the concept of total equivalent dose was introduced as the amount of radiation that, when absorbed by a biological system, results in the same biological effect as one unit with low specific energy loss ($-dE/dx$), as previously defined. To better quantify what low specific energy loss means, the equivalent dose is written as:

$$H = DQ$$

Where D is the absorbed dose and Q is a quality factor specific to the radiation type. For electrons, X- and gamma-rays, $Q \approx 1$; for alpha particles $Q \approx 20$, for neutrons $Q \gg 1$ and varies significantly with energy. The SI unit used to measure the dose equivalent is the Sievert (Sv).

Chapter 2

Detection of Ionizing Radiation

2.1 Introduction

In Chapter 1 we have seen that all types of radiation interact, to some extent, with matter. These interactions ultimately generate charge within the media. The amount of charge is generally proportional to the energy of the incident radiation; therefore it is reasonable to assume that by collecting the amount of charge developed by a radiation interaction (for example, through an electric field applied to the material) is possible to reconstruct the energy of the radiation itself, or *detect* the radiation itself. This is clearly important, because it allows, for example, to calculate exposure, dose, damage, etcetera.

Almost all materials respond to radiation; however, for some materials the response can only be observed long term as damage to their physical and chemical structure; in others the response can be seen (or measured) on a very short time scale. The latter are materials that make good candidates to work as radiation detectors.

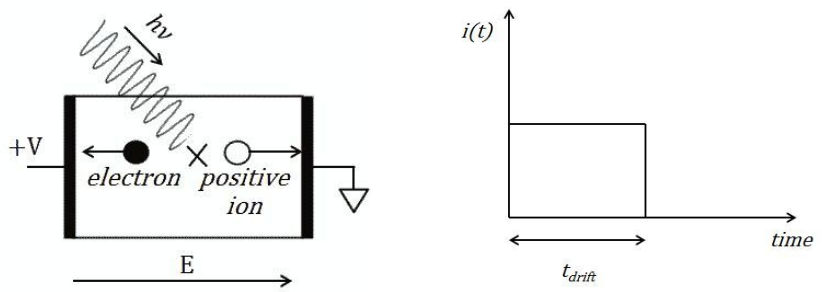
There are two ways by which radiation can be measured in detectors. In some materials the ionization charge can be collected directly by electric means, such as an electric field coupled to collecting electrodes. These

materials tend to be semiconductors. In other materials instead, the charge cannot be collected as it recombines almost immediately, however, from the recombination process an amount of light proportional to the ionization is generated; this can also be used to measure the radiation energy. There are two main distinctions within radiation detectors based on how the charge is utilized to detect radiation: the process of collection of charge generated by radiation is called *direct detection*, as opposed to the process where charge undergoes further transformation (generally into light); this process is called *indirect detection*.

2.2 Direct and Indirect Detection

In order to illustrate the concept of direct detection introduced above, let us consider a simple two-dimensional model of a direct detector material subject to radiation (this could be any of the radiations indicated earlier), and let us apply an electric field at two of its sides (Figure 2.1(a)). When the radiation interacts, it suddenly generates electrons and positive ions. While the overall charge is still zero, the electrons will migrate to the positive side, and the ions will move towards ground, inducing a current signal on the respective electrodes. If the velocity at which charges move (named *drift velocity*) is independent on the sign of the charge, the induced signal will disappear once the charges have reached the electrodes and have been neutralized, after a time t_{drift}. The time taken to collect all charge at the electrodes in called *collection time*. The resulting induced current waveform is shown in Figure 2.1(b). The integral of this waveform is the total charge collected at the electrodes, Q. In a real scenario, the situation repeats continuously, with different amounts of charge deposited each time. If the rate at which the radiation particles reach the detector is sufficiently low, the induced current will look like a random time distribution[1] of pulses with an average proportional to the source strength and amplitudes following the radiation energy distribution. On the other hand, if the rate is high enough, a net current can be observed. Thus, two regimes can be identified: a *pulsed*

[1] It can be shown that the temporal distribution follows Poisson statistics.



(a) Left (b) Right

Figure 2.1: Two-dimensional model of direct detection (left) and induced current response (right) at one electrode.

mode and a *current mode* regime. The former is found in most radiation detection applications, while the latter is more characteristic of radiation beam measurements at accelerators, where current mode detectors are used to measure the beam intensity, and dosimetry. Our application is concerned mostly on the first mode of operation.

The case of indirect detection is conceptually as simple as direct detection, but involves more steps than just ionization and charge collection, with the final product being light in the visible or UV range. This case is of great interest, as most of the available commercial detectors are based on indirect detection. Indirect detection involves excitation of atoms by the radiation into singlet, triplet, or other excited states introduced in the forbidden band of a bandgap. The decay from the excited states produces an amount of light proportional to the number of atoms excited. The light can be emitted over a wide range of time intervals (called *relaxation* time) and over one or more wavelengths (depending on the type and number of excited states allowed). Clearly, for pulse mode applications, the shorter the relaxation time the better. Since the ultimate product of the radiation interactions is light, for these types of detectors further conversion is needed

in order to transduce light onto electric signals. This is normally done via optoelectronic devices such as Silicon diodes, photomultiplier tubes and others. Indirect conversion devices are made of organic compounds (liquid, solid, gaseous), inorganic salt-like crystals and other gaseous materials.

Independent of the process that generated a given response to radiation in a material, it is necessary to understand first how that information is used to characterize the radiation. The technique through which a radiation detector is used to detect and study radiation is called *spectroscopy*. We will see that spectroscopy (from merely counting photons to giving information about quantity, type and intensity of radiation) is the predominant tool for studying all ionizing radiation phenomena.

2.3 Fundamentals of Spectroscopy

We stated several times that a detector responds to radiation via charge or light generation and that, in many cases, the amount of charge or light is proportional to the energy content of the radiation, and therefore representative of it. In some detectors, however, the response does not result in a proportional signal, but it is rather a constant output that only determines whether there was detection or not. A well-known example of this is the Geiger-Mueller (G-M) counter [7]. This is one of the oldest radiation detectors, introduced by Geiger and Mueller in 1928, and certainly one of the most popular, in virtue of its simplicity. In a G-M counter, a chamber is filled with a gas that has the property, under an intense electric field, of generating carrier multiplication when a free ion transverses it. After several multiplications, limiting effects come into play, so that only a predetermined amount of free charge can be sustained, independent of the original number of free ions. In general, for every discharge, 10^9 to 10^{10} ion-pairs are created, corresponding to 0.1 to 1 nC (a very large number), assuming one unit charge per ion. The G-M counter emits an audible signal for every discharge. While this does not allow knowing what the energy of the radiation was, it certainly permits counting rates, thus, gaining insights

on exposure levels. The general method is called photon counting, and is the precursor to spectroscopy.

2.3.1 Definition of Spectrum

When the detector allows an output proportional to the energy deposited, additional information can be obtained. At this point, we can imagine another conceptual experiment that will serve to illustrate how the information available from a detector is collected and processed. We know that charge or light is available at the output of a detector, and that in the case of light a transducer transforms it into charge. Thus, we can visualize the detector as a charge source. We will discuss later the best methods of charge collection, but for now, we can just assume that the charge is further processed into a voltage. This voltage will be read out by appropriate means, for example, an analog-to-digital converter (ADC) and a computer for visualization. Figure 2.2 shows such a simple arrangement that will be called a *spectroscopy system*. We will discuss more in detail the specific

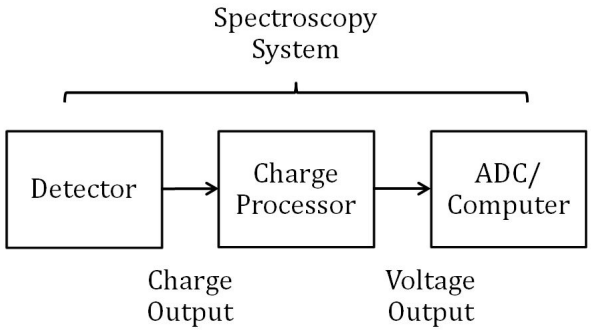


Figure 2.2: Conceptual spectroscopy system.

detectors and the charge processor later; the main concern at this point is how to measure the available information in order for it to be meaningful.

For this purpose, let us recall that the amount of charge is proportional to the energy deposited in the detector, and this is a signature of the radiation, therefore, it is important to correctly correlate the charge to the energy deposited in the detector. We stated that the detector material is a gas, a semiconductor or some type of organic or inorganic compound. All of these materials can be characterized by an energy bandgap or excited levels. The energy of these states is usually lower than the energy of the radiation, therefore it is expected that several ionic pairs be produced upon interaction. In fact, each detector can ultimately be characterized by how many ionic pairs are produced per electronvolt of radiation. This parameter, called *ionization energy* (eV/pair), and indicated by the Greek letter ϵ, allows precise determination of the amount of energy required to produce one ionic pair (or how many pairs are produced per unit energy), independent on the physical phenomena that create the charges. For indirect detectors, it is preferable thinking in terms of eV/photon rather than eV/-pair. Once ϵ is known, the overall output S_{out} of the spectroscopy system will be given simply by:

$$S_{out} = G \cdot \frac{E_{radiation}}{\epsilon}$$

Where G is constant and represents the overall gain of the system. G is in general not dimension-less and includes the specifics of the detectors and the charge processing steps. The output quantity may be a voltage, a current, charge or a binary form.

It is clear at this point that the most meaningful way to know the radiations energy content at this point, is to count how many times the measure of S_{out} falls within a certain interval ΔS_{out} centered around a value $\overline{S}_{out}$. The resulting plot is conveniently normalized in a way to represent directly the energy of the radiation instead of the actual value of S_{out}, and is a probability plot, called spectrum of the radiation. This is not conceptually different than the energy spectrum introduced in Chapter 1. It is rather an application of the same concept to specific cases of radioactive sources. For example, in the case of ideal materials and spectroscopy systems, the

spectrum for a gamma ray source will look approximately like a series of lines precisely centered on the energies present in the radiation. Figure 2.3 shows what a ^{60}Co spectrum would look like. The lines are centered on 1.17 MeV and 1.33 MeV and have exactly the same height. This is expected, because a ^{60}Co source has two emission lines with equal probability at those energies. Clearly, this is an ideal situation. In reality, spectra look

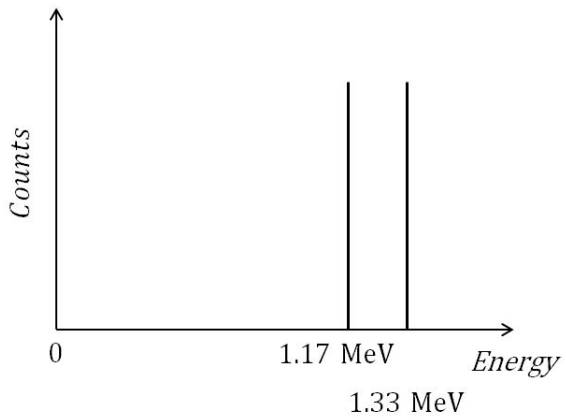


Figure 2.3: Ideal spectrum of a ^{60}Co source.

much different, but if the source of these differences is known, a great deal of information about an unknown source can be gathered nevertheless.

2.3.2 Real Spectra

There are many reasons why spectra taken with real spectroscopy systems are different than the ideal spectrum shown in Figure 2.3. The causes for these differences are found in the detectors, the spectroscopy system and the environment in which the measurement is conducted. For example, lead shielding will influence the shape of the spectrum for any X- and gamma-ray source; air will alter the spectrum of alpha particles. Even in vacuum, surrounding materials may change the spectrum due to scatter. At the

same time, detectors are not perfect and the charge generation process itself is subject to the laws of statistic, causing broadening of the energy lines. Electronics has noise, and this may broaden the lines as well, the same way statistics do. Finite detector size will allow Compton-scattered photons to escape the active volume, leaving only partial peaks behind, and so forth.

It would be impossible to gather all sources of imperfection in a single place, because they depend on too many factors. What is reasonable, however, is to give a sense of the general factors affecting spectral shape that are found in all spectral measurements.

Resolution

Although the energy released by a radioactive decay is a constant determined by the difference in the energy of excited states and ground states, the amount of charge (or light) generated within a detector contains a certain degree of randomness due to the statistical nature of the processes involved. Such statistics are well described by Poisson statistics and correctly modeled by a normal distribution such as [8]:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Where P is the probability of occurrence of the quantity x, σ is the variance of the normal distribution and μ is the average value of x. This equation can be re-written for our purposes as:

$$S(E) = \frac{A}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(E-\overline{E})^2}{2\sigma^2}\right)$$

where E is the energy, $\overline{E}$ is often referred to in this form as the *centroid* and the term outside of the exponential function is the peak value of the curve, related to its area A. Note that it is assumed that $S(E)$ can be greater than unity. The form is called Gaussian distribution. Its shape and the relevant parameters are shown in Figure 2.4.

Clearly, a detector with small σ will allow distinguishing nearby peaks much better than detectors with large values of σ. In spectroscopy, rather than using σ to characterize the detector's ability to separate nearby peaks, another parameter is introduced, called *energy resolution*. The energy reso-

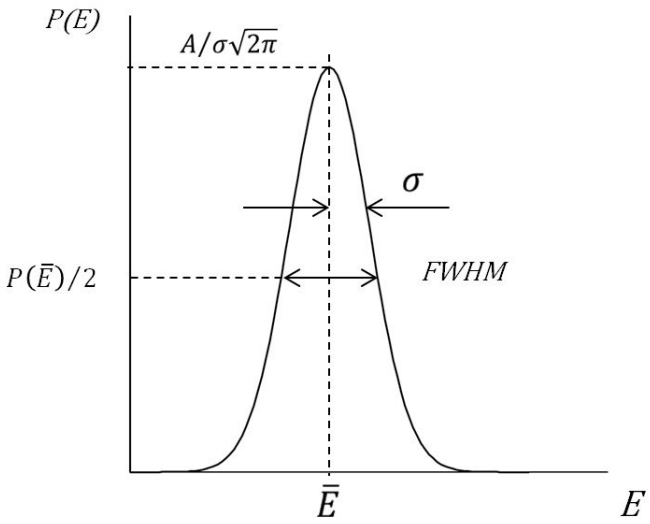


Figure 2.4: Gaussian distribution.

lution is defined by the value of the ratio of the full-width-at-half-maximum (FWHM— see Figure 2.4), equal to 2.35σ, to the value of the centroid, or:

$$R = \frac{\text{FWHM}}{\overline{E}} \cdot 100$$

The spectrum in Figure 2.5 shows, as an example, how the ideal spectrum of a ^{60}Co source changes when collected by a detector with a resolution of 0.42% and a different detector with 6% resolution (both specified at 1.17 MeV).

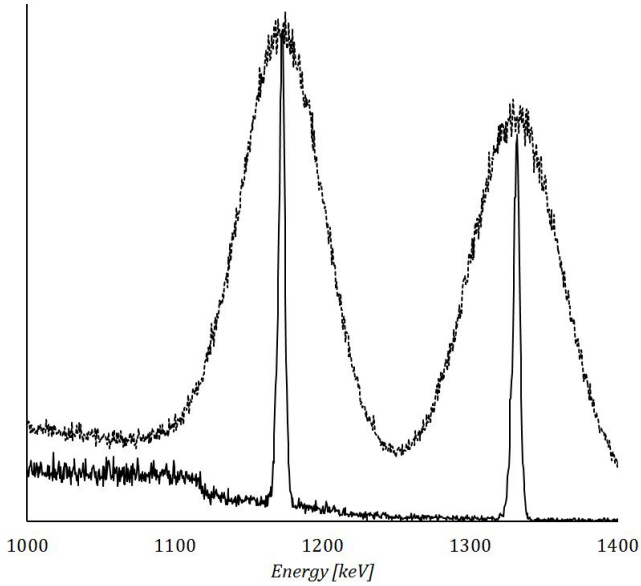


Figure 2.5: Spectrum of a ^{60}Co source taken using a high-purity Germanium detector (solid line) and a NaI scintillator (dotted line).

Compton-scattering effects and finite detector size

So far, we assumed all of the energy of the incident photon was deposited in the detector; this is equivalent to assuming an infinite volume of detector. We know that some interaction types (Compton and pair-production) are such that the energy is not deposited in a single site. If the detector has finite size, it is possible that the photon may escape the detector volume after one or more interactions, and that not all of its energy be recorded in the detector. This is especially true for Compton interactions. Let us

recall the Compton formula introduced in Chapter 1:

$$h\nu' = \frac{h\nu}{1 + \dfrac{h\nu}{m_e c^2}(1 - \cos\theta)}$$

the energy E_d left in the detector is:

$$E_d = h\nu - h\nu' = h\nu \left(\frac{\dfrac{h\nu}{m_e c^2}(1 - \cos\theta)}{1 + \dfrac{h\nu}{m_e c^2}(1 - \cos\theta)} \right)$$

At the extremes ($\theta = 0$, $\theta = \pi$), we have:

$$E_d|_{\theta=0} = 0$$

$$E_d|_{\theta=\pi} = h\nu \left(\frac{2\dfrac{h\nu}{m_e c^2}}{1 + 2\dfrac{h\nu}{m_e c^2}} \right)$$

All angles are scattered, therefore, in addition to the full-energy peaks of Figure 2.5, there will also be a continuum of energy between these two values. More peaks (called *backscatter peaks*) can also be seen overlapped to this continuum due to partial energy deposition of Compton scattered events in the surrounding materials.

The spectrum in Figure 2.6, obtained from a high-purity Germanium detector (HPGe), is a good example containing all of the features introduced above. In this spectrum, where the vertical axis is a logarithmic scale, we can recognize four energy lines. From lowest to highest energy, they are ^{137}Cs at 662 keV, the two lines of ^{60}Co and the ^{40}K line at 1.46 MeV. We can easily distinguish the Compton edges for Cesium and Cobalt (the Potassium edge is masked by the presence of the Cobalt lines), as well as a backscatter peak roughly around 200 keV, as expected. Note that the

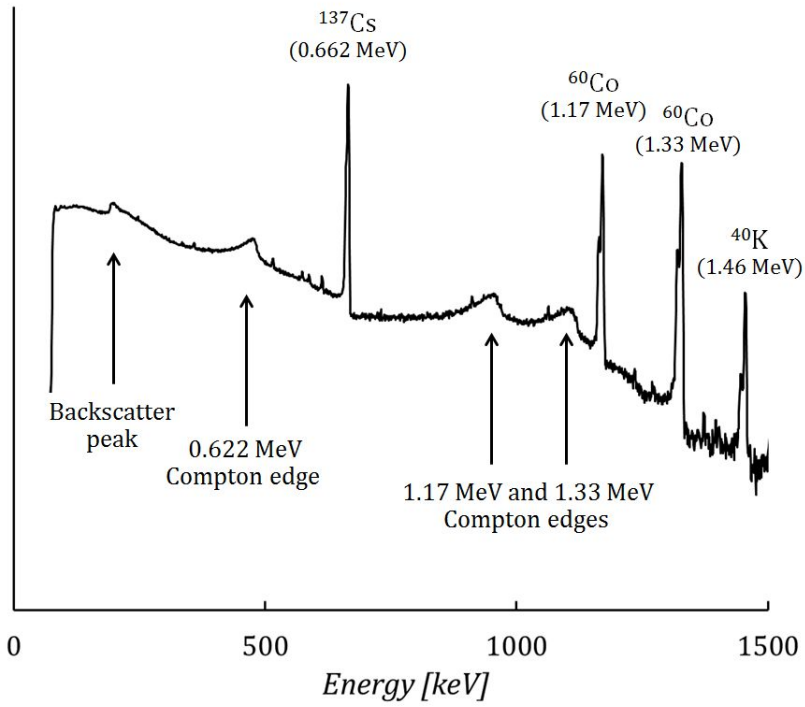


Figure 2.6: HPGe spectrum illustrating common spectral features.

spectrum does not go all the way to zero, but has a low-energy cut-off. This is also common and due to two main factors: at the lowest energies the photons are stopped by the encapsulation material around the detector (if present) and the electronics that read the charge from the detector have intrinsic noise that limits the lowest attainable threshold in some cases. The latter is a very important potential limitation and will be the subject of a later chapter.

2.4 Detectors

Up until now, we have seen the ways radiation interacts with matter, and how materials respond to it. A detector is made of radiation detecting material, and everything that has been seen so far will directly apply. However, a real detector involves more than the basic interaction principles. A basic knowledge of radiation detection would not be complete at this point without knowing some specifics about radiation detectors.

There exist several types of detectors, and an all-encompassing view of them would be outside of the main scope of this thesis. We intend therefore to give a coarse view of the main types of detectors and focus on the details of what is directly relevant to the work contained here.

2.4.1 Solid state detectors

Solid state detectors are direct detection (or conversion) type and made of semiconductor materials. Their main characteristic is that solid state detectors are, basically, diodes operated in reverse bias mode, or depletion mode. This allows the material to sustain relatively elevated fields in regions almost depleted of free carriers, therefore when one interaction occurs in such a region, the charge can be swept directly to the electrodes and collected by electronics means for processing. Being semiconductor diodes in nature, these detectors can be studied exactly like conventional semiconductor devices in terms of energy bandgap, leakage, carrier mobility, impurities etc. The main difference with solid state devices for electronics applications is in the purity requirements imposed by this specific application. This is why, with the exception of a few experimental cases [Si-drift reference], it is not possible to integrate signal processing electronics directly onto the detectors themselves.

Composition

There are several types of semiconductors used for radiation detectors. The performance of the detector depends on a number of factors determined by

Material	Z	Density (g/cm^3)	Bandgap (eV)	Ionization Energy (eV/pair)
Silicon	14	2.3	1.12	3.6
Germanium	32	5.3	0.72	2.96
CdTe	48/52	6	1.52	4.4
HgI_2	80/53	6.4	2.13	4.3
CZT ($Cd_{0.8}Zn_{0.2}Te$)	48/30/52	6	1.64	5

Table 2.1: Characteristic parameters of semiconductor detector materials

the specific material. Historically, the first solid state detectors used in radiation detection were based on crystals of Germanium or Silicon. Today, these detectors still constitute the standard for radiation detection, although a number of new materials and compounds have been successfully introduced over the last two decades. The reason for research on new materials is mostly practical: in order for Germanium detectors to function, they need to be cooled to temperatures near liquid nitrogen (77 K); for silicon some cooling may also required in order to obtain best performance. The search for new materials has mostly focused on semiconductors that can work at room temperature, removing thereby the need for cooling apparatuses. Another reason is that compound semiconductors (those obtained from a combinations of group III-V or II-VI on the periodic table), have higher atomic number, therefore better cross-section for the radiation. Regardless of material type, the principles governing the radiation interactions in solid state detectors are the same as highlighted previously for direct detection (Figure 2.1(a)); these detectors, can thus be characterized, from the instrumentation point of view, by a specific set of parameters dictated by the particular semiconductor used. For example, Table 2.1 shows a sample set of such parameters [9, 10].

Other parameters instead, depend on the specific geometry adopted for

the detector itself.

Geometries

The geometries are generally of two types: coaxial and planar (often dictated by the primary lattice directions). Both are used in the construction of germanium-based detectors, while all other detector types are found mostly in the planar geometry, with different configurations for the electrodes (monolithic, pixelated, strips, grids and point-like) to which the depletion voltages V_1 and V_2 are applied. Figure 2.7 shows such geometries. In coaxial detectors, the outer surface of the cylinder constitutes one

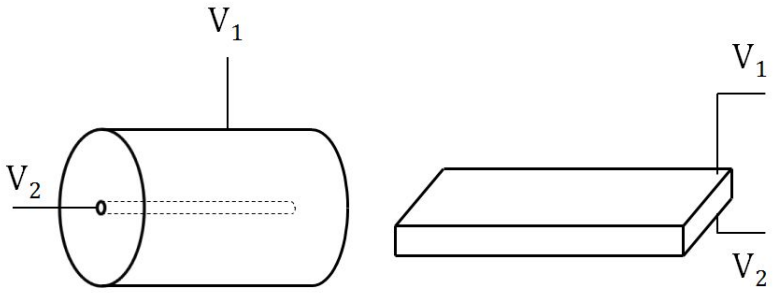


Figure 2.7: Semiconductor detector coaxial and planar geometry (note that planar geometries can be also found with circular shape).

electrode, while the second electrode is bored along the cylinder axis. Sizes can range, for the coaxial geometries, from a few 10 mm diameter by 10 mm length to about 10 cm by 12 cm. For planar geometries, from a few mm on the side to about 15 cm, with thicknesses ranging from 0.1 mm to about 1 cm.

Performance

By virtue of extremely low values of ionization energy, semiconductor detectors generate relatively large amounts of charge for every electronvolt. The

energy resolution is the variance on such charge, therefore, in the Poissonian approximation, the larger the charge, the better the resolution. In fact, if the detector response is linear (true for semiconductor detectors), the average amount of charge generated cane be expressed as $Q_0 = KN = K\frac{E}{\epsilon}$, where K is a proportionality constant. The standard deviations is, thus, $\sigma = K\sqrt{N}$, and the FWHM$= 2.35 \cdot K \cdot \sqrt{N}$. In this case, the expression for the resolution given previously becomes:

$$R_{Poisson} = \frac{\text{FWHM}}{Q_0} = \frac{2.35 \cdot K \cdot \sqrt{N}}{K \cdot N} = \frac{2.35}{\sqrt{N}} = 2.35 \cdot \sqrt{\frac{\epsilon}{E}}$$

Measurements on semiconductor detectors indicate an attainable resolution far better than anticipated by Poisson statistics, but still dependent on $N^{-\frac{1}{2}}$, indicating that for these materials there is a certain degree of correlation that cannot be ignored. In order to account for this discrepancy, a factor F (called Fano factor) has been introduced as:

$$F = \frac{\text{Observed variance}}{\text{Poisson variance}}$$

and the actual resolution is expressed as:

$$R_{actual} = 2.35 \cdot \sqrt{\frac{F}{N}}$$

Some typical values for the Fano factor are shown on Table 2.2.

For these reasons, semiconductor detectors have the best achievable resolution (with the exception of some special materials that respond to phonons rather than photons at extremely low temperatures). Drawbacks for these detectors are cost, generally much higher than other detectors, relatively limited detector size achievable, necessity of cooling for some, and added challenges in readout instrumentation design.

Material	Fano factor
Germanium	0.129
Silicon	0.143
CZT	0.089

Table 2.2: Typical Fano factor values for some semiconductor detector materials

2.4.2 Scintillators

Most of the radiation detectors commercially available are based on scintillators. This is due to their low cost, availability in large form-factors (up to several m^3) and acceptable performance. Scintillators do not require collecting electrodes, as the response to radiation is light, but they do require light-to-charge transducers (or *photoconverters*) in order to convert their output in a signal that can be easily processed by electronics. This is a mature technology; however, presently there is still significant investment in the understanding of these devices and in material research for better performance. Significant effort is also spent in research on more advanced photoconverters, in order to ease some drawbacks associated with the use of these materials. The work presented in this thesis, in fact, is meant to add to this effort.

As we mentioned previously, the physical process thanks to which light is generated upon interaction with radiation, is a process where the kinetic energy of the photons excites the material components into a number of excited states; the decay from such states produces a fluorescence pulse of light with duration between a few ns to a few µs[2] . The wavelength of this light is usually in the visible to UV-range. Depending on the materials, the states are either energy levels of molecules (in organic scintillators), or

[2]Longer responses are also observed; however, materials with ms to s responses are useful for dosimetry mostly, and not for radiation detection per se. The longer decay processes are referred to as *phosphorescence*.

energy states determined by a lattice (inorganic scintillators) and/or impurities purposely introduced, known as activators. It is instructive to take a look at scintillators, based on their organic or inorganic composition, given their different behavior.

Organic Scintillators

Scintillation in materials based on organic molecules depends on molecule states, therefore it does not require a lattice structure to exist. Organic scintillators, in fact, can be found in solid, liquid and gaseous form, where the solid state can be polycrystalline. Anthracene ($C_{14}H_{10}$) and Stilbene ($C_{14}H_{12}$) are the most commonly found pure organic crystals; liquid scintillators are organic chains such as Benzene or mineral oils with atomic ratios of Hydrogen to Carbon between 1.2 and 2. They are found under several trade names, depending on vendor (EJ-301/NE-213/BC-501A etc.)[11, 12]. Their main prerogative is to be toxic; therefore careful considerations must be employed in designing their containers. Polycrystalline scintillators are usually organic molecules dissolved in a solvent that, once hardened, acts as a plastic matrix for the organic molecules. The good optical properties of the plastic matrix allow for these scintillators to be made in virtually any shape and large volumes. These materials are also rather inexpensive.

Organic scintillators usually exhibit a very fast and prompt light emission in response to radiation, in the form of exponential shape with a decay constant in the ns range. An interesting feature of this class of materials, is that the temporal shape of their response often depends in a very subtle way on the type of radiation (gamma- and X-rays, heavy charged particles, beta and neutrons): while the instantaneous response is virtually unchanged, the decay time constant depends on the nature of the interacting particle as shown in Figure 2.8, where the time response to a neutron (black line) can be distinguished from the response to a gamma-ray (gray line) on a logarithmic vertical scale, by a slight change in decay time, more peak roundness as well as several secondary peaks occurring from about 100 ns to 200 ns after the main peak. By integrating the waveforms it

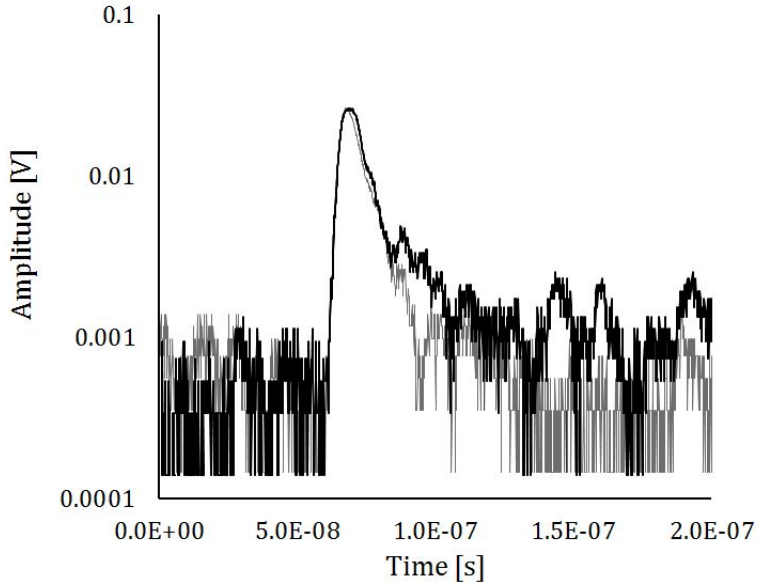
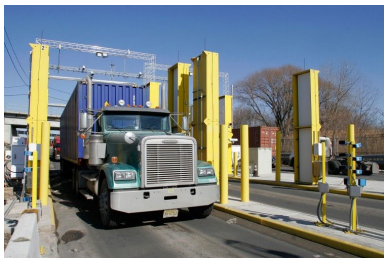


Figure 2.8: Gamma-ray (gray line) vs. neutron (black line) response of an organic scintillator.

is possible to separate neutrons from gamma-rays by weighting the integral on different time intervals. The technique is known as pulse shape discrimination, or PSD [13].

Performance of Organic Scintillators

Organic scintillators have low atomic numbers, being composed mostly of C, O and H. This makes them good candidates for neutron detection (via proton recoil), but their performance for gamma- and X-ray detection is rather poor, as most of the interactions happen via Compton scattering. The gamma response is typically a Compton continuum with no peaks at the energies of interest, so that it is not even appropriate speaking of energy resolution for this radiation. Their main research applications are, in fact,



(a)

(b)

Figure 2.9: Standard radiation portal monitor (a) used to scan commercial traffic for radiation sources and advanced concept of portal monitor based on video tracking and gamma-ray imaging (b).

for neutron and particle detection and gamma/neutron discrimination in mixed radiation fields. More practical applications are in radiation portal monitors in virtue of their low cost and high attainable volumes. In these detectors, plastic m^2 slabs are set on the side of a lane where trucks are directed at appropriately low velocity. The large detector surface allows indeed for high sensitivity and the ability to detect sources of radiation within the cargo being transported (see Figure 2.9(a)), with some limitations [Lo Presti]. These installments require dedicated infrastructures and have significant impact on the flow of traffic; more recently research by the author and colleagues at the Oak Ridge National Laboratory proposed more transparent approaches based on different detectors and video tracking techniques [RST references] that allow reaching similar performance in highway traffic up to five lanes and to 120 km/h (Figure 2.1(b)).

Inorganic Scintillators

Generally speaking, inorganic scintillators are non-conductive materials with crystalline structures that show well-defined energy bands. Just like in semiconductors or insulators, electrons in the valence band are bound to lattice sites, whereas electrons in conduction band are free to move about

the lattice. When radiation interacts with the crystal, it promotes electrons from the valence band into the conduction band. Once these electrons decay back to the valence band, they produce light. However, the energy difference between the two bands is usually quite large, therefore the light emitted falls outside of the range detectable by most photoconverters; plus, this process is rather inefficient. The addition of a small concentration of impurities within the lattice allows the appearance of energy states within the forbidden gap. Now, the electrons within the conduction band quickly migrate until they find such impurities and decay into the new states by emitting lower wavelength light. The wavelength of the light so emitted falls within the visible range up to the UV range, therefore it is easier to detect by conventional means. The impurities, for their function, are called *activators*. The process is shown in simplified form in Figure 2.10. One

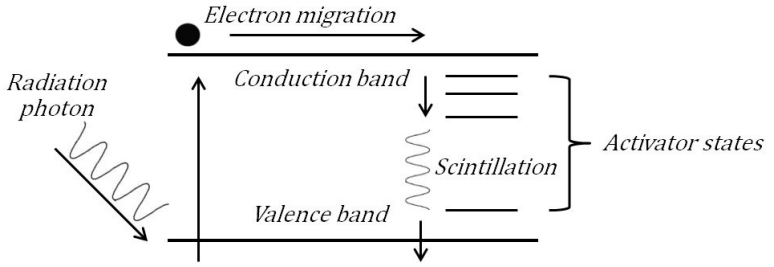


Figure 2.10: Schematic of the scintillation light mechanism within inorganic scintillators.

thing is immediately evident from these considerations: just for the sheer fact that the bandgaps involved in the front-end process (the promotion of electrons into the conduction band), it is intuitive to assume that, compared to semiconductor detectors, inorganic scintillators will require more energy to generate the same amount of free carriers, or, for the same incident energy, fewer carriers are created. This immediately implies that also inorganic scintillators have poorer resolution than semiconductor de-

tectors[3]. If we define the equivalent of the semiconductor ionization energy ϵ as the energy required to create a photoelectron in the photodetector reading the scintillator, we can experimentally find that, for NaI scintillators (the industry's workhorse), $\epsilon \approx 25$ eV/photon versus 2.96 eV/pair for Germanium: a dramatic difference. This, in practice, means a difference in resolution from about a tenth of a percent for HPGe to ~6% for NaI at the 662 keV gamma line of ^{137}Cs. However, scintillators have much higher atomic numbers, therefore better stopping power overall than semiconductor detectors.

Composition of Inorganic Scintillators

The most popular inorganic scintillators are made of Alkali-Halides crystals with special activators (Eu, Na, Tl), but it is not uncommon finding materials based on garnets, glass and other inorganics. Table 2.3 is a collection of inorganic scintillators properties that are of interest when designing scintillator detectors based on inorganic materials [14, 15]. Elements in parenthesis indicate the activator type.

Performance of inorganic scintillators

Inorganic scintillators have generally very good stopping power due to their relatively high atomic number. This means that for a given volume, they potentially offer the best efficiency of detecting radiation, especially gamma- and X-ray. Their cost, although generally higher than plastics, is reasonable, especially compared to semiconductors. This is the reason why the majority of commercial radiation detectors are based upon inorganic scintillators (NaI(Tl) and CsI activated with Sodium or Thallium are by far the most used). The resolution is generally acceptable to good (~3% to 12% at the ^{137}Cs peak). It should be noted that for some of these scintillators, resolution cannot be solely anticipated based on the statistics of

[3]There are also other processes that work against resolution: quenching, non-uniformity in light production, optical coupling with light converters, etc.

	Density (g/cm^3)	Peak Emission Wavelength (nm)	Refractive Index	Decay Time Constant(s) (μs)	Light Yield (ph/MeV)
Alkali-Halides					
NaI(Tl)	3.67	415	1.85	0.23	38000
CsI(Tl)	4.51	540	1.80	0.68+3.34	65000
CsI(Na)	4.51	420	1.84	0.46+4.18	39000
SrI_2	4.55	435	2.05	1.2	115000
Li(Eu)	4.08	470	1.96	1.4	11000
Other Inorganics					
BGO	7.13	480	2.15	0.3	8200
$LaBr_3$(Ce)	5.29	380	1.82	0.026	63000
CLYC	3.31	370	1.81	0.05+1	20000
$CeBr_3$	5.1	380	2.09	0.019	60000
$CdWO_4$	7.9	470	2.3	1.1+14.5	15000
ZnS(Ag)	4.09	450	2.36	0.2	—
CaF_2(Eu)	3.19	435	1.47	0.95	30000
GSO	6.71	440	1.85	0.056+0.4	9000
YAP	5.37	370	1.95	0.027	18000
YAG	4.56	550	1.82	0.088+0.3	17000
LSO	7.4	420	1.82	0.047	25000
Typical plastic scintillator					
NE102A	1.03	423	1.58	0.002	10000

Table 2.3: Inorganic scintillators properties. (Note that ZnS(Ag) properties are for alpha particles only)

the light produced, coupling and quantum efficiency of the optical transducer: the number of photons generated in response to a radiation event is in some cases non-uniform within the volume. Because of this, slightly different light intensities and resolutions are obtained for the same energy deposited in different locations within the material. When summed, the resulting response is inevitably degenerated. For example, a material such as CsI(Tl), whose light output is comparable to $LaBr_3$(Ce), will not exhibit the same resolution as the latter, as the non-uniformity of CsI(Tl) is more pronounced. In this example, CsI(Tl) has resolution that, at best, reaches 5.8% at 662 keV, while $LaBr_3$(Ce) shows better than 3% resolution at the same energy.

2.4.3 Other Scintillators

There are a number of inorganic scintillators that are found in gaseous or liquid form under normal operating conditions. The light production mechanism for these scintillators is similar to that found in organics: as the radiation passes through the media, it leaves a number of excited molecules behind; their decay to the ground state produces light. Another important behavior of these scintillators is that for sufficiently high energies, the radiation creates ionization, so that, in addition to light, free electrons can be collected. Pure noble gases such as xenon and argon show this behavior in both liquid and gaseous phase. These elements have recently gained in popularity because they have interesting properties for applications such as neutrino physics, as it will be extensively seen later on.

2.4.4 Optical Transducers

The main characteristic of scintillators is that they respond to radiation with a proportional amount of light. Optical transducers are thus needed whenever a scintillator is part of an instrument, in order to convert light into an electrical signal. While in optics the amount of light to measure is generally very generous, to the point that photodetectors respond to it via continuous currents (usually in the μA range), the same cannot be said for

radiation detection. To illustrate this, let us consider a very bright scintillator, say $SrI_2(Eu)$ from Table 2.3, and let us suppose the very optimistic case (never true) of 100% conversion efficiency (i.e. every photon is directly converted into an electron inside the photodetector). For the total capture of a 10 MeV signal we would have about $1.15 \cdot 10^{12}$ electrons (or 184 pC) over 6 μs[4]. This corresponds to a current waveform of ~31 mA with an exponential shape qualitatively similar to those shown in Figure 2.8, over a 6 μs duration. While detecting this single pulse would not be a problem at all, even with an event rate of 10^6 particles/sec, the average current would only be in the nA range. This is an extreme case but confirms some important facts: the detection of radiation is usually meaningful on an event-by-event basis, and the amounts of light to be detected are both small and limited in time. Radiation detection via average current measurements is meaningful only when extreme high rates and energies are present (such as in accelerator beam lines).

In radiation detection applications, the amount of light to convert is orders of magnitude smaller than indicated in the example because of several factors:

- Geometrical mismatch between photodetector and scintillator
- Optical coupling between scintillator and photodetector
- Quantum efficiency of light-to-charge conversion
- Photodetector noise.

