Analog and Digital Electronics for Detectors

Helmuth Spieler
Physics Division, Lawrence Berkeley National Laboratory*
Berkeley, California 94720, U.S.A.

1 Introduction

Electronics are a key component of all modern detector systems. Although experiments and their associated electronics can take very different forms, the same basic principles of the electronic readout and optimization of signal-to-noise ratio apply to all. This chapter provides a summary of front-end electronics components and discusses signal processing with an emphasis on electronic noise. Because of space limitations, this can only be a brief overview. A more detailed discussion of electronics with emphasis on semiconductor detectors is given elsewhere[1]. Tutorials on detectors, signal processing and electronics are also available on the world wide web[2].

The purpose of front-end electronics and signal processing systems is to

1. Acquire an electrical signal from the sensor. Typically this is a short current pulse.
2. Tailor the time response of the system to optimize
   (a) the minimum detectable signal (detect hit/no hit),
   (b) energy measurement,
   (c) event rate,
   (d) time of arrival (timing measurement),
   (e) insensitivity to sensor pulse shape,
   (f) or some combination of the above.
3. Digitize the signal and store for subsequent analysis.

Position-sensitive detectors utilize the presence of a hit, amplitude measurement or timing, so these detectors pose the same set of requirements.

Generally, these properties cannot be optimized simultaneously, so compromises are necessary. In addition to these primary functions of an electronic

*This work was supported by the Director, Office of Science, Office of High Energy and Nuclear Physics, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

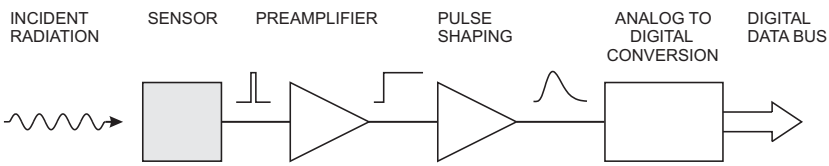


Figure 1: Basic detector functions: Radiation is absorbed in the sensor and converted into an electrical signal. This low-level signal is integrated in a preamplifier, fed to a pulse shaper, and then digitized for subsequent storage and analysis.

readout system, other considerations can be equally or even more important. Examples are radiation resistance, low power (portable systems, large detector arrays, satellite systems), robustness, and – last, but not least – cost.

2 Example Systems

Figure 1 illustrates the components and functions of a radiation detector system. The sensor converts the energy deposited by a particle (or photon) to an electrical signal. This can be achieved in a variety of ways. In direct detection – semiconductor detectors, wire chambers, or other types of ionization chambers – energy is deposited in an absorber and converted into charge pairs, whose number is proportional to the absorbed energy. The signal charge can be quite small, in semiconductor sensors about 50 aC ($5 \cdot 10^{-17}$ C) for 1 keV x-rays and 4 fC ($4 \cdot 10^{-15}$ C) in a typical high-energy tracking detector, so the sensor signal must be amplified. The magnitude of the sensor signal is subject to statistical fluctuations and electronic noise further "smears" the signal. These fluctuations will be discussed below, but at this point we note that the sensor and preamplifier must be designed carefully to minimize electronic noise. A critical parameter is the total capacitance in parallel with the input, *i.e.* the sensor capacitance and input capacitance of the amplifier. The signal-to-noise ratio increases with decreasing capacitance. The contribution of electronic noise also relies critically on the next stage, the pulse shaper, which determines the bandwidth of the system and hence the overall electronic noise contribution. The shaper also limits the duration of the pulse, which sets the maximum signal rate that can be accommodated. The shaper feeds an analog-to-digital converter (ADC), which converts the magnitude of the analog signal into a bit-pattern suitable for subsequent digital storage and processing.

A scintillation detector (Figure 2) utilizes indirect detection, where the absorbed energy is first converted into visible light. The number of scintilla-

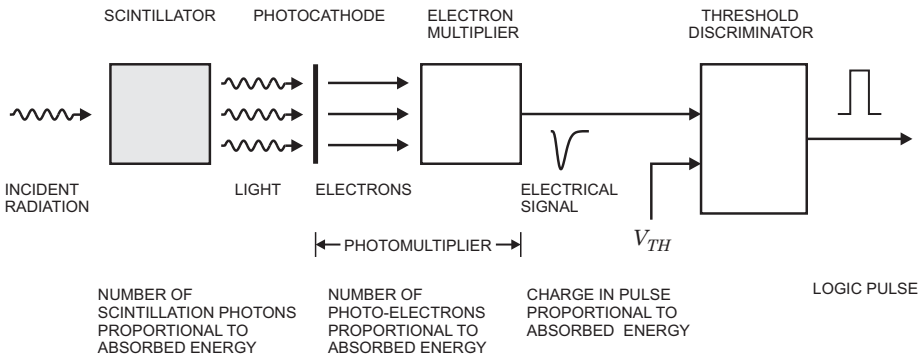


Figure 2: In a scintillation detector absorbed energy is converted into visible light. The scintillation photons are commonly detected by a photomultiplier, which can provide sufficient gain to directly drive a threshold discriminator.

tion photons is proportional to the absorbed energy. The scintillation light is detected by a photomultiplier (PMT), consisting of a photocathode and an electron multiplier. Photons absorbed in the photocathode release electrons, whose number is proportional to the number of incident scintillation photons. At this point energy absorbed in the scintillator has been converted into an electrical signal whose charge is proportional to energy. Increased in magnitude by the electron multiplier, the signal at the PMT output is a current pulse. Integrated over time this pulse contains the signal charge, which is proportional to the absorbed energy. Figure 2 shows the PMT output pulse fed directly to a threshold discriminator, which fires when the signal exceeds a predetermined threshold, as in a counting or timing measurement. The electron multiplier can provide sufficient gain, so no preamplifier is necessary. This is a typical arrangement used with fast plastic scintillators. In an energy measurement, for example using a NaI(Tl) scintillator, the signal would feed a pulse shaper and ADC, as shown in Figure 1.

If the pulse shape does not change with signal charge, the peak amplitude – the pulse height – is a measure of the signal charge, so this measurement is called pulse height analysis. The pulse shaper can serve multiple functions, which are discussed below. One is to tailor the pulse shape to the ADC. Since the ADC requires a finite time to acquire the signal, the input pulse may not be too short and it should have a gradually rounded peak. In scintillation detector systems the shaper is frequently an integrator and implemented as the first stage of the ADC, so it is invisible to the casual observer. Then the system appears

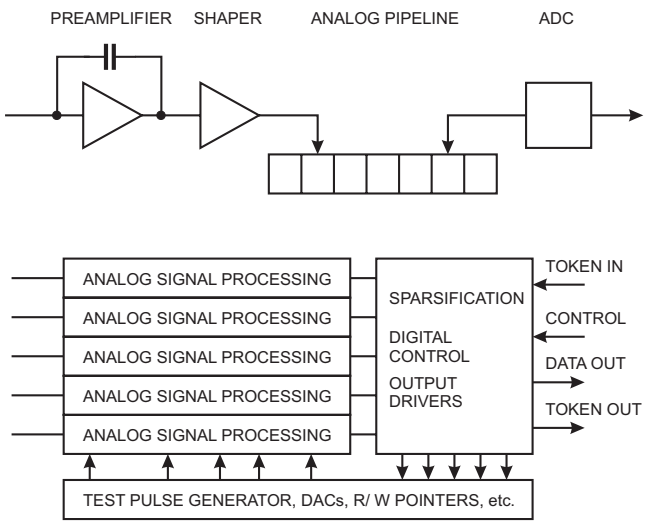


Figure 3: Circuit blocks in a representative readout IC. The analog processing chain is shown at the top. Control is passed from chip to chip by token passing.

very simple, as the PMT output is plugged directly into a charge-sensing ADC.

A detector array combines the sensor and the analog signal processing circuitry together with a readout system. The electronic circuitry is often monolithically integrated. Figure 3 shows the circuit blocks in a representative readout integrated circuit (IC). Individual sensor electrodes connect to parallel channels of analog signal processing circuitry. Data are stored in an analog pipeline pending a readout command. Variable write and read pointers are used to allow simultaneous read and write. The signal in the time slot of interest is digitized, compared with a digital threshold, and read out. Circuitry is included to generate test pulses that are injected into the input to simulate a detector signal. This is a very useful feature in setting up the system and is also a key function in chip testing prior to assembly. Analog control levels are set by digital-to-analog converters (DACs). Multiple ICs are connected to a common control and data output bus, as shown in Figure 4. Each IC is assigned a unique address, which is used in issuing control commands for setup and *in situ* testing. Sequential readout is controlled by token passing. IC1 is the master, whose readout is initiated by a command (trigger) on the control bus. When it has finished writing data it passes the token to IC2, which in turn passes the token to IC3. When the last chip has completed its readout

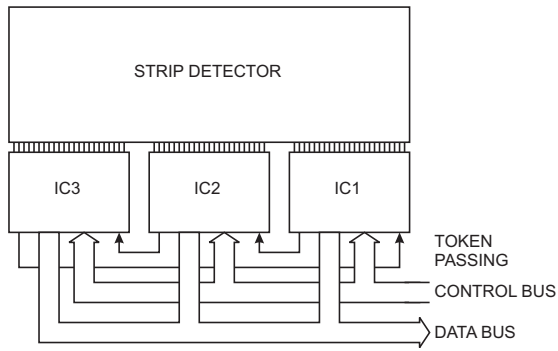


Figure 4: Multiple ICs are ganged to read out a silicon strip detector. The right-most chip IC1 is the master. A command on the control bus initiates the readout. When IC1 has written all of its data it passes the token to IC2. When IC2 has finished it passes the token to IC3, which in turn returns the token to the master IC1.

the token is returned to the master IC, which is then ready for the next cycle. The readout bit stream begins with a header, which uniquely identifies a new frame. Data from individual ICs are labeled with a chip identifier and channel identifiers. Many variations on this scheme are possible. As shown, the readout is event oriented, *i.e.* all hits occurring within an externally set exposure time (*e.g.* time slice in the analog buffer in Figure 3) are read out together. For a concise discussion of data acquisition systems see ref. [3].

In colliding beam experiments only a small fraction of beam crossings yields interesting events. The time required to assess whether an event is potentially interesting is typically of order microseconds, so hits from multiple beam crossings must be stored on-chip, identified by beam crossing or time-stamp. Upon receipt of a trigger the interesting data are digitized and read out. This allows use of a digitizer that is slower than the collision rate. It is also possible to read out analog signals and digitize them externally. Then the output stream is a sequence of digital headers and analog pulses. An alternative scheme only records the presence of hit. The output of a threshold comparator signifies the presence of a signal and is recorded in a digital pipeline that retains the crossing number.

Figure 5 shows a closeup of ICs mounted on a hybrid using a flexible polyimide substrate[4]. The wire bonds connecting the IC to the hybrid are clearly visible. Channels on the IC are laid out on a $\sim 50\,\mu$m pitch and pitch adapters fan out to match the $80\,\mu$m pitch of the silicon strip detector. The space be-

tween chips accommodates bypass capacitors and connections for control busses carrying signals from chip to chip.