There are several photodetectors available, each one with its own different mechanisms and principles. Depending on the photodetector, one can privilege size over function, or sacrifice noise for simplicity; the choice of a particular type of photodetector is mostly dictated by the application, therefore we will present the most common types.

[4]Here we assume 100% of the photodetector charge is collected within five times the decay constant of the scintillating light.

Photomultiplier Tubes (PMT)

Photomultiplier tubes, or in short, PMTs, are probably the oldest photodetector for radiation detection applications. They are based on vacuum tube technologies, and were initially developed before World War II. Despite this, they are still in great demand, and only recently new devices based on semiconductors have come to rival PMT performance. The fundamental effects PMTs are based upon are the photoelectric effect and the secondary emission [16]. In the photoelectric effect, photons impacting an alkaline metal liberate a proportional amount of electrons that become available to conduction processes once ejected in vacuum. Secondary emission is a process through which charges are ejected by metals struck by other charges in vacuum. Depending upon incident energy, more electrons can be emitted than those that struck in the beginning, allowing charge multiplication gain. A PMT is nothing more than a vacuum tube where a suitable alkaline metal is deposited on the end exposed to light (called photocathode); secondary metals (dynodes) are distributed within an electric field to allow secondary multiplication. At the opposite end (anode), the total charge is collected. A simple drawing of a PMT is shown in Figure 2.11. Given the importance

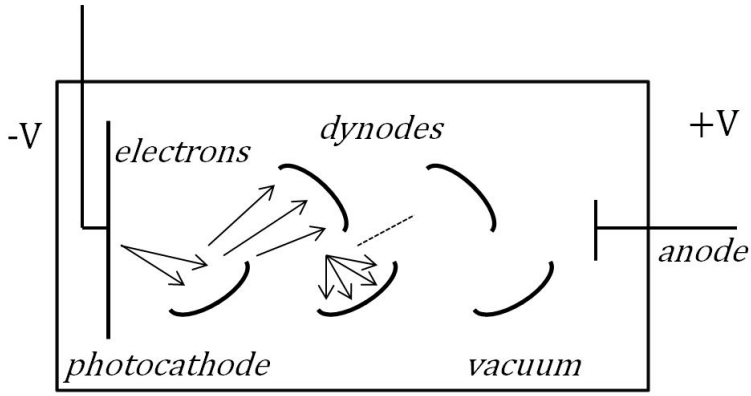


Figure 2.11: Simple schematic of a photomultiplier tube.

of this device, it is worth looking at it into some detail.

Photocathode

The photocathode performs the function of converting the photon's energy into charge. This happens through photoelectric effect. The process is conceptually simple: as a photon interacts with the photocathode, its energy is absorbed by the alkaline metal and transferred to an electron. The electron then migrates from the interaction site to the surface of the photocathode and, if it has sufficient energy to jump the vacuum-metal energy gap (*work function*), is ejected into the surrounding vacuum.
Clearly, some general considerations can be made. Obviously, the photon has to have sufficient energy to excite an electron with enough energy to overcome the work function. Since the typical value of photon energies emitted by scintillators is in the 3 eV range (blue light), comparable to the work function for most metals (semiconductors instead have work functions of about 1.5-2 eV), it is important that the energy lost to the migration process be as low as possible. This is normally accomplished by having very thin photocathodes (just a few nm) obtained by depositing the photocathode layer directly on the vacuum side of the glass tube (known also as the entrance window). This is a major limitation, because at these thicknesses, metals remain partially transparent to light. Another limitation is what is called spontaneous thermal electron emission. The average thermal kinetic energy of the electron is $\frac{kT}{q}$, where k is Boltzmann's constant, T the temperature and q the electronic charge. The value of the kinetic energy at room temperature is 0.025 eV. The actual values at any temperature follow a statistical distribution; therefore there is always a finite chance that a few electrons have enough energy to overcome the work function. Thus, a spontaneous electronic emission can be observed from a photocathode. Typically, this rate is in the 100 electrons/(s·m^2) at room temperature. In semiconductors such rate is much higher in virtue of the reduced work function. Typical numbers for semiconductors are in the 10^6-10^8 electrons/(s·m^2).

The most important parameter regarding photocathodes is their sensitivity to light and how efficiently they generate photoelectrons. We may be familiar with a definition of such sensitivity as current produced per unit light flux (A/lumen). While this is an excellent way of quantifying performance for constant illumination, we know this is not the case for radiation detection. A more significant parameter is what is known as quantum efficiency (QE), defined as:

$$QE = \frac{\text{number of emitted photons}}{\text{number of incident photons}}$$

The QE is a function of photon wavelength for a given photocathode: at long wavelengths the photons will not have enough energy to create electrons capable of overcoming the work function; at short wavelengths, the photon will be too energetic to efficiently interact with the thin layer of photocathode material. Typical QE values can be found from most PMT manufacturers. For completeness, we report in Figure 2.12 the typical QE for the three main type of photocathode currently in use (known as bialkali), compared to the more advanced type (super bialkali, or SBI) and a future formulation (ultra bialkali or UBA) to be introduced by Hamamatsu Corporation, a major PMT manufacturer based in Japan. It should be noted that while these responses match the peak emission of most organic and inorganic scintillators, they have rather poor response in the UV range (noble gas and liquid scintillators). To address that, manufacturers produce families of photocathodes for UV light; peak efficiencies for those are in the 15-25% range.

Dynodes

Once photoelectrons are emitted by the photocathode, they are focused and accelerated by an electric field in order to acquire enough energy to impact a secondary emission material and create charge gain. The same considerations made for photoelectric emission, can be made for secondary emission, with the exception that now the impacting particle is the accelerated electron rather than a photon. Normally, the photoelectrons leave the

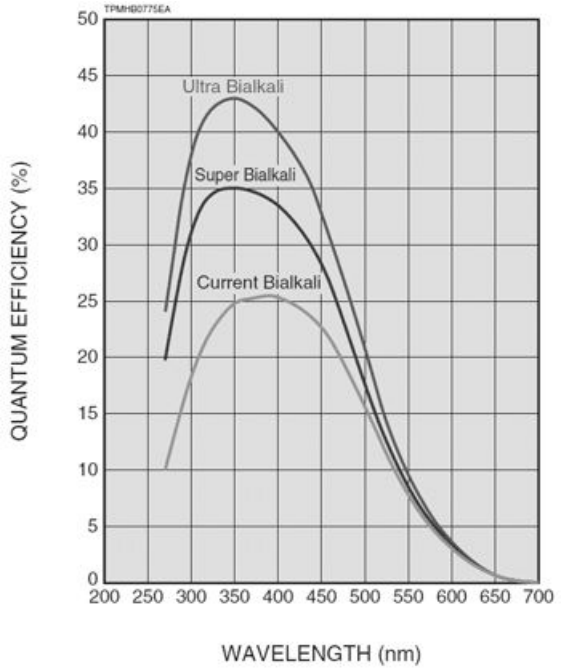


Figure 2.12: Typical QE curve for Hamamatsu's standard bialkali, super bialkali and ultra bialkali [17].

photocathode with energies at or below 1 eV. The energy required to expel an electrode in a secondary emission process is of the order of 2-3 eV. Thus, by accelerating the photoelectron through a 100 V potential, theoretically, 30-50 secondary electrons can be created. This is a significant gain, considering the process can be repeated on several stages (typically 6 to 12). In practice, typical gain values are 10^5–10^6. The parameter that accounts for the multiplication phenomenon for a single dynode is defined as:

$$\delta = \frac{\text{number of secondary electrons}}{\text{number of incident electrons}}$$

Since the emission process has a certain amount of statistics associated, δ is an average value with standard deviation $\sigma = \sqrt{\delta}$. The relative variance is then $\frac{1}{\delta}$. With N identical stages at the same potential, the overall PMT gain can be expressed easily as $G = \delta^N$. Since the statistic is Poissonian, it is interesting to notice that the variance in gain is:

$$\frac{1}{\delta} + \frac{1}{\delta^2} + \cdots + \frac{1}{\delta^N} \longrightarrow \frac{1}{\delta - 1}$$

If $\delta \gg 1$, the statistical gain fluctuations are dominated by the very first dynode, where the number of electrons is smallest. For this reason, the first dynode is usually at the highest difference in potential and/or is made of a material with higher δ.

Timing and linearity of PMTs

The ability of a photomultiplier to quickly respond to light detection is often important. Typical time responses for PMTs are in the ns to a few ns. Linearity is also a fundamental characteristic. Fortunately, PMT responses remain linear between a single photoelectron and several thousand photoelectrons. For normal radiation detection applications, there is no concern for PMT linearity.

PMT limitations

There are very few shortcomings in using PMTs, nevertheless, they are important in some applications. The presence of free-drifting charge within the vacuum implies that the gain is affected by magnetic fields; the need for biasing voltage requires the design of high-voltage dividers to properly bias the anode, photocathode and the dynodes. This often requires some power, as the biasing voltages are in the 1-2 kV range, with currents up to a few mA. Finally, PMTs are always in the few 1000s cm^3 in volume. This size may be a problem in some applications. The content of metals in PMTs is relatively high. Metals contain trace amounts of naturally occurring isotopes; this is often a problem in very low-background applications such as dark matter and neutrino studies.

2.4.5 Semiconductor Photodiodes (PD and APD)

A simple silicon diode is capable of detecting light in a way similar to which it detects radiation, because, as previously mentioned, the photons emitted in the scintillation process have an energy of 3-4 eV, while the semiconductors bandgap is about 1-2 eV. Since the conversion process is internal, there is no need for the photo-charge to leave the surface of the semiconductor; therefore the process can be more efficient (i.e. better QE, better energy resolution) than the photoelectric effect found in PMTs. With the exception of newer devices we will see extensively later, there is no charge amplification after the photo-generation process. Also, photodiodes are essentially reverse-biased junctions that can have significant capacitance (typically from a few 100 pF to a few nF). Detector capacitance competes with the readout electronics in collecting charge, and increases the effects of electronics noise; therefore, the overall signal-to-noise is inherently lower in photodiodes. This poses interesting challenges on the electronic readout design that are completely absent in PMTs. Despite these challenges, excellent resolutions (about 4.4% at 662 keV) have been achieved for CsI(Tl) with special photodiodes as readout elements [18] (the best resolution with PMTs is 6%).

Some gain can be achieved in silicon diodes if the design allows for Zener-type avalanches. Such devices are called avalanche photodiodes (or APD in short). APDs have a thick intrinsic layer and are biased around the breakdown point of the reverse I-V characteristic (about 100-200 V reverse bias). Once electrons are produced, they are accelerated by the electric field present in the intrinsic layer and collide with atomic electrons, generating more carriers. Gains in the 100-200 can be observed. On one hand, the internal gain and the thicker intrinsic layer (lower device capacitance) lead to potential signal-to-noise improvements. In practice, the high biasing field within the intrinsic layer produces a sizeable amount of leakage current (up to a few μA for cm^2 areas). This current is a powerful source of noise and often negates the potential improvements.

Practical sizes for photodiodes and APDs are in the few 10's cm^2, as opposed to several thousand cm^2 for PMTs. In order to obtain larger areas (the current leakage for a monolithic large area would be prohibitive), several single devices have to be individually instrumented and tiled to form the desired surface. This increases complexity and, potentially, power.
Despite the difficulties just mentioned, photodiodes and APDs find their application as readout elements for compact, low power, portable detectors and whenever readout mass needs to be minimized for background radiation considerations.

In recent years another photosensor has been developed that has the potential for being a valid replacement for PMTs. This new device is the focus of this work and will be the subject of the next chapter.

Chapter 3

Silicon Photomultipliers in Radiation Detection

3.1 Introduction

From the discussions in the previous chapter, it is clear that radiation detection via scintillation light involves very low levels of light. Thus, the photosensor has to be either very efficient, or provide internal gain, so that the charge photogenerated is amplified to obtain high signal-to-noise ratios (SNR). PMTs do this well, but they have some shortcoming; silicon-based devices address most PMT shortcomings, but have no- to modest gain, so that the instrumentation design is complex due to poor SNR. In the late 90s new devices were envisioned that could offer the high gain of PMTs and some of the advantages of silicon photodetectors. Such devices are known as *silicon photomultipliers*, or SiPMs in short. These devices are cleverly based on small (~10 μm-sized) elementary cells that respond to a single light photon through an avalanche in a very similar way to Geiger-Mueller counters. If several such elementary cells are constructed on a larger area, they are able to detect the energy deposited in the scintillator in a proportional way. To intuitively show this, let us observe that, as mentioned

previously, the amount of light produced in response to a radiation event is proportional to the radiation energy itself. This means that for an energy E, N photons are nominally generated such that $E = KN$. This light is generally spread into a 4π solid angle into the scintillator crystal. The light will directly hit the photodetector or will reach it after being reflected off the crystal boundaries[1]. If $N = 65,000$ photons are produced by a 1 MeV event into 4π, such as in the case of NaI(Tl), the light per unit solid angle is $65,000/(4\pi) = 5,175$ photons/sr; on the other hand, this is also the number of photons in a 10 mm x 10 mm area, 10 mm away from the interaction point. Thus, at this distance, the photon surface density is less than one photon for every square micron. The QE and the optical coupling of the scintillator to the photodetector further reduce this number. At these densities, an array of ~10 μm square elementary elements properly spaced would have a chance of intercepting every single photon (similar considerations are valid also for the reflected light). If every element of the array responds with the same amount of charge (Geiger mode) to each single photon, the sum charge will then be proportional to the number of photons generated, that is, the original energy deposited in the detector. Since the Geiger discharges produce a large amount of charge for every single photon, we have the potential for achieving high overall gains. Figure 3.1 shows a conceptual geometry of how such an arrangement could be implemented.

The light entering the transparent window interacts within the depletion region (usually an epitaxially grown intrinsic layer), generating electron-hole pairs. The electrons drift towards the avalanche regions, where they are able to start Geiger avalanches (this is determined by the thickness of those regions and the intensity of the electric field). The top electrodes (usually highly doped implants) collect the sum charge. A resistor (made internally to the device with a polycrystalline silicon layer) is placed at the output in order to limit the maximum charge generated in the avalanches and avoid possible damage to the device itself. The structure depicted in

[1]Here we suppose the crystal is encapsulated in a reflector or a diffuser material. This is common practice in constructing scintillator detectors.

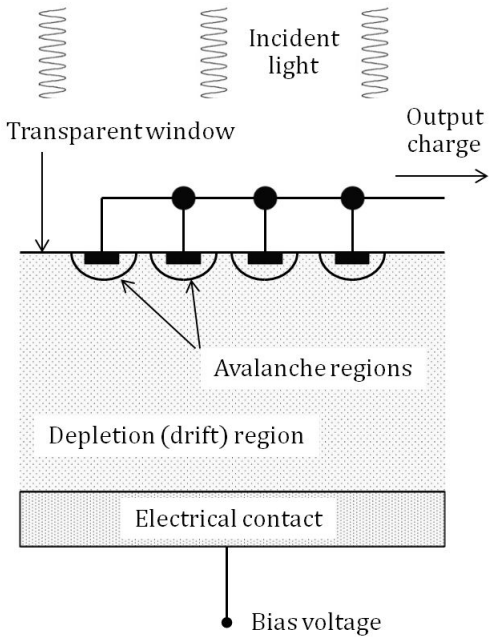


Figure 3.1: Conceptual geometry of SiPM cells

Figure 3.1 was actually the one employed for the construction of the first SiPMs, and successfully demonstrated gains of 10^4 [19]. In recent years, optimized structures have been designed that allow gains in excess of 10^6, comparable or higher than those found in PMTs.

3.2 Design of Silicon Photomultipliers

Recently, viable SiPM designs have evolved into more sophisticated structures than the one shown in Figure 3.1. The new designs have significantly improved the overall performance of the devices to the point that they are now seriously considered as PMT replacement in many applications. Fig-

ure 3.2 shows the typical design of a modern SiPM cell [20]. The cell size is

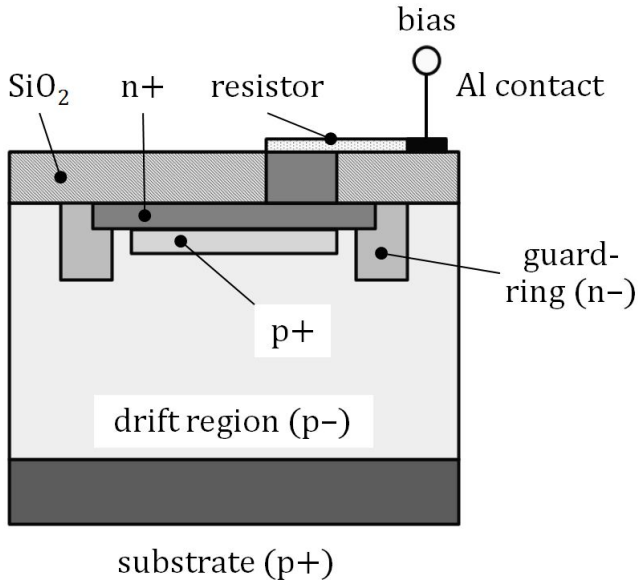


Figure 3.2: Modern SiPM cell design.

optimized at about 40 μm × 40 μm. This size seems to be optimal: while a small cell size is desirable to be sure that at high photon density the device response remains linear (multiple photons hitting the same cell at the same time would generate the same amount of charge as a single photon), the percentage of active area within a very small size cell would be insufficient to lead to an acceptable overall QE. Several cells are connected in parallel through the Al contact running across the device to constitute *pixels*. The maximum achievable surface depends on process yield for the most part. A standard size is 6 mm × 6 mm, although it is not impossible to achieve 1 cm^2 [21]. The drift region is formed by the p^- epitaxial layer, just a few micron thick. The electric field in this region is fairly low. The majority of the field is sustained within a thin depletion layer formed between the n^+

and p^+ contacts. In this region, the electric field is concentrated to values up to a few V/cm, and the conditions for Geiger discharges are present. In order for this to be true, the device is held a few volts above the breakdown voltage (usually 1 to 5 V; this extra voltage is called *overvoltage*). The n^- guard ring maintains field uniformity within the Geiger region and the polysilicon resistor provides protection against over discharge.

A further improvement from this structure is to implement the Geiger region on top of a mesostructure, so to minimize crosstalk between cells [22]. This allows for greater cell density and better QE.

Each cell can be identified by a meaningful set of parameters. These parameters, ultimately, influence the overall SiPM performance.

Cell efficiency

The cell efficiency for generating charge is the product of the probability of a free carrier to be created by a photon within the drift region (QE of Silicon) and the probability that a carrier triggers a Geiger avalanche within the Geiger region, ϵ_G [19]. The QE is typical of silicon and it also depends on how transparent the structure is at a given wavelength. The probability ϵ_G instead, depends on where within the geometry the electric field is high enough to create a Geiger discharge; ϵ_G can be increased by improving the geometry and the field, thus it is a function of overvoltage. Overall, we can write:

$$\epsilon = \mathrm{QE} \cdot \epsilon_G$$

The value for ϵ has been measured to be as high as 0.8 at 670 nm. Note that this is not the overall QE of the device, as the arrangement of cells within a pixels, and pixels within devices leaves a percentage of inactive area, reducing directly the overall QE.

Light Gain

The gain of a single cell, G_C, is equal to the number of charges created by a single Geiger avalanche, usually 10^5–10^6, limited by the quenching resistor. In SiPMs, there is excellent gain uniformity from cell to cell for a given bias, with gain dependence on the bias itself through ϵ_G [20]. In the case of many photons (N_{ph}) illuminating the device, we expect $\epsilon \cdot N_{ph} \cdot G_C \cdot \frac{A_{cell}}{A} = N_{ch}$ charges to be created (assuming no losses due to multiple photons hitting the same cell). The ratio $\frac{A_{cell}}{A}$ is called geometric efficiency and depends on how much area is filled by active cells compared to the total area. If we recall that $\epsilon = QE \cdot \epsilon_G$, then the *light gain*, or effective gain, G_e is $G_e = \text{QE} \cdot \frac{A_{cell}}{A} \epsilon_G \cdot G_C = \eta \epsilon_G G_C$. In this expression, η is an effective quantum efficiency derived from the intrinsic QE and a geometrical factor, and is constant for a given device.

G_e and G_C can be independently measured. For example, G_C can be

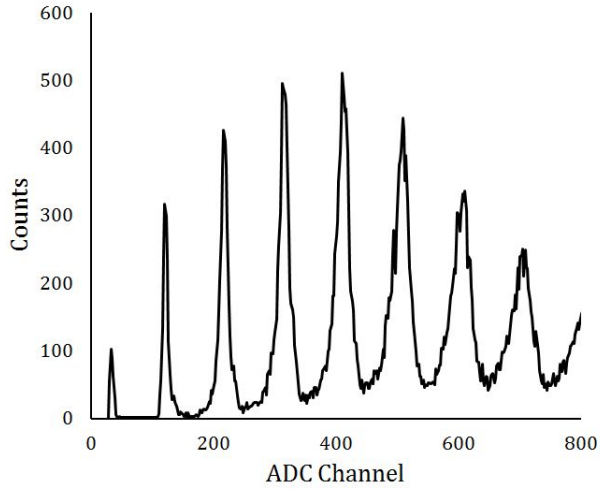


Figure 3.3: Histogram of SiPM response to very low levels of light. The peak around the origin is the electronic noise of the data acquisition system. The other peaks correspond to n photoelectrons (here $1 \leq n \leq 7$).

easily measured by observing the response of the device to very low light levels, down to a single photon, or by observing that a thermally-generated electron is no different than a photo-generated electron, and that this is the minimum attainable signal (no less than one unit charge can be generated within the device!). G_e can be measured quite easily by illuminating the cell with a monochromatic beam of known intensity. From these measurements, ϵ_G can be determined. It is interesting to point out that the statistical characteristics of the cell's response to low numbers of photons (1, 2, 3, ...) are quite remarkable and indicative of the great gain uniformity for these devices: measured responses have Gaussian shapes with resolutions in the few percent, allowing easy determination of the peak positions. An example is given in Figure 3.3.

Dark Noise

Since SiPMs are based on reversed-biased Silicon p–n junctions, they are expected to present a degree of thermally-generated carriers both in the drift region as well as in the Geiger region. These carriers contribute signals that are virtually identical to those generated by the light, and this presents a significant problem: while the devices offer great potential for detecting low levels of light, this can be completely negated by the presence of thermally generated carriers. The rate of thermal generation is in the 10^5 Hz/mm^2, but it is strongly dependent on temperature (for example, such rate decreases to a few 10's Hz/mm^2 at -100 ℃) [23]. At room temperature SiPMs are, in fact, quite useless for very low light detection, especially in large areas, unless coincidence and segmentation techniques are used. Often the dark noise is characterized either as a leakage current or, more commonly, as dark count rate (DCR) in Hz/mm^2.

Timing

The Geiger discharges develop over a very short time, about 10^{-10} s, therefore, the response of SiPMs is quite fast as well [19]. This offers timing capabilities that are quite superior to those of PMTs. The time resolution

is found to be well below a ns and dependent on the number of cells fired as $1/\sqrt{N_{cell}}$. This is why one of the major drivers of SiPM research is the medical field, in PET applications [24].

Cell recovery

The response time only accounts for the instant between a photon detection and the time when the charge is presented at the device output. It does not account for how long a cell takes to restore the charge generated. Once the charge has been generated, the cells, just like capacitors, have to somehow recover to the bias voltage level before they can function again. The recovery process happens through the quenching resistor in each cell, therefore there is a time constant involved. Components of this time constant are, in the most optimistic case, the quenching resistor and the cell's capacitance. If the cell size is chosen accurately, so that the capacitance is in the 100's fF, such time constant is usually found to be about 10 ns or less. It is not unusual, however, to see commercial devices that have recovery times one order of magnitude longer [25].

Capacitance

Reverse-biased junctions behave from the electrical point of view like capacitors. This is perhaps the most striking difference between SiPMs and PMTs. Photomultiplier tubes have virtually no capacitance, independent of size, due to the fact that the anode is simply a small plate or wire in vacuum, and there is a tangible distance between it and the other components. SiPMs cells on the other hand, are very small, but they also have very thin junctions. A single cell has low capacitance, but the overall device, composed of several thousand cells, presents capacitance values of the order of 30 to 100's pF/mm^2. This in general is not a problem thanks to the high signal gain of SiPMs, but under some conditions it can become a significant hurdle in instrumenting large areas of SiPM.

Optical crosstalk

In the absence of light, one would expect to detect only single photoelectron responses due to dark counts. In reality, carriers generated by an avalanche emit photons in the near-infrared when accelerated by the electric field. These photons can travel from the original cell to other cells, and occasionally initiate secondary discharges. Since the process evolves on a very short time scale, the result is to observe, even in the absence of incident light, signatures for 2, 3 or more photoelectrons besides the expected single photoelectron response [23]. The phenomenon is called optical crosstalk and is clearly a problem when dealing with low light levels. The optical crosstalk depends mainly on device geometry (cell distance and optical isolation) and overvoltage. Optical crosstalk is easily measured by histogramming the SiPM response in a dark environment. The plot in Figure 3.4 is a good example of such response, showing the single photoelectron peak as well as peaks for 2, 3, 4 and 5 photoelectrons due to crosstalk.

Afterpulse

Occasionally, charge carriers are trapped after photo-generation by impurities within the cell and are released at a later time [26]. This creates avalanches that are correlated to dark counts or detected photons, and has the effect of increasing the total dark count rate. Afterpulse is dependent mostly on the technology the devices are based upon. Afterpulse measurements are not trivial. Usually, a light pulse is used to illuminate a SiPM; a trigger is generated whenever a light pulse is produced. An electronic gate of variable length is opened upon a trigger signal and events are recorded within the gate duration. Without afterpulse, no events should be observed after the main light trigger; therefore by this procedure the average number of afterpulses and their characteristic time constant can be determined.

Spectral Response

It is the quantum efficiency of silicon that makes SiPMs quite desirable. Typical values are 85% around 700 nm, extending from about 400 nm to 900

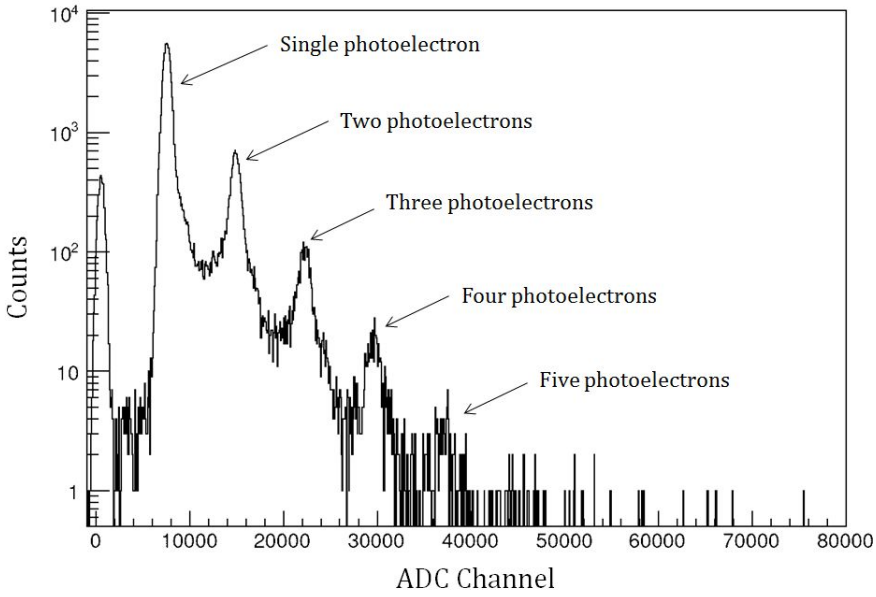


Figure 3.4: Histogram of SiPM dark response showing optical crosstalk effects. The peak around the origin is the electronics readout noise.

nm. This compares quite favorably to PMTs, where the peak value is about 35~40% for the most advanced photocathodes, with narrower wavelength spans. However, what really matters from the device point of view is the overall photon detection efficiency (PDE), that is, how efficient the device ultimately is in detecting photons. As mentioned previously, there are windows and dead areas; therefore SiPMs do not achieve the true potential offered by silicon. Typical PDE values, directly from Hamamatsu's catalog [27] (a major photosensor manufacturer), are shown in Figure 3.5. As it can be seen, typical values are about 25% (up to 35%) at 450 nm peak wavelength. The efficiency depends also on overvoltage.

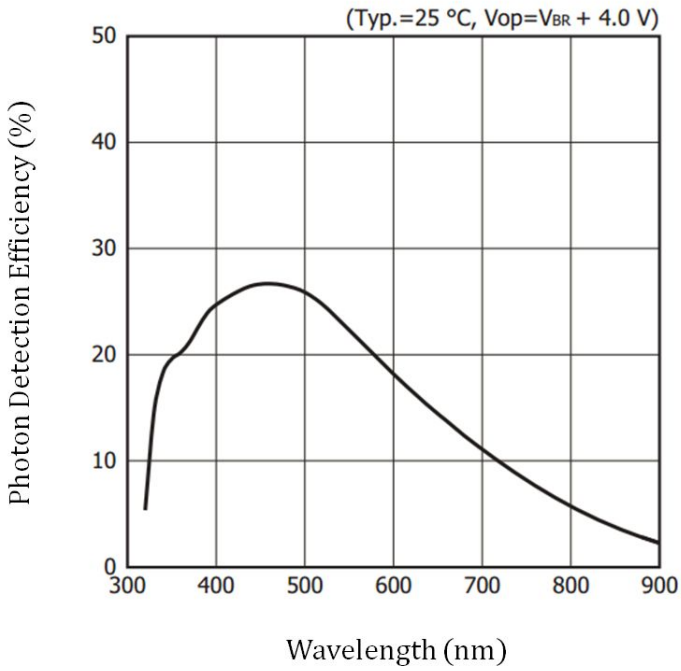


Figure 3.5: Typical photon detection efficiency for a Hamamatsu S12572 – 015C SiPM.

3.3 Commercial Efforts

Thanks to the continuous interest by the medical imaging community, in the last three to four years there has been significant progress on SiPM technology by manufacturers. Medical imaging applications require that photosensors have fast response in a compact, low-power package, and are immune to magnetic fields. SiPMs definitely address these requirements, but as seen previously have potentially significant drawbacks. Manufacturers are or have been addressing these shortcomings by improving specifically the low dark count rate, cell density (or *fill factor*), cost, optical crosstalk,

afterpulse, and parameter uniformity to the point where SiPMs are quickly replacing PMTs in small animal imaging and pre-clinical studies. It should be noted that device capacitance is not among the parameters being improved. The reason for this and the consequences to instrumentation will be addressed in the following chapters.

Recently, the four major SiPM manufacturers, KETEK, SensL, FBK

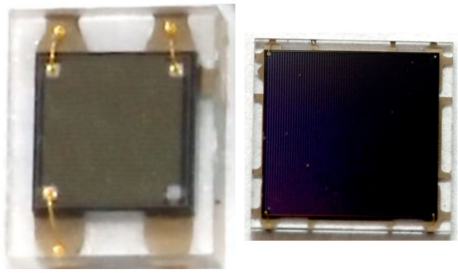


Figure 3.6: Comparison between a standard SiPM (left) and one employing Through Silicon Via technology (right).

and Hamamatsu have all presented very interesting solutions that have improved SiPM performance (especially QE) significantly. The new devices being introduced have QE as high as 60% at 420 nm. The dark count rate is down to less than 100 kHz/mm^2, with optical crosstalk probability reduced down to a maximum of 3% from the original values of 20%.
Afterpulsing is almost down to negligible levels (1% vs. 8%) even at the highest values of overvoltage. Cost and parameter uniformity have been improved by scaling production to large wafer sizes (8 inches). Thanks to the use of modern CMOS technologies, fill factors have increased as well. All four manufacturers are now integrating the use of Through Silicon Via (TSV) technology to reach 85% fill factors for the next generation devices. TSV technology allows the connections of cells to their substrates without using wire bonds. This enables a great reduction of inactive areas between

cells increasing consequently the amount of active area per total device area.
Figure 3.6 shows this situation for two SensL devices of comparable size.

Chapter 4

Readout Design Criteria for SiPM-Based Detectors

4.1 Introduction

Many novel detectors have been designed with the promise of being revolutionary, based on some sets of their anticipated performance. However, in many cases, their widespread adoption has been consistently hindered once they have been introduced into a real system. This often happens because the researchers who design detectors are not the same that design systems. One cannot determine the actual performance of a new detector until he or she has dealt with the instrumentation required to read it out. In other words, detectors cannot be considered as stand-alone devices, but have to be designed in a system context. For example, avalanche photodiodes seemed to be quite promising: they have enough gain to ensure good SNR, they have very high QE and leverage well-known silicon technologies. This is all good in theory, however, in practice, their high leakage current, and the need for very stable bias voltages introduce stringent requirements on the instrumentation, and often a PMT is preferred in their place. Another example is recent advancements in novel neutron detectors based on silicon

pillar structures acting as diodes or high aspect-ratio holes in a diode silicon substrate, filled with neutron-sensitive materials [28], such as those shown in Figure 4.1 [29]. The filler can be ^{6}Li or ^{10}B. When capturing a neutron, the filler emits alpha and Lithium particles that are detected within the silicon, thus signaling the presence of a neutron. In principle, these devices could work exceptionally well, with good detection efficiency ($> 20\%$) and large areas. In reality, the associated leakage and device capacitance make the instrumentation design extremely challenging for large areas.

SiPMs are no different. They are presently thought as complete replace-

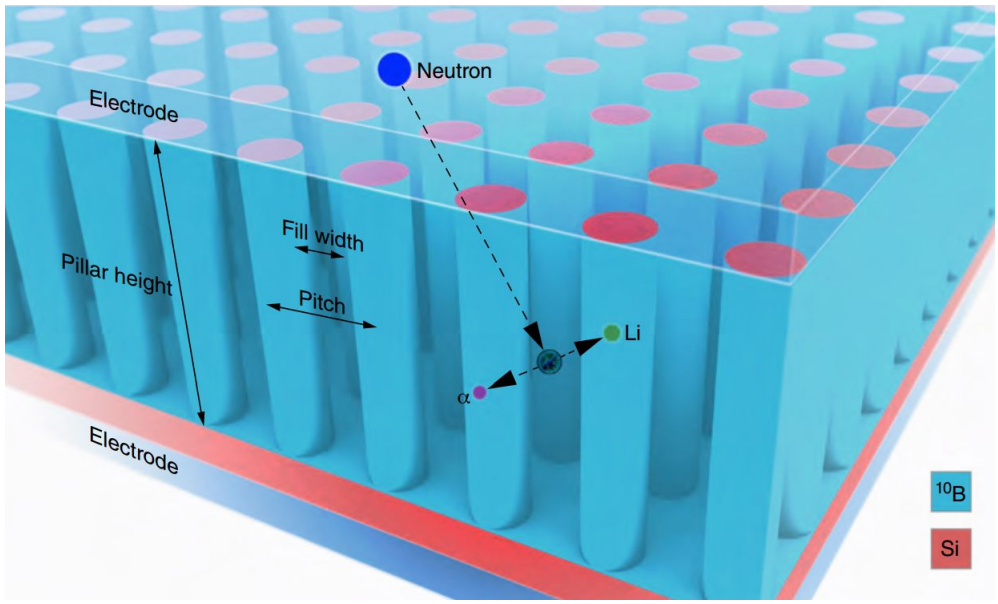


Figure 4.1: Si-pillar neutron detector. A different version has holes in lieu of pillars and uses ^{6}Li instead of ^{10}B as neutron-sensitive filler. (Image courtesy of Rebecca J. Nikolic, Lawrence Livermore National Laboratory)

ments for PMTs, but we will show that, as soon as they are thought of as part of a system, this no longer holds true under all conditions, and

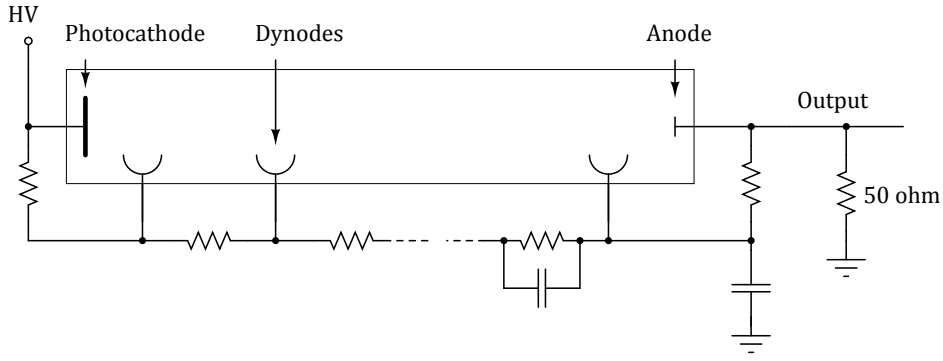
the instrumentation thought process needs to be revised. To understand this fact, some fundamentals of light detector instrumentation have to be introduced. We will examine the traditional PMT readout instrumentation first in order to draw parallels to SiPM instrumentation when appropriate, or to highlight discrepancies and differences in all other cases.

4.2 PMT Readout Instrumentation

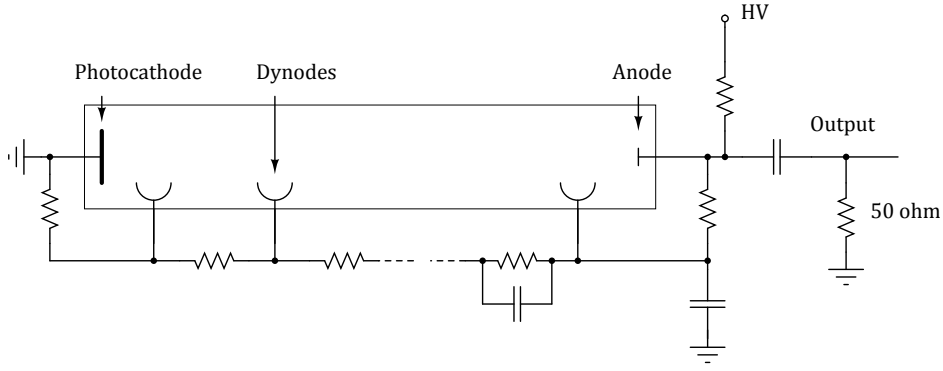
Photomultiplier tubes are the easiest photodetector to design instrumentation for. Their high gain and the absence of leakage currents of any kind, plus their extremely low output capacitance, make them almost ideal devices. From what said in Chapter 2, PMTs require an electric field inside the vacuum tube that is able to accelerate photoelectrons between dynodes, all the way to the anode. This is accomplished very easily to fist order, by providing a resistive divider network from the photocathode to the dynodes, all the way to the anode. The photocathode has to be held at a lower potential than everything else in order to allow electron drift in the right direction, therefore two possibilities exist: apply a negative voltage to the photocathode while grounding the anode, or apply a positive voltage to the anode while keeping the photocathode grounded. This is what is commonly done, with the choice for polarity dictated mostly by the necessity of whether or not an output AC coupling is tolerable, or a grounded photocathode is necessary (for example, if the PMT face is in contact with the metallic structure of the detector encapsulation, it must be at ground potential to prevent either electrical shock or long term photocathode damage due to ionic currents migrating through the face of the glass).

Often capacitors are added to the last dynodes because the charge required by these stages may be relatively high, to the point of altering momentarily the DC bias conditions, leading to non-linearity for large signals.

It is not uncommon to find different arrangements for the dynode bias. When an independent control of the dynodes bias is desired, the divider is implemented through taps on a Cockcroft-Walton power supply [30], rather than via resistive dividers.



(a)



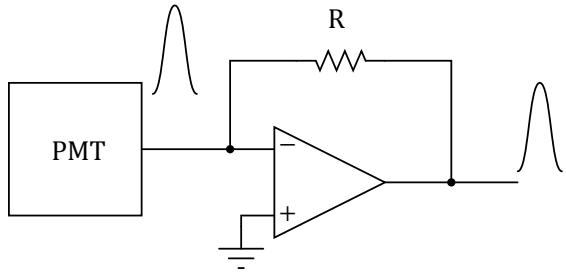
(b)

Figure 4.2: PMT biasing schemes. Negative high voltage (a) and positive high voltage (b).

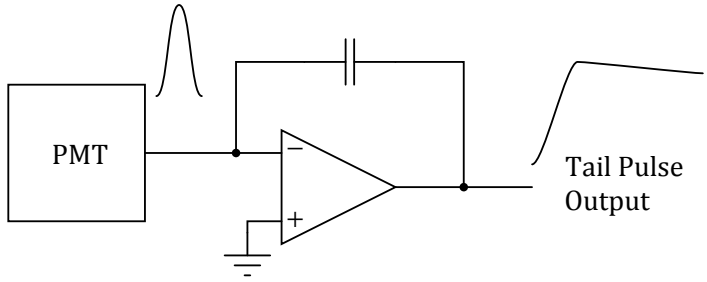
Figure 4.2 shows biasing schemes for negative and positive bias. Typical bias values are 800 to 3,000 volts.
Once a bias is established, the charge collected at the anode has to be read out by appropriate electronics circuits. The most common method is to simply terminate the anode output on a 50 ohm resistor to ground, so the charge can develop a voltage across it. Of course, any resistor value would work; however, 50 ohm is the preferred choice because it matches the impedance of the most popular coaxial cables, thus allowing signal transfer over considerable lengths without any additional effort.
Since the PMT has virtually no electronic noise associated, the lower limit in detectable voltage on the output resistor is either set by the amplitude of single photoelectrons, the statistics associated to the charge multiplication, or the noise of the subsequent readout system.

More sophisticated ways of reading out PMTs that involve active devices are also used. In the most popular, the output resistor is replaced by a virtual ground of an operational amplifier. The voltage gain is developed across a feedback capacitor or resistor (see Figure 4.3). These are normally known as charge or current amplifiers respectively. The current amplifier offers the advantage of allowing for larger resistor values, thus higher voltage gains, without incurring the penalty of slower signals due to cable and parasitic capacitances. The charge amplifier, on the other hand, integrates the signal on the feedback capacitor, generating what is known as a tail-pulse response at its output. This allows the direct use of standard signal processing blocks[1] at the output to complete the acquisition chain. An additional benefit of using a charge amplifier is that the achievable gain is quite elevated, therefore lower voltage bias is necessary to achieve the same output amplitude. Lower bias means longer PMT operating life as well as lower design complexity.
While these circuits add their own noise components, the signal-to-noise ratio is hardly affected in the case of PMTs (this fact will be clarified later

[1]Such as conventional NIM electronics. These systems implement standard signal processing functions based on charge amplifier signals.



(a)



(b)

Figure 4.3: PMT biasing schemes. Negative high voltage (a) and positive high voltage (b).

on), therefore from the SNR point of view the readouts are completely equivalent; the choice of a particular scheme over another is dependent on experimental needs, not SNR considerations.

Given that PMTs and SiPM have comparable gains and quantum efficiencies, one would expect being able to directly migrate PMT-specific designs to SiPM applications. Careful studies have proven that this holds true only for some applications. The type of application in which traditional PMT designs criteria apply to SiPM design is driven by the amount of light expected to be collected per unit area. We will see that when the specific light intensity is high enough, traditional techniques apply. When the intensity is very low, new techniques have to be introduced. Both cases are equally important in radiation detection problems.

4.3 Definition of "light regime"

We already mentioned the fact that light production in a scintillator is a statistical process: upon material excitation by radiation, photons are emitted by the decay from excited states; light emission follows a statistical process in which the population of excited particles decays to the ground state within an average time called the decay or relaxation time of the scintillator excited levels. If the excited particles are a very large number, the statistical aspect of the decay process is very small, so that the light produced perfectly follows the average response of the scintillator, as illustrated in Figure 4.4. In other words, at any given time, the amount of photons generated is equal to what is expected from the exponential decay of the scintillator. On the other hand, a low energy event will generate a relatively low number of photons, and the associated statistics may not be negligible. In fact, the solid line in Figure 4.4 shows this situation: instead of a continuous response, photons are produced in clusters during the scintillator's relaxation time as can be seen in the small peaks found on the trailing edge of the pulse. This holds true all the way to one photon produced, as the extreme case of Figure 4.5, showing a scintillator response

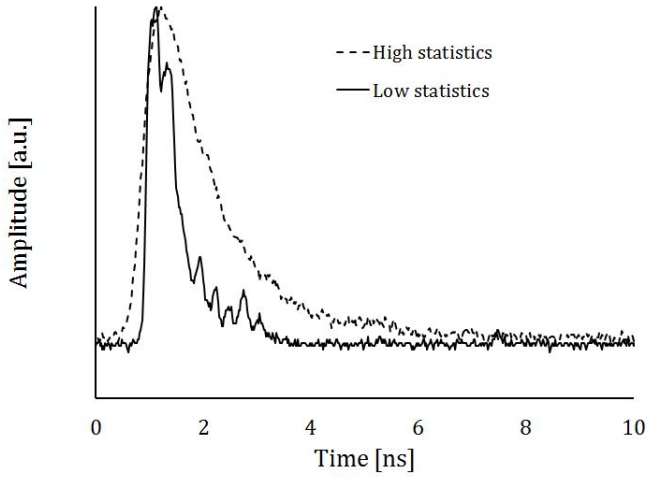


Figure 4.4: Typical scintillator response (arbitrary units) to a low energy event with low statistics (solid line) in comparison to a high statistics curve (dashed line). The vertical scale is normalized to the peak

of only a few tens of photons. The plots just shown highlight an important property of the scintillator time response: if many photons are in play, the time response has an exponential behavior with the decay time constant we are already familiar with. The nature of this time constant was discussed in Chapter 2 and it was defined as the relaxation time of the scintillator. As we enter a regime of low photons, instead, the response is made up of clusters of photons distributed in time according to a probability law described by a function that is identical to the scintillator response to large number of photons. To first approximation, if τ_0 is the only time constant, we have, for this probability (of arrival):

$$P(t) = A\mathrm{e}^{-\frac{t}{\tau_0}}$$

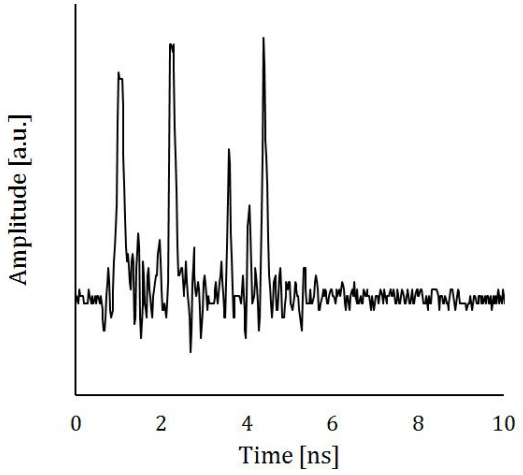


Figure 4.5: Scintillator response to a very low (a few tens photons produced) energy event.

With the constraint that the integral must be equal to the total number of photons N. This leads to $A = N/\tau_0$, or:

$$P\left(t\right) = \frac{N}{\tau_0}\,\mathrm{e}^{-\frac{t}{\tau_0}}$$

with

$$\int P(t)dt = N$$

The formula states the probability of a photoelectron to arrive between $\bar{t}$ and $\bar{t} + \Delta t$ after an interaction has occurred. It does not give the actual overall shape of the response. That can be obtain by convolving $P(t)$ with the shape of the single photoelectron response (ideally a delta function). The first expression will allow us to identify two extreme regimes.

4.3.1 Photon starved regime

In a situation where only a few photons are produced, the use of $P(t)$ must be modified to account for the fact that there is only a small discrete number of photons in play. We can assume that the probability that a photon be emitted after it populates an excited sate is equal for all photons[2]. The overall signal will be the overlap of individual photon decays, each obeying the probability distribution:

$$p_i\left(t_i\right) = a\mathrm{e}^{-\frac{t_i}{\tau_o}} = \frac{1}{\tau_0}\,\mathrm{e}^{-\frac{t_i}{\tau_o}}$$

Since only N photons are generated, the signal will now be sum of all single photon responses, each with its own arrival time t_i. The response of the scintillator is no longer represented by a continuous function, but rather by a discrete sum of the type:

$$f\left(t\right) = \sum\nolimits_i \frac{b}{\tau_0}\,\mathrm{e}^{-\frac{t}{\tau_0}} * s(t)$$

where $*$ indicates a convolution and $s(t)$ is the single photoelectron time response. Since the sum of all responses must be equal to the total charge Q generated by the radiation event, we have, more specifically:

$$f\left(t\right) = \sum_{i=1}^{N}\left(\frac{b}{\tau_0}\,\mathrm{e}^{-\frac{t}{\tau_0}}\right) * s(t) = b\sum_{i=1}^{N}\left(p_i\left(t_i\right) * s(t)\right) = bN = Q$$

that leads to $b = Q/N = q$, or one electronic charge[3] ($=1.6 \cdot 10^{-19}$ C).

This result is very important, because the ability of distinguishing the single photons is now exclusively dependent on the single photon response shape and on the individual time of arrival of the single photoelectrons. Under

[2]Since this probability depends on τ_0, the relaxation time of the scintillator states, it must be an intrinsic parameter independent on the number of photons.

[3]Note that for simplicity we are neglecting all effects for which more than one photoelectron is generated.

ideal conditions, the single photoelectron shape is a delta function with no width. With real detectors, such as PMTs or SiPM, this is typically a fast exponential response, combined with the noise of the electronics system that reads the detector output. The time scale over which the phenomena occur is only a fraction of the relaxation time of the scintillator. We will refer to this situation as *photon starved regime*.

4.3.2 High signal-to-noise regime

When the incident energy is large enough, $N \gg 1$. In this case the expression found for $P(t)$ describes the time response as well because the time scale is that of the scintillator relaxation time, and the single photoelectron response is always approximated by a delta function. The output waveform will then have the form:

$$f\left(t\right) = \frac{Nq}{\tau_0}\,\mathrm{e}^{-\frac{t}{\tau_0}}$$

We will define this situation as *high SNR*. In this case, any electronics noise will be convolved with this expression, not with the single photon responses. While this noise is still present, it is now often negligible compared to the overall fluctuation on the number of charges generated within the scintillator (its resolution), that is, the variance associated with N.

We can actually set a boundary to distinguish between the two regimes just introduced. The number N comes with its variance, σ_{res}, due to the charge generation process. After readout, any electronics noise convolved with the scintillator response, adds another component to the original variance, σ_{noise}. The total variance on N is then:

$$\sigma_N^2 = \sigma_{res}^2 + \sigma_{noise}^2$$

Hence, a high SNR regime is identified when:

$$\sigma_{noise} < \sigma_{res}$$

This condition, while not very meaningful in PMTs because in this case is almost always true, is fundamental in designs involving SiPMs, due to

their potential for high electronics noise contributions from to their high capacitance values.

4.4 SiPM DAQ Design in Different Regimes

The first question that bears to mind is: why is capacitance so important for electronics noise? The answer can be given by looking at a simple situation. Let us consider the generic case of detector readout (this does not need necessarily to be a photodetector), where an amplifier system is connected to a generic detector. For noise purposes, the detector can be modeled as a current pulse of area Q and an equivalent capacitance C_D; the amplifier system can be modeled as a noiseless amplifier, its own input capacitance, C_i, a series, and a parallel noise source [31]. Figure 4.6 shows a schematic of these models.

It is common practice to indicate the noise sources through their spectral

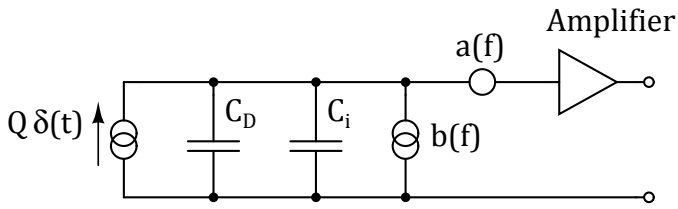


Figure 4.6: Detector readout system representation.

power densities $a(f)$ (V^2/Hz) and $b(f)$ (A^2/Hz); by using a voltage source in series to the input of the amplifier and a current source in parallel, all noise in the system is described. The series component, $a(f)$, generally contains a frequency-independent component called white noise.

This is often the thermal noise in the channel of a field-effect transistor (FET), or the noise in the collector of a bipolar transistor (BJT). Another noise component in $a(f)$ has a $|f|^{-1}$ dependence in order to account for low frequency noise in active and passive (such as dielectric materials) devices within the amplifier. The parallel component, $b(f)$, accounts for shot noise

associated to currents at the amplifier input. These currents can be gate or base bias currents of transistors as well as leakage current in the detector; they can be included in a single term because the corresponding spectral densities are uncorrelated. For most applications, $b(f)$ is independent of frequency. The overall signal at the output of the preamplifier can be easily calculated as:

$$v_{out}(t) = \frac{QA_0}{(C_D + C_i)} H(t)$$

With A_0 we indicate a constant such that the maximum of $H(t)$ over all times is 1. $H(t)$ is thus the normalized response the amplifier has to a delta function. The noise contributions, instead, have to be calculated quite differently because noise is a stochastic variable that needs to be characterized through its spectral density [32]; therefore statistical methods have to be used. With this in mind, we indicate with $A(s)$ the Laplace transform of the amplifiers response $A(t)$ and we calculate the mean square value of the noise at the preamplifier output by integrating the single spectral contributions (the power spectral densities) over all frequencies:

$$\overline{v_{n,a}^2} = \int_0^\infty a(f) |A(f)|^2 df$$

and

$$\overline{v_{n,b}^2} = \int_0^\infty \frac{b(f)}{(2\pi f)^2 (C_D + C_i)^2} |A(f)|^2 df$$

We can focus initially on the case where $b(f)$ has a negligible contribution. We will refer to this approximation as the series noise approximation or SNA in short. Under this assumption, we can indicate with $v_{out,0}$ the amplitude of the output signal corresponding to 1 eV energy.

$$v_{out,0} = \max\left\{ \frac{1\text{eV} \cdot qA_0}{\epsilon (C_D + C_i)} H(t) \right\} = \frac{qA_0}{\epsilon (C_D + C_i)}$$

By using this consideration, we can express the equivalent energy fluctuation in the detector given by the noise as:

$$\overline{E_{eq}^2} = \frac{1}{v_{out,0}^2} \int_0^\infty a\,(f)\,|A\,(f)|^2\,df$$

When the $1/f$ component in $a(f)$ can be neglected, we can assume $a(f) = a$, constant.

$$\overline{E_{eq}^2} = \frac{\epsilon^2\,(C_D + C_i)^2}{q^2}\,a\left[\frac{1}{A_0^2}\int_0^\infty |A\,(f)|^2\,df\right] = \frac{\epsilon^2\,(C_D + C_i)^2}{q^2}\,a\int_0^\infty |H\,(f)|^2\,df$$

This is the noise that establishes the condition of having a unity SNR at the output of the preamplifier. In terms of charge, this is defined as the equivalent noise charge, ENC_e, or:

$$ENC_e^2 = a\,(C_D + C_i)^2 \int_0^\infty |H\,(f)|^2\,df = a\,(C_D + C_i)^2\,R_D$$

In this expression there is a term that depends on the white noise spectral density of the amplifier and the total input capacitance, plus a term that depends on the normalized frequency response of the amplifier only, R_D, known also as residual delta function, as introduced in [33].
The equation is in direct support of the statement made several times regarding the fact that a larger detector capacitance means increased electronics noise, as highlighted by the direct proportionality of ENC_e to the capacitance. This will help in defining a high SNR versus a photon starved regime in scintillator applications.
It is also customary to represent ENC in terms of Coulomb as $ENC = qENC_e$. In general, ENC represents the noise contributed by the readout circuit: a quantity with infinitely narrow variance would appear at the preamplifier output as a Gaussian distribution with variance $\sigma_{noise}^2 = ENC^2$.

4.4.1 Validity of the series noise approximation

The magnitude of the parallel term relative to the series term determines when the series noise approximation can be used. By separately solving the expression for $\overline{v_{n,b}^2}$ [31], and assuming constant value (no $1/f$ noise present) for a, we see immediately that at low frequency the parallel term will overwhelm the series component, but at high enough frequencies, the series component will always dominate, thus there is a lower limit to the validity of the white noise approximation. Such limit is ultimately determined by the value of the parameters a, b as well as the total input capacitance. An expression for this frequency is given in [31]:

$$f_c = \frac{1}{2\pi}\sqrt{\frac{b}{a}}\left(\frac{1}{C_D + C_i}\right)$$

This particular frequency is defined as *noise corner* of the system.

We can gain a better understanding of the applicability of the series noise approximation by quantifying the noise corner. This can be done by knowing expressions for the parameters a and b. A thorough discussion on the nature of these parameters is given elsewhere [34]. Suffice to say that both parameters depend on the type of device used to implement the amplifiers input (BJT, JFET, MOSFET), and that a may implicitly be dependent on C_i, while b is not. Table 4.1 gives a general expression for a and b in these devices.

In the table, k is Boltzmann's constant ($1.38 \cdot 10^{-23}$ J K^{-1}), q the electron's charge, T is the temperature and g_m is the transistor's transconductance (expressing the ratio between a variation in voltage applied to the input and the subsequent variation in transistor output current). The values of the parameter Γ are 0.5 for BJTs and 0.7 for FETs. The transconductance is the parameter through which a may depend on C_i. Typical values of g_m are 40 mA/V for every 1 mA of current in BJTs, and 0.01 mA/V to 100 mA/V for JFETs and MOSFETs at different currents. Typical values for I_b are 0.1 μA, while I_g is in the pA range. With these numbers is mind,

Device	a [V^2/Hz]	b [A^2/Hz]	Note on g_m	Max. f[Hz]
BJT	$\frac{4kT\Gamma_B}{g_m}$	$2qI_B$	$\frac{qI_c}{kT}$	>10 GHz
JFET	$\frac{4kT\Gamma_F}{g_m}$	$2qI_G$	Size-dependent	A few 100 MHz
MOSFET	$\frac{4kT\Gamma_F}{g_m}$	-	Size-dependent	GHz

Table 4.1: Expressions for a and b in common active devices.

we find that:

$$f_{c,BJT} = \frac{6.3 \cdot 10^{-5}}{(C_D + C_i)} \tag{4.1}$$

$$f_{c,JFET} = \frac{2.7 \cdot 10^{-8}}{(C_D + C_i)} \tag{4.2}$$

$$f_{c,MOS} = 0 \tag{4.3}$$

It should be noted that there are other factors affecting the expressions above that have been neglected, especially in MOSFETs, where the actual f_c can be in the MHz range due to a high degree of $1/f$ noise (a term we ignored from the beginning). In order to gain meaning, the noise corner frequencies should be compared to the frequencies a device can operate and the frequency form of $H(f)$; these limits ultimately determine the whole amplifier's bandwidth. Typical values of maximum operating frequencies for transistors are shown in the last column of Table 4.1. As for bandwidth limitations introduced by $H(f)$, it depends on the application. In some applications it is convenient to adopt bandwidths that go from 50 kHz all the way to 20 MHz; in other applications, no filtering is imposed and the

bandwidth is limited only by the amplifier's design.

The most important observation that can be gathered is that JFETs, though slower, have relatively low corner frequencies, thus are most suitable for low frequency measurements, while BJTs and MOSFETs will be suitable for high frequency applications where high detector capacitance is in play (these are also intrinsically faster devices).

4.4.2 Resolution-equivalent transconductance

With an expression for the amplifier's noise contributions, we can now extend the definition of the limit at which a system operates in either high SNR or photon starved mode to identify practical parameters. Let us recall that the condition for being in high SNR regime was found to be $\sigma_{noise} < \sigma_{res}$. The detector resolution, σ_{res}, is usually given in electronvolts.

$$\overline{E_{eq}^2} = \frac{\epsilon^2 \left(C_D + C_i\right)^2}{q^2} aR_D < \sigma_{res}^2$$

using the series noise approximation. We can substitute for a and find:

$$\overline{E_{eq}^2} = \frac{\epsilon^2}{q^2} \frac{4kT\Gamma}{g_m} \left(C_D + C_i\right)^2 R_D < \sigma_{res}^2$$

This formula can be used to determine the amplifier's active device parameters to ensure work in the desired regime. If we solve for the transconductance:

$$g_m > \frac{4kT\Gamma}{\sigma_{res}^2} \frac{\epsilon^2}{q^2} \left(C_D + C_i\right)^2 R_D$$

For a given detector capacitance, in order for the front-end noise to be acceptable, we must ensure a transconductance higher than the limit above. This automatically becomes a design parameter, as the transconductance for a given device is set by its drain or collector current, thus it is a contribution to the overall system's power dissipation. This fact is extremely important in applications where large numbers of channels are involved.

It is useful to define the resolution-equivalent transconductance (RT) as:

$$g_S = \frac{4kT\Gamma}{\sigma_{res}^2}\frac{\epsilon^2}{q^2}(C_D + C_i)^2 R_D$$

In other words, this is the minimum necessary transconductance the amplifier's front-end must have in order to minimize its noise contribution to the detectors intrinsic resolution.

We can generalize the use of this parameter to applications that go beyond the use of scintillators. In fact, although we started by examining different light regimes, we never assumed the charge came from a light conversion process.
As an example, let us apply our considerations to the case of high-purity Germanium detectors (HPGe). The intrinsic statistical broadening introduced on the energy by the detector is proportional to the square root of the energy through the Fano factor and the ionization energy [35] ϵ $(= \sqrt{F\epsilon E})$. By substituting in the expression we obtained for g_S, we obtain:

$$g_{S,HPGe} = \frac{4kT\Gamma}{\sigma_{res}^2}\frac{\epsilon^2}{q^2}(C_D + C_i)^2 R_D = \frac{4kT\Gamma}{q^2 FE}\epsilon\,(C_D + C_i)^2 R_D$$

The value of R_D is unknown until a filter is selected. Let us assume the simplest situation of an RC differentiator followed by an RC integrator with the same time constant $\tau_0 = RC$.
In this case ($A_0 = \mathrm{e}^{-1}$), being:

$$R_D = \int |H(f)|^2\, df = \mathrm{e}^2 \int_0^\infty \left[\frac{2\pi f\tau_0}{\left(1 + 4\pi^2 f^2 \tau_0^2\right)}\right]^2 df = \frac{\mathrm{e}^2}{8\tau_0},$$

we have:

$$g_{S,HPGe} = \frac{kT\Gamma \mathrm{e}^2}{2q^2 FE}\epsilon\,(C_D + C_i)^2\frac{1}{\tau_0}$$

When we read out such a detector with, for example, $C_i = 10$ pF and $\tau_0 = 15$ μs, we find the curves in Figure 4.7. These curves tell us that the

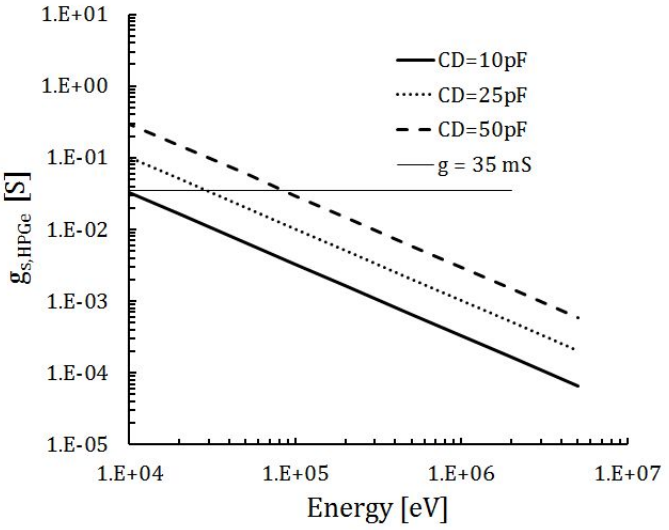


Figure 4.7: Resolution-equivalent transconductance in presence of white series noise for HPGe detectors of different sizes.

minimum acceptable trasconductance is determined by the minimum energy we want to detect and varies with detector size (C_D). It is important to notice that choosing C_i is equivalent to selecting the preamplifier's input device. Sometimes there is no device that for a set C_i is able to generate the transconductance needed. In those cases (such as for many applications using coaxial HPGe detectors), the actual resolution will be dominated by the amplifier's noise at the lowest energies. The horizontal line in Figure 4.7, in fact, shows a reasonable value for a suitable JFET with $C_i = 10$ pF. The transconductance values attainable for such device are of 35 mS. This line intercepts the $g_{S,HPGe}$ lines at 10 keV, ~50 keV, and about 100 keV energy point, in excellent agreement with practical experience. This can also be stated by saying that above the horizontal line the resolution is set by the noise in the preamplifier, while in all other cases the resolution is limited by the curves generated for g_S.

For scintillators, depending on the device that converts the scintillation light into charge, different cases can be identified. We will explore those cases later.

The RT concept can be extended to the case where the parallel or $1/f$ noise components are non-negligible. In this case, we need to remember that the overall noise contribution from the noise sources in Figure 4.6 can be found as the sum of the single power spectral densities, and that the overall r.m.s. contribution (square) is given by the integration of the single spectral contributions over the whole frequency range:

$$\overline{v_n^2} = \int_0^\infty a\left(f\right)\left|A\left(f\right)\right|^2 df + \int_0^\infty \frac{b\left(f\right)}{\left(2\pi f\right)^2\left(C_D + C_i\right)^2}\left|A\left(f\right)\right|^2 df$$

With $a(f) = a + a_f/f$ and $b(f) = 2qI$, where I is a current (for example, the collector current of the amplifier's input transistor), and a_f is the $1/f$ noise constant, thus:

$$\overline{v_n^2} = a\int_0^\infty \left|A\left(f\right)\right|^2 df + a_f\int_0^\infty \frac{\left|A\left(f\right)\right|^2}{f} df + \int_0^\infty \frac{2qI}{\left(2\pi f\right)^2\left(C_D + C_i\right)^2}\left|A\left(f\right)\right|^2 df$$

By arguments similar to those given for the first term on the right side of the equation, we find:

$$\overline{E_{eq}^2} = \frac{\epsilon^2\left(C_D + C_i\right)^2}{q^2}\left[\frac{4kT\Gamma}{g_m}\int_0^\infty \left|H\left(f\right)\right|^2 df + a_f\int_0^\infty \frac{\left|H\left(f\right)\right|^2}{f} df + \right.$$
$$\left. + \frac{qI}{2\pi^2\left(C_D + C_i\right)^2}\int_0^\infty \frac{\left|H\left(f\right)\right|^2}{f^2} df\right] < \sigma_{res}^2$$

Once $H(f)$ is chosen, one can calculate the integrals and find:

$$\overline{E_{eq}^2} = \frac{\epsilon^2\left(C_D + C_i\right)^2}{q^2}\left(\frac{4kT\Gamma}{g_m}R_D + a_f R_f + \frac{qI}{2\pi^2\left(C_D + C_i\right)^2}R_S\right) < \sigma_{res}^2$$

or

$$\frac{\epsilon^2 \left(C_D + C_i\right)^2}{q^2} \frac{4kT\Gamma}{g_m} R_D + \frac{\epsilon^2 \left(C_D + C_i\right)^2}{q^2} a_f R_f + \frac{\epsilon^2 q I}{2\pi^2 q^2} R_S < \sigma_{res}^2$$

We already commented that the residual R_D was inversely proportional to the filter's time constant τ_0. For convenience we can explicitly indicate the proportionality in all residual terms. It results that R_f does not depend on τ_0 while R_S is directly proportional to τ_0 [36]:

$$\frac{\epsilon^2 \left(C_D + C_i\right)^2}{q^2} \frac{4kT\Gamma}{g_m} \frac{1}{\tau_0} r_D + \frac{\epsilon^2 \left(C_D + C_i\right)^2}{q^2} a_f r_f + \frac{\epsilon^2 q I}{2\pi^2 q^2} \tau_0 r_S < \sigma_{res}^2$$

Here we have $r_D = R_D \cdot \tau_0$, $R_f = r_f$ and $r_S = R_S/\tau_0$. Note that when considering a particular detector resolution, the components of the left side of the equation do not contain any energy dependence; therefore, although in a less meaningful fashion than the case of series noise approximation, one can always compute a resolution-equivalent transconductance to compare to the preamplifier's input transconductance. The difference now is that the amplifier's transconductance alone no longer establishes the noise limits and other considerations need to be made in order to minimize the additional contributions.

4.4.3 High SNR design for SiPMs and similarities to conventional PMT architectures

Before adapting our reasoning to the design of SiPM readouts, it is helpful to observe how the concept of resolution-equivalent transconductance applies to all light detectors when used in conjunction with scintillators. The results are very different from case to case, with surprising similarities to the semiconductor detector cases.

Silicon photodiodes

The RT concept is an excellent way of showing how difficult the front-end design is when using silicon photodiodes coupled to scintillators. Also in

this case, under the series noise approximation, we would like to write:

$$g_S = \frac{4kT\Gamma}{\sigma_{res}^2} \frac{\epsilon^2}{q^2} (C_D + C_i)^2 R_D$$

But it is not clear what σ_{res} and ϵ mean; in order to understand that, we need to look at the charge generation process in detail. The scintillator responds to radiation by generating a number of photons N_E for every MeV of radiation. These photons bounce around the crystal until collected at the photodiode that acts as a black body for the light. Some photons are lost to refraction index mismatch at the crystal/photodiode interface; of the photons that reach the bulk of the photodiode, only a fraction determined by the quantum efficiency of the device are converted to photoelectrons. In this chain of events, we can identify an equivalent of the ionization energy, ϵ_{Scint}, as the product of the quantum efficiency QE, the interface losses, T_i, and N_E:

$$\epsilon_{Scint} = \frac{10^6}{N_E \, QE \, T_i}$$

in eV per photoelectron. With this definition, if we want to find the number of photoelectrons produced in the photodiode for a given event of energy E, we just calculate:

$$N_{ph} = \frac{E}{\epsilon_{Scint}}$$

The constant ϵ_{Scint} has the same meaning as the ionization energy in semiconductors, so we have a convenient parameter we can use in our formula. The main difference is the value of ϵ_{Scint}, which is much higher that what is found for semiconductors (i.e. it takes more energy to produce a carrier). As for σ_{res} in the original equation, this is the scintillator resolution in eV r.m.s., according to our formulation. Such value can either be calculated or measured experimentally. In any case, from statistical arguments and from analogies with semiconductors, we know that the variance in the charge generated is $\sqrt{\epsilon E}$ (the Fano factor is this case is close to unity). At this

point we can write g_S as:

$$g_S = \frac{4kT\Gamma}{q^2 E} \epsilon_{Scint} \left(C_D + C_i\right)^2 R_D$$

Just like was done for the case of a HPGe detector, a numerical example under the same assumptions is in order. By choosing the same type of filter, we have:

$$g_S = \frac{kT\Gamma \mathrm{e}^2}{2q^2 E} \epsilon_{Scint} \left(C_D + C_i\right)^2 \frac{1}{\tau_0}$$

If we choose NaI(Tl) as scintillator, from its properties, the number of photons generated by a 1 MeV radiation event is expected to be about 38,000 (see Table 2.3). This is clearly not the actual number detected by the photodiode and converted into charge. We can expect that part of the light will be reflected at the scintillator/photodiode interface due to refraction index mismatch, and not all of the light that makes it through the interface will be absorbed in the Silicon, due to non-unity quantum efficiency. Overall, it is safe to assume that only 70% of the photons are converted into charges. By substituting these numbers in the previous equations, we find $\epsilon_{Scint} = 40$ eV/photoelectron. The capacitances of photodiodes range from 20 pF to a few nF. The resulting g_S plot is shown in Figure 4.8.

From the plot we see that only the smallest photodiodes are suitable for readout, and also in that case the preamplifier noise will be inconsequential only above 20 to 40 keV.

Avalanche photodiodes

The advantage APDs offer over photodiodes is their internal gain. We mentioned earlier that this gain is obtained by internal multiplication, therefore, from the statistical point of view it does not affect the statistics on the number of photoelectrons produced before multiplication. This allows us to use the same parameters used above, with the exception of the fact that the extra gain must be accounted for somewhere in the equation. Since we are observing the signals at the amplifier's input, it is intuitive to apply this

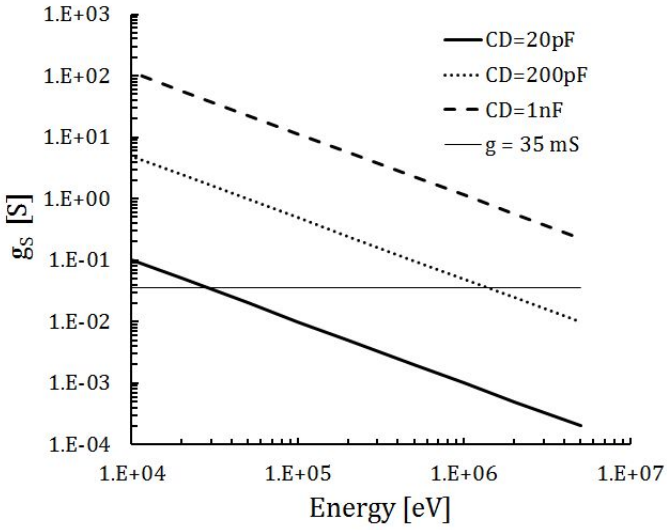


Figure 4.8: Resolution-equivalent transconductance in presence of white series noise for a NaI(Tl) detector read out by different sizes photodiodes.

gain directly to the total charge seen, therefore we have:

$$g_S = \frac{4kT\Gamma}{q^2 E} \frac{\epsilon_{Scint}}{G^2} (C_D + C_i)^2 R_D$$

Having used G to indicate the avalanche gain. With this in mind, and remembering typical APDs have capacitances in the range from 10 pF to 300 pF, and gains in the hundreds, we find the plot in Figure 4.9. This case paints a much more favorable situation than the previous examples. With the exception of the largest APD, it does not matter what the amplifier transconductance is: with some internal gain, the electronics noise does not matter anymore. Unfortunately, this is purely academic, as APDs are devices with abnormally high leakage current. This, unfortunately, invalidates the series noise approximation because the parallel term we afforded to neglect so far, is actually the dominant noise source when APDs are used

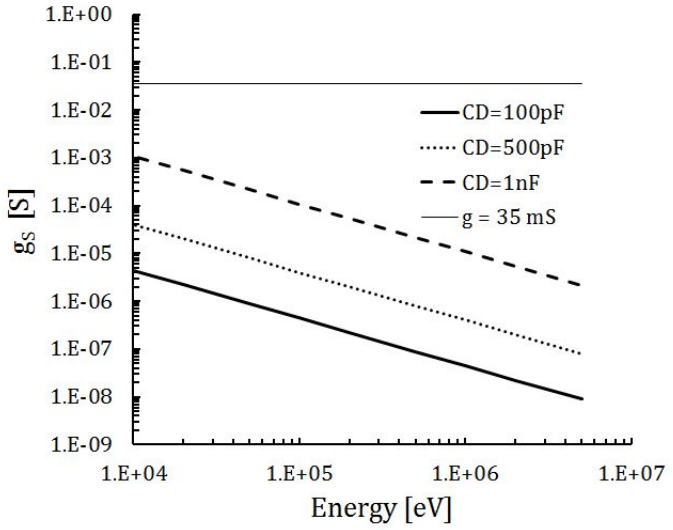


Figure 4.9: Resolution-equivalent transconductance in presence of white series noise for a NaI(Tl) detector read out by different sizes APDs.

[37].

However, the demonstration that internal gain introduces a benefit is an important result with deep consequences to the design of front-end circuits.

Photomultiplier tubes and Silicon photomultipliers

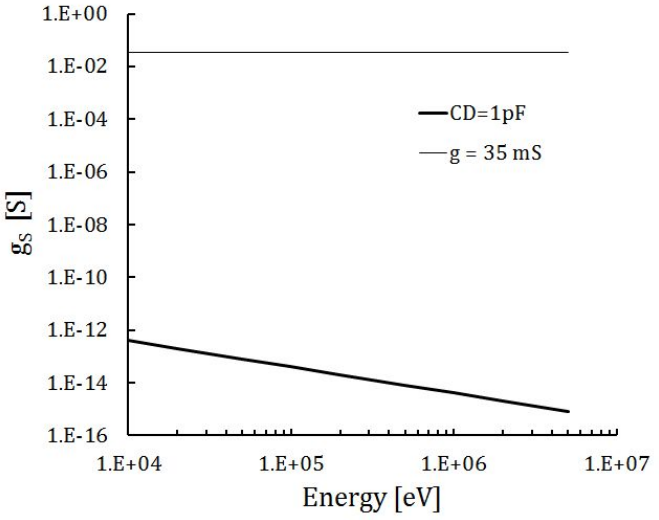
Both PMTs and SiPMs have very low leakage current and virtually no $1/f$ noise, so we can safely assume the series noise approximation holds. The gain available from a PMT or a SiPM is of the order of 10^6, therefore we expect no contribution whatsoever from the readout circuits, for any value of g_m and energy. In fact, typical g_S plots for these transducers look like the ones in Figure 4.10.

There are still some significant differences between PMTs and SiPMs: the former presents a detector capacitance which is mostly constant with photocathode area; the latter has a capacitance that is dependent on the area of SiPM desired, and comparably higher to that of a PMT. For the plots of Figure 4.10 we assumed a value of 35 pF/mm^2 for SiPM capacitance. We calculated g_S for a 1 inch and a 3 inches sensitive area. This is comparable to the area of the most popular PMTs. One could invert the formula and ask what is the maximum area one can read before electronic noise becomes a problem in SiPMs at a nominal energy of, say, 10 keV. For a 35 pF/mm^2 capacitance, this would be about $5,000$ cm^2, corresponding to 17.5 μF of detector capacitance. We can certainly state that there is no practical difference in instrumenting SiPMs compared to PMTs for most applications.

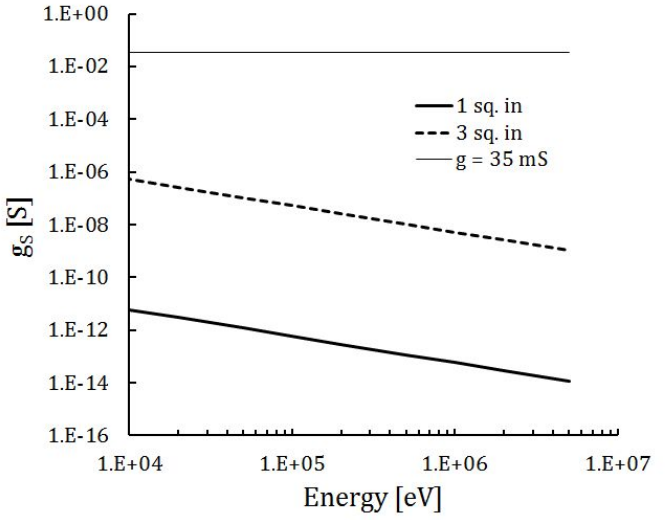
This means that most circuits currently used for reading out PMTs can be directly used on SiPMs, with the only constraints of minding the effects of the somewhat increased leakage current (dark count rate) and bandwidth-limiting effects on the amplifier due to the increased capacitance. These factors only drive the type of architecture one can adopt for the amplifier. In the case of very high dark count rate, one has to make sure the equivalent current ($= DCR \cdot q$) does not create a large value for the parallel noise component (the term $qIR_S/(2\pi^2(C_D + C_i)^2)$). In some cases, one may need to use either a resistor to ground or a current amplifier such as a common base (or common gate for FETs) configuration to convert charge into voltage. In some other cases, a charge preamplifier will work just fine so long that a wide bandwidth is not needed (> 50 MHz).

The obvious question is then: is the previous analysis really needed if the conclusions seem obvious, that is, SiPMs are equivalent to PMTs?

In truth, this seems to be the current wisdom. However, as it will be demonstrated in the next paragraph, there is a whole class of applications in which the common wisdom does not hold true.



(a)



(b)

Figure 4.10: g_s plots for photomultiplier tubes (a) and for comparable areas of SiPM (b).