3 Detection Limits and Resolution

The minimum detectable signal and the precision of the amplitude measurement are limited by fluctuations. The signal formed in the sensor fluctuates, even for a fixed energy absorption. In addition, electronic noise introduces baseline fluctuations, which are superimposed on the signal and alter the peak amplitude. Figure 6 (left) shows a typical noise waveform. Both the amplitude and time distributions are random. When superimposed on a signal, the noise alters both the amplitude and time dependence, as shown in Figure 6 (right). As can be seen, the noise level determines the minimum signal whose presence can be discerned.

In an optimized system, the time scale of the fluctuations is comparable

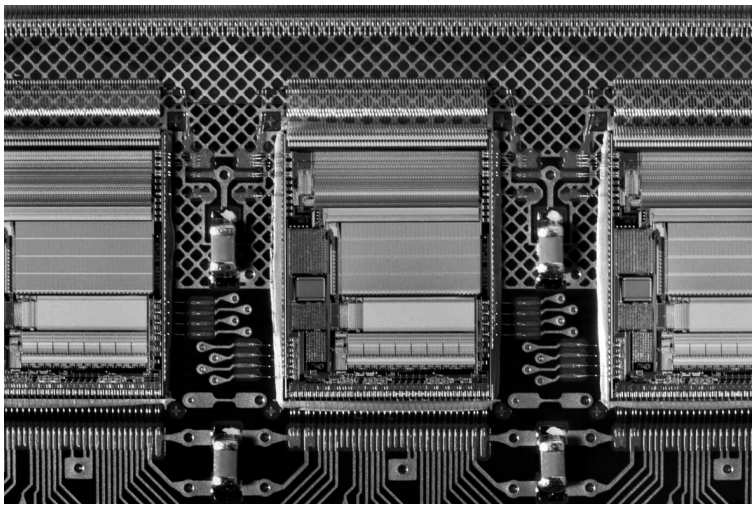


Figure 5: Closeup of ICs mounted on a hybrid utilizing a flexible polyimide substrate. The high-density wire bonds at the upper edges connect via pitch adapters to the 80 μm pitch of the silicon strip detector. The ground plane is patterned as a diamond grid to reduce material. (Photograph courtesy of A. Ciocio.)

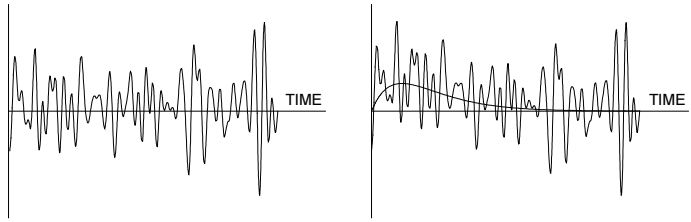


Figure 6: Waveforms of random noise (left) and signal + noise (right), where the peak signal is equal to the rms noise level ($S/N = 1$). The noiseless signal is shown for comparison.

to that of the signal, so the peak amplitude fluctuates randomly above and below the average value. This is illustrated in Figure 7, which shows the same signal viewed at four different times. The fluctuations in peak amplitude are obvious, but the effect of noise on timing measurements can also be seen. If the timing signal is derived from a threshold discriminator, where the output fires when the signal crosses a fixed threshold, amplitude fluctuations in the

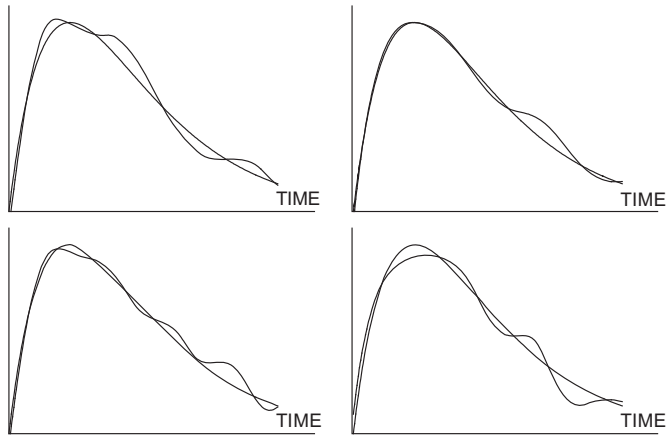


Figure 7: Signal plus noise at four different times, shown for a signal-to-noise ratio of about 20. The noiseless signal is superimposed for comparison.

leading edge translate into time shifts. If one derives the time of arrival from a centroid analysis, the timing signal also shifts (compare the top and bottom right figures). From this one sees that signal-to-noise ratio is important for all measurements – sensing the presence of a signal or the measurement of energy, timing, or position.

4 Acquiring the Sensor Signal

The sensor signal is usually a short current pulse $i_s(t)$. Typical durations vary widely, from 100 ps for thin Si sensors to tens of μs for inorganic scintillators. However, the physical quantity of interest is the deposited energy, so one has to integrate over the current pulse

$$E \propto Q_s = \int i_s(t)dt \ . \tag{1}$$

This integration can be performed at any stage of a linear system, so one can

1. integrate on the sensor capacitance,
2. use an integrating preamplifier ("charge-sensitive" amplifier),
3. amplify the current pulse and use an integrating ADC ("charge sensing" ADC),
4. rapidly sample and digitize the current pulse and integrate numerically.

In high-energy physics the first three options tend to be most efficient.

4.1 Signal integration

Figure 8 illustrates signal formation in an ionization chamber connected to an amplifier with a very high input resistance. The ionization chamber volume could be filled with gas or a solid, as in a silicon sensor. As mobile charge carriers move towards their respective electrodes they change the induced charge on the sensor electrodes, which form a capacitor C_d. If the amplifier has a very small input resistance R_i, the time constant $\tau = R_i(C_d + C_i)$ for discharging the sensor is small, and the amplifier will sense the signal current. However, if the input time constant is large compared to the duration of the current pulse, the current pulse will be integrated on the capacitance and the resulting voltage at the amplifier input

$$V_i = \frac{Q_s}{C_d + C_i} \ . \tag{2}$$

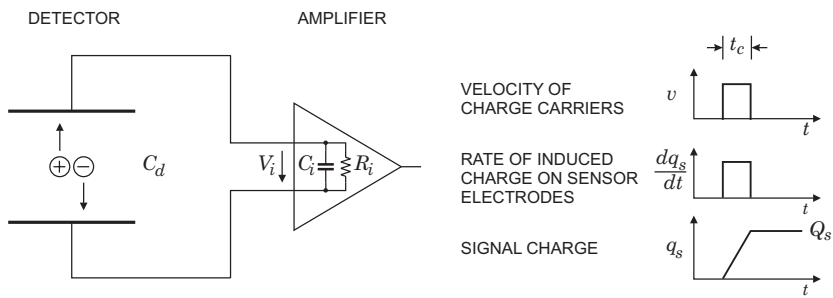


Figure 8: Charge collection and signal integration in an ionization chamber.

The magnitude of the signal is dependent on the sensor capacitance. In a system with varying sensor capacitances, a Si tracker with varying strip lengths, for example, or a partially depleted semiconductor sensor, where the capacitance varies with the applied bias voltage, one would have to deal with additional calibrations. Although this is possible, it is awkward, so it is desirable to use a system where the charge calibration is independent of sensor parameters. This can be achieved rather simply with a charge-sensitive amplifier.

Figure 9 shows the principle of a feedback amplifier that performs integration. It consists of an inverting amplifier with voltage gain $-A$ and a feedback capacitor C_f connected from the output to the input. To simplify the calculation, let the amplifier have an infinite input impedance, so no current flows into the amplifier input. If an input signal produces a voltage v_i at the amplifier input, the voltage at the amplifier output is $-Av_i$. Thus, the voltage difference across the feedback capacitor $v_f = (A+1)v_i$ and the charge deposited on C_f is

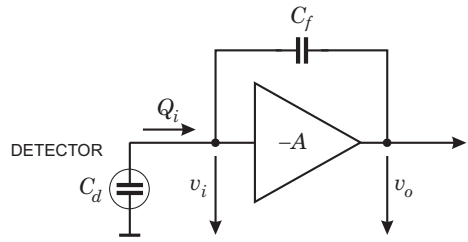


Figure 9: Principle of a charge-sensitive amplifier

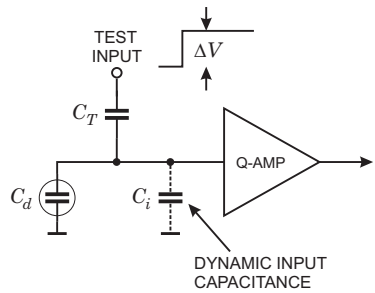


Figure 10: Adding a test input to a charge-sensitive amplifier provides a simple means of absolute charge calibration.

$Q_f = C_f v_f = C_f(A+1)v_i$. Since no current can flow into the amplifier, all of the signal current must charge up the feedback capacitance, so $Q_f = Q_i$. The amplifier input appears as a "dynamic" input capacitance

$$C_i = \frac{Q_i}{v_i} = C_f(A+1) \; . \tag{3}$$

The voltage output per unit input charge

$$A_Q = \frac{dv_o}{dQ_i} = \frac{Av_i}{C_i v_i} = \frac{A}{C_i} = \frac{A}{A+1} \cdot \frac{1}{C_f} \approx \frac{1}{C_f} \quad (A \gg 1) \; , \tag{4}$$

so the charge gain is determined by a well-controlled component, the feedback capacitor.

The signal charge Q_s will be distributed between the sensor capacitance C_d and the dynamic input capacitance C_i. The ratio of measured charge to signal charge

$$\frac{Q_i}{Q_s} = \frac{Q_i}{Q_d + Q_i} = \frac{C_i}{C_d + C_i} = \frac{1}{1 + \dfrac{C_d}{C_i}} \; , \tag{5}$$

so the dynamic input capacitance must be large compared to the sensor capacitance.

Another very useful byproduct of the integrating amplifier is the ease of charge calibration. By adding a test capacitor as shown in Figure 10, a voltage step injects a well-defined charge into the input node. If the dynamic input capacitance C_i is much larger than the test capacitance C_T, the voltage step at the test input will be applied nearly completely across the test capacitance C_T, thus injecting a charge $C_T \Delta V$ into the input.

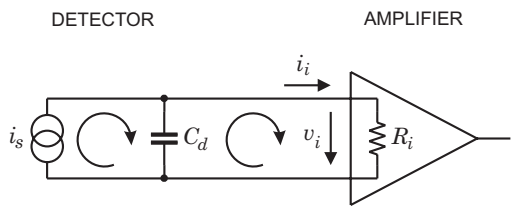


Figure 11: Charge integration in a realistic charge-sensitive amplifier. First, charge is integrated on the sensor capacitance and subsequently transferred to the charge-sensitive loop, as it becomes active.

4.2 Realistic charge-sensitive amplifiers

The preceding discussion assumed that the amplifiers are infinitely fast, so they respond instantaneously to the applied signal. In reality this is not the case; charge-sensitive amplifiers often respond much more slowly than the time duration of the current pulse from the sensor. However, as shown in Figure 11, this does not obviate the basic principle. Initially, signal charge is integrated on the sensor capacitance, as indicated by the left hand current loop. Subsequently, as the amplifier responds the signal charge is transferred to the amplifier.

Nevertheless, the time response of the amplifier does affect the measured pulse shape. First, consider a simple amplifier as shown in Figure 12.