Photon starved design: innovative solutions

We have seen that there is a significant difference, when reading out a scintillator, between its response to large numbers of photons and its response to single photons (the photon starved regime). This was shown in Figures 4.4 and 4.5.
In all cases where the response to single photons is of interest, interesting conclusions can be reached, thanks to the formulation derived in the previous paragraphs. A question bears at this point: why is it so important being able to detect single photoelectrons? First, this is not such an exotic requirement as one might think. New detection systems tend to move towards large areas (in the m^2 range) [38, 39, 40]. When this happens, the photons produced by a single event are spread out to a very large detector area, so it is possible that only a few photons from an event hit a single photodetector. Second, there are experiments where the number of photons is intrinsically low because either the energy is low or the efficiency at a given scintillation wavelength is low (such as the case when using UV-scintillating materials).
In all these cases, we will see, the classic PMT/SiPM duality breaks down.

The reason for this lies in the fact that in high SNR regimes, the parameter σ_{Scint} is the resolution of the scintillator itself, and it is set by the fact that the statistics are dominated by the high flux of photons. In the photon-starved regime instead, the resolution is no longer set by the fluctuation on the information, as the information is composed of single photons, and each single photon has exactly the same shape and amplitude as any other. The uncertainty on the single photon waveform is determined by any noise added by the readout electronics. If we want to detect every single photoelectron reliably, we need to ensure an electronic noise low enough to set the appropriate threshold on the single photon signals. Generally, this threshold is required to be at least at half of the peak height of a single photon signal, which correspond to a noise of one tenth a single photoelectron (the threshold is always set at five times the σ to minimize accidental triggering). By repeating the steps used to come to the general derivation

for g_S, we must consider that now, for a single photoelectron:

$$v_{out}(t) = \frac{QA_0}{(C_D + C_i)} H(t) = \frac{qGA_0}{(C_D + C_i)} H(t)$$

Consequently:

$$v_{out,0} = \max\left\{ \frac{qGA_0}{(C_D + C_i)} H(t) \right\} = \frac{qGA_0}{(C_D + C_i)}$$

For the electronic noise we have:

$$\overline{v_{out}^2} = \int_0^\infty a(f) |A(f)|^2 df = aA_0^2 \int_0^\infty |H(f)|^2 df = aA_0^2 R_D$$

We can find the equivalent noise charge as the amount of charge that corresponds to a voltage at the output equal to $\sqrt{\overline{v_{out}^2}}$:

$$v_{out,ENC} = \frac{qGA_0}{(C_D + C_i)} ENC_e = \sqrt{aA_0^2 R_D}$$

If we require that $ENC_e < 0.1$ photoelectrons and we solve for g_m:

$$g_m > \frac{4kT\Gamma}{0.1^2 G^2} \frac{1}{q^2} (C_D + C_i)^2 R_D$$

Finally:

$$g_S = \frac{4kT\Gamma}{0.1^2 G^2} \frac{1}{q^2} (C_D + C_i)^2 R_D$$

If we want to use our simple filter, we need to keep in mind that the signal corresponding to single photoelectrons is generally very fast (of the order of a few ns to a few hundred ns for the entire duration), so the choice of filter time constant should be accordingly adequate. Here we assume 100 ns for simplicity. The formula becomes:

$$g_S = \frac{kT\Gamma \mathrm{e}^2}{2q^2 0.1^2 G^2} (C_D + C_i)^2 \frac{1}{\tau_0}$$

We could have arrived at the same result by looking at the general formula for g_S and observing that σ_{Scint}/ϵ is simply the ENC_e and that instead of q we should have used Gq as minimum quanta of charge at the amplifier input.

It is not surprising that g_S is no longer dependent on energy. In fact, in the photon-starved regime it does not make sense talking about energy, as it implies a continuum of values. We are dealing with a discrete set of allowed values from 1 to a few photoelectrons. In this case it makes sense to plot g_S as a function of filter time τ_0 for a given SiPM capacitance (or area).

In Figure 4.11 we find the familiar fact that there is a minimum value of transconductance we need to obtain in order for the single photoelectrons to be detectable.

We see that if we want to read a 1 inch^2 area under these circumstances,

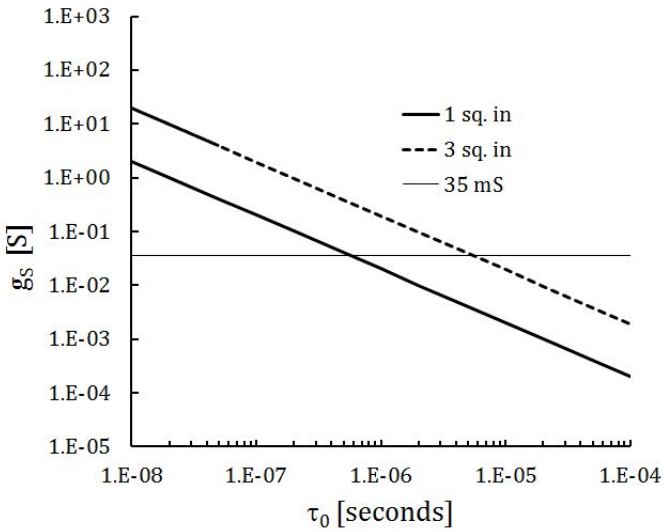


Figure 4.11: Minimum g_S values for two different SiPM sizes.

it will be impossible to do so below a 600 ns time constant for the filter.

We can always lower the limit by using longer peaking times, however, if the rate of single photons is particularly high (true in many instances for SiPMs), we will not be able to distinguish single photons after the filtering process, unless sophisticated coincidence and gating techniques are also used. In this case the obvious choice would be to reduce the area read by the single channel, thereby increasing the total number of channels. This would reduce the sensor capacitance and allow shorter τ_0. Another way would be to increase g_m of the amplifier. This can be done by increasing the drain or collector current of the active device setting g_m, or using integrated CMOS devices or BJTs, as they offer better values of g_m for the same current as JFETs. Either way, we see that the choice will have a significant impact on overall system power, an important parameter especially for large area systems.

Another interesting way to look at the dependence of g_S on system parameters is to plot it as a function of sensor capacitance, C_D. This, for a given peaking time and transistor transconductance, allows a good view of what will the maximum instrumented area be for each channel. For example, the plot in Figure 4.12 shows that when we use a transistor with 80 mA/V transconductance, we can easily measure up to 20 nF worth of SiPM area (this would be about 600 mm^2) with a filter time of 1 μs, but we would not be able to do the same thing with a faster (250 ns) filter. A graph of this kind is a very useful tool for designing front-end circuits.

In light of this new research, it is important to recognize this important limitation in the use of SiPMs in single photon counting or photon-starved regime and how important was to provide new design criteria and solutions for readout electronics in these applications.

The following chapters will focus on two case studies that take full advantage of the theory developed. We will detail designs in both high SNR and photon starved regime, highlighting for each case how the newly developed guidelines drive the specific choices.

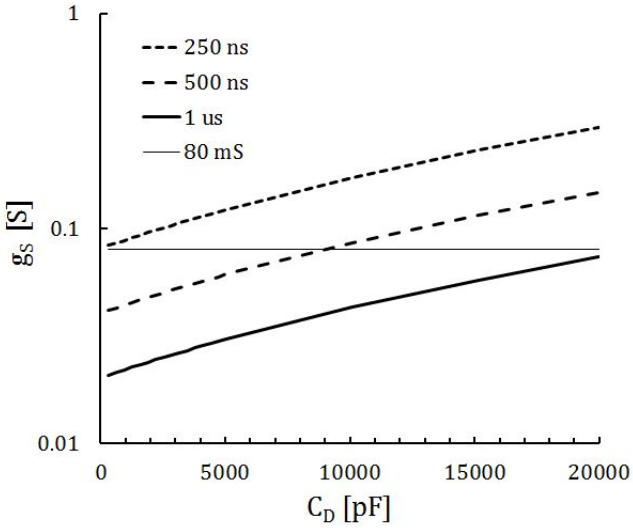


Figure 4.12: Values of g_S vs. SiPM capacitance for different filter time constants.

Chapter 5

Compact Radon Detection Platform

5.1 Introduction

The introduction of silicon photomultipliers in radiation detection enables designing systems with characteristics not achievable with conventional photomultiplier tubes in the past, so long that the criteria introduced in the previous chapter are understood and applied. An example of this are portable systems. The use of PMTs is such systems means incorporating a bulky and potentially brittle device into what has to be a compact and rugged system. The requirements for bias voltages in the kilo volt range also do not help: several cubic centimeters need to be dedicated to the task of generating such a high voltage, often from a single-cell battery voltage of the order of 3 to 4 V. While the task is not complicated, it still requires the use of a relatively large number of components. The components themselves and connectors able to withstand high voltages are bulky and often expensive. The smallest PMT still occupies a volume of about 10 cm^3, not including the necessary base to distribute the bias voltage, for a mere 2 cm^2 of active area! For comparison, a similarly sized SiPM would require less

than one fiftieth the PMT volume. This alone is a strong motivation for choosing SiPMs as baseline readout device in most portable applications. On the flip side, the SiPM gain is sensitive to temperature variations, therefore some sort of stabilization has to be implemented. To a lesser degree, however, the same consideration applies for PMTs in mobile applications.

In this chapter, we are going to design the first portable detector system for Radon detection based on SiPM technology. This will allow benchmarking of the theory developed in the previous chapter as well as serve as a demonstration of the advantages brought by SiPM technology to the portable detection system design.

5.2 Origins of Radon

Several agencies in the world, including the World Health Organization and the US Environmental Protection Agency, have classified radon, an odorless and colorless gas, as human carcinogen that can increase the risk of lung cancer if breathed into the lungs [41]. The carcinogenic nature of radon is due to the fact that the element is an alpha source. As introduced in Chapter 1, alpha particles readily stop (i.e. deposit all of their energy) in human tissue. When radon is introduced into the lungs, its alpha emissions damage the lung tissue with cancerous effects. In fact, radon is the number one cause of lung cancer deaths among non-smokers in the United States, and the number two for smokers [42]. Each year in the US there are an estimated $20,000$ deaths from lung cancer.

Radon is generated in the decay chain of ^{238}U, the most common uranium isotope, present at 99.284% abundance in uranium ore. The isotope has a very long half-life, about 4.5 billion years. It decays to the stable element ^{206}Pb through a long decay chain (Figure 5.1 [43]).

Many branches in the chain decay via alpha or beta emissions, but ^{222}Rn is the first gaseous element with a short 3.8 days half-life, decaying via 5.6

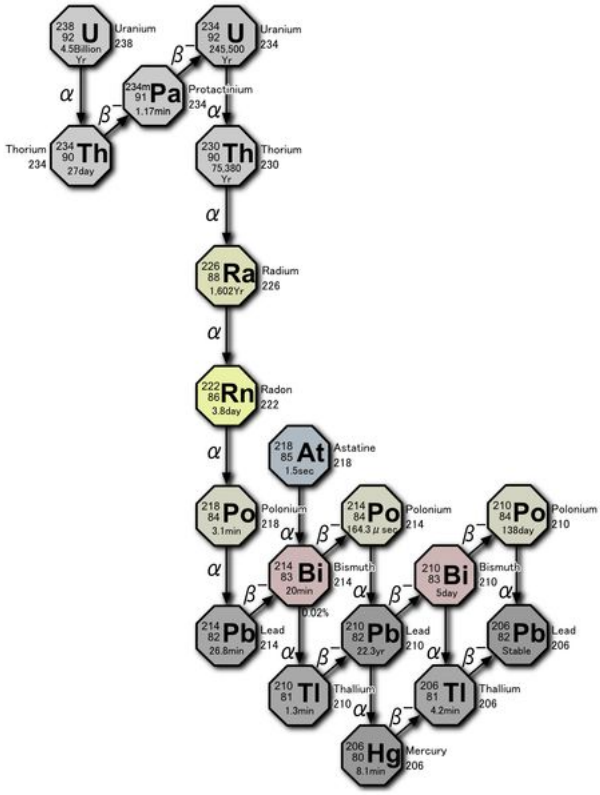


Figure 5.1: ^{238}U decay chain.

MeV alpha emissions. These characteristics make radon a health concern: it can propagate, accumulate in areas with little ventilation and, finally, can be breathed into the lungs. Clearly, the abundance of radon is strongly dependent on the geology; in fact, its concentrations show strong geographical variations: generally a strong presence is found in granitic or shale formations, while lower concentrations are found in water and sandy soil [44]. Typical radon concentrations are in the 1 pCi/l (or 0.037 Bq/l). From health physics studies, the threshold of concern for radon concentrations in

the US has been set by the EPA to 4 pCi/l (0.148 Bq/l) [45].

Fortunately, radon mitigation for the home is rather simple and inexpensive. If a high radon concentration is found in a house, simple improvements in ventilation will quickly fix the problem. For this reason, agencies recommend that all dwellings be periodically tested for radon, and if high levels of radon are found, to provide mitigation [46]. Hence, radon testing is very important. There are several ways homes can be tested for radon. Testing methods vary from using activated charcoal sponges that trap radon atoms, to optical detection of alpha tracks in plastic films [47]. In these cases, a specimen is sent to a laboratory for analysis. There are also monitoring methods based on alpha particle spectroscopy and counting. These methods allow for real time measurements of concentrations directly in-situ. A comprehensive description of these methods is given in [48, 49]. A quick synopsis of the techniques is given in the following paragraph.

5.2.1 Radon monitoring and measurement techniques

The main difference in techniques is found, as hinted in the previous paragraph, in the nature of the measurement. A first family of measurements requires some form of sampling into a removable material and a successive laboratory analysis. This type of procedure is a typical testing procedure: it is representative of what happened cumulatively in a certain time frame, and it gives no indication whatsoever if the concentration values were seen as constant during a certain period or if they were due to a spike during the collection time. Typical collection times for these techniques can be as high as several months. In other words, this type of measurement gives a worst-case scenario in that it identifies the total radon dose a person would have received if he/she stayed in the environment for the whole duration of the collection campaign.
Often it is more indicative to have a measurement of how a concentration changes as a function of time, as this is more representative of a real scenario. For these purposes, the second family of measurements allows continuous monitoring of the concentrations, so one can establish patterns

and relations between concentrations and other factors. Devices belonging to this second family are also known as continuous area monitors (CAM). The principle behind CAM devices is rather simple, mostly based on alpha counting techniques. As we know, alpha particles are a very detectable form of radiation, and the detection can be performed via gaseous, liquid or solid state detectors (both semiconductor- or scintillation-based). In fact, there is a great variety of CAM detectors. Some employ a semiconductor detector such as a silicon barrier diode to directly detect alphas present in a measurement chamber. Others use ionization chambers where the radon decays in the air are directly part of the charge generation process. Depending on how air enters the measurement volume, some detectors may be faster than others. In any case, CAM detectors tend to be expensive (several hundreds to thousands of dollars), while the simpler accumulation-based detectors cost only a few tens of dollars (although the laboratory analysis may make up for the cost difference).

In adopting SiPMs and a scintillator as detector elements, it is possible to reduce system costs quite significantly: SiPMs are devices based on well-established silicon processes normally used for making integrated circuits. This is not true for devices such as silicon barrier diodes, as the purity levels required are not compatible with integrated circuit technologies. PMT technology has reached maturity and costs cannot be lowered any further. In small quantities, SiPM prices are still comparable to PMT prices, but since SiPMs can be produced on standard silicon substrates using conventional IC technologies, they have the potential for becoming inexpensive, as already indicated by some manufacturers. The necessity for high voltage bias in PMTs also has important consequences: costs and physical space required for high voltages above a few hundred volts are at least 50 times and 10 times respectively higher than the equivalent for voltages in the few 10's of volts. Finally, the low noise requirements for semiconductor-based detectors may also add complexity and cost to the front-end design. We now know that this is avoidable if the SiPM works in high SNR regime.

These reasons alone are a primary driver for investigating the possibility of

designing a radon detection system based on SiPMs and scintillators. In the next section we will investigate the tradeoffs associated with this choice and determine the feasibility of such design.

5.3 System requirements and goals —Preliminary design

The first step in designing a new radon detection system (or any system) is to establish some reasonable requirements and identify major constraints.

As previously stated, the EPA recommends that all dwellings be periodically tested for the presence of radon. In order for this to be possible, two main requirements can be set: the radon detection system should be first of all as portable as possible and as affordable as possible.
Periodical testing means the instrument may be utilized either as a continuous monitor in a given room, or to make spot-measurements only periodically, therefore it is unthinkable to baseline an instrument that takes more than a few 100 cm^3 of space and requires constant maintenance (such as battery replacement or continuous calibrations). Therefore, more requirements are identified: the device must be as maintenance-free as possible, and must be able to function for a long period of time without needing special attention. Of course, "long" period of time does not mean anything unless specified. Currently, the standard radon measurement requires the measuring device to be left unattended and undisturbed in an area for at least two days [50]. This is due to the fact that in order to reach the statistical significance needed to complete a measurement, a certain number of events have to be collected. Given the activity related to typical radon concentrations (pCi/l), this is accomplished in the time indicated above.
If there were a way of increasing the number of events per unit time, there would be no other reasons why the measurement could not take less time. This is another important point: finding a technique that allows for quicker collection times. The practical impact on the system design is not only obvious time savings, but also less stringent requirements on how long the

Application Requirement	System Requirement
Portability	Total size and weight constrained, low power
Cost	Choice of inexpensive detectors and electronics
Ease of operations	Functions without operator intervention or maintenance
Measurement duration	Shorter than two days for conventional applications

Table 5.1: Application requirements and how they translate into system design considerations (i.e. system requirements)

instrument must be able to operate on its own power source.

From the simple hand waving arguments above, an initial set of requirements can be compiled. Table 5.1 shows what the application requirements are and how they translate into system design requirements.

We now need to quantify the arguments. As indicated earlier, portability means that the instrument cannot be heavy, its size must be contained and must possess its own internal power source. Commercial systems are available in several form factors, depending on the underlying principles and the performance level. The simplest traditional monitors are comparable to standard smoke detectors in size and weight [51]; they offer limited performance (the reading levels become statistically significant after an average of 7 days) at costs in the $100. Professional systems are larger and heavier (4000 cm^3 and > 1 kg), cost in the $\$1,000$ but offer the possibility of completing measurements within 48 hours [52]. Recently, new radon detectors have hit the market. They are targeted to homeowners at costs of about $250 [53, 54]. Their size is comparable to a digital multimeter, weight only 130 g and can execute measurements at 2.7 pCi/l with a 20% precision in

Volume [liters]	Decays per second
0.1	0.00148
0.2	0.00296
0.3	0.00444
0.4	0.00592
0.5	0.0074
0.6	0.00888
0.7	0.01036
0.8	0.01184
0.9	0.01332
1	0.0148
1.1	0.01628
1.2	0.01776
1.3	0.01924
1.4	0.02072
1.5	0.0222
1.6	0.02368
1.7	0.02516
1.8	0.02664
1.9	0.02812
2	0.0296
2.1	0.03108
2.2	0.03256
2.3	0.03404
2.4	0.03552
2.5	0.037
2.6	0.03848
2.7	0.03996
2.8	0.04144
2.9	0.04292
3	0.0444

Table 5.2: Number of decays per second in different volumes for a specific activity of 0.4 pCi/l.

Time [s]	Decays in 0.1 l/s	Decays in 1 l/s	Decays in 3 l/s
1	0.00148	0.0148	0.0444
2	0.00296	0.0296	0.0888
5	0.0074	0.074	0.222
10	0.0148	0.148	0.444
20	0.0296	0.296	0.888
50	0.074	0.74	2.22
100	0.148	1.48	4.44
200	0.296	2.96	8.88
500	0.74	7.4	22.2
1000	1.48	14.8	44.4
2000	2.96	29.6	88.8
5000	7.4	74	222
10000	14.8	148	444

Table 5.3: total events for a 0.4 pCi/l of volume exchange.

1 week or 10% in a month. There is clearly room for a comparably-sized device capable of measuring similar concentrations in a few hours at similar cost. Based on these observations, we can then set a goal for a new design to be in the few hundreds cm^3, weight a few hundred grams and cost a few hundred dollars.

The key in meeting such ambitious goals is clearly in the right choice for the technology adopted.

The very first thing to verify is whether the goals are allowed by the underlying physics. We can do that by starting from some assumptions. Without even considering what detector we are going to use, let us imagine a simple scenario where air is introduced into a predetermined volume; let us assume that all of the decays within the volume can somehow be detected. We know the mitigation threshold is set at 4 pCi/l, therefore we ask our-

selves how long it would take to detect $1/10^{\text{th}}$ such activity as a function of measurement volume. Since 0.4 pCi corresponds to 0.0148 Bq, or 0.0148 decays per second, we can quickly build Table 5.2 that indicates how many decays per second exist in a given volume.
If, as assumed above, all particles are detected, we would count, for example, 3 events every 20 liters of air. On the other hand, we can also assume we can recycle the entire volume of air every second, so to keep the activity constant, and calculate how many events are detected as a function of time, as reported in Table 5.3 Since these are radioactive decays, they are

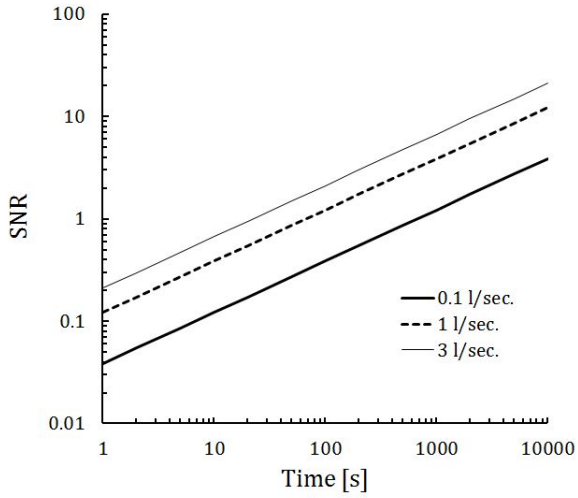


Figure 5.2: Signal-to-noise ratio for different flows at a 0.4 pCi/l activity.

subject to Poissonian statistics that, as a reminder, determines the total uncertainty on a number of events N to be $\sqrt{N}$. Thus, we can build a plot showing a signal-to-noise ratio defined as the inverse of the uncertainty. Such plot is shown in Figure 5.2.
From the plot we can see after how long a unity SNR is reached, or how long an acquisition should last in order to achieve a certain SNR. For example, at 1 l/s an SNR of 10 can be reached in about $10,000$ seconds. This is less

than 3 hours.
The important conclusion in this case is that there are no obvious physics reasons why a good detection would require more than a few hours.

This, of course, is a rough estimate: in a real system only a fraction of the decays is actually collected by a detector, and one has to deal with non-ideal geometries. By selecting the system components we can eventually refine the estimates and determine the limits this technique can achieve.

5.4 System Design

The choice of the detector will determine what fraction of the total events is collected within the volume, because the detector determines, along many other things, most of the geometrical constraints. In order to operate this choice correctly, we need to consider the main properties of alpha radiation. As introduced in Chapter 2, alpha radiation is composed by helium nuclei (2 protons), therefore it is quite massive. It will tend to stop within short distances inside materials (see Figure 1.3 for absorption length in air). This means that the measurement volume does not have to be too deep beyond 3~4 cm: anything beyond that will only contribute to losses of useful events. Having established this, there is clearly no physical limitation to the surface area of the detector; therefore we can resort to our design constraints in order to pose reasonable limits to the detector size and type. These limits are practicality in obtaining certain sizes, cost of the material and cost of related sensors (such as SiPMs if a scintillator is used). For economic reasons, we would like to use scintillators because they usually are inexpensive. We need to make sure the detector of choice will be blind to other radiation (gamma, neutron), so not to alter the measurements. For similar reasons, we need to limit the surface area to about 1 inch^2 in order to avoid costly choices and unnecessarily growing overall system size.
A quick literature survey indicates that the ideal solution is one of the very first scintillators used in alpha spectroscopy (since the 1930's): zinc-sulfide silver activated. This material has practically no sensitivity to gamma

Parameter	Value
Light Output (% of Anthracene [1])	300
Peak Emission Wavelength (nm)	450
Decay Time (ns)	200
Density (mg/cm^2)	3.25
Polyester Film Thickness (mm)	0.25
Refraction Index	2.4
Cost	\$1/cm^2

Table 5.4: Main characteristics of the scintillator ZnS(Ag) on Mylar backing.

rays or neutrons (unless activated with elements that have finite neutron cross-section) and is available as polycrystalline powder that, with the appropriate binders, can be painted on a transparent backing (usually Mylar) with the thickness of choice. The interesting thing to notice is that the detection efficiency for alpha particles (as measured by detected particles/incident particles) is nearly 100% for thicknesses between 10 μm and 30 μm [55]. The only drawback for this material is the fact that it is not transparent due to its polycrystalline form, and its refraction index is high (2.4). We anticipate, however, that, given the alpha particle energy (5.6 MeV), the overall brightness of the events will be sufficient to compensate for these drawbacks.

ZnS(Ag) is readily available from scintillator manufacturers in sheets typically measuring 150 mm x 150 mm, although larger surfaces are possible upon request. Its trade name is EJ-440. The scintillator is deposited on a clear plastic sheet that serves as optical coupling mechanism to a photodetector. The thickness of the deposited layer is chosen to optimize alpha particle detection (thus, as mentioned above, it is between 10 and 30 μm).

[1] It is common referring to Anthracene to measure light output of certain scintillators. The light output for Anthracene is $17,400$ photons/MeV.

The Table 5.4 supplies a list of the common parameters for this scintillator. A complete datasheet is available from [56]. The low cost of this scintillator clearly does not drive the choice for its size. This will be determined by the cost of the light readout. We mentioned previously that the cost for cm^2 of PMTs and a SiPMs is quite comparable, with the outlook for SiPMs to become much cheaper in the near future. It is therefore worth taking advantage of the benefits offered by the latter.

A market survey highlights immediately several vendors that offer array of SiPMs in the few cm^2. One specific vendor offers a 4×4 array of 9 mm^2 cells for about $400 in small quantities (SensL model Array SB-4 [57]) and is readily available for purchase. The array, depicted in Figure 5.3, is conveniently mounted on a low-profile ceramic carrier with thru-hole pins to enable easy design and installation on a PC board. The pixel characteristics are shown in Table 5.5, with the whole array characteristics in Table 5.6.

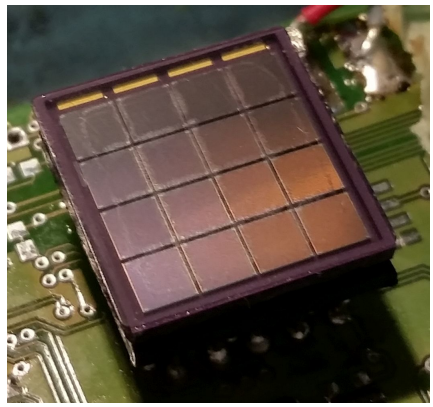


Figure 5.3: SensL array SB-4, silicon photomultiplier.

The noticeable parameters are the PDE (slightly higher than standard PMTs) and the peak sensitivity wavelength of 420 nm: an excellent match to our scintillator. The SiPM will be coupled to the scintillator sheet via

Parameter	Value
Pixel area	3.16×3.16 mm^2
Pixel thickness	450 μm
Pixel active area	3×3 mm^2
Breakdown voltage (VBr)	24.5 ± 0.5 V
Overvoltage range	1 to 5 V
Peak sensitivity wavelength	420 nm
Microcell recovery time	130 ns
Gain	3×10^6
Number of cells per pixel	4774
PDE	31%
Dark Current	2.8 μA
Temperature dependence of VBr	< 20 mV/°C

Table 5.5: Pixel characteristics for the SensL Array SB-4.

Parameter	Value
Array active area	13.4×13.4 mm^2
Pixel pitch	3.36 mm
Package size	15.81×15.31 mm^2
Package height	1.5 mm
Package type	Alumina Al_2O_3

Table 5.6: Full array characteristics for the SensL Array SB-4.

Interface	Amplitude reflection coefficient	Transmitted photons [photons/MeV]
ZnS(Ag)/Mylar	0.179	21,428
Mylar/BC-630	0.06	20,143
BC-630/SiPM	0.02	19,740

Table 5.7: Interface losses and light transmission through the scintillator/Mylar/grease/SiPM interfaces, assuming 52,200 photons are generated within the scintillator for every MeV of incident energy and 50% of them reach the scintillators exit window. Note that the scintillator is supposed to be transparent at this point. This assumption will be re-assessed later.

optical grease.

The ZnS(Ag) scintillator has a refraction index of 2.4, and, as mentioned above, it is deposited on a plastic film (n=1.67). Since the SiPM light entry window is encapsulated in clear epoxy (n=1.54), most of the light losses will be concentrated in the ZnS(Ag)/Mylar interface, where the strongest mismatch exists. In fact, from the Fresnel equations for normal incidence, we find that the reflected light at this interface is 18% of the incident light. We will simply couple the remaining optics with standard silicon-based optical grease such as the Saint-Gobain BC-630 [58] with a refraction index of 1.47. This is a standard material, widely used in the field.
With these choices, the total amount of light transmitted through the interfaces is shown in Table 5.7. Note that in building this table we assumed that, of the light generated within the scintillator, only 50% is directed towards the exit window in the direction of the SiPM and the rest is lost. In reality some of the light in the opposite direction bounces back at the ZnS(Ag)/Air interface due to the strong refraction index mismatch for that interface (2.4/1).

As can be seen from the table, about 75% of the amount of light in the direction of the SiPM reaches it (or 38% of the total). Since the SiPM has

a QE of 30% at these wavelengths, we have in the end about 6,000 photoelectrons generated for each MeV of energy deposited. Each alpha particle (5.6 MeV) will then produce in the worst scenario an average of 33,600 photoelectrons. In principle, we are in a situation where signal abounds. However, the analysis holds under the assumption that the scintillator is completely transparent.
ZnS(Ag) is basically a white polycrystalline powder or paint, therefore its opacity is expected to be non-negligible. No indication was found in the literature of what the transparency coefficients are for this material, so we had to set up a simple experiment to estimate the total losses due to opacity.

5.4.1 Estimation of ZnS(Ag) light output

Knowing the order of magnitude of the actual light output of the scintillator is key in determining what type of readout electronics needs to be designed.
Given the scarcity of information available on the opacity of ZnS(Ag), it was deemed useful to set up an experiment to estimate the actual light output of our scintillator in response to the 5.6 MeV alpha events, based on a calibrated setup. For this purpose, we used a conventional Hamamatsu photomultiplier tube (R11265U-100) coupled with a ZnS(Ag) screen matching the PMT window size (23 mm × 23 mm). The advantages in using a PMT in this case were that the device was readily available and could be integrated immediately with the bench top instrumentation in the laboratory.
The arrangement is shown in 5.4. Initially, the PMT was calibrated by coupling it with a CsI(Tl) crystal. By exciting the crystal with a 662 keV ^{137}Cs source, energy spectra as a function of bias voltage were recorded. By knowing the light output of the crystal and its coupling to the PMT, it is possible, to first order, estimating the number of photons that reach the PMT window. The Cs(Tl) scintillator produces 65,000 photons/MeV (see Table 2.3), therefore 43,000 at the 662 keV energy. The light coupling (from n= 1.79 to n= 1.47) introduces a loss of 1.8 % due to reflection.

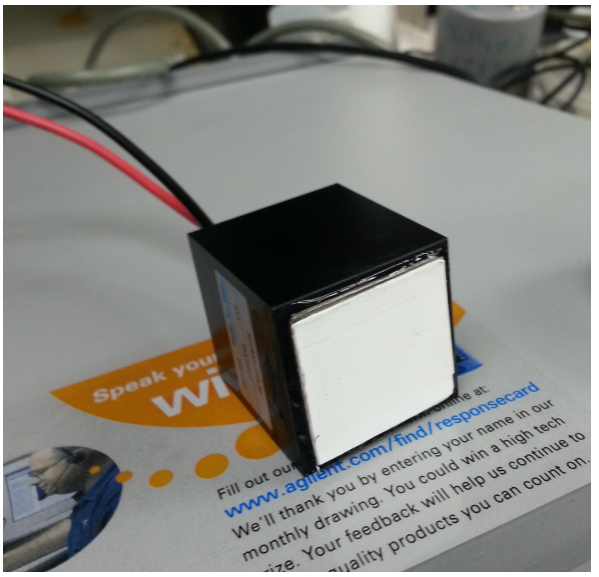


Figure 5.4: ZnS(Ag)/PMT setup to measure the actual light output of the scintillator in photons/MeV.

The QE of the photocathode at the wavelength of interest (550 nm) is 15 %. This corresponds to about 6,300 photoelectrons to be amplified by the PMT.
Since the PMT amplification and whatever means of signal processing used will be common to all measurements, we can, in the case of ZnS(Ag), attribute any difference in amplitude to the difference in light generated by the radiation in the different scintillator. Part of the difference is attributable to the different refraction index, emission wavelength, and to the coupling, and has been already calculated and shown in Table 5.7. The remaining margin will be indicative of opacity.
With these premises, we histogrammed the PMT output at two bias voltage values (950 V and 1,000 V), and carefully calculated the 662 keV peak position. Later, we replaced the ^{137}Cs source and the scintillator with a 5.6

MeV alpha source and the ZnS(Ag) sheet. In this case we recorded again a spectrum and the peak position of the alpha interactions. By doing this, we have two comparable data points; one expressing the known amount of charge at the photocathode for 662 keV in CsI(Tl), the other expressing the unknown amount of charge for 5.6 MeV alpha particles in ZnS(Ag). The unknown amount has been calculated at this point as 750 photoelectrons at the photocathode[2]. If we remove from this number the total known contributions due to the interfaces, we can state that because of opacity, the actual photon yield is about 1,000 photons for a 5.6 MeV signal, or about 180 photons/MeV.
The influence of opacity is clearly the dominant factor, and reduces brightness considerably. The comforting observation is that, for the energy of interest, there are still many photons that can be detected.

The choices made so far make sense from the system requirements point of view. We need to verify they also allow the desired detection performance, at least from a modeling point of view.

5.4.2 Detector model

According to the arguments made in the previous paragraph, the detector will be enclosed in a measurement chamber in which the air flow is kept at the appropriate rate, as can be calculated from Figure 5.2. Based on the fact that the mean free path of alpha particles in air is in the few cm range (anything beyond that will not contribute to the measurement), we can choose to concentrate the air within a volume that is 1~2 cm deep, so that all alpha particles travel for a distance shorter than the mean free path. By doing this there is a higher chance for the particles to reach the detector.
This has been confirmed in the SRIM [59] Monte-Carlo simulation plotted in Figure 5.5. Here, a 5.6 MeV alpha beam crosses a 2 cm air volume with an absorber at the end. Figure 5.5(a) shows the ionization losses along the

[2]An average between the values obtained at the two bias voltages was taken.

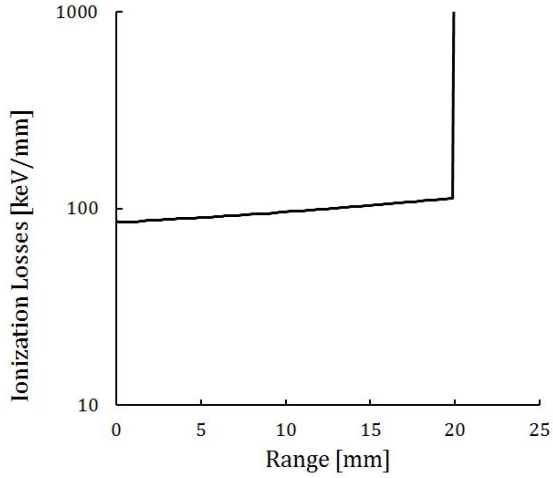
path to be about 100 keV/mm, so we never incur the abrupt Bragg losses shown in Figure 1.3 (note however the sharp increase of ionization losses at the absorber). This allows assuming that almost no alpha particles will be lost to air absorption. A further confirmation of that is shown in Figure 5.5(b) where the direct Monte-Carlo simulation output of the alpha tracks is shown. A simple integration indicates that all of the alpha particles in the simulation reach the absorber.

As for the energy distribution within the detector, we know that the particles of interest are mono-energetic at 5.6 MeV and we know, from Figure 5.5 that the particles reaching the detector will not all deposit the full alpha energy, but also a continuum of lower energies, since each particle loses about 100 keV each millimeter in air. The resulting energy distribution in the detector will be the overlap of the distributions of alpha particles with a continuum of energy between 5.6 MeV down to 5.6 MeV less the highest possible energy loss in air. Figure 5.6 shows the ionization losses of the alpha particles entering the detector with initial conditions given by the simulation results of Figure 5.5, and a zoom into the particle tracks in a region located ± 15 μm around the incident beam starting at the entrance window of the scintillator (30 μm thick).

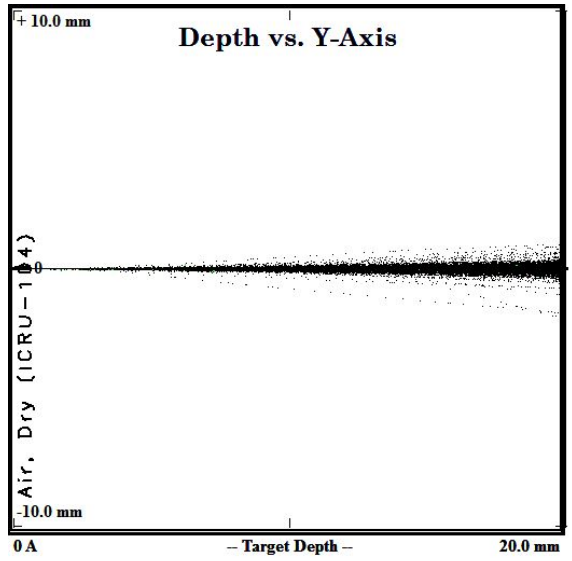
From the simulation, we can anticipate the energy spectrum collected by the detector to be mostly a continuum of energies, without prominent peaks (although we cannot exclude the presence of a peak completely). In order to have a clear measurement of how many particles have been detected, we will simply need to, literally, count each event in the detector independent of its energy. Not requiring energy discrimination in the electronics is an advantageous simplification, as it can be accomplished by simply counting anything above a pre-determined threshold.

The sizing of the chamber's cross-section is somewhat arbitrary: it will need to match the SiPM's active area at a minimum (13.4×13.4 mm^2), likely be somewhat larger so that alpha particles from off-axis directions can contribute to the measurement as well, but not so large that the particles are never detected.

We decided to start by considering a measurement volume that can accom-

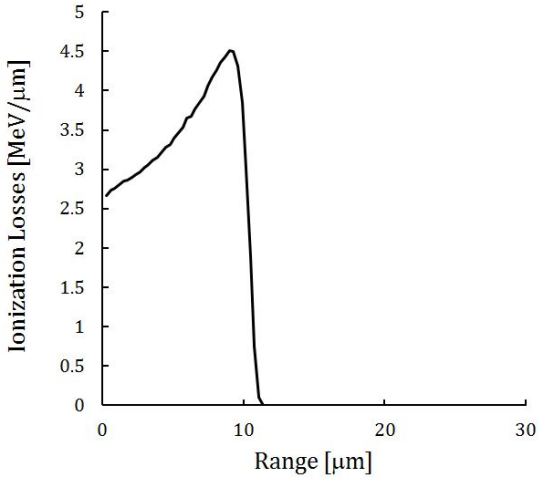


(a)

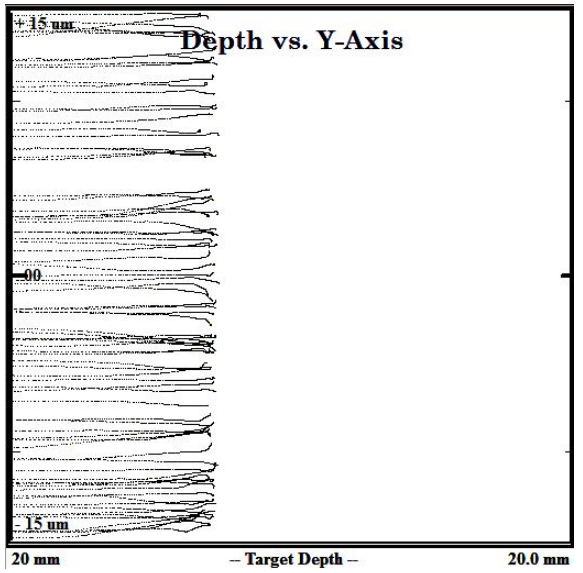


(b)

Figure 5.5: Ionization losses of 5.6 MeV alpha particles in air (a). An absorber is put at a 20 mm distance. Alpha tracks in the same volume of air (b).



(a)



(b)

Figure 5.6: Alpha particle ionization losses in the detector after transit in the air volume (a). Tracks within the detector (b).

modate our SiPM, but at the same time can also accommodate a standard one square inch PMT, in case comparison measurements between the two sensor types are needed in the future. The measurement volume was thus set to be 26.7 mm ×26.7 mm for the surface where the detector is mounted, times 17 mm depth. With these dimensions set, we can finally determine how many decays reach the SiPM area.
An upper bound can be determined by completely neglecting alpha particle absorption in air, and observing that the radioactivity of the air within the measurement volume is homogeneous and constant with an activity a (in Bq). Under these assumptions, all of the alpha particles created within the volume will eventually hit one of the walls; therefore, the number of particles that hit the photodetector area will be proportional to the ratio of the detector area to the total area. In other words, we can estimate the number of particles hitting the SiPM, N (in particles per second) as the fraction:

$$N = a \cdot \frac{179.56\text{mm}^2}{3251.6\text{mm}^2} = 0.055a$$

To confirm this, we ran a Geant4 [60] simulation under the same assumptions, except for the fact that the influence of air on alpha particles was considered. We set up a geometry representative of the measurement chamber and filled it with 1 atm of air with uniform, isotropic emission of 5.6 MeV alphas. The total alpha particles in the chamber resulted to be $2 \cdot 10^5$; the alphas detected in an area equal to the SiPM's were $1.3 \cdot 10^4$; therefore the calculated fraction of particles reaching the PMT was 0.065, in good agreement with the geometrical calculations.

Thanks to these quick analyses, we can determine some realistic performance figure for the instrument. As in the scenario considered when building the plot of Figure 5.2, we would like our minimum detectable activity to be 0.4 pCi/l ($1/10^{\text{th}}$ the EPA threshold for remediation). This means 0.0148 Bq within the measurement volume. Of these, $8.14 \cdot 10^{-4}$ decays will reach the detector every second (using the least optimistic estimate for the fraction). In order to have a 10% precision we need a SNR of 10, thus, we have to count enough decays to ensure that $N/\sqrt{N} = 10$. This

means 100 counts, which is easily done in 122, 850 s, or 34 hours on average.

The result obtained compares very favorably to commercial hand-held devices that are currently achieving the same level of precision in one month of continuous measurement. The underlying result is also that a level comparable to the EPA threshold can be detected in just a little longer than 3 hours.

It is worth pointing out that, as originally stated, the half-life of radon decays is only 3.8 days, so it is not advisable to simply compress, say, 1 liter of air into the measurement chamber and wait for the measurement to complete. If we were to adopt such a technique, we would need to compensate for the spontaneous decays, plus, the measurement would take longer than anticipated (although it is not unreasonable to do so for particularly high radon concentrations). We can easily bypass this potential issue by designing a simple ventilation system around the measurement volume that allows for a constant air exchange equal to 1 liter per second.

5.4.3 Ventilation system design

There are several small (1 inch or so) DC fans capable of low power (less than 1 W) and flow rates in the liter per second. Such a fan, coupled with proper air transport, can be used to ensure the desired flow rates within the measurement volume. Of course, we need to keep in mind that the SiPMs are light sensitive, therefore precautions must be taken to prevent light from propagating into the measurement chamber.
The simple proposed geometry in the model cutout of Figure 5.7 could work, provided the transport losses are kept to a minimum. A fan is connected to the converging duct at the inlet side. The first section of the tube runs into the measurement volume after three right-bends. This way, any light coming from the outside should be removed completely. The exhaust tube after the measurement volume folds five times before reaching the exit port. The diameter of the tubes is 0.5 inch, or 12.7 mm.
It is easy to estimate for such a system the pressure drop along the pipes

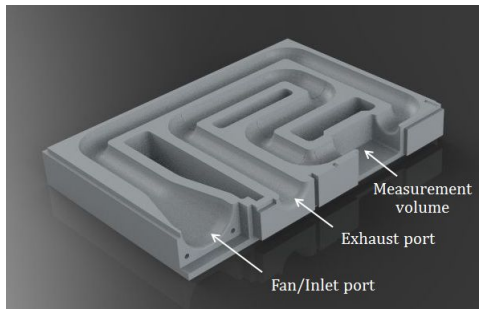


Figure 5.7: 3D model of the tubing system used for the radon detection system.

by using standard calculators (such as the one in [61]). For these simple calculations, losses in elbows can be expressed as equivalent-lengths of pipe from empirical tables or calculators. In our case, the eight elbows and the two sudden enlargements (the measuring volume), are equivalent to losses introduced by 370 mm of extra pipe. The overall length of straight pipe is 408 mm as measured on the model. Thus, the pressure losses from inlet to outlet for the whole system are equivalent to those of a 0.5 inch diameter pipe that is 778 mm in length. From the calculator indicated above, the total expected pressure drop for this piping system is about 0.086 mbar. We can safely assume that the overall air flow will be preserved within such system. This simple calculation could also be verified directly on the model by CFD analysis, or by measurement on a prototype, but neither option was available. The important message from this estimate is that with such a pipe system design the flow imposed at the inlet is conserved without major losses, therefore we have a simple way of maintaining a constant air flow within the measurement volume, and our design considerations for the detector apply.

5.4.4 Electronics readout design

In the detector design section, we concluded that each alpha particle will generate about $1,000$ photoelectrons and this indicates we are in a high SNR situation. In this case, we can use conventional techniques to design the electronics instrumentation, based on the arguments introduced in Chapter 4. The simplest circuit we can think of is a 50 ohm resistor to ground. This would match cable impedances and allow for easy handling of the information. However, the resistor will be directly in parallel with the whole 13.6 nF sensor capacitance, introducing a time constant of $50 \times 13.6 \cdot 10^{-9} = 680$ ns, therefore we would expect a fairly long exponential current waveform with an area equal to the total charge produced. The total amount of charge generated by an alpha particle and amplified by the SiPM (assuming a gain of 10^6), would be 0.16 nC. This means the voltage drop on the 50 ohm resistor would look like the waveform in Figure 5.8, with peak amplitude of 11.8 mV. While this may be an acceptable solution

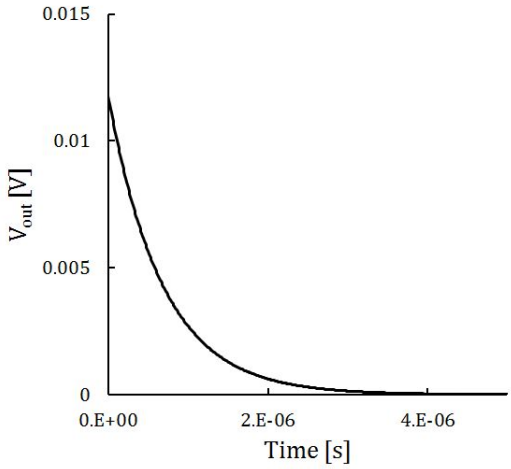


Figure 5.8: 5.6 MeV output waveform on a 50 ohm load resistor.

in some case, it falls short of two characteristics in our application. First, it

is a full-bandwidth signal, so the noise content is expected to be relatively high: it is always good practice to limit the bandwidth to what is really needed and nothing more. Second, the waveform is highly asymmetrical. This has implications when attempting to identify the presence of an event via counting pulses at a comparator output: the fast edge requires fast triggering circuits; these circuits will respond equally fast to the slow edge, creating possible oscillatory conditions due to the noise present on the tail. The spurious triggering can be mitigated by using hysteresis, but since the full bandwidth noise is relatively high, the hysteresis would need to be relatively large, and a combination of large hysteresis and appropriately low threshold may not be possible.

A more elegant solution would allow for only as much bandwidth as is needed, a more symmetric waveform, and lower noise. A charge amplifier and a properly designed filter would accomplish these tasks quite nicely, but there would be some complexity due to the high SiPM capacitance. The effect of such a high capacitance on the charge amplifier would be to introduce an undesired bandwidth limitation in a similar way to having the 50 ohm resistor. The situation could be improved by having more current in the amplifier, but this would mean an increased power demand on the system. Overall, this solution will likely require the use of many components.

On the other hand, a current amplifier would offer lower input impedance for a relatively low standing current, making the current waveform more symmetric. In fact, the input impedance of a bipolar transistor in a common base configuration is, to first approximation, the inverse of its g_m, with $g_m = 40$ mS for each mA of collector current. This, in practice, means 25 ohm per mA. With additional feedback, the input impedance can be made even lower. This means that the SiPM's original signal can be preserved without influence from the preamplifier, and the same amount of charge would be available in a shorter waveform (in this case, as short as the longest response between the scintillator relaxation time, 200 ns, and the SiPM cell relaxation time, 130 ns). With such a short pulse we would

also have full control on the system bandwidth by properly designing the later amplification stages.

The circuit of Figure 5.9, inspired by the circuits described in [62], achieves a very low input impedance, good bandwidth and low noise; plus, it easily allows implementing any amplification and signal conditioning in later stages; all with extreme simplicity. The transistor Q_1 would normally be

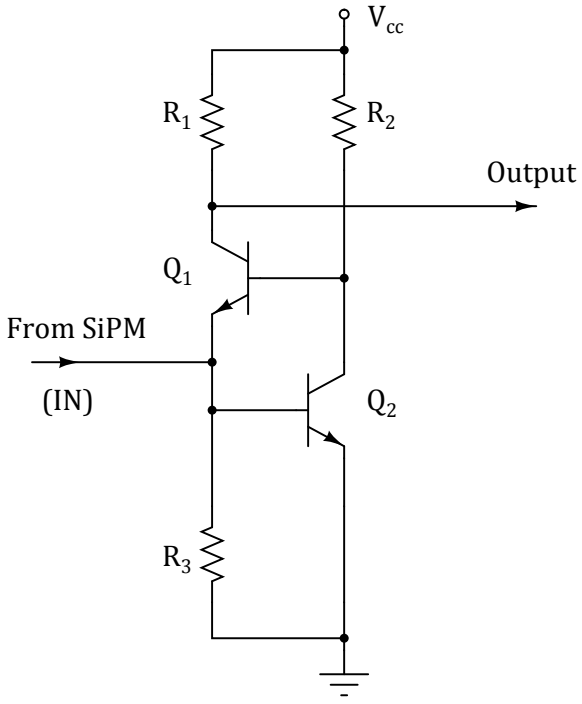


Figure 5.9: Fed-back common base architecture for SiPM readout.

a common base configuration, but in this case, the transistor Q_2 implements a feedback loop around Q_1 that has the effect of lowering its input impedance. The current output is then taken on the collector of Q_1. The output impedance of the circuit is R_1 in parallel with the impedance seen

at the collector of Q_1, which is normally $\gg R_1$.
In order to keep power low (this is a portable application after all), we selected the parameters to guarantee low power levels and good performance at the same time (both for signal and noise).
By indicating with I_1 and I_2 the bias currents in Q_1 and Q_2 respectively, and by neglecting the base currents, we can easily see that (assuming the base-to-emitter voltages equal to V_{BE} for both transistors):

$$I_1 = \frac{V_{BE}}{R_3}$$
$$I_2 = \frac{V_{CC} - 2V_{BE}}{R_2}$$

We see immediately that, if $V_{CC} > 2V_{BE}$, the current I_2 can flow in the desired direction. Without too many constraints, we can arbitrarily choose $V_{CC} = 2.8$ V, and set the currents, as $I_1 = 0.35$ mA and $I_2 = 0.81$ mA. The values for these currents set the bipolar transistors transconducance.
It can be shown that the noise is set mostly by Q_2 and Q_1 does not need much current to implement its feedback effect. The choice made for the currents optimizes power consumption and, as we will see later, output impedance. The current values imply a power consumption of 3.3 mW total. Consequently, the resistor values are $R_3 = 2.05$ kohm, $R_2 = 1.65$ kohm. As for R_1, its value sets the voltage on the collector of Q_1. If, in the later stages of the design we draw the signal current from this point with low impedance, such as a virtual ground, the expected voltage swing will be minimal; therefore, we only need to select the value of R_1 to ensure a high enough output impedance and some headroom from the voltage rail. A value of R_1 of 1 kohm will accomplish this by setting the collector of Q_1 at 2.4 V. The bipolar transistors we selected are standard BFR92A, available in surface mount package. These transistors are the preferred choice for many applications due to their versatile characteristics and very low cost [63].

The stages that follow the preamplifier are chosen to ensure that the signal

has a shape that can be easily handled for the purpose of identifying events in the detector, and to make sure the available dynamic range is fully exploited. As we mentioned previously, the detector signal has a fast initial rise and a slower exponential decay. For the choices made, the fast rise will be limited by the bandwidth of the preamplifier, while the exponential decay will be identical to the scintillator relaxation time (in this case, that time is slower than the cell recovery time constant of the SiPM).
As we mentioned earlier, such low symmetry in the signal imposes somewhat stringent requirements on the pulse detection circuits. With little effort and without much impact on power consumption, it is possible to introduce stages that properly shape the signal into something more manageable. Typically, these are active filters, so they accomplish amplification as well.

In order to properly design these stages we must first make some signal-to-noise considerations. In fact, on one hand, the amplitude of the maximum expected signals sets the maximum output, while the baseline noise sets the minimum amplitude that can be determined.

The noise limits can easily be found by considering our equivalent transconductance method. Let us remember that, in high SNR, we have:

$$g_S = \frac{4kT\Gamma}{q^2 E} \cdot \frac{\epsilon_S}{G^2} (C_D + C_i)^2 \, R_D$$

In the sole presence of the series component. In this case, we have $C_D =$ 13.6 nF, ϵ_S is $(180)^{-1}$ MeV/photon, based on arguments made for Table 5.7 and opacity. The corresponding g_S curve is shown in Figure 5.10, when we use the same filter assumed for the calculations made in Chapter 4.

In Figure 5.10 the horizontal line is the transconductance of our input transistor (Q_2) with a collector current of 0.8 mA. We see that under these conditions the electronic noise becomes a problem only below 20 keV. This value is what may set our threshold and determine our system efficiency, unless other sources of noise force a higher limit.

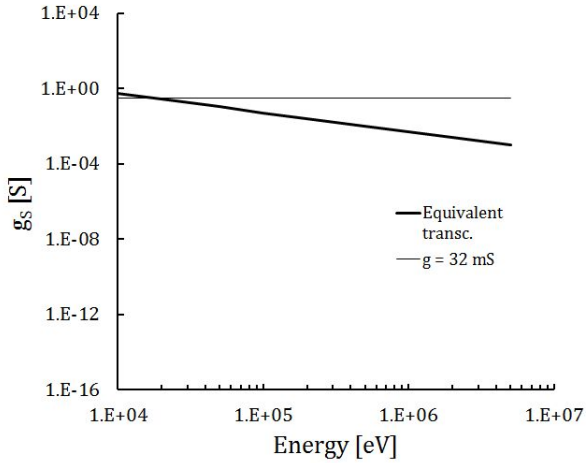


Figure 5.10: Equivalent transconductance plot for the 13.6 nF SensL sensor and the preamplifier of Figure 5.9.

Noise analysis

The transconductance method sets guidelines for the system's detection threshold under the hypothesis of white noise only. The sensor may have other sources of noise, such as shot noise related to leakage current, for example. The way the sensor noise is processed, depends also on the specific filter we adopt. The equivalent circuit in Figure 5.11 highlights the main noise sources in our preamplifier. The detector capacitance is C_D. The current source $i_n(t)$, with power spectral density $\overline{i_n^2} = 2qI$, accounts for the noise associated with the circuit's input current and the detector leakage current. The voltage source $v_n(t)$ with power spectral density $\overline{v_n^2} = \dfrac{4kT\Gamma}{g_{m,2}}$ accounts for the input noise of the transistor Q_2, the dominant noise contribution within the preamplifier. Also, Q_1 has a similar noise source at its base, but we can anticipate its contribution to be minimal because this source sees less amplification in its path compared to $v_n(t)$.

From Figure 5.11 we can recognize the familiar topology shown previously in Figure 4.6, hence we can adopt the same qualitative arguments to obtain the general expression for the output noise as found in Chapter 4. The series, $1/f$ and parallel noise contributions are evident.

$$\overline{v_n^2} = a\int_0^\infty |A(f)|^2\,df + \int_0^\infty \frac{|A(f)|^2}{f}df + \int_0^\infty \frac{2qI}{(2\pi f)^2 (C_D + C_i)^2}|A(f)|^2\,df$$

Once again, $A(f)$ is the filter's transfer function in the Laplace domain. Assuming a simple two-pole filter we can further show that the resolution can be expressed by (see Chapter 4):

$$\overline{\sigma_n^2} = \frac{\epsilon^2 (C_D + C_i)^2}{q^2}\cdot\frac{4kT\Gamma}{g_{m,2}}\cdot\frac{1}{\tau_0}r_D + \frac{\epsilon^2 (C_D + C_i)}{q^2}\cdot a_f r_f + \frac{\epsilon^2}{q^2}\cdot\frac{qI\tau_o}{2\pi^2}r_S$$

In this expression, the terms r_x represent the solutions of the integral terms (our residual functions), when the dependence on the peaking time is made explicit. C_i is the input capacitance and is, for our detector choice, negligible. g_m is the transconductance of Q_2. I is the input current and is mostly identifiable with the sensor's leakage.
In our bipolar design the $1/f$ terms is negligible, so we can simplify our expression:

$$\overline{\sigma_n^2} = \frac{\epsilon^2 (C_D + C_i)^2}{q^2}\cdot\frac{4kT\Gamma}{g_{m,2}}\cdot\frac{1}{\tau_0}r_D + \frac{\epsilon^2}{q^2}\cdot\frac{qI\tau_o}{2\pi^2}r_S$$

Of the two terms, one will dominate over the other, depending on detector characteristics and peaking time. During the design phase we will consider these aspects in selecting an appropriate filter stage.

Circuit design

To make sure the design is sound, and can indeed be built, we ran some simple circuit simulations with the components values set previously. The

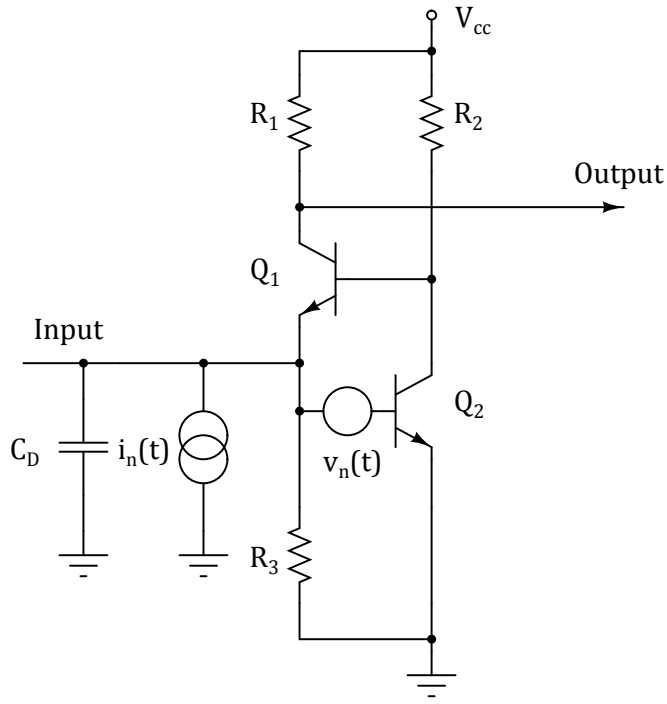


Figure 5.11: Noise equivalent circuit for the preamplifier in Figure 5.9.

initial bias point simulations simply confirm the values calculated and the power consumption that was anticipated during the initail design step.
The parameters we are most interested in are the input impedance and the dynamic response of the circuit. Since these parameters are more difficult to calculate, we rely on simulations (and later on measurements) to gain some insight. An initial simulation was set up to determine the value and frequency behavior of the circuit's input impedance, expected to be $(g_{m,1}g_{m,2}\mathrm{R}_1)^{-1}$. This is done by injecting a small-signal, AC-swept current, into the circuit input and measuring the corresponding AC voltage. The ratio between the voltage and the current will give the input impedance as function of frequency. Figure 5.12 shows the plot of input resistance

vs. frequency obtained from the simulation just described. We see that

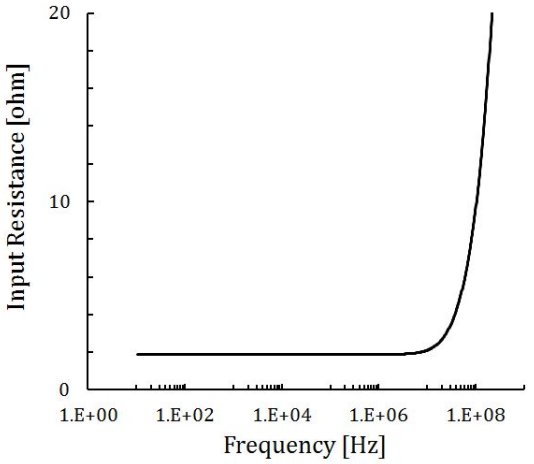


Figure 5.12: Input impedance vs. frequency (simulated data).

the impedance behaves like a resistor for all frequencies of interest, with a nominal value of 1.88 ohm. Thanks to the feedback provided by Q_2, the impedance is more than an order of magnitude lower than what could be obtained from a single common-base transistor operated at the same total current (in that case, the impedance would be 21 ohm).

As for the circuit response, we need to make a few simple assumptions at this point. The intrinsic signal from the SiPM sensor, as observed several times already, will have a fast edge followed by a slower, 130 ns trailing edge. The light signal, on the other hand, will be similar, except for the fact that the trailing edge will be set by the relaxation time of the scintillator, 200 ns. The total charge delivered for a 5.6 MeV signal will be 0.16 nC; this is also the area of the current signal coming from the sensor. It is easy to simulate such signal in a first order approximation. All we need to do is to use a triangular waveform with a fast (a few ns) leading edge and a 200 ns trailing edge. The peak value of this signal is such that the area over the

time axis is 0.16 nC. Thus, the current peak is 1.6 mA. As for the output signal, if we were to read the voltage signal developed on R_1, we would have a swing of 1.6 V. Given the biasing point conditions, this amplitude swing may bring some transistors outside of their linear regions. In order to avoid that, we can use the preamplifier as current buffer instead of voltage amplifier, by reading out the signal current in the collector of Q_1 directly as our output variable. This, is easily done by AC-coupling an operational amplifier's virtual ground at the collector of Q_1. Another way of doing this would be to read the current off Q_1 via another common base transistor or a current mirror. For this design, we will stick with the simplest solution, that is, using an operational amplifier's inverting input. This particular choice allows also building some filtering into the operational amplifier itself, thereby saving in power and overall component count.
The benefits of introducing filtering (or shaping) on a signal are discussed in depth in [36, 64, 65]. For our purposes we highlight the fact that shaping reduces the overall bandwidth to what is necessary to correctly process the information of interest and nothing more. This way, the total output r.m.s. noise (equal to the integral of the noise spectral density over the bandwidth) is reduced compared to having a wideband output. This leads to better signal-to-noise ratio, improving the identification of signal pulses. Normally, the filtering is implemented through the use of standard active filtering blocks whose response is designed to provide a Gaussian-like voltage signal in response to a step at the input. The characteristics of the pulse vary with the number of poles used to implement the filter (also known as the filters order). The simplest filters have two coincident poles and one zero (second order filters), and it is not uncommon to find up to eight order filters (eight coincident poles, one zero). When the poles are real, the shape is exponential and does not represent a semi-Gaussian shape; when the poles are complex (or coincident), the response is Gaussian.

In the previous paragraph we highlighted the dependence of the noise on capacitance and peaking time. As we found out, the integrals of the transfer function $A(f)$ produce terms dependent on a time constant τ_p, characteristic of the filter. Such time constant is conveniently referred to as the

time at which the maximum of the filters time response to a step function occurs. This allows for immediate visual awareness. We also showed that the white term depends on τ_p^{-1}; the $1/f$ term is independent of τ_p and that the parallel noise term is directly proportional to τ_p. The constants from the integration (r_x) are the parts that determine the weight of each of the three types of noise. As a general rule it can be shown that the area of the filter determines the low frequency noise contributions, while the derivative (or the characteristics of the transition edges) determines the weight for the series noise.

In our situation the chief requirements are low power and simplicity, therefore, the obvious choice is the simplest filter we can make with the fewest possible components. For this reason we selected a simple second order filter.

As for the peaking time, this is usually chosen based on the desired level of noise as well as the expected rate of events: a longer peaking time will, in fact, increase the chance of overlapping events and enhance the parallel noise. In our simple system the expected event rate is rather low under most conditions (we expect to measure at most a few tens of nCi activity), but we would like to be able to use calibration sources at the μCi level, so that eventual calibration procedures will only take minutes to accumulate a meaningful statistic rather than hours. This binds the highest expected incoming rate to less than one event for every ten microseconds. By choosing a peaking time of 10 μs we can certainly handle the rate during calibration. The peaking time sets then the system bandwidth and the noise and will need to be optimized once the circuit is prototyped to obtain the best noise performance. The resulting preliminary circuit is shown in Figure 5.13.

The first operational amplifier, OA_1, transforms the current pulse (ideally a delta) into the step function required for the filter to present the desired exponential response, via integration on the feedback capacitor $\mathrm{C}_{F,1}$ chosen to be 250 pF. This introduces a transimpedance gain equal to $1/C_{F,1}$, or 10^9 times the amount of charge generated by the SiPM (thus the voltage step is expected to be in the mV range). The resistor $R_{F,1}$ simply restores the output voltage of OA_1 to the baseline value and does not affect the subsequent shaping process, so long as the time constant $\mathrm{R}_{F,1} \cdot \mathrm{C}_{F,1}$ is

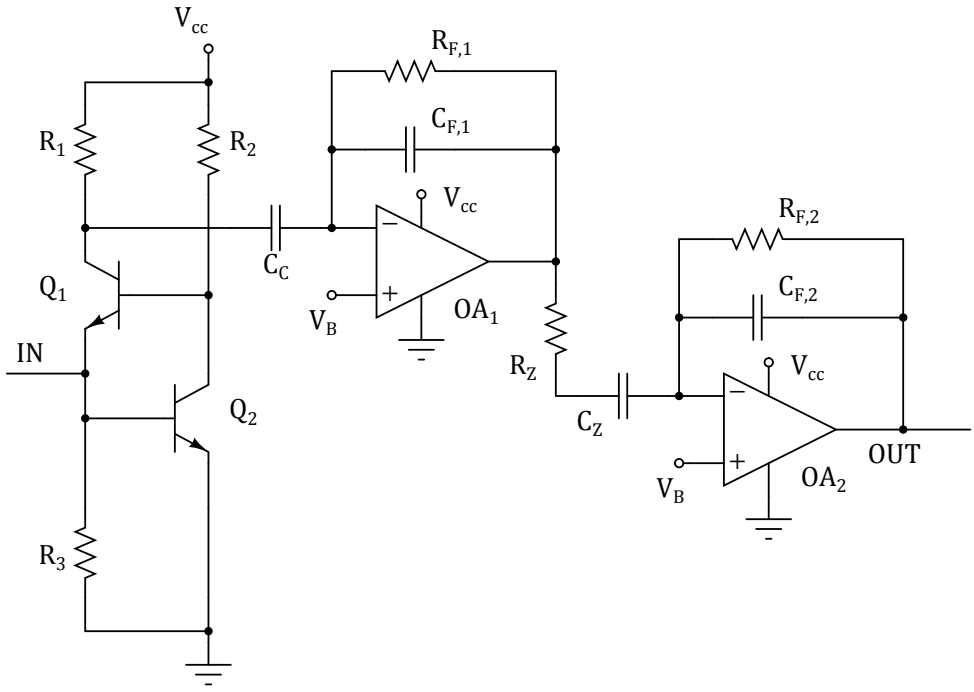


Figure 5.13: Preamplifier, first stage and second stage of readout.

$\gg \tau_p$. Note that the stage has been AC-coupled to the input amplifier via the capacitor C_C. For this stage we selected $C_C = 10$ μF and $R_{F,1} = 1$ Mohm. The second stage is the actual filter. The transfer function is rather easy to calculate in term of Laplace formalism, by remembering that, for an operational amplifier with negative feedback the gain equals the ratio between the impedance located from the output to the virtual ground, and the impedance from virtual ground to the input. Thus, we can write as transfer function for OA_2:

$$T(s) = \frac{sC_Z R_{F,2}}{(1 + sC_Z R_Z)(1 + sC_{F,2} R_{F,2})}$$

In this equation, s is the Laplace complex variable. We recognize immediately one zero at the numerator and two poles at the denominator. The inverse transform is in general the weighted sum of two exponents. When the poles coincide, we have the desired semi-Gaussian shape, thus we set $\mathrm{R}_Z\mathrm{C}_Z = \mathrm{R}_{F,2}\mathrm{C}_{F,2} = \tau$. In this special case, the inverse transform is:

$$\mathcal{L}^{-1}\left\{\frac{s\mathrm{C}_Z\mathrm{R}_{F,2}}{(1+s\tau)^2}\right\} = \frac{\mathrm{C}_Z\mathrm{R}_{F,2}}{\tau^3}\cdot(t-\tau)\,\mathrm{e}^{-\frac{t}{\tau}}$$

This is also the response of the stage to a delta at its input. When a charge pulse is collected by the current amplifier, it is integrated by the first stage and becomes a step (we will assume of unity amplitude for simplicity), therefore, the output response to this step will simply be the integral of the pulse response above, with the initial condition of being zero at $t = 0$:

$$v_{0,step}(t) = \frac{\mathrm{C}_Z\mathrm{R}_{F,2}}{\tau}\cdot\left(\frac{t}{\tau}\right)\cdot\mathrm{e}^{-\frac{t}{\tau}}.$$

From this expression, we can calculate any characteristic of the output: from the voltage gain to the peaking time. The peaking time has been set to 10 μs; the gain, defined as the ratio of the output peak to the input step size, is set by $\mathrm{C}_Z\mathrm{R}_{F,2}/\tau^2$. For $\mathrm{C}_Z = 1$ nF ($\mathrm{R}_Z = 10$ kohm) and $\mathrm{R}_{F,2} = 40$ kohm ($\mathrm{C}_{F,2} = 250$ pF), the value is about $1,450$. Since only millivolts are expected after the integration stage, this is a perfectly reasonable value to choose. As operational amplifiers, we chose a high accuracy, micropower, CMOS dual operational amplifier, TSV732, made by ST-Microelectronics. The detailed description is given in [66]. What is important to highlight for our purposes is the bandwidth (900 kHz), the power consumption (60 μA at 5 V supply) and the fact that the device is able to swing rail-to-rail input and output. This is important for our low power design, where the supply voltage is only 2.8 V. The voltage V_B in Figure 5.13 is simply a DC baseline shift to account for the fact that the nominal baseline zero is not an actual ground in our single-supply design. The value for V_B was chosen to be 1.4 V, the midpoint of the DC bias, but could be optimized later on to

a different value to improve our use of the operational amplifier's dynamic range.
The plot in Figure 5.14, shows the calculated response of the filter stage only to an ideal unity step at the input, with a 10 μs peaking time. The

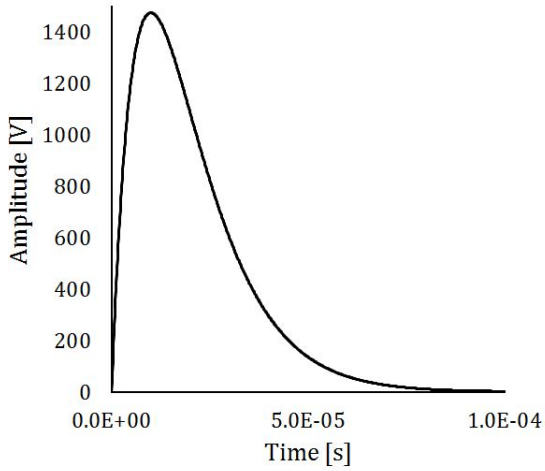


Figure 5.14: Calculated filter time response to a unity voltage step for a peaking time of 10 μs.

expected response of the filter will be somewhat different, as the integrator stage does not produce a perfect step at its own output. In fact (see Figure 5.13), the charge restoration mechanism is implemented by the resistor $R_{F,1}$; this means the voltage developed on $C_{F,1}$ upon an integration will not remain constant in time, but will slowly decay exponentially to the baseline with a time constant equal to $R_{F,1} \cdot C_{F,1}$=1 ms. This has the effect of modifying the filter's output by introducing a zero. The output will undershoot the baseline for some time, rather than being strictly unipolar. The modified shape is normally undesirable because if a new event happens during the pulse tail, its peak value would be contaminated by the tail value over imposed to it. Of course, this could be compensated through

what is called a pole-zero cancellation process [67], where a stage with a pole equal to this zero and a further, longer zero is introduced in the chain to remove the original zero, but given the low expected rate, and the fact that we do not need the peak value of the pulses, we deemed this process not necessary.
The PSpice simulation in Figure 5.15, finally shows the response of the circuit to a mid-range energy event. The tail introduced by the non-ideality of the charge restoration in the charge amplifier is clearly visible. What is

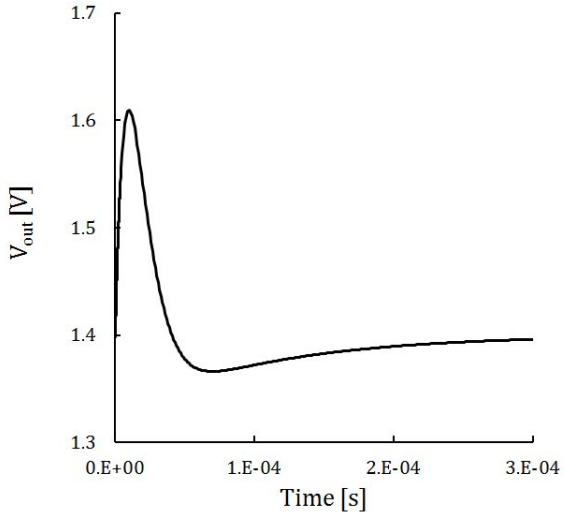


Figure 5.15: PSpice simulation of the circuit output response to a mid-range energy event.

really striking about this circuit is its ability of presenting the same response as the input load varies. With a resistor or a charge preamplifier connected directly to the photosensor, we would expect at a minimum a change in shape, due to the added time constant formed by the detector capacitance and the impedance of the readout electronics. This time constant would vary with the load, so that the peak position would shift to longer peaking

time for larger loads. In the case of charge preamplifier, we would expect also some signal loss due to the capacitive divider at its input formed by the detector capacitance and the dynamic input capacitance of the charge amplifier. None of these effects are visible in our circuit (see Figure 5.16.

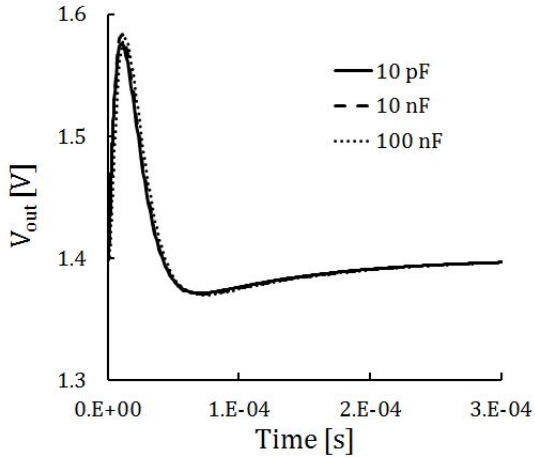
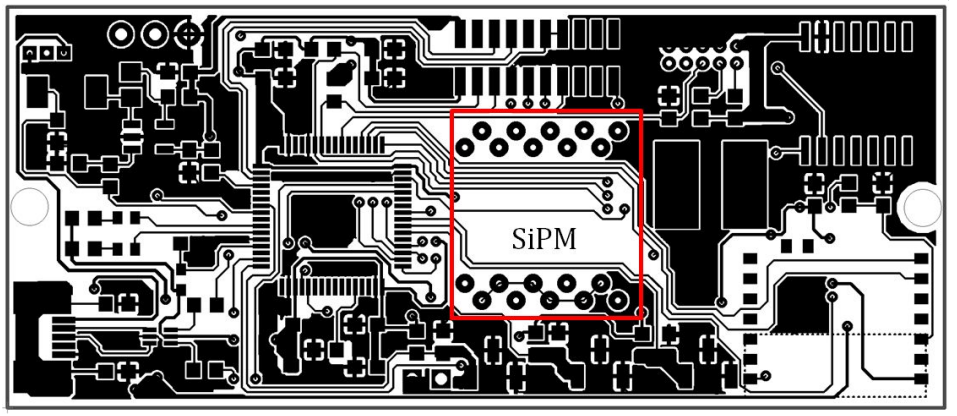


Figure 5.16: PSpice simulation of the circuit output response to a small-signal event from sensors with 10 pF, 10 nF and 100 nF capacitance.

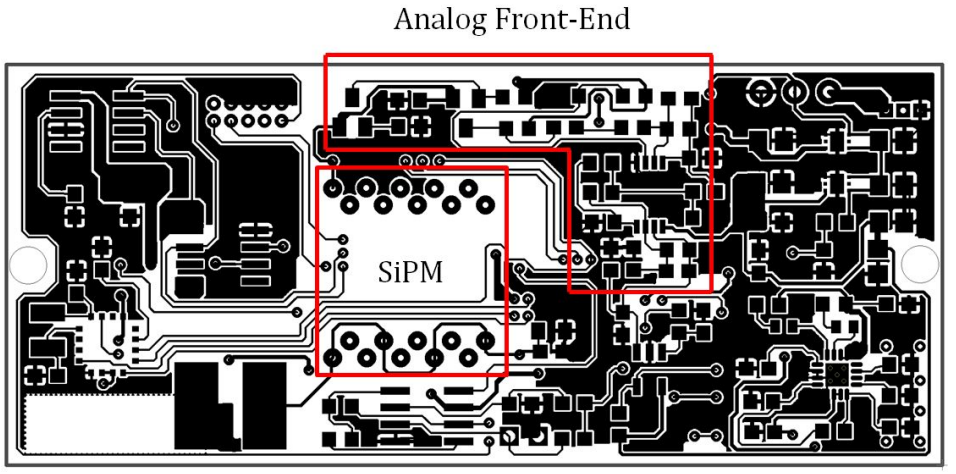
5.4.5 Physical design

With the design complete, the next obvious step is to implement the physical design into a test board and perform some validation tests.
Given the simplicity of this design, rather than implementing the readout only, we decided to incorporate all of the system features in the test board directly, and bypass individual prototyping of said features. The layout was kept modular in nature, to preserve the ability to incrementally test each function independent of the others. At this time we will focus solely on the analog readout tests. The remaining features will be addressed later on.

The layout, done with Eagle Layout Editor [68], is shown in Figure 5.17. The board has 2 layers and measures 7 cm $\times$ 3 cm. The analog front-end



(a)



(b)

Figure 5.17: Prototype board, top (a) and bottom (b) layer.

section and the SiPM location can be seen. The remaining sections on the board include DC-DC converters to generate the necessary bias to the circuits and the SiPM from a Li-polymer battery, a comparator that triggers on events, a microcontroller to coordinate between sensors and the external world, and communication circuits (USB, Bluetooth and NFC). A picture of the assembled board is shown in Figure 5.18.