The gain element shown is a bipolar transistor, but it could also be a field effect transistor (JFET or MOSFET) or even a vacuum tube. The transistor's

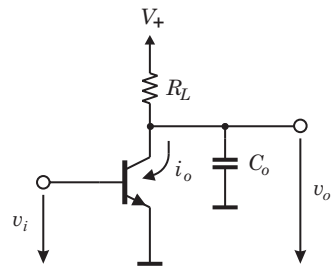


Figure 12: A simple amplifier demonstrating the general features of any single-stage gain stage, whether it uses a bipolar transistor (shown) or an FET.

output current changes as the input voltage is varied. Thus, the voltage gain

$$A_v = \frac{dv_o}{dv_i} = \frac{di_o}{dv_i} \cdot Z_L \equiv g_m Z_L \ . \tag{6}$$

The parameter g_m is the transconductance, a key parameter that determines gain, bandwidth, and noise of transistors. The load impedance Z_L is the parallel combination of the load resistance R_L and the output capacitance C_o. This capacitance is unavoidable; every gain device has an output capacitance, the following stage has an input capacitance, and in addition the connections and additional components introduce stray capacitance. The load impedance is given by

$$\frac{1}{Z_L} = \frac{1}{R_L} + \mathbf{i}\omega C_o \ , \tag{7}$$

where the imaginary $\mathbf{i}$ indicates the phase shift associated with the capacitance. The voltage gain

$$A_v = g_m \left(\frac{1}{R_L} + \mathbf{i}\omega C_o \right)^{-1} \ . \tag{8}$$

At low frequencies where the second term is negligible, the gain is constant $A_v = g_m R_L$. However, at high frequencies the second term dominates and the gain falls off linearly with frequency with a 90° phase shift, as illustrated in Figure 13. The cutoff (corner) frequency, where the asymptotic low and high frequency responses intersect, is determined by the output time constant $\tau = R_L C_o$, so the cutoff frequency

$$f_u = \frac{1}{2\pi\tau} = \frac{1}{2\pi R_L C_o} \ . \tag{9}$$

In the regime where the gain drops linearly with frequency the product of gain and frequency is constant, so the amplifier can be characterized by its gain–bandwidth product, which is equal to the frequency where the gain is one, the unity gain frequency $\omega_0 = g_m/C_o$.

The frequency response translates into a time response. If a voltage step is applied to the input of the amplifier, the output does not respond instantaneously, as the output capacitance must first charge up. This is shown in the second panel of Figure 13.

In practice, amplifiers utilize multiple stages, all of which contribute to the frequency response. However, for use as a feedback amplifier, only one time constant should dominate, so the other stages must have much higher cutoff frequencies. Then the overall amplifier response is as shown in Figure 13, except that at high frequencies additional corner frequencies appear.

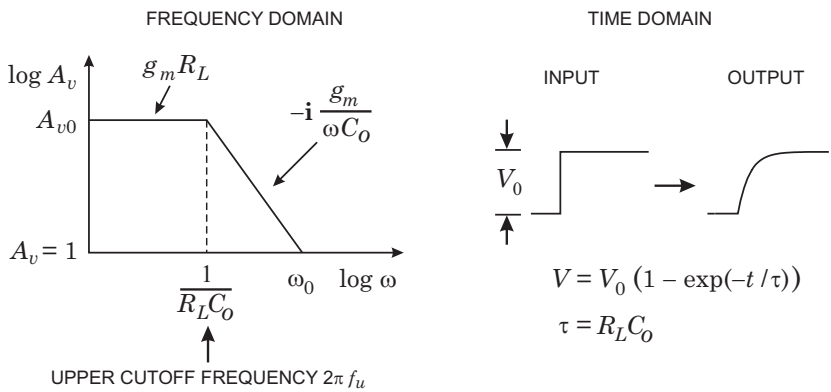


Figure 13: The time constants of an amplifier affect both the frequency and the time response. Both are fully equivalent representations.

We can now use the frequency response to calculate the input impedance and time response of a charge-sensitive amplifier. Applying the same reasoning as above, the input impedance of an inverting amplifier as shown in Figure 9, but with a generalized feedback impedance Z_f, is

$$Z_i = \frac{Z_f}{A+1} \approx \frac{Z_f}{A} \qquad (A \gg 1)\ . \tag{10}$$

At low frequencies the gain is constant and has a constant 180° phase shift, so the input impedance is of the same nature as the feedback impedance, but reduced by $1/A$. At high frequencies well beyond the amplifier's cutoff frequency f_u, the gain drops linearly with frequency with an additional 90° phase shift, so the gain

$$A = -\mathbf{i}\frac{\omega_0}{\omega}\ . \tag{11}$$

In a charge-sensitive amplifier the feedback impedance

$$Z_f = -\mathbf{i}\frac{1}{\omega C_f}\ , \tag{12}$$

so the input impedance

$$Z_i = -\frac{\mathbf{i}}{\omega C_f} \cdot \frac{1}{-\mathbf{i}\frac{\omega_0}{\omega}} = \frac{1}{\omega_0 C_f} \equiv R_i\ . \tag{13}$$

The imaginary component vanishes, so the input impedance is real. In other words, it appears as a resistance R_i. Thus, at low frequencies $f \ll f_u$ the input of a charge-sensitive amplifier appears capacitive, whereas at high frequencies $f \gg f_u$ it appears resistive.

Suitable amplifiers invariably have corner frequencies well below the frequencies of interest for radiation detectors, so the input impedance is resistive. This allows a simple calculation of the time response. The sensor capacitance is discharged by the resistive input impedance of the fedback amplifier with the time constant

$$\tau_i = R_i C_d = \frac{1}{\omega_0 C_f} \cdot C_d \ . \tag{14}$$

From this we see that the rise time of the charge-sensitive amplifier increases with sensor capacitance. As noted above, the amplifier response can be slower than the duration of the current pulse from the sensor, but it should be much faster than the peaking time of the subsequent pulse shaper. The feedback capacitance should be much smaller than the sensor capacitance. If $C_f = C_d/100$, the amplifier's gain–bandwidth product must be $100/\tau_i$, so for a rise time constant of 10 ns the gain–bandwidth product must be 10^{10} radians = 1.6 GHz. The same result can be obtained using conventional operational amplifier feedback theory.

The mechanism of reducing the input impedance through shunt feedback leads to the concept of the "virtual ground". If the gain is infinite, the input impedance is zero. Although very high gains (of order 10^5 to 10^6) are achievable in the kHz range, at the frequencies relevant for detector signals the gain is much smaller. The input impedance of typical charge-sensitive amplifiers in strip detector systems is of order kΩ. Fast amplifiers designed to optimize power dissipation achieve input impedances of 100 to 500 Ω [5]. None of these qualify as a "virtual ground", so this concept should be applied with caution.

Apart from determining the signal rise time, the input impedance is critical in position-sensitive detectors. Figure 14 illustrates a silicon-strip sensor read out by a bank of amplifiers. Each strip electrode has a capacitance C_b to the backplane and a fringing capacitance C_{ss} to the neighboring strips. If the amplifier has an infinite input impedance, charge induced on one strip will capacitively couple to the neighbors and the signal will be distributed over many strips (determined by C_{ss}/C_b). If, on the other hand, the input impedance of the amplifier is low compared to the interstrip impedance, practically all of the charge will flow into the amplifier, as current seeks the path of least impedance, and the neighbors will show only a small signal.

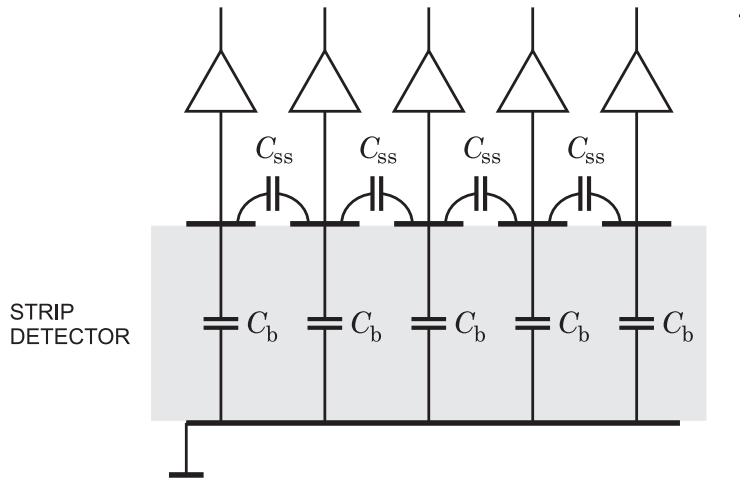


Figure 14: To preserve the position resolution of strip detectors the readout amplifiers must have a low input impedance to prevent spreading of signal charge to the neighboring electrodes.

5 Signal Processing

As noted in the introduction, one of the purposes of signal processing is to improve the signal-to-noise ratio by tailoring the spectral distributions of the signal and the electronic noise. However, for many detectors electronic noise does not determine the resolution. For example, in a NaI(Tl) scintillation detector measuring 511 keV gamma rays, say in a positron-emission tomography system, 25 000 scintillation photons are produced. Because of reflective losses, about 15 000 reach the photocathode. This translates to about 3000 electrons reaching the first dynode. The gain of the electron multiplier will yield about $3 \cdot 10^9$ electrons at the anode. The statistical spread of the signal is determined by the smallest number of electrons in the chain, *i.e.* the 3000 electrons reaching the first dynode, so the resolution $\Delta E/E = 1/\sqrt{3000} = 2\%$, which at the anode corresponds to about $5 \cdot 10^4$ electrons. This is much larger than the electronic noise in any reasonably designed system. This situation is illustrated in Figure 15 (top). In this case, signal acquisition and count rate capability may be the prime objectives of the pulse processing system. The bottom illustration in Figure 15 shows the situation for high resolution sensors with small signals, for example semiconductor detectors, photodiodes or ionization chambers. In this

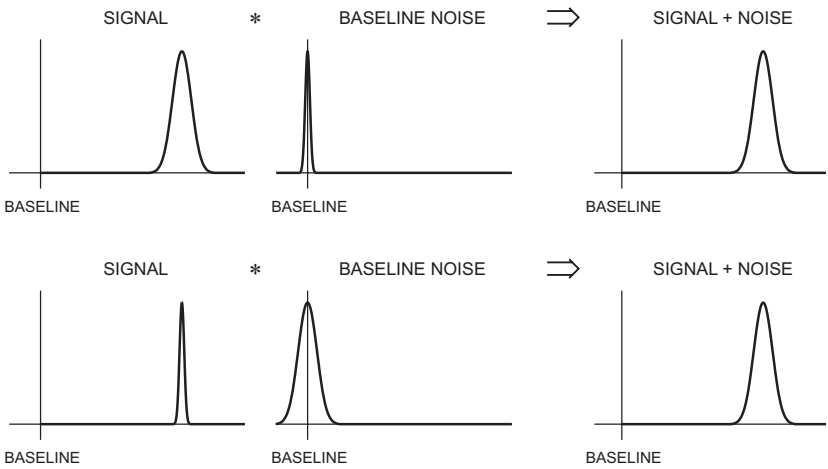


Figure 15: Signal and baseline fluctuations add in quadrature. For large signal variance (top) as in scintillation detectors or proportional chambers, the baseline noise is usually negligible, whereas for small signal variance as in semiconductor detectors or liquid Ar ionization chambers, baseline noise is critical.

case, low noise is critical. Baseline fluctuations can have many origins, external interference, artifacts due to imperfect electronics, etc., but the fundamental limit is electronic noise.