We conducted several tests on the circuit. First of all, we verified the func-

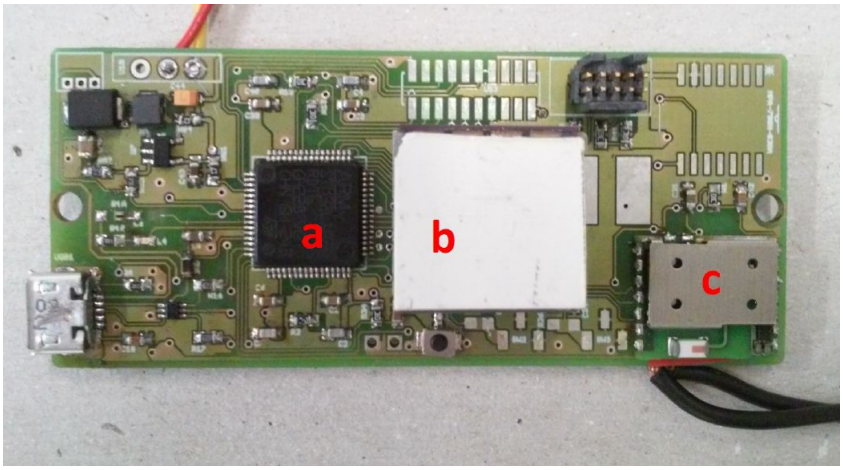


Figure 5.18: Picture of the fully assembled board. The microcontroller (a), SiPM coupled with a ZnS screen (b) and Bluetooth module are shown. The front-end electronics are located on the bottom layer. (Photo courtesy of Benedetta Nodari, University of Bergamo.)

tionality in response to ideal electrical pulses and compared the results to simulations. In order to do so, a charge pulse was injected into the circuit without the SiPM sensor to simulate a charge event. The easiest way of doing this is to apply a voltage step to a capacitor connected in series with the circuit input. The amount of charge to be delivered is arbitrary, but should be representative of a real event and within the limits of the circuit dynamic range. For sake of testing, we chose 3.4 MeV-equivalent. With a

1 nF injection capacitor (C_{inj} in Figure 5.19), we find that the step needs to be 96 mV in amplitude.

This is the most basic test to verify circuit functionality, but it is also

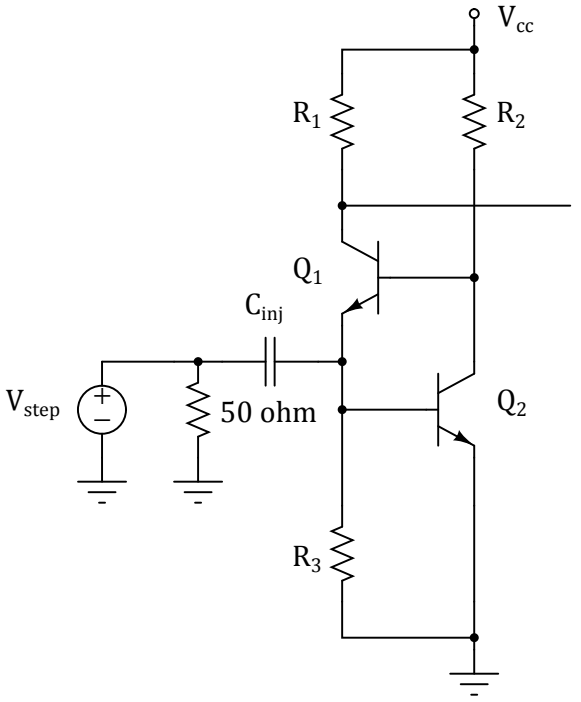


Figure 5.19: : Schematic of the test pulse injection mechanism.

a simple way of finding the correlation between simulations and actual response (at least within the boundaries of the linear region of the circuit). Note that the capacitor C_{inj} also simulates the load condition set by the detector, as in the absence of signal, it is directly connected to ground through the voltage generator V_{step}. The result of this test is shown in Figure 5.20, where the measured output is compared to the results from the simulation. The match between simulated and measured data is excellent. Another important measurement is to analyze the behavior of the circuit

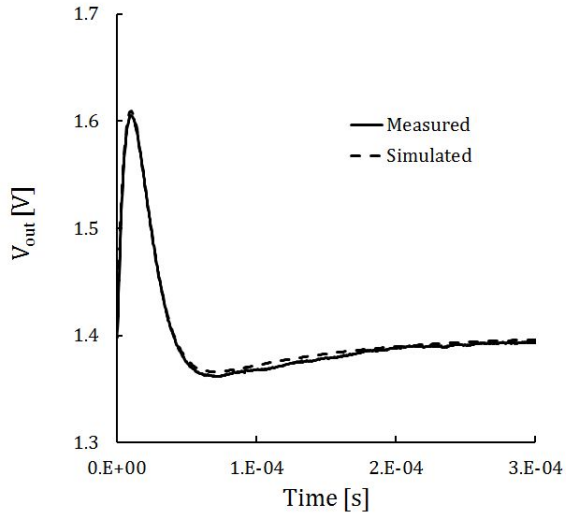


Figure 5.20: Comparison between simulated response (dashed line) and measured response (solid line) under the same conditions.

to saturation. This is necessary, as energetic cosmic rays interactions may be close to the dynamic range limits. There will be a small number of such pulses; however, we need to be sure their presence will not introduce any unnecessary behavior. Such behavior is shown in Figure 5.21, where the response to a large event is recorded. The region around the peak is expanded for convenience. We see that the overall response does not show any unexpected or undesired behavior. We should add that, although not shown in Figure 5.21, the tail portion of the pulse behaves as well.
Before moving to performing measurements with the sensor coupled to the circuit, it is useful to repeat the measurements above with a larger capacitance to simulate the sensor presence, and once again, compare simulations and real data. For a 13.6 nF capacitor, the plot in Figure 5.22 shows the good agreement between the simulated and the measured data. Perhaps, what is more interesting is to compare the normalized peak output of the

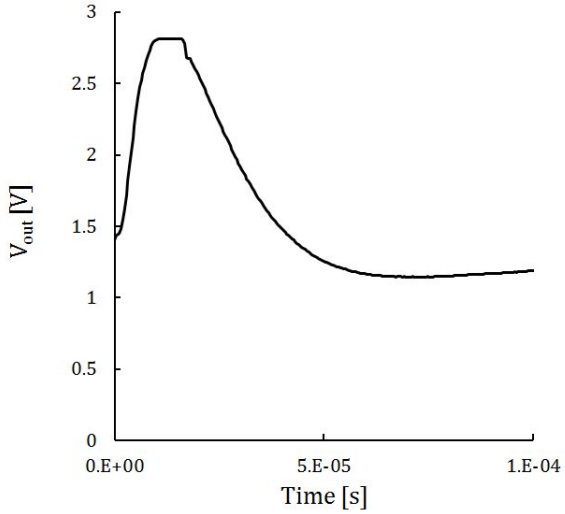


Figure 5.21: Response to a large charge event.

measured signal under the two different load conditions (1 nF and 13.6 nF). This is done in Figure 5.23 and confirms experimentally the simulation results previously depicted in Figure 5.16.

Another interesting parameter is the circuit's linearity. This would be a major parameter in a spectroscopy system; however, in our case we only count events, regardless of their energy. For completeness, the simulated linearity of the current buffer is plotted in Figure 5.24, and a measurement of the overall non-linearity is shown in Figure 5.26. Note that in Figure 5.24 we plot the peak current from the input current amplifier (i.e. the current in the collector of Q_1) versus the total charge generated in the SiPM, before internal amplification. This way, it is easy to determine equivalent energy once an accurate energy calibration is available. The value at zero input is the DC current in Q_1 (350 μA). Clearly, the input stage favors electron collection, however, the fact that the current saturates is not a complete indication of the overall linearity of the circuit. In fact, since the second stage

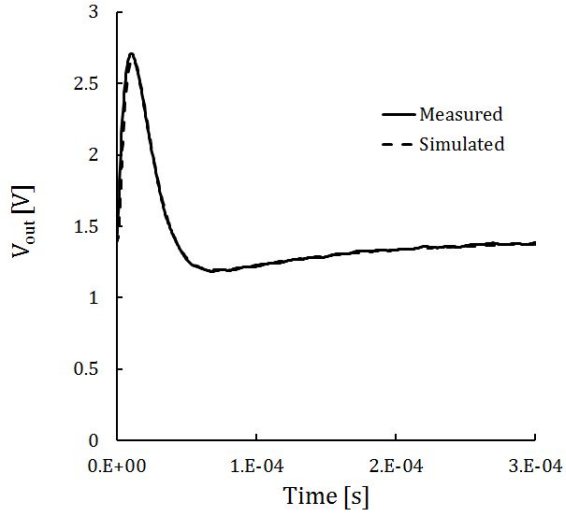


Figure 5.22: Comparison between simulated response (dashed line) and measured response (solid line) for a detector-like input load (13.6 nF).

is an integrator, it is able to read different amounts of charge from saturating current signals, depending on how long the signal remains saturated. Therefore, there is still a dependence between the time the signal spends at saturation and its energy content (or its intended amplitude). This is illustrated in the simulation of Figure 5.25. Here, the collector current of Q_1 and the corresponding response of the circuit show how the time spent by a saturating current signal beyond the circuit range does not show as a saturated signal at the output of the shaper (note the different time and amplitude scales in Figure 5.24). In light of the previous observations, we ran a simulation to estimate the overall dynamic range, as well as a set of measurements under the same conditions as simulated. Figure 5.26 shows the comparison between simulated and measured behaviors. It is noticeable how the simulated data differs from the measured data in the non-linear regions. This is expected, as often PSpice models for discrete design do not

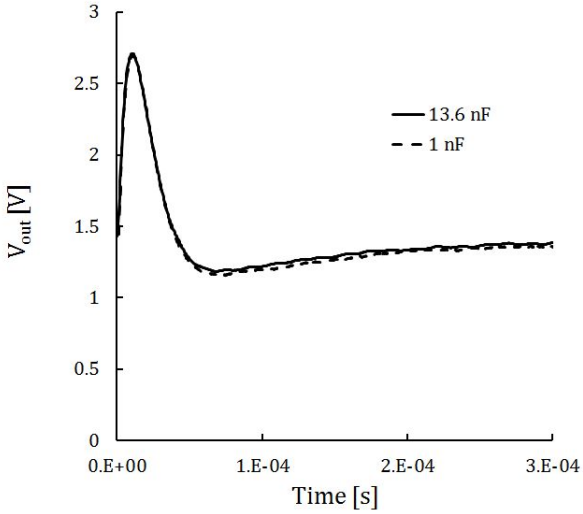


Figure 5.23: Comparison between the measured response with a 1 nF and 13.6 nF detector capacitance.

address these regions well. Finally, the electronics noise of the circuit for a 13.6 nF capacitive load was measured with an r.m.s. meter and found to be about 1.6 mV r.m.s. over the circuit bandwidth. This value represents about 0.06% of the full dynamic range. If we were to set our gain such that the full dynamic range corresponds to 3.5 MeV, the electronics noise contribution would only be 2.2 keV. In reality, as it will be shown in the following paragraph, this is not the only noise source present in the circuit once the detector is connected.

5.4.6 Detector tests

Proving that the circuit works satisfactorily is only half of what is needed to have a working solution. In fact, it is imperative to understand how the circuit integrates with a detector and how the system obtained behaves.

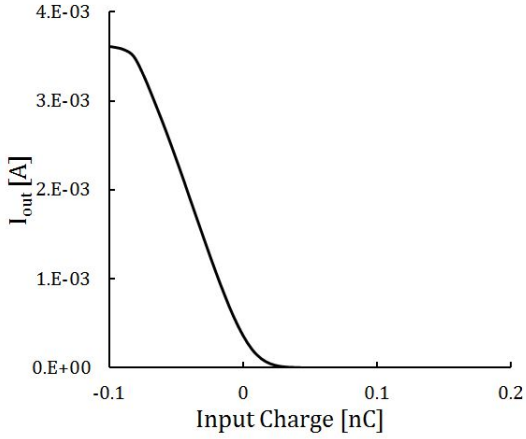


Figure 5.24: Simulated linearity of the current buffer.

For example, we will ultimately count the number of events seen by the detector by adding a comparator at the circuit's output. The comparator threshold will be set above the noise at the output of our circuit to ensure a low number of counts when no source is present. We know that the circuit presents an intrinsically low electronics noise, but we also know that the detector, once connected, will introduce a large amount of leakage current and, consequently, may introduce high parallel noise. This may change the assumptions one would make on where to set the threshold, based on the circuit noise only. To address that, we connected the SensL sensor to our prototype circuit and ran a few measurements. The contribution of the parallel noise was immediately evident, in that the noise floor increased by about 2 orders of magnitude. This is a signal that we are dominated by the parallel noise associated to the detector leakage. Thus, it is important to know how the noise affects the measurements. To estimate that, we performed a measurements of the response to an alpha source of the same ZnS(Ag) detector used in our system, coupled to a PMT (Hamamatsu R11265U-100). The PMT output was read from a 50 ohm resistor to ground via a conventional spectroscopy system (spectroscopy amplifier

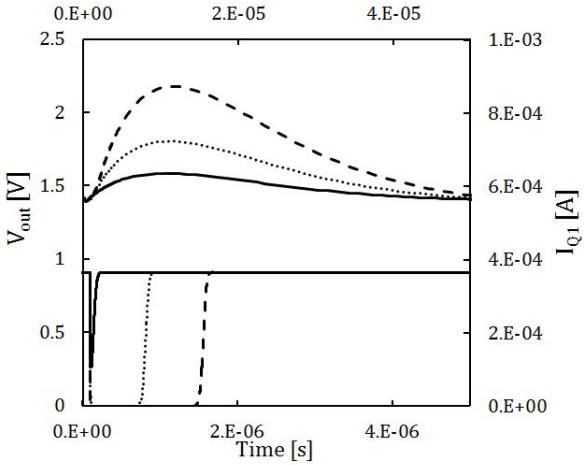


Figure 5.25: Effects of saturation on a charge amplifier. The dynamic range is increased despite saturation occurring before the charge amplifier. (Note: the solid line curve was expanded by a factor 8 to improve its visibility on the graph)

at 8 µs peaking time); the measurement was executed in a dark box with the PMT biased at -1130 V to ensure a gain close to 10^6, with a quantum efficiency of 35%. The low energy threshold was set just above the electronics noise. A second measurement without source was taken afterwards in order to allow background subtraction. We obtained the spectral response, over a 60 seconds acquisition time, shown in Figure 5.27, solid line.

We repeated the measurements above by coupling the ZnS(Ag) detector to our SiPM. In this case, the bias of 27 V, should lead, according to the datasheet [57], to a gain of $4 \cdot 10^6$, a quantum efficiency of about 30 %, and a dark current of about 100 μA. Under these conditions, the two optical sensors result sufficiently matched to allow for comparisons to be made. The measured spectrum is shown by the dotted line in Figure 5.27. In order to correctly compare the two spectra, we need to compensate for the different dynamic ranges and gains. A quick Gaussian interpolation shows the peak

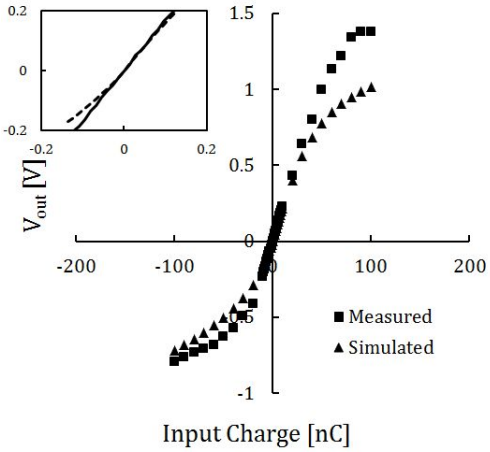


Figure 5.26: Measured circuit. The inset shows a detail of the linearity for small signals.

of the PMT response around channel 119, and that of the SiPM around channel 32. In order to match the gains, we need to stretch the response of the SiPM by a factor $119/32 = 3.7$, but we also need to divide its amplitude by the same factor in order to keep the total counts recorded the same. The result of this operation is the spectrum indicated by a dashed line in Figure 5.27. As can be seen, the spectra are a very good match, with the exception of the high-energy tail, where non-linearity effects may be taking place; also, a difference is in the effective low energy threshold (below this threshold the electronics noise would introduce unwanted counts). The total counts in 60 seconds for the PMT measurement were $130,000$; for the SiPM measurement $114,100$. This is only a 12.2% loss in total counts. For our prototype, we believe a 12% loss is quite acceptable compared to the savings in complexity. Further optimization of the circuit peaking time and currents may reduce this loss.
Another optimization parameter is the SiPM bias voltage. In fact, as the overvoltage increases, the quantum efficiency and gain increase as well, so

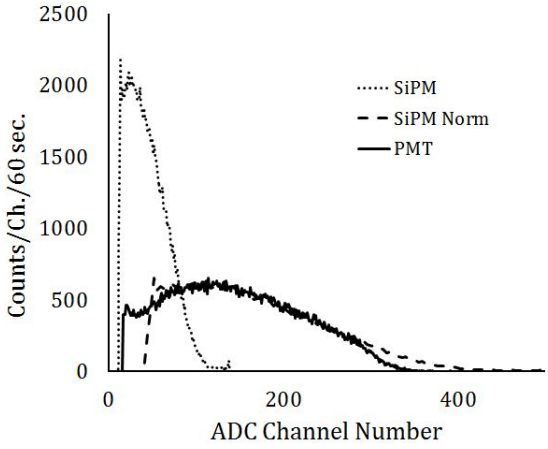


Figure 5.27: Comparison between our system response and the response recorded using a traditional PMT-based spectroscopy system. Both spectra are backround-subtracted.

that more light is collected. However, the dark current increases too (note that this increase is not necessarily due to the gain increase, but rather to the increase in overvoltage). We need to investigate if there is a dominant effect, or if an optimum point exists. To this purpose, we acquired the background-subtracted spectra in Figure 5.28 at different bias voltages, by adjusting the low energy threshold each time and by taking any saturated events into account. From the plots, we can clearly see the effects of the dark current on the threshold. The apparent decrease in counts in the linear region is due to the fact that as the gain is increased, the spectra move toward lower energies where fewer counts are present, but the peak is compressed into the non-linear region. The overall counts are summarized in Figure 5.29. From the figure we can clearly see a peak in total counts recorded around the 27 V bias. This means that initially the gain and QE increase has a beneficial effect on efficiency, but after the 27 V bias point the effects of the additional dark current on the noise start dominating so

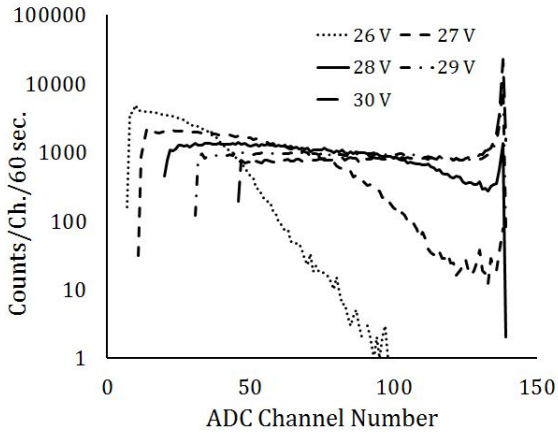


Figure 5.28: Background-subtracted spectra at different values of bias voltage.

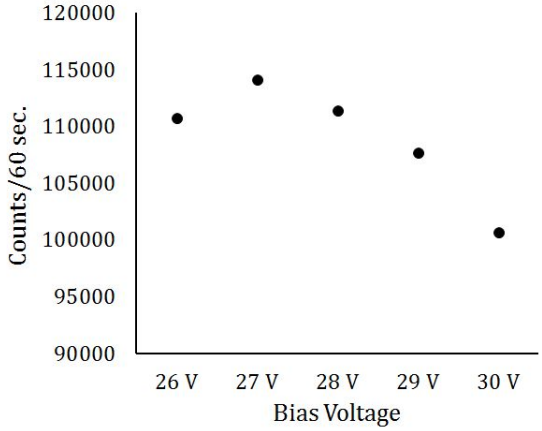


Figure 5.29: Counts recorded in 60 seconds as a function of SiPM bias voltage.

that counts are lost to the higher threshold.

5.4.7 Design improvements

So far, we have shown that the design is sound and, for a 12% trade-off in detection efficiency compared to PMT-based designs, has very desirable features for a portable instrument: very compact footprint, low sensor bias voltage, ruggedness, large detecting surface. There are however some possible future improvements that may increase efficiency even further or decrease power consumption. We set the bipolar transistors currents to values that optimize the series noise component. We later noticed that the dark current is the main parameter to consider when setting the detection threshold. This means that not all of the g_m in the transistor Q_2 is actually needed, because the thermal noise of that transistor does not dominate. Thus, we could reduce that current and save power. Another important observation is we are reading out the full area of the SiPM with a single channel, to optimize power consumption. As a drawback, we take a 12% hit in detection efficiency due to the large leakage. If we were to subdivide the sensor readout in more channels, each reading out a 1/4 of the total area or a quarter of the total dark current, it would be reasonable to expect better low energy threshold. This approach would require using four times the area and four times the power, but depending on which requirement among power or detection efficiency is preferred, it could become a desirable option. Figure 5.30 shows a comparison between the background-subtracted spectrum acquired originally from the full SiPM array through a single channel of readout and that acquired through 1/4 of the total SiPM area. We see the improvement of a factor 2 in the threshold. The counts in the original spectrum were about $114,100$; in the second spectrum, we counted about $40,000$: a clear detection efficiency improvement overall.

System integration

The readout electronics, in the preliminary design previously examined has been integrated with a suite of sensors as the application requirements man-

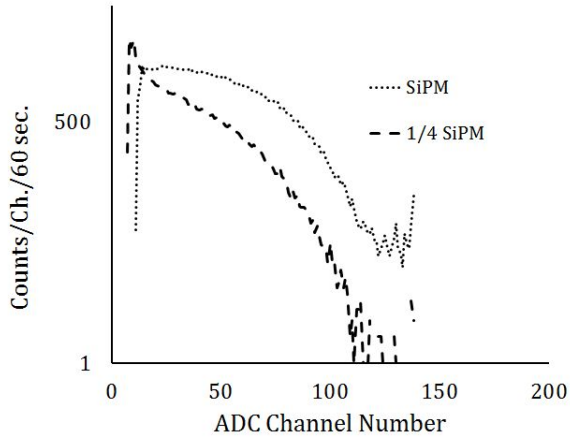


Figure 5.30: Comparison between counts recorded by the full SiPM area, and only one quarter of the same area.

date. Although this is outside of the scope of this work, we will give a quick overview of the whole system. The system has been conceptualized according to the block diagram in Figure 5.31. Here, the power supply section provides the necessary voltages to bias the analog front-end, the comparator, the SiPM as well as any other of the system components. The voltages are obtained from a single-cell Li-ion battery with a 300 mAh capacity ($33 \times 28 \times 3.5$ mm). The battery can be recharged either via micro-USB connector or energy harvesting RFID devices, according to the most modern charging standards. The suite of sensors includes pressure, humidity and temperature, in order to be able to correlate radon readings with environmental conditions. Communications are implemented via Bluetooth and RFID ISO-15963-compliant module that offers the possibility of external readout at zero power for the on-board systems. The fully assembled system is shown in Figure 5.32. Preliminary system measurements were done in the laboratory on a time span of about 167 hours (roughly 7 days). Radon counts and data from humidity, temperature and pressure sensors

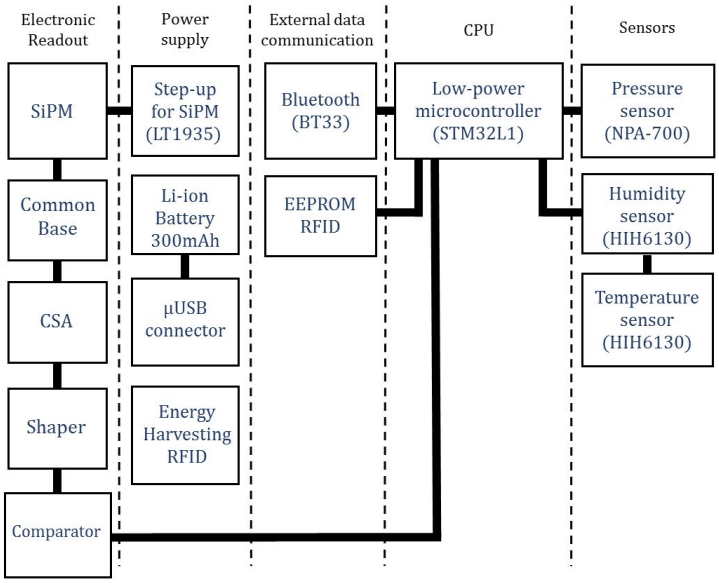


Figure 5.31: System block diagram. The components part numbers are shown in the respective box where applicable.

were extracted. The resulting plots are shown in Figure 5.34. Note that the count rate is of the order of 100 Bq/m^3, or about 10 counts per minute in the detector[3]. The statistics associated with this count rate are of the order of $\sqrt{10} \approx 3$ counts per minute. Thus, we expect significant variance in the data if plotted as a series of one minute-long acquisitions. For this reason, we applied some smoothing after collection. The resulting plot is shown in Figure 5.33, where the patterns present in the original data series are still clearly visible. Of course, the data can be smoothed further by, for example, opting for hour-long averages. This is a practical

[3]This number corresponds to 2.7 pCi/l, quite comparable to the rates we expect to measure in many practical scenarios

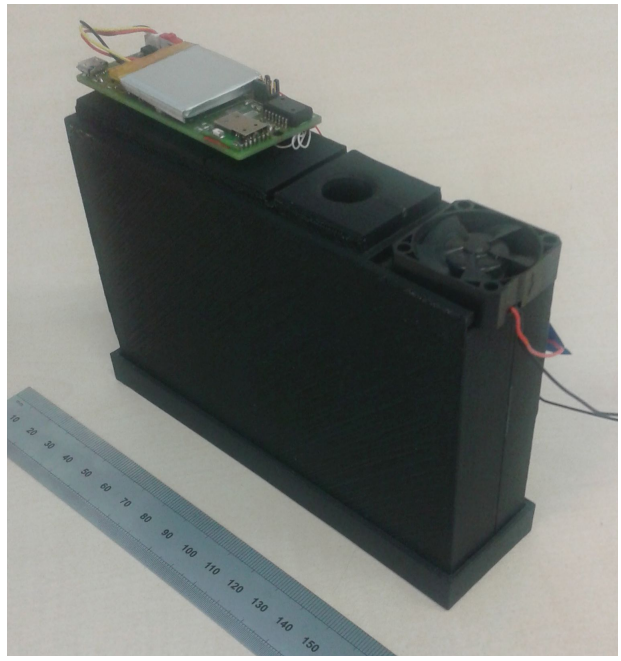


Figure 5.32: Fully assembled system. The covers that prevent light from reaching the SiPM are not shown.

system choice to be optimized after more measurements have been done. Some attempts to establish if a correlation between the radon dataset and humidity, temperature and pressure exists were made, but no obvious correlation was found (0.3, -0.6 and 0.3 for the respective correlation factors). This is expected, as the variations in environmental values are rather small. The most interesting result is the weak correlation between the temperature curve and the radon counts. A strong correlation would indicate that either the radon concentration depends directly on temperature (not substantiated by research), or that some of the electronics readout settings drift with temperature (most likely the SiPM gain and/or the low-energy threshold). Obviously, if temperature behavior were a concern, a more

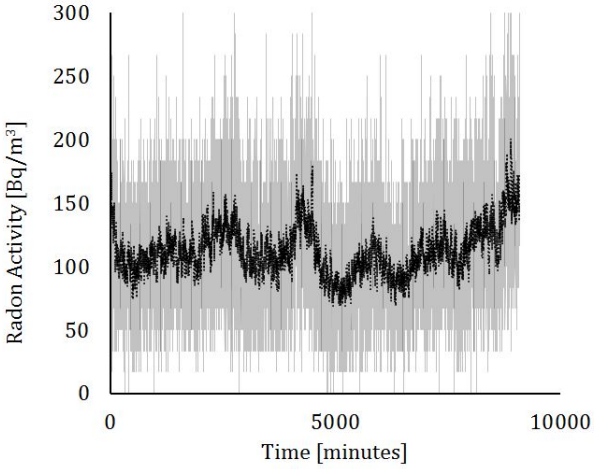


Figure 5.33: Acquired radon activity data before (black line) smoothing and after (gray line) smoothing.

in-depth analysis would be necessary.

5.5 Conclusions

In this chapter we have shown an application of the design criteria for SiPM-based systems to a very simple case where a relatively large SiPM is used to read out a scintillator. Even in this simple application the electronics noise assumes great importance: if it were not for the high dark current, the SiPM would behave the same way as a traditional photomultiplier tube. The effect of noise is such that for the given SiPM area the same low energy threshold cannot be reached compared to using a PMT. This highlights how, despite becoming a very popular device, SiPMs should not be considered exactly equivalent to more traditional solutions, but some care must be taken in designing readout circuits. From this design we learned a number of facts that can be very useful for future designs or for improvements to the current

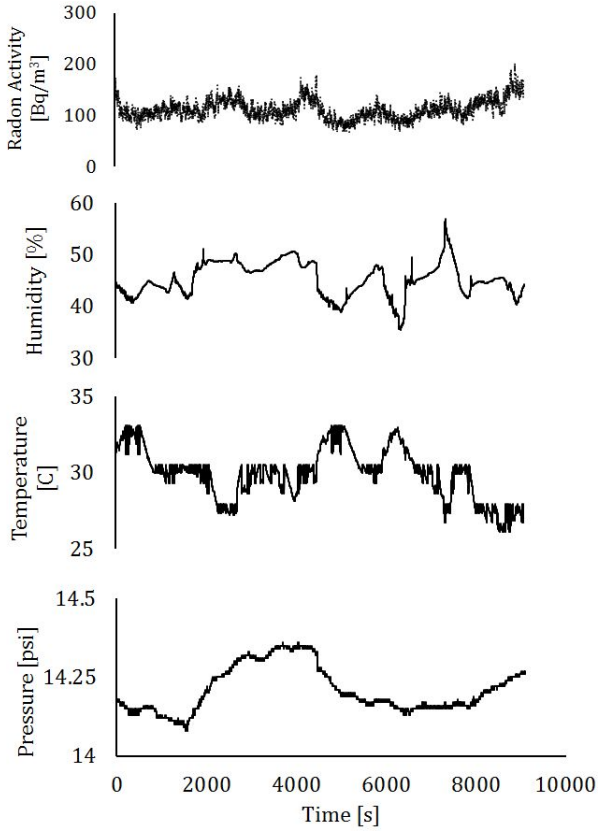


Figure 5.34: Smoothed radon activity data and environmental data.

design:

- A better low energy threshold can be reached at the expenses of consuming more power if the readout is segmented. Large SiPMs lend themselves very well to segmentation because they are made of small cells individually accessible to the readout.

- Using brighter scintillators allows for a larger instrumented area for

each channel. We have seen that the low energy threshold depends on the amount of dark current; therefore, if more photons are produced per unit energy, the resulting SNR is better and the low-energy threshold can be effectively lowered.

- There is an optimum bias point for the SiPM where the effects of increased quantum efficiency and increasing dark current balance to offer the best achievable SNR.

As for the project in particular, we only exposed a general preliminary design and more things need to be accomplished. Future plans include optimization of SiPM segmentation (i.e. how many more channels of readout can the power budget afford in relation to what detection efficiency improvement can be obtained), some noise versus peaking time analysis (is there a peaking time where the dark noise contribution is less?), measurements with known sources to verify the expected vs. measured detection efficiency, with consequent determination of calibration curves.

Chapter 6

Square Meter-class, Low-Background Design for a Neutrino-less Double-Beta Decay Experiment

6.1 Introduction

In this chapter we will look at a class of detectors that fully benefit from the theory developed previously in Chapter 4. These detectors have, generally speaking, large areas of light sensors as well as stringent requirements on radioactive backgrounds induced by the materials the detectors themselves are made of, including the instrumentation. The reason for this is the rarity of the physics events the experiment is designed to observe: this makes larger detectors very desirable, but requires that the events within the instrument not be contaminated by unnecessary backgrounds. These requirements drive the whole design: low intrinsic radioactivity demands for extremely clean materials and for a minimization of the overall mass,

but large sensor area drives the design in the opposite direction, at least as far as the amount of necessary material is concerned. This poses significant limitations on the instrumentation in term of availability of components, sensors etc. that meet the specifications. These are common threads to this kind of experiment. For a specific experiment there will also be further constraints dictated by the specific implementation (power is an example).

6.2 $0\nu\beta\beta$ decay

The reason why many large area experiments are becoming so important is found in the future directions taken by the field of Nuclear Physics [69]. In recent years, researchers have made significant progress towards the understanding of the nature of the strong and weak interactions and have identified new science to improve such understanding. The new science will lay the path towards new important discoveries. One new area is the investigation of neutrino properties to support the New Standard Model of fundamental interactions.
The neutrino is the only known fundamental particle that has no electric charge; it is presently unknown whether neutrinos are their own antiparticle, or possess distinct anti-particle states (antineutrino). In the former case, they would be Majorana particles, while in the latter they would be Dirac particles. The distinction is important, because if neutrinos were Majorana particles, their interactions would violate important conservation laws (in this case, the total lepton number conservation). These laws are important ingredients in a possible explanation for the predominance of matter over antimatter in the early universe. If confirmed that neutrinos are Majorana particles, there would be strong evidence in support of this theory. On the other hand, if neutrinos were Dirac particles, then other explanations for the matter/antimatter imbalance would be more likely.
Another important unknown property of neutrinos is their mass. This is a very important parameter in Cosmology. Their mass is bound on the lower limit by a value of 0.02 eV dictated by oscillation data and by a number hundred times larger on the upper limit from some decay observations.

From Cosmology, that limit would be 0.2–0.5 eV [70]. It is obvious there needs to be a more conclusive test on the neutrino mass.
In simple words, the future of many Nuclear Science experiments is to answer the fundamental questions: are neutrinos their own antiparticle? What is the mass of neutrinos?
A single process has been identified as the best tool to answer both questions. It is known as neutrino-less double-beta decay, or $0\nu\beta\beta$ [71, 72]. In this process, two neutrons are converted into two protons with the emission of two electrons and no antineutrinos. The dual is also a $0\nu\beta\beta$ process (two neutrons into two anti-protons with emission of two positrons and no neutrinos). The absence of neutrinos (or antineutrinos) would imply the lepton number violation and prove that neutrinos are Majorana particles. The same observations would provide information on the neutrino mass as well.
Another fascinating phenomenon is the possibility that the $0\nu\beta\beta$ could proceed with the exchange of a heavier Majorana particle postulated by other New Standard Model candidates [73].

Currently, there are at least three large international experimental collaborations addressing the $0\nu\beta\beta$ decay, each using different (and sometimes complementary) techniques to conduct their observations. CUORE is a US/Italian/Spanish collaboration that intends to search for $0\nu\beta\beta$ in the isotope ^{130}Te; the Majorana collaboration (Canada, Japan, Russia, USA) uses ultra-pure ^{76}Ge detectors and EXO, a collaboration composed by 138 scientists from seven countries, focuses on the unique approach of using the isotope ^{136}Xe. The EXO collaboration has already demonstrated the existence of $2\nu\beta\beta$ decay[1] in ^{136}Xe in a 200 kg-experiment called EXO-200, giving also a competitive limit on the $0\nu\beta\beta$ decay [74]. The common thread to these experiments, as anticipated, is the need for a large volume (driven by the rarity of $0\nu\beta\beta$ decay) of detecting material and extremely low backgrounds for the same reason. In the near future, the US Department of

[1] This process is very similar to $0\nu\beta\beta$ decay, with he exception that neutrinos are observed.

Energy may fund a ton-scale experiment to search for $0\nu\beta\beta$ decay.

6.3 The nEXO experiment

The EXO collaboration is among the candidates that might be funded to search for $0\nu\beta\beta$ decay. As mentioned in the previous paragraph, the collaboration has already built an medium-scale experiment (a few 100 kg) and successfully recorded $2\nu\beta\beta$ decay events in the isotope ^{136}Xe. Unlike the materials used by the other collaborations, ^{136}Xe is a gas at room temperature and becomes a liquid at about 161 K. Its gaseous form is not so appealing at the moment because the density is too low to efficiently have enough atoms per unit volume, but its liquid form, with a density of 2.9 g/cm^3 compares favorably to crystalline materials such as germanium (5.3 g/cm^3) and Tellurium (6.24 g/cm^3), the candidate materials for the other two experiments. The natural abundance of the ^{136}Xe isotope is also good. In Table 6.1 we show the isotopic abundance in materials that, according to predictions, have a likelihood of presenting $0\nu\beta\beta$ decay.

Xenon compares very well to any other candidate material. The abundance is important because large quantities of material will be needed, and the cost of an isotope is directly dependent on its abundance. There are also other reasons why Xenon is an ideal material: there is no need to grow crystals; it can be continuously purified and is easily transferrable to storage or to a different detector. Being a noble gas, Xe is also easier and safer to enrich in a centrifuge, as no chemistry is involved in the process.
The expected $0\nu\beta\beta$ decay theoretical spectral signature in Xenon is summarized in Figure 6.1 [75]. In the figure, the number of events per year is plotted versus electron energy. The detector resolution was assumed to be 1.6 %, a conservative number for liquid Xenon. The first, large peak is made of events from the regular double-beta decay (where neutrinos or antineutrinos are emitted). The small peak just outside of the main peak tail corresponds to the $0\nu\beta\beta$ decays. The area of the small peak is about $1,000$ times smaller than the area of the large peak; this means that, in

Candidate → Decay Product	Natural Isotopic Abundance (%)
^{48}Ca →^{48}Ti	0.187
^{76}Ge →^{76}Se	7.8
^{82}Se →^{82}Kr	9.2
^{96}Zr →^{96}Mo	2.8
^{100}Mo →^{100}Ru	9.6
^{110}Pd →^{110}Cd	11.8
^{116}Cd →^{116}Sn	7.5
^{124}Sn →^{124}Te	5.64
^{130}Te →^{130}Xe	34.5
^{136}Xe →^{136}Ba	8.9
^{150}Nd →^{150}Sm	5.6

Table 6.1: Candidate $0\nu\beta\beta$ decay materials and their isotopic abundance (courtesy of G. Gratta, Stanford University).

this scenario, about 35 events are expected each year from the $0\nu\beta\beta$ decays, out of $40,000$, highlighting both the importance of minimizing any additional background in the region of interest (from 2.3 keV to 2.6 keV) and increasing the fiducial detector volume to increase the number of useful events.

6.4 Experiment design

The design of the nEXO experiment is based upon experience gained by building a smaller scale concept instrument, called EXO-200. This detector, operated between May 2011 and February 2014, was a 200 kg–^{136}Xe time projection chamber (TPC). A time projection chamber is a detector where interactions occur within a homogeneous volume. The ionization products from the interaction are drifted towards a collecting electrode where they can be mapped based on their time of arrival. Thanks to this time map,

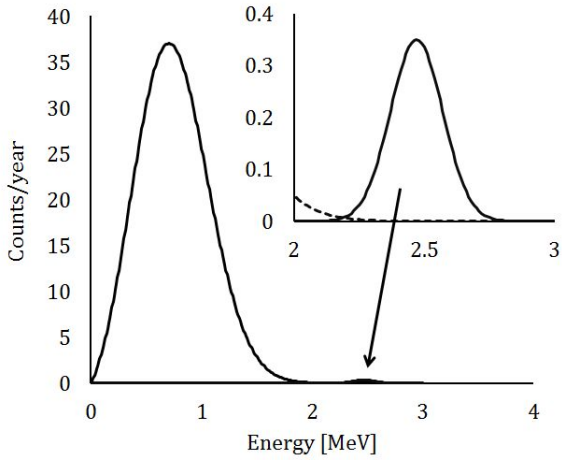


Figure 6.1: Expected signature of $0\nu\beta\beta$ decay in ^{136}Xe. The large peak is the omnipresent two-neutrino double-beta decay; the smaller peak is the $0\nu\beta\beta$ decay signature.

the tracks generated by the ionizing radiation can be reconstructed. Liquid Xe also has the attractive characteristic of being a scintillator; by collecting light as well as charge, the overall energy resolution can be improved. In EXO-200, the ionized tracks were collected on two instrumented low-background copper grids (wire planes), while the light was collected by APDs placed just under the wire planes at opposite ends of the detector chamber. Figure 6.2 shows a cutout of the EXO-200 TPC [76].

The deployment of this preliminary experiment allowed gaining great insight into the design of future large scale experiments, as well as producing databases of acceptable materials (those with low intrinsic radioactivity) and generating interesting scientific data.