6 Electronic Noise

Consider a current flowing through a sample bounded by two electrodes, *i.e.* n electrons moving with velocity v. The induced current depends on the spacing l between the electrodes (following "Ramo's theorem"[6], [1]), so

$$i = \frac{nev}{l} \ . \tag{15}$$

The fluctuation of this current is given by the total differential

$$\langle di \rangle^2 = \left(\frac{ne}{l} \langle dv \rangle \right)^2 + \left(\frac{ev}{l} \langle dn \rangle \right)^2 \ , \tag{16}$$

where the two terms add in quadrature, as they are statistically uncorrelated. From this one sees that two mechanisms contribute to the total noise, velocity and number fluctuations.

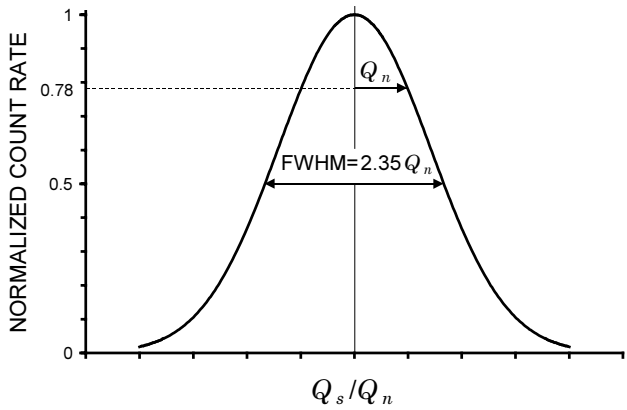


Figure 16: Repetitive measurements of the signal charge yield a Gaussian distribution whose standard deviation equals the rms noise level Q_n. Often the width is expressed as the full width at half maximum (FWHM), which is 2.35 times the standard deviation.

Velocity fluctuations originate from thermal motion. Superimposed on the average drift velocity are random velocity fluctuations due to thermal excitations. This "thermal noise" is described by the long wavelength limit of Planck's black body spectrum where the spectral density, *i.e.* the power per unit bandwidth, is constant ("white" noise).

Number fluctuations occur in many circumstances. One source is carrier flow that is limited by emission over a potential barrier. Examples are thermionic emission or current flow in a semiconductor diode. The probability of a carrier crossing the barrier is independent of any other carrier being emitted, so the individual emissions are random and not correlated. This is called "shot noise", which also has a "white" spectrum. Another source of number fluctuations is carrier trapping. Imperfections in a crystal lattice or impurities in gases can trap charge carriers and release them after a characteristic lifetime. This leads to a frequency-dependent spectrum $dP_n/df = 1/f^\alpha$, where α is typically in the range of 0.5 – 2. Simple derivations of the spectral noise densities are given in [1].

The amplitude distribution of the overall noise is Gaussian, so superimposing a constant amplitude signal on a noisy baseline will yield a Gaussian amplitude distribution whose width equals the noise level (Figure 16). Injecting a pulser signal and measuring the width of the amplitude distribution yields the noise level.

6.1 Thermal (Johnson) noise

The most common example of noise due to velocity fluctuations is the noise of resistors. The spectral noise power density *vs.* frequency

$$\frac{dP_n}{df} = 4kT \ , \tag{17}$$

where k is the Boltzmann constant and T the absolute temperature. Since the power in a resistance R can be expressed through either voltage or current,

$$P = \frac{V^2}{R} = I^2 R \ , \tag{18}$$

the spectral voltage and current noise densities

$$\frac{dV_n^2}{df} \equiv e_n^2 = 4kTR \qquad \text{and} \qquad \frac{dI_n^2}{df} \equiv i_n^2 = \frac{4kT}{R} \ . \tag{19}$$

The total noise is obtained by integrating over the relevant frequency range of the system, the bandwidth, so the total noise voltage at the output of an amplifier with a frequency-dependent gain $A(f)$ is

$$v_{on}^2 = \int_0^\infty e_n^2 A^2(f) df \ . \tag{20}$$

Since the spectral noise components are non-correlated (each black body excitation mode is independent), one must integrate over the noise power, *i.e.* the voltage squared. The total noise increases with bandwidth. Since small bandwidth corresponds to large rise-times, increasing the speed of a pulse measurement system will increase the noise.

6.2 Shot noise

The spectral density of shot noise is proportional to the average current I:

$$i_n^2 = 2eI \ , \tag{21}$$

where e is the electronic charge. Note that the criterion for shot noise is that carriers are injected independently of one another, as in thermionic emission or semiconductor diodes. Current flowing through an ohmic conductor does not carry shot noise, since the fields set up by any local fluctuation in charge density can easily draw in additional carriers to equalize the disturbance.

7 Signal-to-noise Ratio *vs.* Sensor Capacitance

The basic noise sources manifest themselves as either voltage or current fluctuations. However, the desired signal is a charge, so to allow a comparison we must express the signal as a voltage or current. This was illustrated for an ionization chamber in Figure 8. As was noted, when the input time constant $R_i(C_d + C_i)$ is large compared to the duration of the sensor current pulse, the signal charge is integrated on the input capacitance, yielding the signal voltage $V_s = Q_s/(C_d + C_i)$. Assume that the amplifier has an input noise voltage V_n. Then the signal-to-noise ratio

$$\frac{V_s}{V_n} = \frac{Q_s}{V_n(C_d + C_i)} \ . \tag{22}$$

This is a very important result – the signal-to-noise ratio for a given signal charge is inversely proportional to the total capacitance at the input node. Note that zero input capacitance does not yield an infinite signal-to-noise ratio. As shown in ref. [1], this relationship only holds when the input time constant is greater than about ten times the sensor current pulse width. The dependence of signal-to-noise ratio on capacitance is a general feature that is independent of amplifier type. Since feedback cannot improve signal-to-noise ratio, eqn 22 holds for charge-sensitive amplifiers, although in that configuration the charge signal is constant, but the noise increases with total input capacitance (see[1]). In the noise analysis the feedback capacitance adds to the total input capacitance (the passive capacitance, not the dynamic input capacitance), so C_f should be kept small.

8 Pulse Shaping

Pulse shaping has two conflicting objectives. The first is to limit the bandwidth to match the measurement time. Too large a bandwidth will increase the noise without increasing the signal. Typically, the pulse shaper transforms a narrow sensor pulse into a broader pulse with a gradually rounded maximum at the peaking time. This is illustrated in Figure 17. The signal amplitude is measured at the peaking time T_P.

The second objective is to constrain the pulse width so that successive signal pulses can be measured without overlap (pileup), as illustrated in Figure 18. Reducing the pulse duration increases the allowable signal rate, but at the expense of electronic noise.

In designing the shaper it is necessary to balance these conflicting goals. Usually, many different considerations lead to a "non-textbook" compromise;

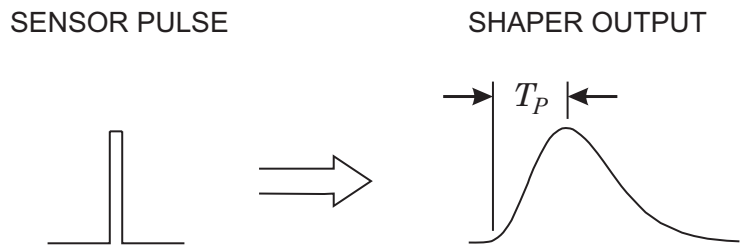


Figure 17: In energy measurements a pulse processor typically transforms a short sensor current pulse to a broader pulse with a peaking time T_P.

"optimum shaping" depends on the application.

A simple shaper is shown in Figure 19. A high-pass filter sets the duration of the pulse by introducing a decay time constant τ_d. Next a low-pass filter with a time constant τ_i increases the rise time to limit the noise bandwidth. The high-pass is often referred to as a "differentiator", since for short pulses it forms the derivative. Correspondingly, the low-pass is called an "integrator". Since the high-pass filter is implemented with a CR section and the low-pass with an RC, this shaper is referred to as a CR-RC shaper. Although pulse shapers are often more sophisticated and complicated, the CR-RC shaper contains the essential features of all pulse shapers, a lower frequency bound and an upper frequency bound.

After peaking the output of a simple CR-RC shaper returns to baseline rather slowly. The pulse can be made more symmetrical, allowing higher signal rates for the same peaking time. Very sophisticated circuits have been developed towards this goal, but a conceptually simple way is to use multiple integrators, as

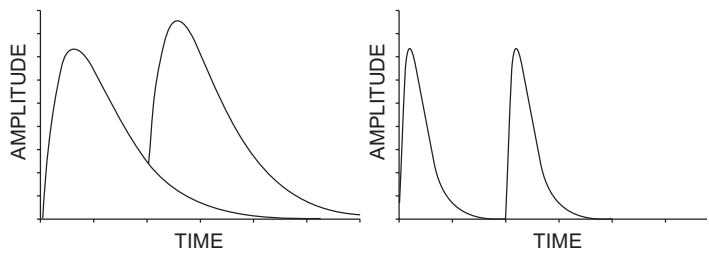


Figure 18: Amplitude pileup occurs when two pulses overlap (left). Reducing the shaping time allows the first pulse to return to the baseline before the second pulse arrives.

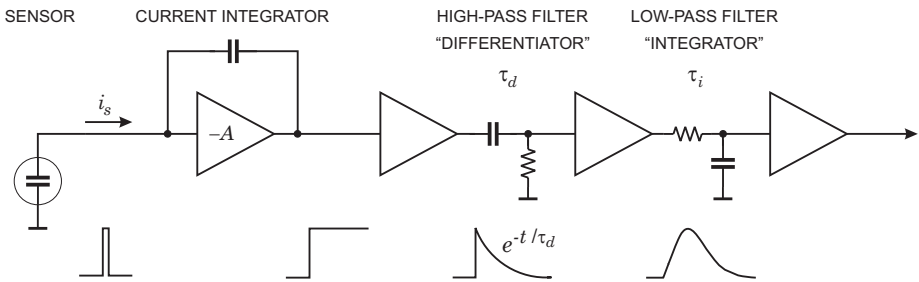


Figure 19: Components of a pulse shaping system. The signal current from the sensor is integrated to form a step impulse with a long decay. A subsequent high-pass filter ("differentiator") limits the pulse width and the low-pass filter ("integrator") increases the rise-time to form a pulse with a smooth cusp.

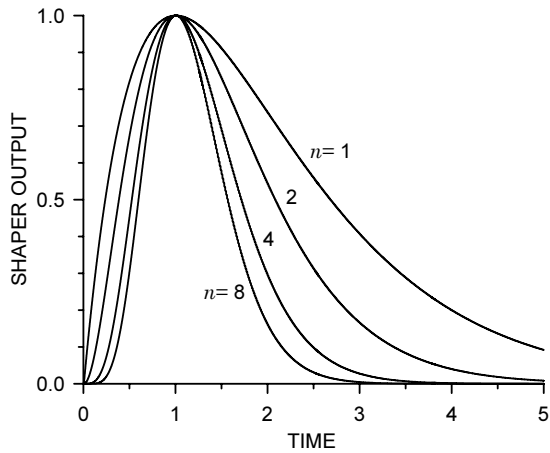


Figure 20: Pulse shape *vs.* number of integrators in a CR-nRC shaper. The time constants are scaled with the number of integrators to maintain the peaking time.

illustrated in Figure 20. The integration and differentiation time constants are scaled to maintain the peaking time. Note that the peaking time is a key design parameter, as it dominates the noise bandwidth and must also accommodate the sensor response time.