One important outcome of the EXO-200 effort was the demonstration that the concept of a homogeneous detector is a key factor in increased performance. This is mostly due to the fact that the mean free path of the

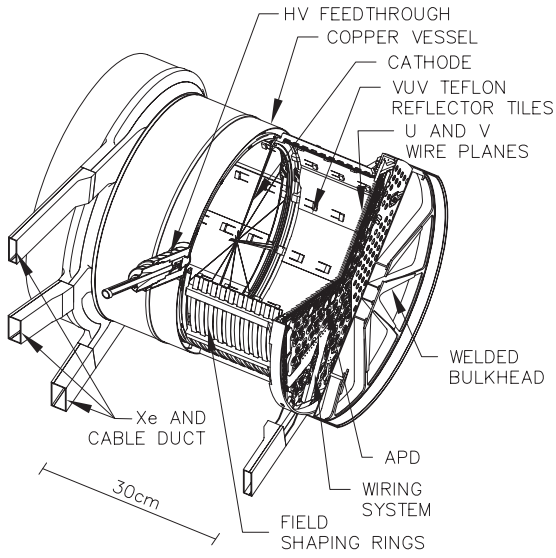


Figure 6.2: Cutout of the EXO-200 TPC.

radiation becomes more confined in an increasingly large detector, so that the interactions are fully contained. If events are fully contained, their identification is easier, and the overall signal-to-background discrimination is enhanced.
The nEXO TPC design is a very active effort that is still in a phase of research and development and will be for the near future.

From the instrumentation point of view, the nEXO TPC will look not so dissimilar from the EXO-200 design, with the main difference being (beside the total volume) that the charge collection plane will be located at the top of the TPC vessel and the scintillation light readout will be performed from the sides of the vessel. The size of the vessel is determined by the total active volume of enriched Xe which, in turn, is determined by the science requirements. That amount has been calculated to be about 5 tons,

therefore, the expected overall vessel height will be 1.3 m, with a diameter of 1.3 m. These choices (and the demanding background concerns) have profound impact on the instrumentation design, presenting unprecedented challenges, but also an excellent opportunity, as we will see later, to apply our new theories on light readout.
The detailed TPC and charge readout designs are outside of the scope of this work, so we will concentrate on the light readout design exclusively.

6.5 Light readout in nEXO

As scintillator, liquid xenon (often referred to as LXe) has an ionization energy of 50 eV/photon for electrons [77]. The light response is typically a single-time constant exponential decay of 70 ns. Its refraction index is 1.69 at 170 K, so from the point of view of radiation detection it looks no different than any other inorganic scintillator. The main and most impactful difference between LXe and ordinary scintillator is the peak emission wavelength which, in the case of LXe, is 178 nm (VUV). This is an important observation because most light sensors are designed to have maximum QE in the 400–500 nm range and have usually poor response in the UV. The first requirement for the nEXO photosensors will be, in fact, to present acceptable QE at this wavelength. The objective is to reach at least 15% QE. Presently, there are only a few PMTs capable of such QE, but PMTs present unacceptable radiopurity. This leaves only a few reasonable choices. Silicon photodetectors have no intrinsic gain; avalanche photodiodes have some gain, but require moderate high voltage and have high leakage, even at the temperatures of LXe. We have seen in the previous chapter how dark current can be the main contributor to electronics noise. Plus, the high capacitance and gain of APDs may not allow reaching good resolutions (see Figure 4.9) as an example). In fact, APDs were used in EXO-200 as light sensing elements because at the time they were the only available option. The cost was lower resolution than desired with fairly elevated low-energy thresholds.
Once again, SiPMs seem to be the best device with their high gain and

extremely low mass (silicon is a radiopure element because the processing technologies are extremely clean and free from radioactive contaminants). The only problem with SiPMs is their availability in versions that have the desired QE at 178 nm. Fortunately, there are several experiments requiring an extension of the SiPM QE into the UV range, and vendors have been progressing quite nicely in that direction [78, 79, 80]. Recent results from nEXO measurements on next-generation SiPMs from Hamamatsu and Fondazione Bruno Kessler (FBK), have reported QEs of the order of 12–16% at 178 nm. As for SiPM capacitance, that could still be a problem, but we need to determine in which regime of light nEXO will be working in order to further elaborate on the potential issue.

Figure 6.3 shows the most recent concept for the nEXO vessel. The ex-

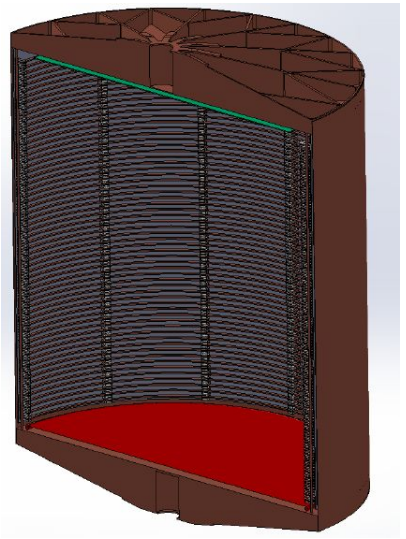


Figure 6.3: Sectional view of the concept design for the nEXO LXe detector chamber.

ternal height is 1.3 m and the base diameter is 1.3 m.
The SiPMs would be mounted in a way to completely cover the internal cylindrical surface, covering an approximately 4 m^2 area, assuming an in-

ternal height of 1.2 m and base diameter of 1.2 m. Our signals of interest have a 2.4 MeV energy, therefore, an event on the detector axis, in the middle of the vessel will produce approximately $50,000$ photons in a 4π sphere, or $3,980$ photon per steradian. A 1 cm^2 control surface on the cylinder surface and in the same horizontal plane as the event corresponds to $0.01/(0.6)^2 = 2.8$ sr, and will consequently see only 110 photons from the original signal in the best case. With a 15% QE, and a 7% loss to the optical coupling, the total number of photons reaching the SiPM multiplication area is only 15 for every square centimeter. This number only gets worse as the control surface moves away from the interaction plane. Clearly, from the SiPM point of view, this is a situation of light-starved regime. In order to recover the full event signature, each photon has to be accounted for, placing the mode of operation to a photon counting mode. This opens another issue: how does the dark count rate affect our ability of measuring such a small number of photons per unit area?
Some limitations on dark noise have to be determined. The dark noise counts look exactly like valid signal counts, except that while the signal counts are time-correlated, the dark counts happen at completely random times. The dark noise rate must be such that the probability of a dark noise event falling within the time valid events are present is acceptably low. We expect all signal events to be found within the relaxation time of the scintillation process[2]. The time constant associated to such time, as mentioned previously, is 70 ns, and the occupancy time of an event can be assumed to be five times that time constant, or 350 ns. Since we expect overall $3,980$ signal photons during the occupancy time, if we set the dark rate contribution to, say, 1% of the total counts within that window, or about 40 counts within 350 ns, the corresponding total dark count rate is 114 million counts per second of dark noise for 4 m^2, or 28.6 counts/sec/mm^2. The nEXO experiment set the dark noise rate specification for the SiPM at 50 counts/sec/mm^2 at LXe temperature, accepting up to 2% dark count

[2]For this to be true, the decay time of the exponential response of a single photoelectron signal should be smaller than the scintillator relaxation time. This may, at times, become a further requirement on the SiPM characteristics.

rate contribution.

Dark rate specifications at low temperature are not a parameter commonly found in manufacturer's datasheets, therefore specific measurements had to be done within the collaboration. Many devices from Hamamatsu, Ketek and FBK showed promising levels of dark noise rates at LXe temperature: the Hamamatsu device measurements reported about 0.6 counts/sec/mm^2, FBK's numbers were around 300 and Ketek's 7 counts/sec/mm^2. Other vendors did not supply devices or were not interested in the application. As for the other device parameters, such as capacitance per unit area, limits have to be found based on what can be achieved at the instrumentation level versus energy resolution requirements.

6.6 Energy resolution in nEXO

As mentioned previously, the nEXO detector is a time projection chamber that takes advantage of the scintillating properties of LXe as scintillator to combine charge and light information to achieve better resolution. From the data in Figure 6.1, and based on physics arguments, the desired overall energy resolution that ensures a minimum background contribution within the energies of interest for $0\nu\beta\beta$ decay (around 2.4 MeV), is about 1%. In order to achieve that, the light channel must be able to efficiently detect single photoelectrons, with an electronics noise as good as 0.1 photoelectrons r.m.s. In the charge readout system, the energy resolution requirement is better than 1% (translating to 240 keV FHWM).

It is clear that from the light readout point of view, rigorous calculations and testing have to be done in order to achieve the noise requirement.

We identified the system as working in a photon-starved regime, therefore we can take advantage of our noise theory, and, assuming a generic

readout like the one in Figure 4.6:

$$g_S = \frac{kT\Gamma e^2}{2q^2 0.1^2 G^2} (C_D + C_i)^2 \frac{1}{\tau_0}$$

A plot of this equivalent transconductance was shown earlier in Figure 4.12 for different values of sensor capacitance. The 0.1 photoelectron threshold could be reached, depending on peaking time, only for relatively small sensors capacitances. Specifically, for a 1 μs peaking time (which is already beyond our 350 ns occupancy time), the electronics noise was found to be tolerable only up to 20 nF for an 80 mS transconductance in the readout circuit. In order to instrument larger areas, a higher value for the transconductance is needed. Higher transconductance values are possible, of course, but they depend on how much current is in the input device. Ultimately, there will be a limit on how much transconductance can be obtained. For example, in the case of bipolar transistors at LXe temperature, the transconductance can be expressed as:

$$g_m = \frac{I_c}{15.4\,\mathrm{mV}}$$

We see that, for example, 80 mS require 1.2 mA. To have an idea of how much power is involved for the nEXO light readout, we can ask ourselves how much current is needed to instrument 4 m^2 of SiPM, while preserving a 0.1 photoelectrons r.m.s. electronics noise. To find this, we set $T = 175$ K, assume $G = 10^6$ and $\tau_0 = 1$ μs, and write (note that C_i is negligible as we expect very high values of C_D):

$$\frac{I_c}{15.4\,\mathrm{mV}} = \frac{kT\Gamma e^2}{2q^2 0.1^2 G^2} (C_D)^2 \frac{1}{\tau_0}$$

or

$$I_c = 2.68 \cdot 10^{-11} (C_D)^2$$

The total system power will be $p_T = V_{cc} \cdot I_c = 2.68 \cdot 10^{11} (C_D)^2 V_{cc}$. We can plot the power as a function of C_D, assuming unity voltage bias ($V_{cc} = 1$ V) in Figure 6.4. To understand the meaning of this graph, we need first to

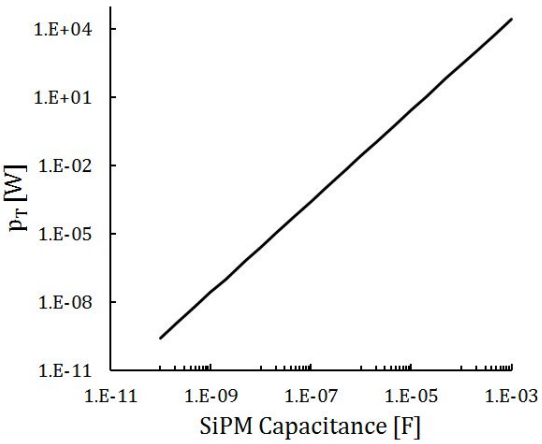


Figure 6.4: Power per unit supply voltage as function of SiPM capacitance.

make assumptions on what is a reasonable amount of capacitance we can expect from the SiPMs, and what power nEXO can tolerate.

The simulation group within nEXO provided guidance for the power allowed to the electronics system. Their thermal simulations indicated that up to 200 W can safely be dissipated by the LXe without appreciable change in properties of the liquid. We assume we can divide this power equally between the charge and light readout systems, and that at least 40% of the power in each readout system should be dedicated to the front-end preamplifier. In the preamplifier, we assume half of the power to be available for setting the transconductance. For this reason in the following calculations we will always assume a power figure of 20 W.
An extensive survey among SiPM vendors indicated that current devices have capacitance between 30 pF/mm^2 and a few 100 pF/mm^2. This places limits to the expected total capacitance for 4 m^2 in the range from a minimum of 120 μF to several hundred micro Farad. Within these limits, the power requirements would be 250 W to 650 W. These numbers are clearly

on the high side. This means the SiPM surface cannot be read as a whole monolithic array, but some other techniques need to be adopted. Also, there is a practical limit on transconductance one can achieve; the solution under the previous assumption would require a g_m of the order of several Siemens. This expresses the trivial observation that not all of the power budget can be spent into a single transistor, and some form of segmentation is needed.

We mentioned that a 0.1 photoelectron r.m.s. noise is desired. We can observe that this noise level can more easily be achieved for smaller capacitances. We can also observe that, once the presence of single photoelectrons has been identified under acceptable noise constraints (for example by a threshold or a peak digitization), a decision has been made, so the stochastic nature of the signals no longer applies. We conclude that the best way to operate in this case is to make "decisions" on the signals on areas for which the desired noise levels are achieved, combining the results of these decisions later in the signal processing chain [81]. Once this is done, the problem becomes determining the largest area of SiPM that can be instrumented by a single channel under the power constraints introduced above. For a 20 W total power dissipation in the input transistor branches, the power density for our 4 m^2 surface is $p_{T,a} = 0.5$ mW/cm^2. This means that every 1 cm^2 of SiPM surface should be instrumented by a single channel consuming 0.5 mW in the front-end. In this case for a unity value of supply voltage, we have:

$$0.5\,\text{mA} = 2.68 \cdot 10^{11}\,(C_D)^2$$

Solving for C_D we find the maximum capacitance that is acceptable for this power consumption to be $C_D = 43$ nF for our centimeter square of surface (the situation is illustrated in Figure 6.5). This number is a capacitance per unit area of 430 pF/mm^2. Since the candidate devices acceptable by the collaboration have of the order of 50 pF/mm^2, we could, under the assumptions made above, instrument as much as about 8.6 cm^2 with a single channel, for a total of approximately $4,652$ channels for our system. In other words, we can either read one square centimeter of devices with

a specific capacitance of 430 pF/mm^2 or, if we have devices with lower specific capacitance, we can read more area with the same power.

We have, so far, calculated the collector current by assuming a power

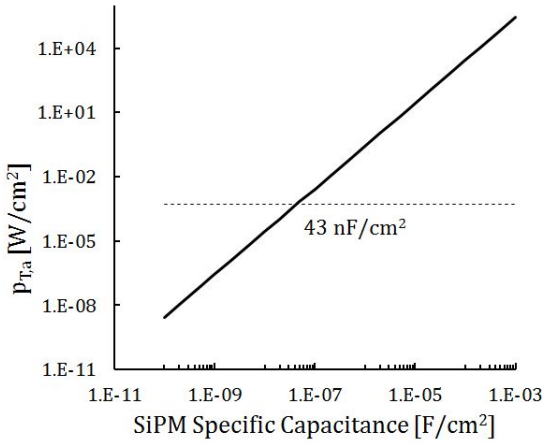


Figure 6.5: Power per unit supply voltage and unit area as function of SiPM capacitance per square centimeter. The horizontal line indicates the 0.5 mW per square centimeter level. An artificial offset has been added to the curves for clarity.

dissipation of 0.5 mW from a unitary bias voltage source. If that source were different in value, the current would need to be reduced or increased to maintain the power constant. Thus, for a bias of 2.8 V we would write:

$$p_{T,a} = 2.68 \cdot 10^{11} \, (C_D)^2 \cdot 2.8 = 0.5 \, \text{mW}$$

That leads to $C_D = 25.8$ nF and after some calculation, a maximum area of 7.3 cm^2 with $7,752$ channels for a 50 pF/mm^2 device.

The other factor that has great influence on the amount of area that can be instrumented is the shaping time. We assumed so far 1 μs peaking time. In our case, however, we previously assumed a 350 ns occupancy time to calculate the acceptable dark noise rate. If we want to adjust the

calculations above for the correct peaking time, we need to observe that the occupancy time may not be the same as the peaking time. In fact, with reference to Figure 6.6, where the dashed line is the LXe intrinsic response, the solid line represents a 50 ns shaping time and the dotted line a 1 μs peaking time, the 50 ns peaking time choice would preserve the dark rate overlap requirements.

If we introduce the 50 ns peaking time into our equations, we find:

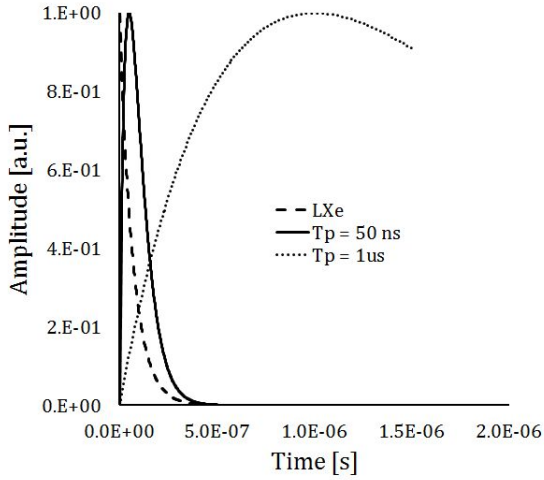


Figure 6.6: Intrinsic response of LXe (dashed line) compared to shaped responses for a 50 ns (solid line) and a 1 μs (dotted line) peaking times.

$$p_{T,a} = 5.37 \cdot 10^{12} \, (C_D)^2 \cdot 2.8 = 0.5 \, \mathrm{mW}$$

With this correction, $C_D = 5.8$ nF/cm^2, or 58 pF/mm^2 for the specific SiPM capacitance. The total number of channels would now be 34, 483.

Another way of looking at the occupancy time is to define it as the time at which a decision on the number of photoelectrons recorded is made. For example, we can imagine triggering on the first single photoelectron and, at

that time, open an observation window to integrate the number of photoelectrons seen within 350 ns. In this case, the peaking time in the previous formulae assumes the characteristic of an observation time and we have:

$$p_{T,a} = 7.67 \cdot 10^{11} \, (C_D)^2 \cdot 2.8 = 0.5 \, \mathrm{mW}$$

In this situation, $C_D = 15.25$ nF/cm^2, or 152.5 pF/mm^2 (13, 115 channels with an elementary channel area of about 3 cm^2 at less than 60% the initial power budget).
We can now state that under the proper circumstances, we can design a readout system where 7, 000 to about 35, 000 channels of electronics (depending on how low the specific sensor capacitance is) read out our 4 m^2 of SiPM within the assigned power budget. Should SiPM technologies improve further over the next few years in the direction of supplying devices with lower specific capacitance, the benefit will be seen directly as a decrease in total number of readout channels and/or power. For example, in the last calculation having 35 pF/mm^2 would mean about 9, 000 channels. As for nEXO, considerations on radiopurity of materials for interconnects, cables and supports limit the maximum number of light readout channels to less than 10, 000. The nEXO specification for SiPM capacitance has been, in fact, set to 35 pF/mm^2.

The important corollary to the numerical examples above is that we can now highlight a method for using the design criteria introduced. In a design, we would first determine the maximum power density and, based on rate and other considerations, the type of filtering to be employed. Once this is done, we would calculate g_S and verify it is compatible with power constraints. If the result (let us remember g_s is only a lower limit) is acceptable, we could design a readout around the value of g_S. If not, we would need segmentation, therefore choose an acceptable value for g_m and set it equal to g_S. From this, extract the maximum acceptable C_D and thus the cluster area and the number of channels.
This procedure is quite simple and applicable under the initial hypothesis made.

6.7 Electronics readout

After the previous considerations, we now have most of the parameters needed to design the front-end circuits. First, we must consider what type of technology to adopt: the radiopurity requirements dictate absolute minimization of the amount of materials used, therefore readouts using discrete components are forbidden. Integrated technologies should accomplish the task very well, in virtue of their low use of materials and the radiopurity of silicon. This will also help locating $10,000$ channels within the physical constraints of the nEXO detector. We based our analysis on using bipolar transistors. This is not in contradiction with the choice of an ASIC technology, as there are many modern technologies offering bipolar transistors with the performance needed in this design. Using bipolars also give us the advantage of being able to prototype on larger scales (for example, at the single 3 cm^2 channel) with discrete components.
The fast nature of the signals and the high capacitance of the sensors demand for a front-end stage with very low and stable input impedance. This will avoid introducing distortions in the charge signal coming from the SiPM sensors. We have experience with this kind of choice, for example from the design used in Chapter 5. In this case, we want the prototype design to be optimized for later implementation in integrated technologies, so some further modifications need to be made to the basic design.
For this circuit we will minimize the number of passive components required, so that its translation to an ASIC will eventually require very little modifications. With similar arguments made during the design of the circuit in Chapter 5, the resulting schematic is shown in Figure 6.7. The schematic is indeed very similar to the one used for the application in Chapter 5, with the exception of the presence of the current mirror Q_3, Q_4 that replaces a resistor and allows the buffered current output to be collected by subsequent stages on a very low capacitive load (allowing thus more bandwidth). The current sources I_1 and I_2 provide bias to the transistors. In case of a discrete design, I_1 and I_2 can be simple resistors; when an ASIC is considered, they can be implemented by NMOS devices biased as resistors or current sources. The bipolar transistors are available in any

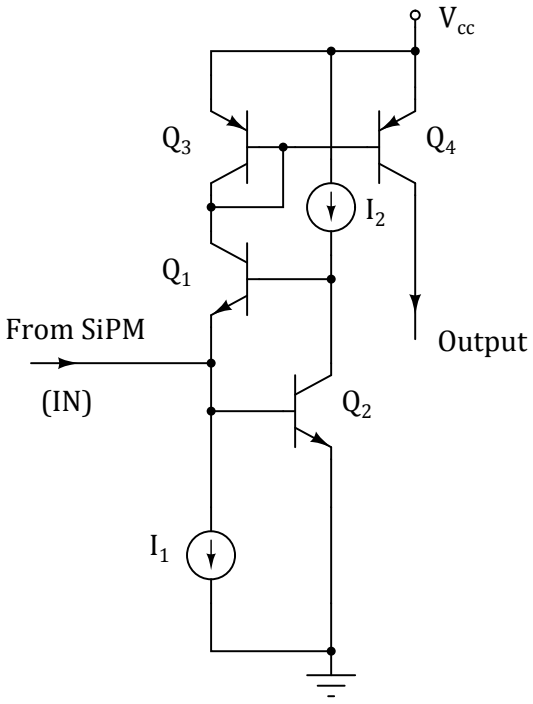


Figure 6.7: Schematic of the prototype readout circuit for the nEXO experiment.

Bi-CMOS technology for integrated designs, but in discrete they should be made of matched transistor pairs, or there may be non-ideal behaviors due to component mismatch. We selected for our prototype a complementary BJT array (Intersil HFA3096 [82]). The devices in the array are guaranteed to be matched and offer an f_T of 8 GHz (NPN) and 5.5 GHZ (PNP).

In order to understand and demonstrate the circuit functionality, we opted for an initial prototype design made of discrete components. In our previous considerations, we concluded we needed a power of 0.5 mW in the input transistor branch; the corresponding collector current is 0.178 mA

out of a 2.8 V supply rail. This allows reaching the desired noise at LXe temperature, where the required g_m is 11.6 mS. We would like the ability to make measurements at room temperature while preserving similar noise performance, therefore the value of g_m (and the collector current) must be scaled to account for the different temperature (300 K instead of 175 K). A conservative scaling leads to a collector current in Q_2 of 0.5 mA (20 mS of transconductance at room temperature). To set this current, we used a 3.01 kohm resistor. I_1 sets the g_m in Q_2, hence the open loop gain of the Q_1–Q_2 feedback. Since the gain is the product of Q_1 and Q_2's transconductances, we do not need $g_{m,1}$ to be as large as $g_{m,2}$, therefore a value of I_1 of the order of 300 μA will suffice (we will than approximate the source I_1 with a 2.32 kohm resistor). The right side of the current mirror is the output branch of the circuit and as such, it must provide sufficient current to drive the load. Currently, its standing current is only 300 μA for the choice just made. We notice, however, that as long as the current signal into Q_4 flows from the emitter to the collector, the power rail can supply enough dynamic current to virtually drive any load. For signal current of opposite sign, only 300 μA are available. This means the circuit, as designed, will perform best with signal currents that exit from the input (i.e. collection of electrons). Should a different arrangement be needed, it would be sufficient to swap NPN transistors for PNP and vice-versa in the schematic. With these choices, the power consumption will be 3 mW. Note that, despite the design was made for room temperature, this number is still within the nEXO power budget, allowing the readout of $9,000$ channels with 27 W of power. In an integrated design this power could be further reduced. In that case, the current mirror would be implemented by using two PMOS transistors; MOSFET current mirrors allow for scaling of branch currents in proportion to the ratio of the areas used for the single transistors, thus, the output DC current could be reduced to, say, 50 μA, while preserving circuit performance.

The circuit output needs to be converted to a voltage without significant bandwidth loss. The best way to do that is either to terminate the output into a resistor to ground or into the virtual ground of an appropriately fast

operational amplifier in trans-resistance amplifier configuration. For a discrete prototype the best approach is to use a resistor, since external means of further amplification are always available on a benchtop scenario. In the final version, we will likely design an integrated low power, fast operational amplifier. The single photoelectrons carry a charge of $1.6 \cdot 10^{-19}$ C in a very short pulse; generally, a few ns rise time and, according to the datasheets, a 100 ns exponential fall. The current signal will be a similar exponential with a 3.2 μA peak value for a 10^6 gain. The signal developed into a 511 ohm resistor in response to an equivalent charge of 10 photoelectrons has been simulated for a 850 pF and a 15 nF total sensor capacitance, and the results are shown in Figure 6.8. Note that the bias current induces a DC voltage offset at the output and that the two waveforms have been offset artificially for convenience. For the choices made for the bias currents, the correct offset is about 250 mV.

The voltage amplitude is 10.5 mV for the 15 nF sensor and 15 mV for

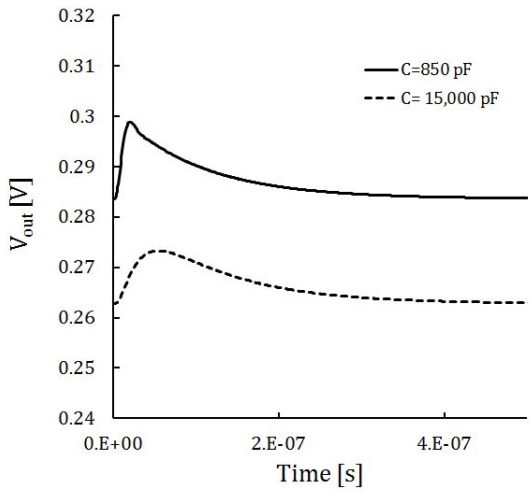
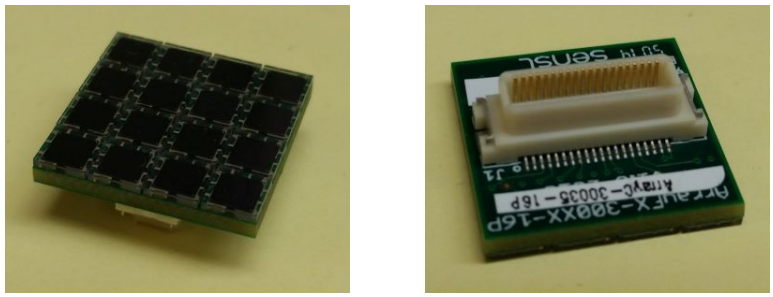


Figure 6.8: Simulated response to 10 photoelectrons for a 850 pF light sensor (solid line) and a 15 nF sensor (dashed line).



(a) (b)

Figure 6.9: 4×4 sensor pixel array board ARRAYC-30035-16P-PCB, top (a) and bottom (b).

the 850 pF device (some amplitude losses due to the extra capacitance are expected). For the single photoelectron we expect one tenth those values. More gain could be introduced, if needed, by increasing the value of the output resistor, at the expenses of a higher DC offset. Clearly, an output operational amplifier could provide the same gain increase without offset, when properly biased, but would require more power.

This prototype circuit was implemented as circuit board. Out of convenience, we selected once again a SiPM made by SensL (ARRAYC-30035-16P-PCB [83]), mostly because of availability. The device is a newer version of the one used in Chapter 5. It is an array of 16 type-C, 3×3 mm cells that have overall better performance [84] than previous generation devices. Their specific capacitance is about 90 pF/mm^2. The array is assembled on a connectorized board (see Figure 6.9). This way, if devices from different manufacturers have to be tested later on, they can be arranged on a similar PC board accessorized with the same connector and be directly plugged into the test board without requiring any further development. In implementing the test board we took the approach of offering the maximum flexibility for the intended measurements: using a connector-daughter board arrangement offers the possibility of testing a wide range of SiPM devices, as just

mentioned; instead of connecting the SiPM outputs to the current amplifier directly we used configurable jumpers, so that from one to all 16 channels of a given SiPM can be tested at any given time (this is useful for investigating the behavior of the electronics noise versus capacitance, or dark current versus number of pixels). We also opted to replicate the channel four times, so that a total of 10.24 cm^2 can be read out by one channel or divided among the four preamplifiers. This philosophy has been well received; in fact, three different member institutions of the nEXO collaboration use the prototype board for all types of tests on candidate SiPMs. A fourth board is also being used by another, non-related, collaboration for experimental measurements on neutron interactions. The layout of the board top layer is shown in Figure 6.10.

In the figure, the SiPM and jumper arrays are evident at the center of the board. The connectors J7, J8, J9 and J10 provide bias to the individual SiPMs; the integrated circuits U1, U2, U3 and U4 in SOIC-16 package are the matched transistor arrays. J5 and J6 are ancillary connectors that allow AC measurement of the input impedance of the current preamplifier, if so desired. The header J1 allows to connect the preamplifier outputs to a buffering (or amplification) output stage (needed when driving long cables); J2 is the output BNC. The few components located on the copper planes constitute the voltage regulator for obtaining the 2.8 V rail, as the main board power is ±5 V. J11 is the power connector (±5 and ground). Although only the top layer is shown for simplicity, the board has four layers, with the two inner layers being power and ground planes. The bottom layer has additional configurable jumper arrays to connect to ground any unused SiPM pixel.

6.8 Light response measurements

To test performance, we set up a few simple tests using the SensL SiPM. The first, possibly most significant test is to verify the response to single photoelectrons under representative conditions. The most critical aspect of this measurement is being able to identify single photoelectrons, especially

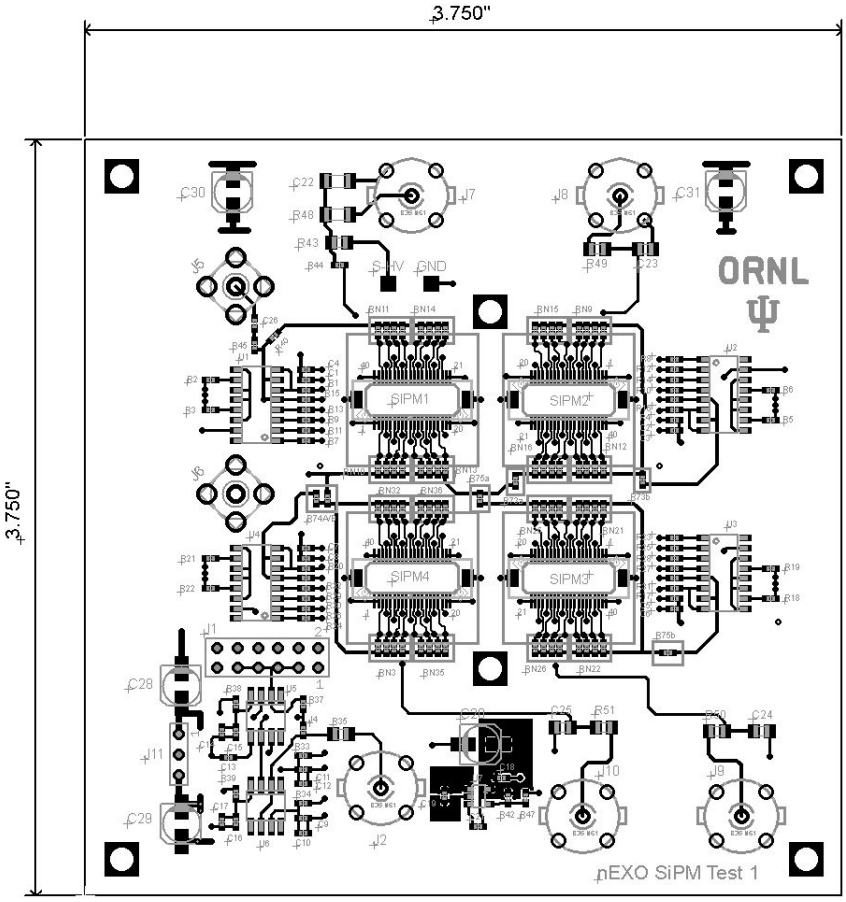


Figure 6.10: Prototype SiPM test board layout (top layer and silkscreen only).

at room temperature where the sensor leakage is considerable: about 7 μA (13.7 Mcps of dark rate). With such a high rate, it may be impossible to distinguish a few signal photons from the dark photons. For this reason, a coincidence measurement is preferred. This is simply set up by introducing

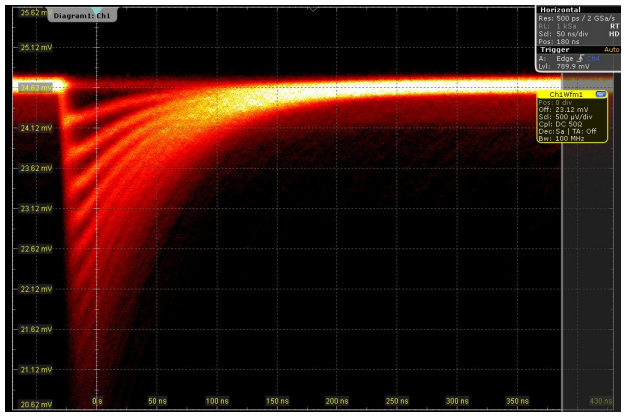


Figure 6.11: Intrinsic SPE response to 470 nm photons for a 3×3 mm SiPM pixel.

a pulsed light source shining a very small amount of light onto the sensor, and by triggering off the light source signal. As light source we used a blue, 470 nm, LED (Lite-on LTL1CHTBK4 [85]). The reason for this is to have the injected photons at a wavelength close to that of peak efficiency of the sensor (420 nm). The LED was driven by a square wave of amplitude and duration such that only "a few" photons are seen by the sensor. The appropriate duration is actually quantified later on during testing. An initial measurement of the direct output of the preamplifier on a 50 ohm load was taken to show the intrinsic response of the SiPM. An oscilloscope screenshot is shown in Figure 6.11. Here, only one single 3×3 mm pixel (850 pF) was connected to the preamplifier.
The very fast leading edge and the discrete nature of the signals are evident in this picture.
For all subsequent measurements, the preamplifier output has been buffered by an operational amplifier (National Instruments LM6171) in order to be able to easily drive long 50 ohm cables. This is functionally equivalent to reading out the preamplifier directly, with the exception of the fact that the system bandwidth is now limited somewhat by the operational amplifier.

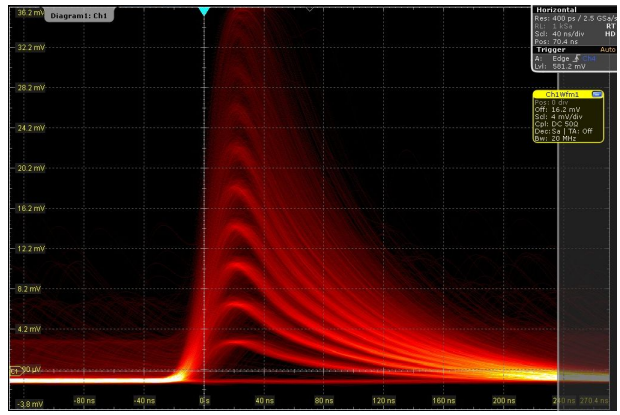


Figure 6.12: Filtered SPE response to 470 nm photons for a 3×3 mm SiPM pixel.

The oscilloscope bandwidth used for these measurements had been set to 20 MHz in order to simulate the filtering or signal shaping that will ultimately be employed. The responses measured and previously shown in 6.11 will look different due to the set bandwidth. Figure 6.12 shows such responses. It is interesting to notice the large amount of single photons present just before the triggering point indicated by the cyan triangle on top of the picture, and how well we are able to select only those generated by the LED, as there is virtually no coloration under the single photoelectron peak, except for the baseline. The corresponding histogram of the amplitudes is shown in Figure 6.13. Some interesting observations can be made. The narrow peak in the origin is the likelihood that when the LED fired, no photons were detected, and represents, in practice, the output baseline (zero in this case). The horizontal scale is, obviously, a match of the vertical scale of Figure 6.12. The peaks represent, from left to right, a single photoelectron, two photoelectrons etc., all the way to about 11 photoelectrons. The single photoelectron amplitude is 6.8 mV (in the figures above there is a ($\times 2$) loss due to terminating the output on a 50 ohm resistor), in good agreement with the simulation in Figure 6.8. The last two peaks saturate: this is due

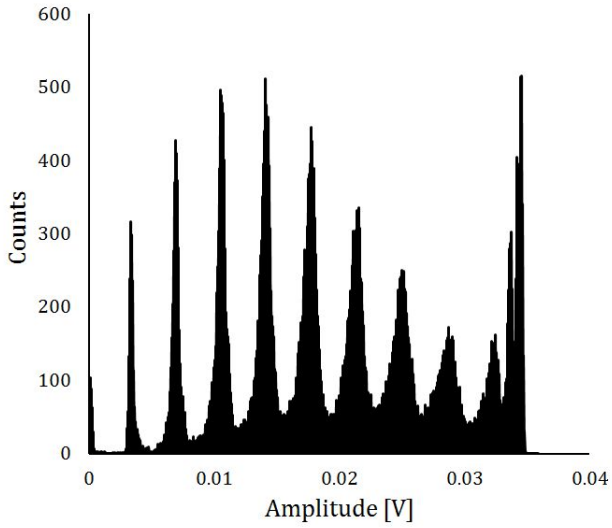


Figure 6.13: Histogram of the SPE response.

to the fact mentioned earlier that the preamplifier is designed to collect electrons, but in this case the sensor is positively biased to collect holes, a feature determined by the manufacturer over which we have no control. The peaks are equally spaced in support of their discrete nature. Finally, the resolution is around 5% for all peaks in the linear region, in terms of FWHM divided by the centroid position. In r.m.s. terms, this is well below the one tenth of a single photoelectron requirement, at a resolution of 2.1%.

Clearly, as sensor area is added, the electronics noise will increase and the SPE resolution will decrease. The response of the whole 16 elements array is depicted in Figure 6.14. We notice immediately the difference in noise, but individual lines can still be seen. In this case, the dark count rate is equivalent to a current of about 7 μA (one dark count every 73 ns average), therefore even with our coincidence technique it was very hard to separate the photoelectrons created by the LED from those of dark noise;

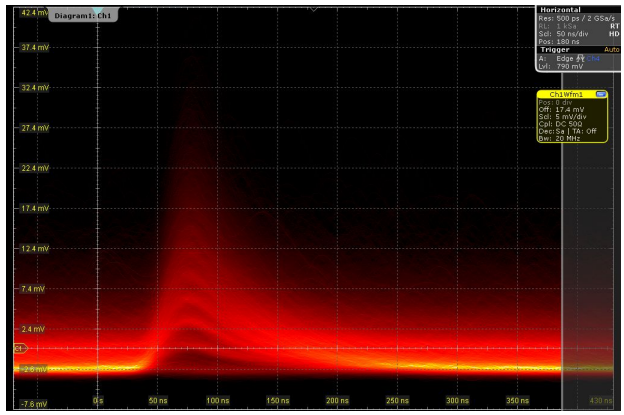


Figure 6.14: SPE response of all 16 pixels in the SiPM array.

these signals will contaminate the measurement to some extent.

The corresponding histogram is in Figure 6.15. In the histogram the effect of the dark noise is evident: the first peak corresponding to the single photoelectron is lower than previously observed and the noise peak in the origin has not been recorded, due to the exceptionally high dead time encountered when acquiring the lowest amplitudes. This would be greatly mitigated at LXe temperatures. The second and third peaks however, can be resolved and a Gaussian fit indicated a 16.8% and 13.1% FWHM resolution, corresponding to 6 to 7% r.m.s. This is quite encouraging and demonstrates the ability of the design to meet requirements even at room temperature. Of course, this is purely speculative, as the dark rate at room temperature would prevent from recognizing signal-originated single photoelectrons in any real scenario. However, it is clear that at -100 ℃ the situation will be much more favorable. A minor difference with the plot of Figure 6.13 is the absence of the higher order peaks (beyond 6 photoelectrons). This is attributable to the partial inability to distinguish peaks but also to the fact that the light distribution from the source on the SiPM may have been slightly different.