Another type of shaper is the correlated double sampler, illustrated in Figure 21. This type of shaper is widely used in monolithically integrated circuits, as many CMOS processes (see Section 11.1) provide only capacitors

and switches, but no resistors. Input signals are superimposed on a slowly fluctuating baseline. To remove the baseline fluctuations the baseline is sampled prior to the signal. Next, the signal plus baseline is sampled and the previous baseline sample subtracted to obtain the signal. The prefilter is critical to limit the noise bandwidth of the system. Filtering after the sampler is useless, as noise fluctuations on time scales shorter than the sample time will not be removed. Here the sequence of filtering is critical, unlike a time-invariant linear filter (*e.g.* a *CR-RC* filter as in Figure 19) where the sequence of filter functions can be interchanged.

This is an example of a time-variant filter. The *CR-nRC* filter described above acts continuously on the signal, whereas the correlated double sample changes filter parameters *vs.* time.

9 Noise Analysis of a Detector and Front-end Amplifier

To determine how the pulse shaper affects the signal-to-noise ratio consider the detector front-end in Figure 22. The detector is represented by the capacitance C_d, a relevant model for many radiation sensors. Sensor bias voltage is applied through the resistor R_b. The bypass capacitor C_b shunts any external interference coming through the bias supply line to ground. For high-frequency signals this capacitor appears as a low impedance, so for sensor signals the "far end" of the bias resistor is connected to ground. The coupling capacitor C_c blocks the sensor bias voltage from the amplifier input, which is why a capaci-

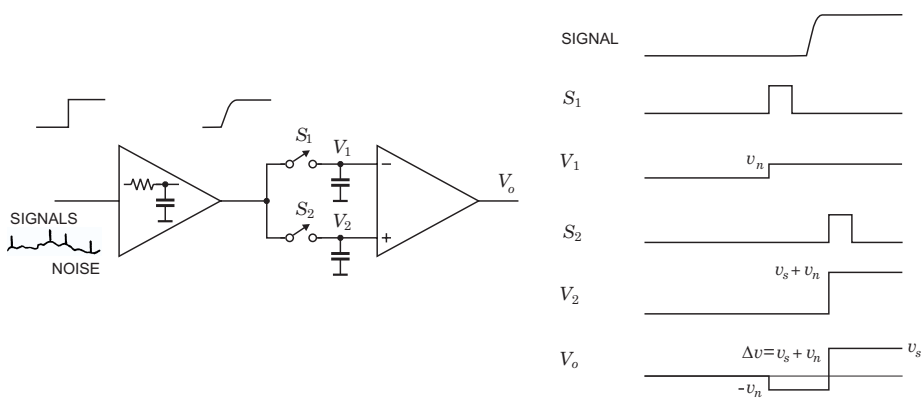


Figure 21: Principle of a shaper using correlated double sampling.

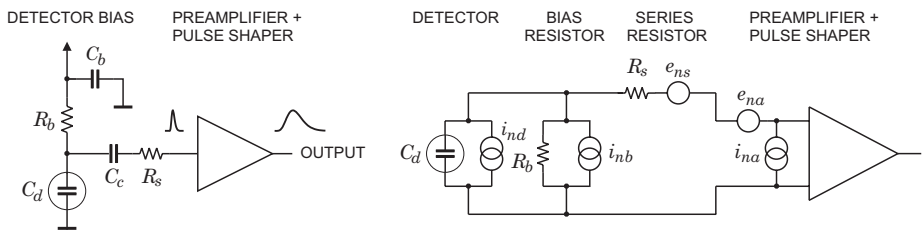


Figure 22: A detector front-end circuit and its equivalent circuit for noise calculations.

tor serving this role is also called a "blocking capacitor". The series resistor R_s represents any resistance present in the connection from the sensor to the amplifier input. This includes the resistance of the sensor electrodes, the resistance of the connecting wires or traces, any resistance used to protect the amplifier against large voltage transients ("input protection"), and parasitic resistances in the input transistor.

The following implicitly includes a constraint on the bias resistance, whose role is often misunderstood. It is often thought that the signal current generated in the sensor flows through R_b and the resulting voltage drop is measured. If the time constant R_bC_d is small compared to the peaking time of the shaper T_P, the sensor will have discharged through R_b and much of the signal will be lost. Thus, we have the condition $R_bC_d \gg T_P$, or $R_b \gg T_P/C_d$. The bias resistor must be sufficiently large to block the flow of signal charge, so that all of the signal is available for the amplifier.

To analyze this circuit we'll assume a voltage amplifier, so all noise contributions will be calculated as a noise voltage appearing at the amplifier input. Steps in the analysis are 1. determine the frequency distribution of all noise voltages presented to the amplifier input from all individual noise sources, 2. integrate over the frequency response of the shaper (for simplicity a CR-RC shaper) and determine the total noise voltage at the shaper output, and 3. determine the output signal for a known input signal charge. The equivalent noise charge (ENC) is the signal charge for which $S/N = 1$.

The equivalent circuit for the noise analysis (second panel of Figure 22) includes both current and voltage noise sources. The "shot noise" i_{nd} of the sensor leakage current is represented by a current noise generator in parallel with the sensor capacitance. As noted above, resistors can be modeled either as a voltage or current generator. Generally, resistors shunting the input act as noise current sources and resistors in series with the input act as noise voltage

sources (which is why some in the detector community refer to current and voltage noise as "parallel" and "series" noise). Since the bias resistor effectively shunts the input, as the capacitor C_b passes current fluctuations to ground, it acts as a current generator i_{nb} and its noise current has the same effect as the shot noise current from the detector. The shunt resistor can also be modeled as a noise voltage source, yielding the result that it acts as a current source. Choosing the appropriate model merely simplifies the calculation. Any other shunt resistances can be incorporated in the same way. Conversely, the series resistor R_s acts as a voltage generator. The electronic noise of the amplifier is described fully by a combination of voltage and current sources at its input, shown as e_{na} and i_{na}.

Thus, the noise sources are

$$\begin{aligned} \text{sensor bias current}: \quad & i_{nd}^2 = 2eI_d \\ \text{shunt resistance}: \quad & i_{nb}^2 = \frac{4kT}{R_b} \\ \text{series resistance}: \quad & e_{ns}^2 = 4kTR_s \\ \text{amplifier}: \quad & e_{na}, \quad i_{na}, \end{aligned}$$

where e is the electronic charge, I_d the sensor bias current, k the Boltzmann constant and T the temperature. Typical amplifier noise parameters e_{na} and i_{na} are of order $\mathrm{nV}/\sqrt{\mathrm{Hz}}$ and $\mathrm{fA}/\sqrt{\mathrm{Hz}}$ (FETs) – $\mathrm{pA}/\sqrt{\mathrm{Hz}}$ (bipolar transistors). Amplifiers tend to exhibit a "white" noise spectrum at high frequencies (greater than order kHz), but at low frequencies show excess noise components with the spectral density

$$e_{nf}^2 = \frac{A_f}{f} \ , \tag{23}$$

where the noise coefficient A_f is device specific and of order 10^{-10} – $10^{-12}\,\mathrm{V}^2$.

The noise voltage generators are in series and simply add in quadrature. White noise distributions remain white. However, a portion of the noise currents flows through the detector capacitance, resulting in a frequency-dependent noise voltage $i_n/(\omega C_d)$, so the originally white spectrum of the sensor shot noise and the bias resistor now acquires a $1/f$ dependence. The frequency distribution of all noise sources is further altered by the combined frequency response of the amplifier chain $A(f)$. Integrating over the cumulative noise spectrum at the amplifier output and comparing to the output voltage for a known input signal yields the signal-to-noise ratio. In this example the shaper is a simple CR-RC shaper, where for a given differentiation time constant the noise is minimized when the differentiation and integration time constants are equal $\tau_i = \tau_d \equiv \tau$. Then the output pulse assumes its maximum amplitude at the time $T_P = \tau$.

Although the basic noise sources are currents or voltages, since radiation detectors are typically used to measure charge, the system's noise level is conveniently expressed as an equivalent noise charge Q_n. As noted previously, this is equal to the detector signal that yields a signal-to-noise ratio of one. The equivalent noise charge is commonly expressed in Coulombs, the corresponding number of electrons, or the equivalent deposited energy (eV). For the above circuit the equivalent noise charge

$$Q_n^2 = \left(\frac{\mathrm{e}^2}{8}\right)\left[\left(2eI_d + \frac{4kT}{R_b} + i_{na}^2\right)\cdot\tau + \left(4kTR_s + e_{na}^2\right)\cdot\frac{C_d^2}{\tau} + 4A_fC_d^2\right] . \tag{24}$$

The prefactor $\mathrm{e}^2/8 = \exp(2)/8 = 0.924$ normalizes the noise to the signal gain. The first term combines all noise current sources and increases with shaping time. The second term combines all noise voltage sources and decreases with shaping time, but increases with sensor capacitance. The third term is the contribution of amplifier $1/f$ noise and, as a voltage source, also increases with sensor capacitance. The $1/f$ term is independent of shaping time, since for a $1/f$ spectrum the total noise depends on the ratio of upper to lower cutoff frequency, which depends only on shaper topology, but not on the shaping time.

Just as filter response can be described either in the frequency or time domain, so can the noise performance. Detailed explanations are given in papers by Goulding and Radeka[7] [8] [9] [10]. The key is Parseval's theorem, which relates the amplitude response $A(f)$ to the time response $F(t)$.

$$\int_0^\infty |A(f)|^2\, df = \int_{-\infty}^{\infty} [F(t)]^2\, dt\ . \tag{25}$$

The left hand side is essentially integration over the noise bandwidth. The output noise power scales linearly with the duration of the pulse, so the noise contribution of the shaper can be split into a factor that is determined by the shape of the response and a time factor that sets the shaping time. This leads to a general formulation of the equivalent noise charge

$$Q_n^2 = i_n^2F_iT_S + e_n^2F_v\frac{C^2}{T_S} + F_{vf}A_fC^2\ , \tag{26}$$

where F_i, F_v, and F_{vf} depend on the shape of the pulse determined by the shaper and T_S is a characteristic time, for example the peaking time of a CR-nRC shaped pulse or the prefilter time constant in a correlated double sampler[1]. As before, C is the total parallel capacitance at the input. The

shape factors F_i, F_v are easily calculated:

$$F_i = \frac{1}{2T_S}\int_{-\infty}^{\infty}\left[W(t)\right]^2 dt\ , \qquad F_v = \frac{T_S}{2}\int_{-\infty}^{\infty}\left[\frac{dW(t)}{dt}\right]^2 dt\ . \tag{27}$$

For time-invariant pulse shaping $W(t)$ is simply the system's impulse response (the output signal seen on an oscilloscope) with the peak output signal normalized to unity. For a time-variant shaper the same equations apply, but $W(t)$ is determined differently. See refs. [7], [8], [9], and [10] for more details.