Our setup does not allow for precise temperature control. However, liquid

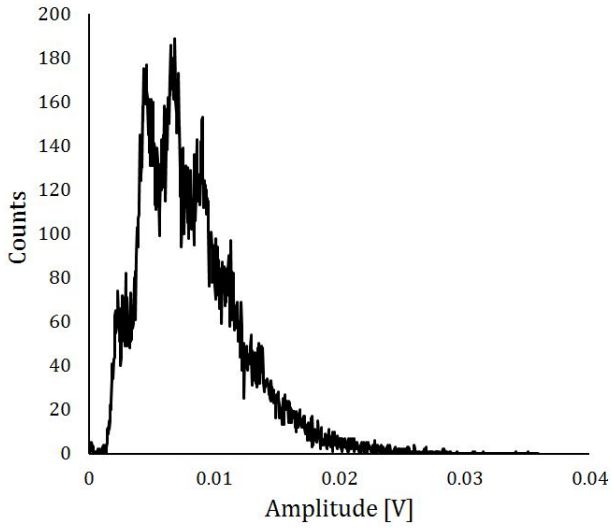


Figure 6.15: Amplitude histogram of the waveforms of Figure 6.14

nitrogen was available, so we used it to flood the circuit box and reach cryogenic temperatures, albeit without much control over them. The temperature change during warming in air was of about 0.1 ℃ per second. Since the SiPM gain does not change appreciably within a 10 ℃ temperature range [datasheets], we decided to measure responses within a ± 5 degrees interval around the target temperature of -100 ℃, corresponding to measurement times of 100 seconds. For this measurement, we also changed the preamplifiers operating point to its intended values, previously calculated for LXe temperatures. We intend, this way, to verify the claim made previously that scaling the collector current of Q_2 in Figure 6.7 compensated for the different temperature used so far. The results are shown in Figures 6.16 and 6.17.

Despite the lower statistics in the measurement, the improvement in the response is dramatic when compared to the histogram at room temperature (Figure 6.15) for the same total capacitance: the separation between

the first photoelectron and the second is pronounced, and the ability of recording the electronics noise in the origin is restored. In other words, the lower temperature reduced the total dark rate to a more manageable level as anticipated. The single photoelectron amplitude in this case was 5.6 mV (again, the histogram shows the amplitude on a back-terminated 50 ohm load) due to a difference in bias voltage to compensate somewhat for SiPM gain changes with temperature. The peak resolutions in FWHM are, from one photoelectron: 15.7%, 12.9%, 9.74%, 8.7% and 7%, for a noise level of 6.7% all the way to 3% r.m.s. respectively.

There is no doubt the system would be capable of recognizing a single

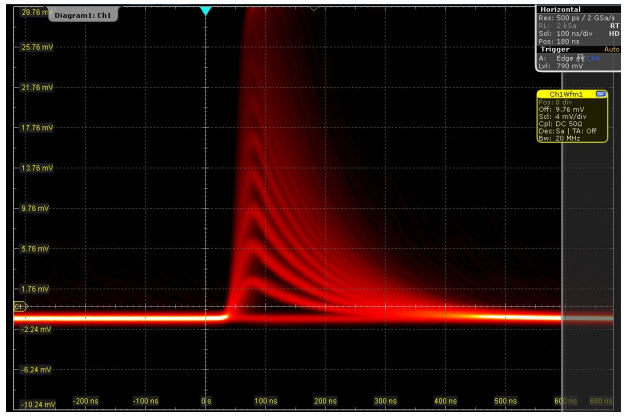


Figure 6.16: SPE response of all 16 pixels in the SiPM array at −100 °C.

photoelectron under these conditions with a 0.1 SPE resolution, setting an electronics threshold at 0.5 SPE (the commonly used value of 5 σ): this conclusively demonstrates the validity of our instrumentation approach.

6.9 The nEXO light readout system concept

The scope of the design process presented so far was to both provide a viable solution to the nEXO instrumentation needs and to benchmark the theory developed in Chapter 4 in the case of true photon-starved regime.

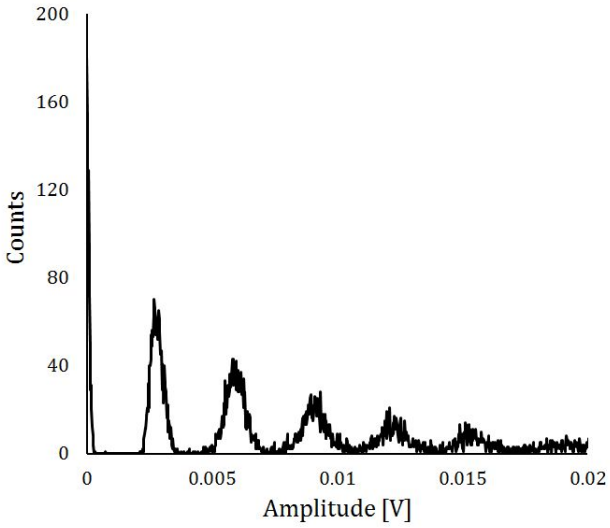


Figure 6.17: Histogram of amplitudes of the SiPM array at −100 ℃.

The discussion, however, would not be complete without giving an overview of the system where the front-end will be employed and will be vital part of. It should be noted that the overall system design is not an effort concentrated within a single institution, but it is rather a concerted effort among several institutions, such as, besides Oak Ridge National Laboratory (ORNL) who directly sponsored this research and co-leads the overall nEXO instrumentation effort, the Instrumentation Division at Brookhaven National Laboratory (BNL), Indiana University (IU), the University of Illinois Urbana-Champaign, TRIUMF in Canada, Stanford University and the Stanford Linear Accelerator Center (SLAC). Ideas were contributed from all institutions, in the full spirit of a scientific collaboration.

As we mentioned before, the 4 m^2 of SiPM sensors will require 7,000 to 10,000 (the nEXO specification limit) channels of instrumentation, with the goal in mind of also minimizing feedthroughs in the cryostat, in order

to meet the radiopurity requirement as well. We will assume for the subsequent discussion $10,000$ channels, knowing we will likely end up doing slightly better.

We already verified that a coincidence technique is a really powerful technique for discriminating against undesired background events when the time of arrival of signals is known. In a real experiment there is no knowledge of when an event of interest occurs, but once an event has occurred the light response will be concentrated within a few tens of nanosecond (the relaxation time of the scintillator). We can thus imagine triggering on every single photoelectron seen by any area of SiPM and observe for a given amount of time if any more photons are seen within the same area. This would be, under the correct assumptions, another way of implementing a coincidence technique. There are of course some limitations: if the area under observation is too large in relation to the dark rate, we would risk counting dark rate pulses as valid event signatures, therefore a limit in area and/or in rate must be identified. In a way, we already observed what the limitations might be in paragraph 6.5 where we set the design limit for the dark count rate (DCR) to 50 Hz/mm^2. In this case, our channel area of a few cm^2 will have a dark rate of the order of 10^4 counts per second. This, compared to the occupancy window previously found leads to a probability of 10^{-5} of finding coincident dark counts within 350 ns and is such that the overall contribution to the dark noise is only 1.75%. Consequently, in nEXO we can safely apply the coincidence technique on a per-channel basis on the whole SiPM surface. This would mean triggering on every single dark count and observing each time how many were seen within the coincidence window. If the counts are less than a certain number it means only dark counts were observed and the process would be restarted at the next dark count available.

This way, we could design a readout system where the current preamplifier instruments the maximum area of SiPM calculated from noise considerations (we can assume that to be 5 cm^2 for sake of argument), followed by an adjustable, nominally 350 ns-long integration window, during which the number of counts is tallied. A discriminator would identify the presence of

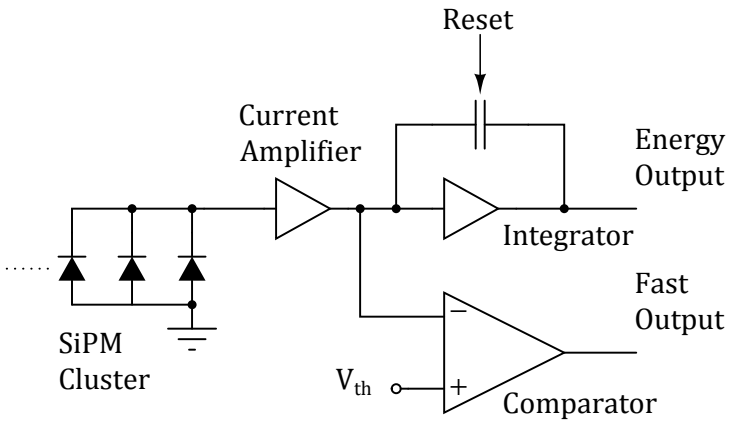


Figure 6.18: First stage (or level 1) of the readout system.

an event and signal it to the subsequent parts of the circuit. Figure 6.18 shows the block diagram for such an arrangement.
The reset signal on the integrator feedback capacitor is generated by a logic block that starts the integration window as soon as an event is seen, and resets it after the energy output has been recorded at the end of the coincidence window time.
In practice this method would generate a very high, steady data rate, estimated in excess of 1.6 Gbps when only 8 bits are used to record the slow output. We can improve this by observing that the light from any event will also have a limited spatial footprint, due to solid angle as well as light absorption within the LXe, therefore we could apply the procedure only to properly-sized adjacent clusters. In this case, we would introduce a second stage (or a level 2 logic system) where each fast channel is inspected within the coincidence window for the presence of an event. If at least n coincident events were seen among m channels, the amplitude from the slow output would be stored in an analog buffer and made available for readout by a multiplexer to be converted by an ADC shared among several channel clusters. If fewer than n coincident events are recorded, the cluster of m

channels would be ignored and re-set.
This way, the data rate would be greatly reduced by an amount dependent on $m \times n$, approaching only hundreds of kbps under the best assumptions. Of course, one could embed the original mode of operation where, for a limited amount of time, all of the dark counts are read. The high data rate would be buffered to allow data transfer on slower and fewer lines. As a consequence, the readout process would not be done in real time in this mode. This may aid calibration procedures, but it would be unwise adopting it as standard mode, given the rarity of our events of interest. In Figure 6.19 we show the overall concept.

The main challenges in implementing the system described are found in translating the prototype design into an ASIC that works at LXe temperatures. Fortunately, the collaboration has experience in designing circuits operating at liquid noble elements temperatures (except for liquid helium) [86]. This greatly mitigates most risks.

Recent research by the collaboration has also shown the possibility of further increasing readout area if single SiPM elements can be connected in series rather than in parallel. In fact, it is well known that the series of n identical capacitors C has an equivalent capacitance of C/n as opposed to a parallel connection that would have a total capacitance equal to nC. As for the signal-to-noise ratio, we observe that the same charge Q generated in a parallel connection will be divided by n in the series connection among the capacitors (in our low impedance readout). The quadratic noise, instead, depends on g_S in our formulas, which in turn depends on the square of the capacitance, or, in the series connection on $1/n^2$. The SNR remains unaltered, but the g_S requirement has also decreased by a factor n with the same SNR. Comparing the two connections schemes we can then conclude that for the same power (or same g_S), the series connection allows for n times more area to be read within the same noise and power budgets. We are currently investigating this possibility. There are two main factors that limit the value of n that can be achieved. Once factor is the SiPM bias voltage also increases by n, the second is a question about how well matched the dark current is among devices. This is the most pressing prob-

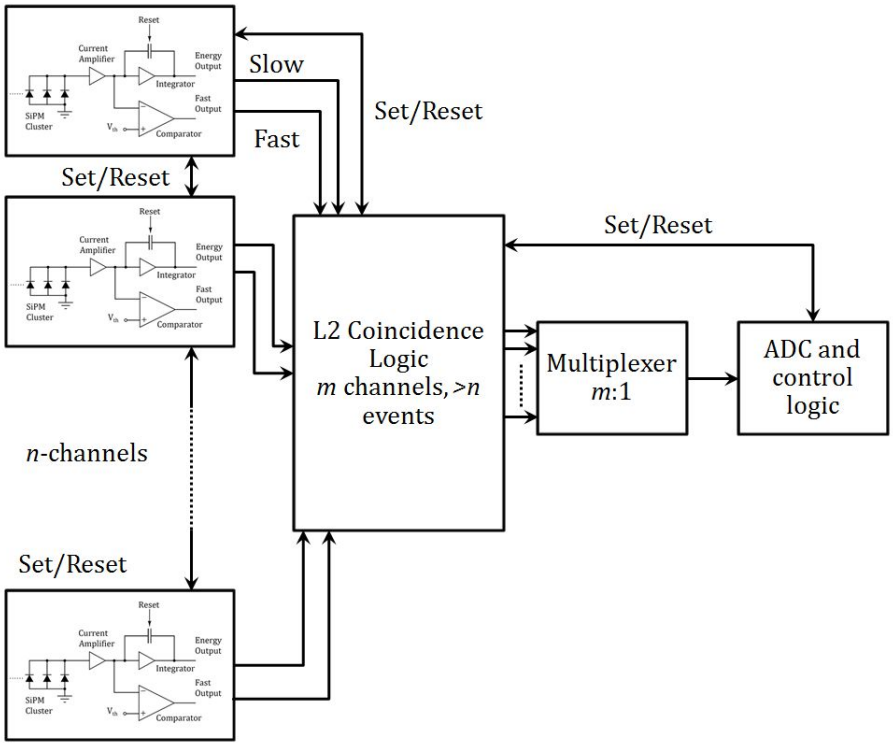


Figure 6.19: Overall readout system concept for the nEXO detector.

lem because devices connected in series must have the same current flowing through them, and if some have substantially different currents, their bias will adjust to generate an average current that satisfies the bias point stability for the whole string. Consequently the bias voltages may differ, with significant consequences to the gain of each device. The resulting string may show very different responses to single photoelectrons that cannot be calibrated. Clearly, the problem will be greatly mitigated at cryogenic temperatures as the currents are very small and any imbalance should not be as problematic. Also, some prior matching among devices may be required. This is the main subject of future tests.

6.10 The future of the nEXO experiment

The nEXO experiment is in the run for becoming the $0\nu\beta\beta$ decay experiment the US Department of Energy may support in the near future. At the moment the collaboration is working on R&D meant to demonstrate a path of risk reduction for all aspects of the effort: radiopurity, high voltage systems, calibration methods as well as electronics readout systems and sensors. In the next one to two years, the collaboration will be required to show progress in all areas. Thanks to the work described here the light readout electronics area is well positioned for a successful demonstration already in the present state, with potential for substantial improvements if the possibility of series connections of SiPMs is demonstrated.

Immediate development plans have as a priority the demonstration that series connection of up to four SiPM areas of a few cm^2 can be achieved. The collaboration is constructing and organizing the thermal environments necessary for such demonstrations. Future plans include the design of an integrated circuit based on the solutions introduced here to demonstrate on a smaller scale (a few 100 cm^2) the feasibility of our architecture and retire most of the electronic design identified risks. Once (and if) nEXO is selected, the collaboration will be ready to begin design and implement a full-scale light readout system meeting the complex nEXO requirements.

Chapter 7

Conclusions

Based on the controversial observation that SiPM photodetectors do not behave as conventional photomultiplier tubes, in this study we developed a new approach to their instrumentation design.
It was shown that there is a regime in which SiPMs do behave as PMTs, but the regime does not extend to all applications. The limits of applicability of conventional techniques, have been demonstrated to be determined by the amount of signal available from the radiation detector that converts radiation quanta into light: when the signal available is limited to only a few electrons per square centimeter, the SiPM behaves more like a semiconductor radiation detector rather than a PMT. In this regime, different considerations need to be made: with only a few photons to detect, the noise introduced by the electronics readout is no longer a negligible contribution and must be taken into account, being in all cases the ultimate limiting factor in sensitivity.

From these observations, new circuit design techniques and architectures were proposed and applied to both light starved and high SNR regimes. It was also noted that even the conventional high SNR designs can benefit from the novel architectures introduced.

A compact, low power radon detector for environmental applications was developed, showing how the state-of-art can be advanced thanks to the availability of SiPMs and the new circuits.
Thanks to the proposed circuit design techniques, the instrumentation for an important nuclear physics experiment could be designed, demonstrating that the experiment will be feasible in the next two to three years. This will enable answering important questions about the role of neutrino in the Standard Physics Model, a very important advancement in our understanding of the universe.

The outcome of the work, however, is not limited to these two case studies: there are several areas of research that will benefit from the notions introduced. The spectrum of applicability is as wide as the radiation detection field itself. If it is true that SiPMs will replace PMTs in future applications, the research presented will be part of the core of knowledge upon which that transition will take place.

Bibliography

[1] Lawrence Berkeley National Laboratory, "The advanced light source: A tool for solving the mysteries of materials." Avalable at `http://www2.lbl.gov/MicroWorlds/ALSTool/EMSpec/EMSpec2.html`; accessed 20-January-2015.

[2] H. A. Bethe and J. Ashkin, *Experimental Nuclear Physics*. New York: J. Wiley, 1953. p. 253.

[3] H. A. Bethe and L. C. Maximon, "Theory of bremsstrahlung and pair production.," *Phys. Rev.*, vol. 93, pp. 768–784, Feb 1954.

[4] E. L. Alpen, *Radiation Biophysics*. San Diego: Academic Press, 1990. p. 61.

[5] J. H. Hubbell and S. M. Seltzer, "Tables of x-ray mass attenuation coefficients and mass energy-absorption coefficients from 1 kev to 20 mev for elements z = 1 to 92 and 48 additional substances of dosimetric interest," 2009. Available at `http://www.nist.gov/pml/data/xraycoef/index.cfm`; accessed 20-October-2015.

[6] A. H. Compton, "The spectrum of scattered x-rays," *Phys. Rev.*, vol. 22, pp. 409–413, Nov 1923.

[7] E. Rutherford and H. Geiger, "An electrical method of counting the number of alpha particles from radioactive substances," in *Proceedings of the Royal Society (London)*, vol. 81 of *A*, pp. 141–161, 1908.

[8] G. F. Knoll, *Radiation Detection and Measurement (3rd Edition)*. New York: J. Wiley, 1999. p. 75.

[9] E. Sakai, "Present status of room temperature semiconductor detectors," *Nuclear Instruments and Methods in Physics Research*, vol. 196, no. 1, pp. 121 – 130, 1982.

[10] D. McGregor and H. Hermon, "Room-temperature compound semiconductor radiation detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 395, no. 1, pp. 101 – 124, 1997. Proceedings of the Fourth International Workshop on GaAs Detectors and Related Compounds.

[11] Eljen. Available at `http://www.eljentechnology.com/index.php/products/liquid-scintillators/71-ej-301`; accessed 20-October-2015.

[12] S. Gobain. Available at `http://www.crystals.saint-gobain.com/uploadedFiles/SG-Crystals/Documents/SGC%20BC501_501A_519%20Data%20Sheet.pdf`; accessed 20-October-2015.

[13] L. M. Bollinger and G. E. Thomas, "Measurement of the time dependence of scintillation intensity by a delayedcoincidence method," *Review of Scientific Instruments*, vol. 32, pp. 1044–1050, 1961.

[14] E. Sakai, "Recent measurements on scintillator-photodetector systems," *Nuclear Science, IEEE Transactions on*, vol. 34, pp. 418–422, 1987.

[15] I. Holl, E. Lorenz, and G. Mageras, "A measurement of the light yield of common inorganic scintillators," *Nuclear Science, IEEE Transactions on*, vol. 35, pp. 105–109, 1988.

[16] A. Einstein, A. Beck, and P. Havas, *The Collected Papers of Albert Einstein. Volume 2: The Swiss Years: Writings, 1900-1909*. Princeton, NJ: Princeton University Press, 1989. p. 86.

[17] Hamamatsu. Available at `http://www.hamamatsu.com/us/en/technology/innovation/photocathode/index.html`; accessed 20-October-2015.

[18] C. Fiorini, A. Longoni, F. Perotti, C. Labanti, P. Lechner, and L. Struder, "Gamma ray spectroscopy with csi(tl) scintillator coupled to silicon drift chamber," *Nuclear Science, IEEE Transactions on*, vol. 44, pp. 2553–2560, 1997.

[19] G. Bondarenko, B. Dolgoshein, V. Golovin, A. Ilyin, R. Klanner, and E. Popova, "Limited geiger-mode silicon photodiode with very high gain," *Nuclear Physics B (Proc. Suppl.)*, vol. 61B, pp. 347–352, 1998.

[20] P. Buzhan, B. Dolgoshein, A. Ilyin, V. Kantserov, V. Kaplin, A. Karakash, A. Pleshko, E. Popova, S. Smirnov, Y. Volkov, L. Filatov, S. Klemin, and F. Kayumov, "The Advanced Study of Silicon Photomultiplier," in *Advanced Technology - Particle Physics* (M. Barone, E. Borchi, J. Huston, C. Leroy, P. G. Rancoita, P. Riboni, and R. Ruchti, eds.), pp. 717–728, Nov. 2002.

[21] P. Buzhan, B. Dolgoshein, L. Filatov, A. Ilyin, V. Kaplin, A. Karakash, S. Klemin, R. Mirzoyan, A. Otte, E. Popova, V. Sosnovtsev, and M. Teshima, "Large area silicon photomultipliers: Performance and applications," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 567, no. 1, pp. 78 – 82, 2006. Proceedings of the 4th International Conference on New Developments in Photodetection BEAUNE 2005 Fourth International Conference on New Developments in Photodetection.

[22] G. Bondarenko, P. Buzhan, B. Dolgoshein, V. Golovin, E. Guschin, A. Ilyin, V. Kaplin, A. Karakash, R. Klanner, V. Pokachalov,

E. Popova, and K. Smirnov, "Limited geiger-mode microcell silicon photodiode: new results," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 442, no. 13, pp. 187 – 192, 2000.

[23] P. Lightfoot, G. Barker, K. Mavrokoridis, Y. Ramachers, and N. Spooner, "Characterisation of a silicon photomultiplier device for applications in liquid argon based neutrino physics and dark matter searches," *JINST*, vol. 3, p. P10001, 2008.

[24] E. Roncali and S. R. Cherry, "Application of silicon photomultipliers to positron emission tomography," *Annals of Biomedical Engineering*, vol. 39, no. 4, pp. 1358 – 1377, 2011.

[25] SensL. Available at `http://sensl.com/download-c-series-um/`; accessed 21-October-2015.

[26] P. Eckert, H.-C. Schultz-Coulon, W. Shen, R. Stamen, and A. Tadday, "Characterisation studies of silicon photomultipliers," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 620, no. 23, pp. 217 – 226, 2010.

[27] Hamamatsu. Available at `http://www.hamamatsu.com/resources/pdf/ssd/s12572-025_etc_kapd1043e.pdf`; accessed 21-October-2015.

[28] Q. Shao, L. F. Voss, A. M. Conway, R. J. Nikolic, M. A. Dar, and C. L. Cheung, "High aspect ratio composite structures with 48.5% thermal neutron detection efficiency," *Applied Physics Letters*, vol. 102, no. 6, 2013.

[29] Lawrence Livermore National Laboratory, "Science and technology review." Available at `https://str.llnl.gov/march-2014/`; accessed 23-October-2015.

[30] J. D. Cockcroft and E. T. S. Walton, "Experiments with High Velocity Positive Ions. (I) Further Developments in the Method of Obtaining High Velocity Positive Ions," in *Proceedings of the Royal Society A*, pp. 619–630, 1932.

[31] E. Gatti and P. Manfredi, "Processing the signals from solid-state detectors in elementary-particle physics," *La Rivista del Nuovo Cimento (1978-1999)*, vol. 9, no. 1, pp. 1–146, 1986.

[32] R. H. Howard, *Principles of Random Signal Analysis and Low Noise Design: The Power Spectral Density and its Applications*. New York: J. Wiley, 2004.

[33] F. S. Goulding and D. A. Landis, *Semiconductor Detector Spectrometer Electronics, from "Nuclear Spectroscopy and Reactions, Part A"*. San Francisco: Academic Press, 1974.

[34] S. Kogan, *Electronics Noise and Fluctuations in Solids*. New York: Cambridge University Press, 2008.

[35] G. F. Knoll, *Radiation Detection and Measurement (3rd Edition)*. New York: J. Wiley, 1999. pp. 416–419.

[36] F. S. Goulding and D. A. Landis, "Signal processing for semiconductor detectors," *Nuclear Science, IEEE Transactions on*, vol. 29, pp. 1125–1141, 1982.

[37] Hamamatsu, "Characteristics and use of Si APD (avalanche photodiode)." Available at `http://neutron.physics.ucsb.edu/docs/Characteristics_and_use_of_SI_APD.pdf`; accessed 23-October-2015.

[38] K. P. Ziock and W. H. Goldstein, "The lost source, varying backgrounds and why bigger may not be better," *AIP Conference Proceedings*, vol. 632, no. 1, pp. 60–70, 2002.

[39] K. Ziock, W. Craig, L. Fabris, R. Lanza, S. Gallagher, B. Horn, and N. Madden, "Large area imaging detector for long-range, passive detection of fissile material," *Nuclear Science, IEEE Transactions on*, vol. 51, pp. 2238–2244, Oct 2004.

[40] R. Raylman, A. Stolin, S. Majewski, and J. Proffitt, "A large area, silicon photomultiplier-based PET detector module," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 735, pp. 602 – 609, 2014.

[41] World Health Organization, *WHO Handbook on Indoor Radon: A Public Health Perspective*. France: World Health Organization, 2009. pp. 7–14.

[42] S. Darby, D. Hill, A. Auvinen, J. M. Barros-Dios, H. Baysson, F. Bochicchio, H. Deo, R. Falk, F. Forastiere, M. Hakama, I. Heid, L. Kreienbrock, M. Kreuzer, F. Lagarde, I. Mäkeläinen, C. Muirhead, W. Oberaigner, G. Pershagen, A. Ruano-Ravina, E. Ruosteenoja, A. S. Rosario, M. Tirmarche, L. Tomášek, E. Whitley, H.-E. Wichmann, and R. Doll, "Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 european case-control studies," vol. 330, no. 7485, p. 223, 2005.

[43] Tosaka, "Decay chain 4n+2, uranium series." From `https://commons.wikimedia.org/wiki/File:Decay_chain(4n%2B2,_Uranium_series).PNG`; accessed 11-May-2015.

[44] Environmental Protection Agency, "EPA map of radon zones by u.s. county," 2015. Available at `http://www2.epa.gov/sites/production/files/2015-07/documents/zonemapcolor.pdf`; accessed 11-May-2015.

[45] Environmental Protection Agency, "A citizen's guide to radon," 2012. Available at `http://www2.epa.gov/sites/production/files/2015-05/documents/citizensguide.pdf`; accessed 26-October-2015.

[46] United States Environmental Protection Agency, "Consumer's guide to radon reduction– how to fix your home," 2013. Available at `http://www2.epa.gov/sites/production/files/2015-05/documents/consguid.pdf`; accessed 26-October-2015.

[47] G. Jönsson, "On the Detection of Alpha Particles by Kodak LR-Film at Various Angles of Incidence," in *Solid State Nuclear Track Detectors: Proceedings of the 11th International Conference*, pp. 275–278, 1981.

[48] NC Radon Program, "Radon testing devices." Available at `http://www.ncradon.org/docs/rntestingdevices.pdf`; accessed 26-October-2015.

[49] A. George, "An overview of instrumentation for measuring environmental radon and radon progeny," *Nuclear Science, IEEE Transactions on*, vol. 37, pp. 892–901, Apr 1990.

[50] United States Environmental Protection Agency, *Protocols for Radon and Radon Decay Product Measurements in Homes*. 1993. EPA 402 R 92 003.

[51] "Safetysiren^{TM} pro series 3." Available at `https://www.safetysirenpro.com/`; accessed 26-October-2015.

[52] "Sun nuclear 1027." Available at `http://radon.sunnuclear.com/1027/1027.asp`; accessed 26-October-2015.

[53] A. Bosi, L. Bidinelli, D. Saguatti, G.-F. Dalla Betta, V. Tyzhnevyi, L. Rovati, G. Verzellesi, G. Batignani, S. Bettarini, F. Forti, F. D' Errico, A. Del Gratta, L. Bosisio, I. Rachevskaia, M. Boscardin, G. Giacomini, M. Calamosca, S. Penzo, and F. Cardellini, "Performance of a radon sensor based on a BJT detector on high-resistivity silicon," in *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*, pp. 266–268, Oct 2012.

[54] "Corentium digital radon meter." Available at `https://usa.corentium.com/`; accessed 26-October-2015.

[55] J. B. Birks, *The Theory and Practice of Scintillator Counting, 1st Edition*. London: Pergamon Press, 2013. pp. 541–551.

[56] Eljen. Available at `http://www.eljentechnology.com/index.php/products/zinc-sulfide-shifting-plastics/100-ej-440`; accessed 26-October-2015.

[57] SensL. Available at `http://www.sensl.com/downloads/ds/DS-ArraySB-4.pdf`; accessed 26-October-2015.

[58] Saint-Gobain. Available at `http://www.crystals.saint-gobain.com/uploadedFiles/SG-Crystals/Documents/Organic%20Product%20Accessories%20Data%20Sheet.pdf`; accessed 26-October-2015.

[59] J. F. Ziegler, J. P. Biersack, and M. D. Ziegler, *SRIM The Stopping and Range of Ions in Matter*. Morrisville, SC: Lulu Press co., 2015.

[60] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytracek, G. Cooperman, G. Cosmo, P. Degtyarenko, A. Dell'Acqua, G. Depaola, D. Dietrich, R. Enami, A. Feliciello, C. Ferguson, H. Fesefeldt, G. Folger, F. Foppiano, A. Forti, S. Garelli, S. Giani, R. Giannitrapani, D. Gibin, J. Gómez Cadenas, I. González, G. G. Abril, G. Greeniaus, W. Greiner, V. Grichine, A. Grossheim, S. Guatelli, P. Gumplinger, R. Hamatsu, K. Hashimoto, H. Hasui, A. Heikkinen, A. Howard, V. Ivanchenko, A. Johnson, F. Jones, J. Kallenbach, N. Kanaya, M. Kawabata, Y. Kawabata, M. Kawaguti, S. Kelner, P. Kent, A. Kimura, T. Kodama, R. Kokoulin, M. Kossov, H. Kurashige, E. Lamanna, T. Lampén, V. Lara, V. Lefebure, F. Lei, M. Liendl, W. Lockman, F. Longo, S. Magni, M. Maire, E. Medernach, K. Minamimoto, P. M. de Freitas, Y. Morita, K. Murakami, M. Nagamatu, R. Nartallo, P. Nieminen, T. Nishimura, K. Ohtsubo, M. Okamura, S. O'Neale, Y. Oohata, K. Paech, J. Perl, A. Pfeiffer, M. Pia, F. Ranjard, A. Rybin, S. Sadilov, E. D. Salvo,

G. Santin, T. Sasaki, N. Savvas, Y. Sawada, S. Scherer, S. Sei, V. Sirotenko, D. Smith, N. Starkov, H. Stoecker, J. Sulkimo, M. Takahata, S. Tanaka, E. Tcherniaev, E. S. Tehrani, M. Tropeano, P. Truscott, H. Uno, L. Urban, P. Urban, M. Verderi, A. Walkden, W. Wander, H. Weber, J. Wellisch, T. Wenaus, D. Williams, D. Wright, T. Yamada, H. Yoshida, and D. Zschiesche, "Geant4a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250 – 303, 2003.

[61] Pipe Flow Claculations. Available at `http://www.pipeflowcalculations.com/pipe-valve-fitting-flow/compressible-gas-flow.php#isothermal-compressible-flow-in-pipe`; accessed 20-June-2015.

[62] F. Corsi, M. Foresta, C. Marzocca, G. Matarrese, and A. Del Guerra, "Current-mode front-end electronics for silicon photo-multiplier detectors," in *Advances in Sensors and Interface, 2007. IWASI 2007. 2nd International Workshop on*, pp. 1–6, June 2007.

[63] NXP Semiconductors. Available at `http://www.nxp.com/documents/data_sheet/BFR92A_N.pdf`; accessed 26-October-2015.

[64] E. Fairstein, "Linear unipolar pulse-shaping networks: current technology," *Nuclear Science, IEEE Transactions on*, vol. 37, pp. 382–397, Apr 1990.

[65] V. Radeka, "Trapezoidal filtering of signals from large germanium detectors at high rates," *Nuclear Instruments and Methods*, vol. 99, no. 3, pp. 525 – 539, 1972.

[66] ST Microelectronics. Available at `http://www.st.com/st-web-ui/static/active/en/resource/technical/document/datasheet/DM00065944.pdf`; accessed 26-October-2015.

[67] C. H. Nowlin and J. L. Blankenship, "Elimination of undesirable undershoot in the operation and testing of nuclear pulse amplifiers," *Review of Scientific Instruments*, vol. 36, no. 12, pp. 1830 – 1839, 1965.

[68] Cadsoft, "Eagle layout editor." Available at `http://www.cadsoftusa.com/`; accessed 26-October-2015.

[69] The 2015 Nuclear Science Advisory Committee, "The 2015 long range plan for nuclear science," 2015. Available at `http://science.energy.gov/~/media/np/nsac/pdf/2015LRP/2015_LRPNS_091815.pdf`; accessed 27-October-2015.

[70] M. Gerbino, M. Lattanzi, and A. Melchiorri, "The ν generation: present and future constraints on neutrino masses from cosmology and laboratory experiments," *ArXiv*. arXiv:1507.08614 [hep-ph].

[71] M. Goeppert-Mayer, "Double-beta disintegration," *Physical Review*, vol. 48, no. 6, pp. 512–516, 1935.

[72] A. Giuliani and A. Poves, "Neutrinoless double-beta decay," *Advances in High Energy Physics*, vol. 2012, pp. 1–38, 2012. Article ID 857016.

[73] M. Drewes, "The phenomenology of right handed neutrinos," *International Journal of Modern Physics E*, vol. 22, no. 08, p. 1330019, 2013.

[74] N. Ackerman, B. Aharmim, M. Auger, D. J. Auty, P. S. Barbeau, K. Barry, L. Bartoszek, E. Beauchamp, V. Belov, C. Benitez-Medina, M. Breidenbach, A. Burenkov, B. Cleveland, R. Conley, E. Conti, J. Cook, S. Cook, A. Coppens, I. Counts, W. Craddock, T. Daniels, M. V. Danilov, C. G. Davis, J. Davis, R. deVoe, Z. Djurcic, A. Dobi, A. G. Dolgolenko, M. J. Dolinski, K. Donato, M. Dunford, W. Fairbank, J. Farine, P. Fierlinger, D. Franco, D. Freytag, G. Giroux, R. Gornea, K. Graham, G. Gratta, M. P. Green, C. Hägemann, C. Hall, K. Hall, G. Haller, C. Hargrove, R. Herbst, S. Herrin, J. Hodgson, M. Hughes, A. Johnson, A. Karelin, L. J. Kaufman, T. Koffas,

A. Kuchenkov, A. Kumar, K. S. Kumar, D. S. Leonard, F. Leonard, F. LePort, D. Mackay, R. MacLellan, M. Marino, Y. Martin, B. Mong, M. Montero Díez, P. Morgan, A. R. Müller, R. Neilson, R. Nelson, A. Odian, K. O'Sullivan, C. Ouellet, A. Piepke, A. Pocar, C. Y. Prescott, K. Pushkin, A. Rivas, E. Rollin, P. C. Rowson, J. J. Russell, A. Sabourov, D. Sinclair, K. Skarpaas, S. Slutsky, V. Stekhanov, V. Strickland, M. Swift, D. Tosi, K. Twelker, P. Vogel, J.-L. Vuilleumier, J.-M. Vuilleumier, A. Waite, S. Waldman, T. Walton, K. Wamba, M. Weber, U. Wichoski, J. Wodin, J. D. Wright, L. Yang, Y.-R. Yen, and O. Y. Zeldovich, "Observation of two-neutrino double-beta decay in ^{136}Xe with the EXO-200 detector," *Phys. Rev. Lett.*, vol. 107, p. 212501, Nov 2011.

[75] S. R. Elliott and P. Vogel, "Double beta decay," *Annual Review of Nuclear and Particle Science*, vol. 52, no. 1, pp. 115–151, 2002.

[76] M. Auger, D. J. Auty, P. S. Barbeau, L. Bartoszek, E. Baussan, E. Beauchamp, C. Benitez-Medina, M. Breidenbach, D. Chauhan, B. Cleveland, R. Conley, J. Cook, S. Cook, A. Coppens, W. Craddock, T. Daniels, C. G. Davis, J. Davis, R. deVoe, A. Dobi, M. J. Dolinski, M. Dunford, W. F. Jr, J. Farine, P. Fierlinger, D. Franco, G. Giroux, R. Gornea, K. Graham, G. Gratta, C. Hägemann, C. Hall, K. Hall, C. Hargrove, S. Herrin, J. Hodgson, M. Hughes, A. Karelin, L. J. Kaufman, J. Kirk, A. Kuchenkov, K. S. Kumar, D. S. Leonard, F. Leonard, F. LePort, D. Mackay, R. MacLellan, M. Marino, K. Merkle, B. Mong, M. Montero Díez, A. R. Müller, R. Neilson, A. Odian, K. O'Sullivan, C. Ouellet, A. Piepke, A. Pocar, C. Y. Prescott, K. Pushkin, A. Rivas, E. Rollin, P. C. Rowson, A. Sabourov, D. Sinclair, K. Skarpaas, S. Slutsky, V. Stekhanov, V. Strickland, M. Swift, D. Tosi, K. Twelker, J. L. Vuilleumier, J. M. Vuilleumier, T. Walton, M. Weber, U. Wichoski, J. Wodin, J. D. Wright, L. Yang, and Y. R. Yen, "The EXO-200 detector, part I: detector design and construction," *Journal of Instrumentation*, vol. 7, no. 05, p. P05010, 2012.

[77] S. Kubota, M. Hishida, M. Suzuki, and J.-z. Ruan(Gen), "Dynamical

behavior of free electrons in the recombination process in liquid argon, krypton, and xenon," *Phys. Rev. B*, vol. 20, pp. 3486–3496, Oct 1979.

[78] SensL, "B-series UV sensitive SiPM and SMT arrays." Available at `http://www.eqphotonics.de/cms/cms/upload/datasheets/PB-SensL%20Product%20Summary_IEEE.pdf`; accessed 26-October-2015.

[79] B. Dolgoshein, R. Mirzoyan, E. Popova, P. Buzhan, A. Ilyin, V. Kaplin, A. Stifutkin, M. Teshima, and A. Zhukov, "Large area UV sipms with extremely low cross-talk," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 695, pp. 40 – 43, 2012. New Developments in Photodetection NDIP11.

[80] N. Yahlali, L. M. P. Fernandes, K. Gonzalez, A. N. C. Garcia, and A. Soriano, "Imaging with sipms in noble-gas detectors," *JINST*, vol. 8, p. C01003, 2013. IWORID 2012.

[81] V. Radeka. private communication, 2014.

[82] Intersil Corp. Available at `http://www.intersil.com/content/dam/Intersil/documents/hfa3/hfa3046-3096-3127-3128.pdf`; accessed 27-October-2015.

[83] SensL. Available at `http://www.sensl.com/downloads/ds/UM-ArraySMT.pdf`; accessed 27-October-2015.

[84] SensL. Available at `http://sensl.com/downloads/ds/DS-MicroCseries.pdf`; accessed 27-October-2015.

[85] Liteon Optoelectronics. Available at `http://optoelectronics.liteon.com/upload/download/DS20-2002-028/S_110_LTL1CHTBK4.pdf`; accessed 27-October-2015.

[86] C. Thorn, G. D. Geronimo, A. D'Andragora, S. Li, N. Nambiar, S. Rescia, E. Vernon, H. Chen, F. Lanni, D. Makowiecki, V. Radeka,

and B. Yu, "Cold electronics development for the LBNE LAr TPC," *Physics Procedia*, vol. 37, pp. 1295 – 1302, 2012. Proceedings of the 2nd International Conference on Technology and Instrumentation in Particle Physics (TIPP 2011).