A shaper formed by a single CR differentiator and RC integrator with equal time constants has $F_i = F_v = 0.9$ and $F_{vf} = 4$, independent of the shaping time constant, so for the circuit in Figure 19 eqn 24 becomes

$$Q_n^2 = \left(2q_eI_d + \frac{4kT}{R_b} + i_{na}^2\right)F_iT_S + \left(4kTR_s + e_{na}^2\right)F_v\frac{C^2}{T_S} + F_{vf}A_fC^2\ . \tag{28}$$

Pulse shapers can be designed to reduce the effect of current noise, *e.g.* mitigate radiation damage. Increasing pulse symmetry tends to decrease F_i and increase F_v, *e.g.* to $F_i = 0.45$ and $F_v = 1.0$ for a shaper with one CR differentiator and four cascaded RC integrators.

Figure 23 shows how equivalent noise charge is affected by shaping time. At short shaping times the voltage noise dominates, whereas at long shaping times the current noise takes over. Minimum noise obtains where the current and voltage contributions are equal. The noise minimum is flattened by the presence of $1/f$ noise. Also shown is that increasing the detector capacitance will increase the voltage noise contribution and shift the noise minimum to longer shaping times, albeit with an increase in minimum noise.

For quick estimates one can use the following equation, which assumes an FET amplifier (negligible i_{na}) and a simple CR-RC shaper with peaking time τ. The noise is expressed in units of the electronic charge e and C is the total parallel capacitance at the input, including C_d, all stray capacitances, and the amplifier's input capacitance.

$$Q_n^2 = 12\left[\frac{e^2}{\mathrm{nA\cdot ns}}\right]I_d\tau + 6\cdot 10^5\left[\frac{e^2\mathrm{k\Omega}}{\mathrm{ns}}\right]\frac{\tau}{R_b} + 3.6\cdot 10^4\left[\frac{e^2\mathrm{ns}}{(\mathrm{pF})^2(\mathrm{nV})^2/\mathrm{Hz}}\right]e_n^2\frac{C^2}{\tau} \tag{29}$$

The noise charge is improved by reducing the detector capacitance and leakage current, judiciously selecting all resistances in the input circuit, and

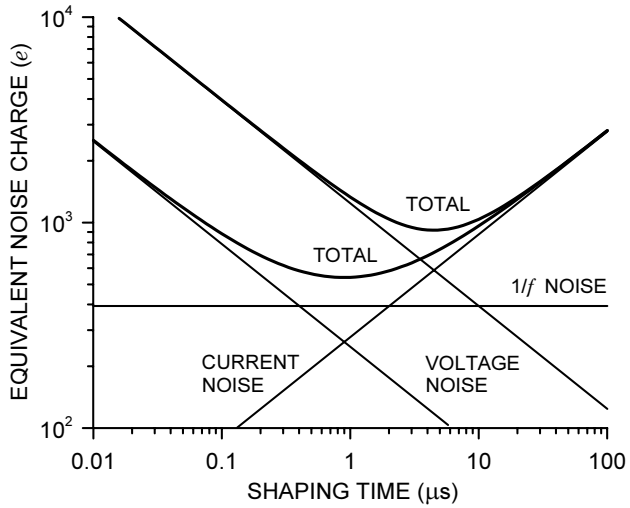


Figure 23: Equivalent noise charge *vs.* shaping time. At small shaping times (large bandwidth) the equivalent noise charge is dominated by voltage noise, whereas at long shaping times (large integration times) the current noise contributions dominate. The total noise assumes a minimum where the current and voltage contributions are equal. The "$1/f$" noise contribution is independent of shaping time and flattens the noise minimum. Changing the voltage or current noise contribution shifts the noise minimum. Increased voltage noise is shown as an example.

choosing the optimum shaping time constant. The noise parameters of a well-designed amplifier depend primarily on the input device. Fast, high-gain transistors are generally best.

In field effect transistors, both junction field effect transistors (JFETs) or metal oxide semiconductor field effect transistors (MOSFETs), the noise current contribution is very small, so reducing the detector leakage current and increasing the bias resistance will allow long shaping times with correspondingly lower noise. The equivalent input noise voltage $e_n^2 \approx 4kT/g_m$, where g_m is the transconductance, which increases with operating current. For a given current, the transconductance increases when the channel length is reduced, so reductions in feature size with new process technologies are beneficial. At a given channel length minimum noise obtains when a device is operated at maximum transconductance. If lower noise is required, the width of the device can be increased (equivalent to connecting multiple devices in parallel). This increases the transconductance (and required current) with a corresponding de-

crease in noise voltage, but also increases the input capacitance. At some point the reduction in noise voltage is outweighed by the increase in total input capacitance. The optimum obtains when the FET's input capacitance equals the external capacitance (sensor + stray capacitance). Note that this capacitive matching criterion only applies when the input current noise contribution of the amplifying device is negligible.

Capacitive matching comes at the expense of power dissipation. Since the minimum is shallow, one can operate at significantly lower currents with just a minor increase in noise. In large detector arrays power dissipation is critical, so FETs are hardly ever operated at their minimum noise. Instead, one seeks an acceptable compromise between noise and power dissipation (see [1] for a detailed discussion). Similarly, the choice of input devices is frequently driven by available fabrication processes. High-density integrated circuits tend to include only MOSFETs, so this determines the input device, even where a bipolar transistor would provide better performance.

In bipolar transistors the shot noise associated with the base current I_B is significant, $i_{nB}^2 = 2eI_B$. Since $I_B = I_C/\beta_{DC}$, where I_C is the collector current and β_{DC} the direct current gain, this contribution increases with device current. On the other hand, the equivalent input noise voltage

$$e_n^2 = \frac{2(kT)^2}{eI_C} \tag{30}$$

decreases with collector current, so the noise assumes a minimum at a specific collector current

$$Q_{n,min}^2 = 4kT\frac{C}{\sqrt{\beta_{DC}}}\sqrt{F_i F_v} \quad \text{at} \quad I_C = \frac{kT}{e}C\sqrt{\beta_{DC}}\sqrt{\frac{F_v}{F_i}}\frac{1}{T_S} \; . \tag{31}$$

For a CR-RC shaper and $\beta_{DC} = 100$,

$$Q_{n,min} \approx 250\left[\frac{e}{\sqrt{\text{pF}}}\right]\cdot\sqrt{C} \quad \text{at} \quad I_C = 260\left[\frac{\mu\text{A}\cdot\text{ns}}{\text{pF}}\right]\cdot\frac{C}{T_S} \; . \tag{32}$$

The minimum obtainable noise is independent of shaping time (unlike FETs), but only at the optimum collector current I_C, which does depend on shaping time.

In bipolar transistors the input capacitance is usually much smaller than the sensor capacitance (of order 1 pF for $e_n \approx 1\,\text{nV}/\sqrt{\text{Hz}}$) and substantially smaller than in FETs with comparable noise. Since the transistor input capacitance enters into the total input capacitance, this is an advantage. Note that capacitive matching does not apply to bipolar transistors, because their

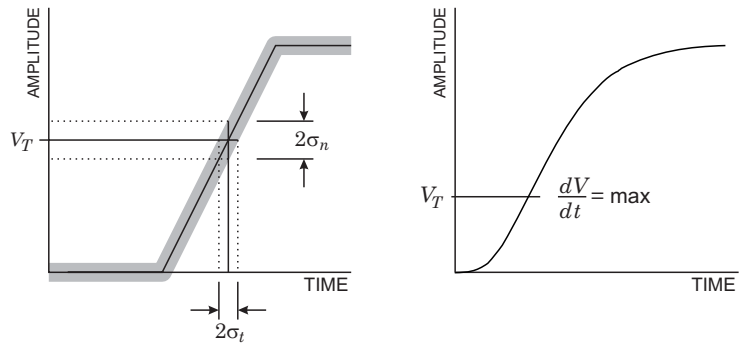


Figure 24: Fluctuations in signal amplitude crossing a threshold translate into timing fluctuations (left). With realistic pulses the slope changes with amplitude, so minimum timing jitter occurs with the trigger level at the maximum slope.

noise current contribution is significant. Due to the base current noise bipolar transistors are best at short shaping times, where they also require lower power than FETs for a given noise level.

When the input noise current is negligible, the noise increases linearly with sensor capacitance. The noise slope

$$\frac{dQ_n}{dC_d} \approx 2e_n \cdot \sqrt{\frac{F_v}{T}} \tag{33}$$

depends both on the preamplifier (e_n) and the shaper (F_v, T). The zero intercept can be used to determine the amplifier input capacitance plus any additional capacitance at the input node.

Practical noise levels range from $< 1\,e$ for CCDs at long shaping times to $\sim 10^4\,e$ in high-capacitance liquid Ar calorimeters. Silicon strip detectors typically operate at $\sim 10^3$ electrons, whereas pixel detectors with fast readout provide noise of 100 – 200 electrons. Transistor noise is discussed in more detail in[1].

10 Timing Measurements

Pulse height measurements discussed up to now emphasize measurement of signal charge. Timing measurements seek to optimize the determination of the time of occurrence. Although, as in amplitude measurements, signal-to-noise ratio is important, the determining parameter is not signal-to-noise, but slope-to-noise ratio. This is illustrated in Figure 24, which shows the leading

edge of a pulse fed into a threshold discriminator (comparator), a "leading edge trigger". The instantaneous signal level is modulated by noise, where the variations are indicated by the shaded band. Because of these fluctuations, the time of threshold crossing fluctuates. By simple geometrical projection, the timing variance, or "jitter"

$$\sigma_t = \frac{\sigma_n}{(dS/dt)_{S_T}} \approx \frac{t_r}{S/N} \ , \tag{34}$$

where σ_n is the rms noise and the derivative of the signal dS/dt is evaluated at the trigger level S_T. To increase dS/dt without incurring excessive noise the amplifier bandwidth should match the rise-time of the detector signal.The 10 – 90% rise time of an amplifier with bandwidth f_u (see Figure 13) is

$$t_r = 2.2\tau = \frac{2.2}{2\pi f_u} = \frac{0.35}{f_u} \ . \tag{35}$$

For example, an oscilloscope with 350 MHz bandwidth has a 1 ns rise time. When amplifiers are cascaded, which is invariably necessary, the individual rise times add in quadrature

$$t_r \approx \sqrt{t_{r1}^2 + t_{r2}^2 + ... + t_{rn}^2} \ . \tag{36}$$

Increasing signal-to-noise ratio improves time resolution, so minimizing the total capacitance at the input is also important. At high signal-to-noise ratios the time jitter can be much smaller than the rise time.

The second contribution is time walk, where the timing signal shifts with amplitude as shown in Figure 25. This can be corrected by various means, either in hardware or software. For a more detailed tutorial on timing measurements see ref. [11].

11 Digital Electronics

Analog signals utilize continuously variable properties of the pulse to impart information, such as the pulse amplitude or pulse shape. Digital signals have constant amplitude, but the presence of the signal at specific times is evaluated, *i.e.* whether the signal is in one of two states, "low" or "high". However this still involves an analog process, as the presence of a signal is determined by the signal level exceeding a threshold at the proper time.

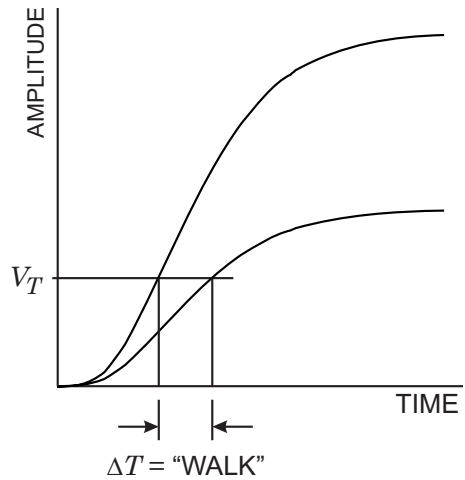


Figure 25: The time at which a signal crosses a fixed threshold depends on the signal amplitude, leading to "time walk".

11.1 Logic elements

Figure 26 illustrates several functions utilized in digital circuits ("logic" functions). An AND gate provides an output only when all inputs are high. An OR gives an output when any input is high. An eXclusive OR (XOR) responds when only one input is high. The same elements are commonly implemented with inverted outputs, then called NAND and NOR gates, for example. The

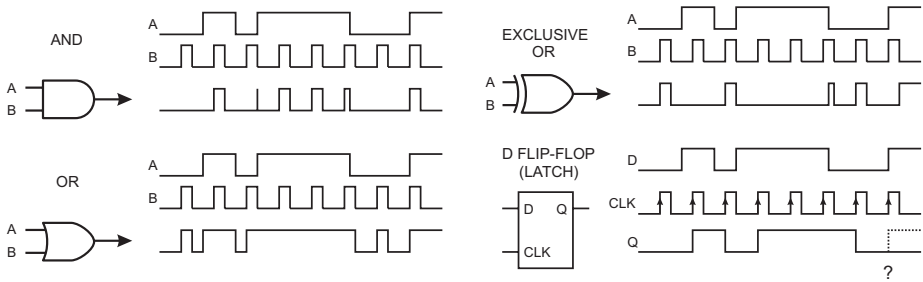


Figure 26: Basic logic functions include gates (AND, OR, Exclusive OR) and flip flops. The outputs of the AND and D flip flop show how small shifts in relative timing between inputs can determine the output state.

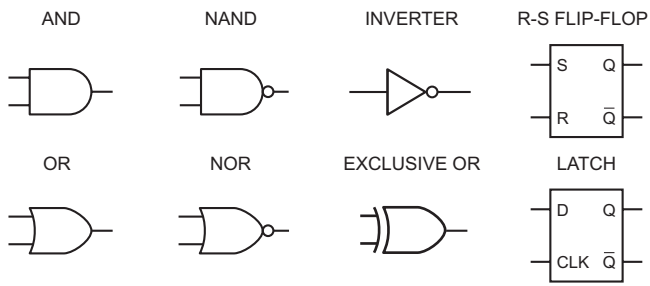


Figure 27: Some common logic symbols. Inverted outputs are denoted by small circles or by a superimposed bar, as for the latch output $\overline{Q}$. Additional inputs can be added to gates as needed. An R-S flip-flop sets the Q output high in response to an S input. An R input resets the Q output to low.

D flip-flop is a bistable memory circuit that records the presence of a signal at the data input D when a signal transition occurs at the clock input CLK. This device is commonly called a latch. Inverted inputs and outputs are denoted by small circles or by superimposed bars, *e.g.* $\overline{Q}$ is the inverted output of a flip flop, as shown in Figure 27.

Logic circuits are fundamentally amplifiers, so they also suffer from bandwidth limitations. The pulse train of the AND gate in Figure 26 illustrates a common problem. The third pulse of input B is going low at the same time that input A is going high. Depending on the time overlap, this can yield a narrow output that may or may not be recognized by the following circuit. In an EX-OR this can occur when two pulses arrive nearly at the same time. The D flip-flop requires a minimum setup time for a level change at the D input to be recognized, so changes in the data level may not be recognized at the correct time. These marginal events may be extremely rare and perhaps go unnoticed. However, in complex systems the combination of "glitches" can make the system "hang up", necessitating a system reset. Data transmission protocols have been developed to detect such errors (parity checks, Hamming codes, *etc.*), so corrupted data can be rejected.

Some key aspects of logic systems can be understood by inspecting the circuit elements that are used to form logic functions. Figure 28 shows simple inverter circuits using MOS transistors. For this discussion it is sufficient to know that in an NMOS transistor a conductive channel is formed when the input electrode is biased positive with respect to the channel. The input, called the "Gate" (G), is capacitively coupled to the output channel connected between the "Drain" (D) and "Source" (S) electrodes. In the NMOS inverter applying a positive voltage to the gate makes the output channel conduct, so the output

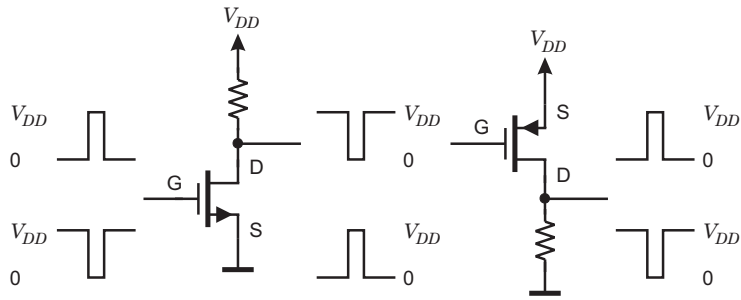


Figure 28: In an NMOS inverter the transistor conducts when the input is high (left), whereas in a PMOS inverter the transistor conducts when the input is low (right). In both circuits the input pulse is inverted, whether the input swings high or low.

level is low. A PMOS transistor is the complementary device, where a conductive channel is formed when the gate is biased negative with respect to the source. Since the source is at positive potential, a low level at the inverter input yields a high level at the output. Regardless of the device and pulse polarity, the output pulse is always the inverse of the input. NMOS and PMOS inverters draw current when in their "active" state. Combining NMOS and PMOS transistors in a complementary MOS (CMOS) circuit allows zero current draw in both the high and low states with a substantial reduction in power consumption. A CMOS inverter is shown in Figure 29, which also shows how devices are combined to form a CMOS NAND gate. In the inverter the lower (NMOS) transistor is turned off when the input is low, but the upper (PMOS) transistor is turned on, so the output is connected to V_{DD}, taking the output high. Since the current path from V_{DD} to ground is blocked by either the NMOS or PMOS device being off, the power dissipation is zero in both the high and low states. Current only flows during the level transition when both devices are on as the input level is at approximately $V_{DD}/2$. As a result, the power dissipation of CMOS logic is significantly less than in NMOS or PMOS circuitry. This reduction in power only obtains in logic circuitry. CMOS analog amplifiers are not fundamentally more power efficient than NMOS or PMOS circuits, although CMOS provides greater flexibility in the choice of circuit topologies, which can reduce overall power.

11.2 Propagation delays and power dissipation

Logic elements always operate in conjunction with other circuits, as illustrated in Figure 30. The wiring resistance in conjunction with the total

load capacitance increases the rise time of the logic pulse and as a result delays the time when the transition crosses the logic threshold. The energy dissipated in the wiring resistance R is

$$E = \int i^2(t) R \, dt \ . \tag{37}$$

The current flow during one transition

$$i(t) = \frac{V}{R} \exp\left(-\frac{t}{RC}\right) \ , \tag{38}$$

so the dissipated energy per transition (either positive or negative)

$$E = \frac{V^2}{R} \int_0^\infty \exp\left(-\frac{2t}{RC}\right) dt = \frac{1}{2} C V^2 \ . \tag{39}$$

When pulses occur at a frequency f, the power dissipated in both the positive and negative transitions

$$P = f C V^2 \ . \tag{40}$$

Thus, the power dissipation increases with clock frequency and the square of the logic swing.

Fast logic is time-critical. It relies on logic operations from multiple paths coming together at the right time. Valid results depend on maintaining minimum allowable overlaps and set-up times as illustrated in Figure 26. Each logic

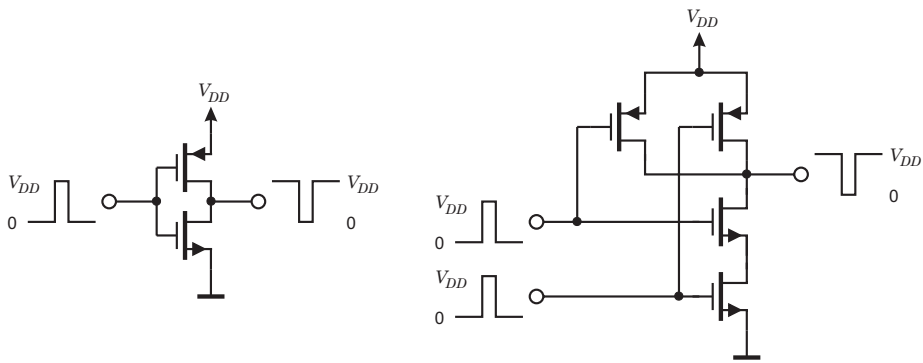


Figure 29: A CMOS inverter (left) and NAND gate (right).

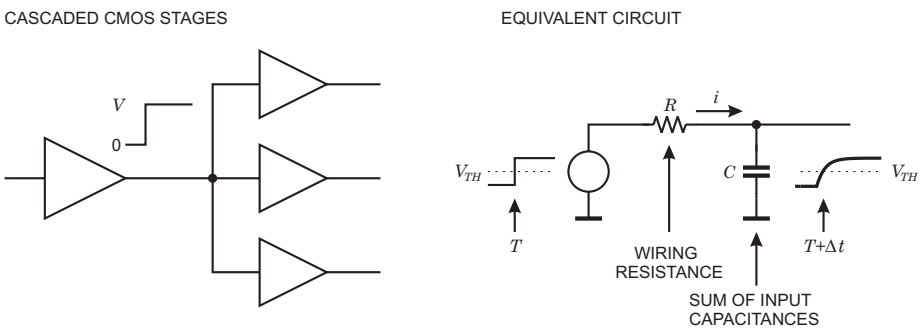


Figure 30: The wiring resistance together with the distributed load capacitance delays the signal.

circuit has a finite propagation delay, which depends on circuit loading, *i.e.* how many loads the circuit has to drive. In addition, as illustrated in Figure 30 the wiring resistance and capacitive loads introduce delay. This depends on the number of circuits connected to a wire or trace, the length of the trace and the dielectric constant of the substrate material. Relying on control of circuit and wiring delays to maintain timing requires great care, as it depends on circuit variations and temperature. In principle all of this can be simulated, but in complex systems there are too many combinations to test every one. A more robust solution is to use synchronous systems, where the timing of all transitions is determined by a master clock. Generally, this does not provide the utmost speed and requires some additional circuitry, but increases reliability. Nevertheless, clever designers frequently utilize asynchronous logic. Sometimes it succeeds . . . and sometimes it doesn't.

11.3 Logic arrays

Commodity integrated circuits with basic logic blocks are readily available, *e.g.* with four NAND gates or two flip-flops in one package. These can be combined to form simple digital systems. However, complex logic systems are no longer designed using individual gates. Instead, logic functions are described in a high-level language (*e.g.* VHDL), synthesized using design libraries, and implemented as custom ICs – "ASICs" (application specific ICs) – or programmable logic arrays. In these implementations the digital circuitry no longer appears as an ensemble of inverters, gates, and flip-flops, but as an integrated logic block that provides specific outputs in response to various input combina-

tions. This is illustrated in Figure 31. Field Programmable Gate or logic Arrays (FPGAs) are a common example. A representative FPGA has 512 pads usable for inputs and outputs, $\sim 10^6$ gates, and $\sim$ 100K of memory. Modern design tools also account for propagation delays, wiring lengths, loads, and temperature dependence. The design software also generates "test vectors" that can be used to test finished parts. Properly implemented, complex digital designs can succeed on the first pass, whether as ASICs or as logic or gate arrays.

12 Analog-to-digital Converters (ADCs)

For data storage and subsequent analysis the analog signal at the shaper output must be digitized. Important parameters for analog-to-digital converters (ADCs or A/Ds) used in detector systems are

1. Resolution: The "granularity" of the digitized output.
2. Differential non-linearity: How uniform are the digitization increments?
3. Integral non-linearity: Is the digital output proportional to the analog input?
4. Conversion time: How much time is required to convert an analog signal to a digital output?
5. Count-rate performance: How quickly can a new conversion commence after completion of a prior one without introducing deleterious artifacts?
6. Stability: Do the conversion parameters change with time?

Instrumentation ADCs used in industrial data acquisition and control systems share most of these requirements. However, detector systems place greater

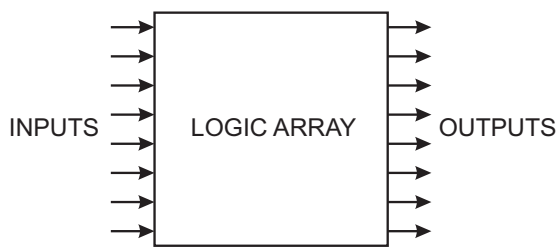


Figure 31: Complex logic circuits are commonly implemented using logic arrays that as an integrated block provide the desired outputs in response to specific input combinations.

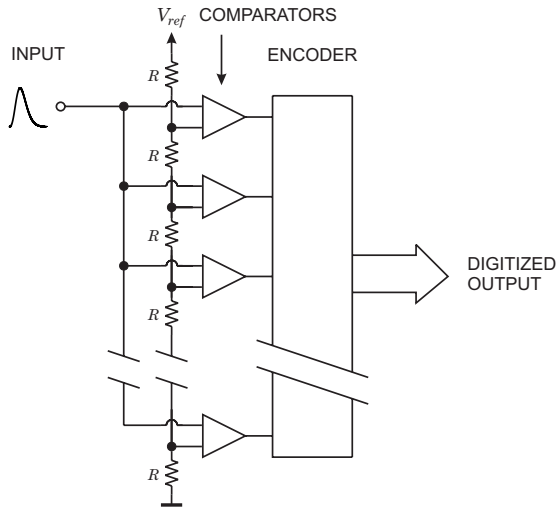


Figure 32: Block diagram of a flash ADC.

emphasis on differential non-linearity and count-rate performance. The latter is important, as detector signals often occur randomly, in contrast to systems where signals are sampled at regular intervals. As in amplifiers, if the DC gain is not precisely equal to the high-frequency gain, the baseline will shift. Furthermore, following each pulse it takes some time for the baseline to return to its quiescent level. For periodic signals of roughly equal amplitude these baseline deviations will be the same for each pulse, but for a random sequence of pulse with varying amplitudes, the instantaneous baseline level will be different for each pulse and affect the peak amplitude.

Conceptually, the simplest technique is flash conversion, illustrated in Figure 32. The signal is fed in parallel to a bank of threshold comparators. The individual threshold levels are set by a resistive divider. The comparator outputs are encoded such that the output of the highest level comparator that fires yields the correct bit pattern. The threshold levels can be set to provide a linear conversion characteristic where each bit corresponds to the same analog increment, or a non-linear characteristic, to provide increments proportional to the absolute level, which provides constant relative resolution over the range.

The big advantage of this scheme is speed; conversion proceeds in one step and conversion times < 10 ns are readily achievable. The drawbacks are component count and power consumption, as one comparator is required per

conversion bin. For example, an 8-bit converter requires 256 comparators. The conversion is always monotonic and differential non-linearity is determined by the matching of the resistors in the threshold divider. Only relative matching is required, so this topology is a good match for monolithic integrated circuits. Flash ADCs are available with conversion rates $> 500\,\mathrm{MS/s}$ (megasamples per second) at 8-bit resolution and a power dissipation of about 5 W.

The most commonly used technique is the successive approximation ADC, shown in Figure 33. The input pulse is sent to a pulse stretcher, which follows the signal until it reaches its cusp and then holds the peak value. The stretcher output feeds a comparator, whose reference is provided by a digital-to-analog converter (DAC). The DAC is cycled beginning with the most significant bits. The corresponding bit is set when the comparator fires, *i.e.* the DAC output becomes less than the pulse height. Then the DAC cycles through the less significant bits, always setting the corresponding bit when the comparator fires. Thus, n-bit resolution requires n steps and yields 2^n bins. This technique makes efficient use of circuitry and is fairly fast. High-resolution devices (16 – 20 bits) with conversion times of order μs are readily available. Currently a 16-bit ADC with a conversion time of $1\,\mu\mathrm{s}$ ($1\,\mathrm{MS/s}$) requires about 100 mW.

A common limitation is differential non-linearity, since the resistors that set the DAC levels must be extremely accurate. For DNL $< 1\%$ the resistor determining the 2^{12}-level in a 13-bit ADC must be accurate to $< 2.4 \cdot 10^{-6}$. As a consequence, differential non-linearity in high-resolution successive approximation converters is typically 10 – 20% and often exceeds the 0.5 LSB (least significant bit) required to ensure monotonic response.

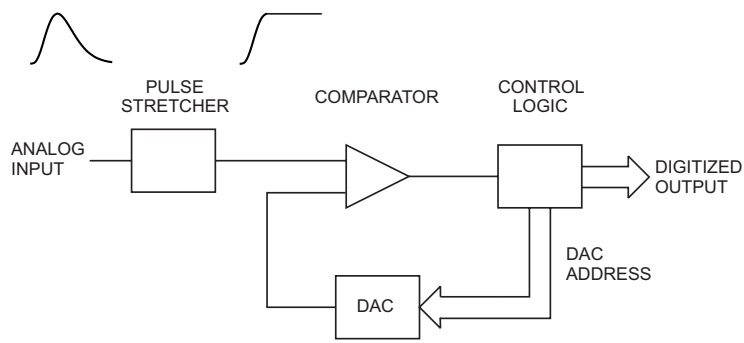


Figure 33: Principle of a successive approximation ADC. The DAC is controlled to sequentially add levels proportional to 2^n, 2^{n-1}, $\ldots 2^0$. The corresponding bit is set if the comparator output is high (DAC output < pulse height).

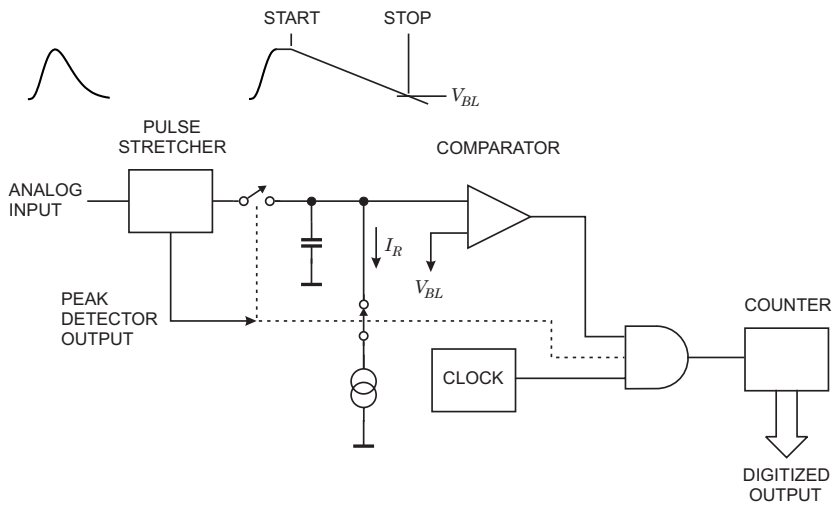


Figure 34: Principle of a Wilkinson ADC. After the peak amplitude has been acquired, the output of the peak detector initiates the conversion process. The memory capacitor is discharged by a constant current while counting the clock pulses. When the capacitor is discharged to the baseline level V_{BL} the comparator output goes low and the conversion is complete.

The Wilkinson ADC[12] has traditionally been the mainstay of precision pulse digitization. The principle is shown in Figure 34. The peak signal amplitude is acquired by a combined peak detector/pulse stretcher and transferred to a memory capacitor. The output of the peak detector initiates the conversion process:

1. The memory capacitor is disconnected from the stretcher,
2. a current source is switched on to linearly discharge the capacitor with current I_R, and simultaneously
3. a counter is enabled to determine the number of clock pulses until the voltage on the capacitor reaches the baseline level V_{BL}.

The time required to discharge the capacitor is a linear function of pulse height, so the counter content provides the digitized pulse height. The clock pulses are provided by a crystal oscillator, so the time between pulses is extremely uniform and this circuit inherently provides excellent differential linearity. The drawback is the relatively long conversion time T_C, which for a given resolution is proportional to the pulse height, $T_C = n \times T_{clk}$, where n is the channel

number corresponding to the pulse height. For example, a clock frequency of 100 MHz provides a clock period $T_{clk} = 10$ ns and a maximum conversion time $T_C = 82\,\mu$s for 13 bits ($n = 8192$). Clock frequencies of 100 MHz are typical, but > 400 MHz have been implemented with excellent performance (DNL $< 10^{-3}$). This scheme makes efficient use of circuitry and allows low power dissipation. Wilkinson ADCs have been implemented in 128-channel readout ICs for silicon strip detectors[13]. Each ADC added only 100 μm to the length of a channel and a power of 300 μW per channel.

13 Time-to-digital Converters (TDCs)

The combination of a clock generator with a counter is the simplest technique for time-to-digital conversion, as shown in Figure 35. The clock pulses are counted between the start and stop signals, which yields a direct readout in real time. The limitation is the speed of the counter, which in current technology is limited to about 1 GHz, yielding a time resolution of 1 ns. Using the stop pulse to strobe the instantaneous counter status into a register provides multi-hit capability.

Analog techniques are commonly used in high-resolution digitizers to provide resolution in the range of ps to ns. The principle is to convert a time interval into a voltage by charging a capacitor through a switchable current source. The start pulse turns on the current source and the stop pulse turns it off. The resulting voltage on the capacitor C is $V = Q/C = I_T(T_{stop} - T_{start})/C$, which is digitized by an ADC. A convenient implementation switches the current source to a smaller discharge current I_R and uses a Wilkinson ADC for digitization,

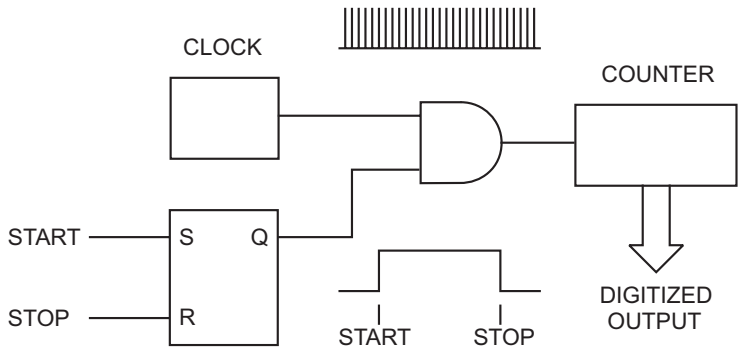


Figure 35: The simplest form of time digitizer counts the number of clock pulses between the start and stop signals.

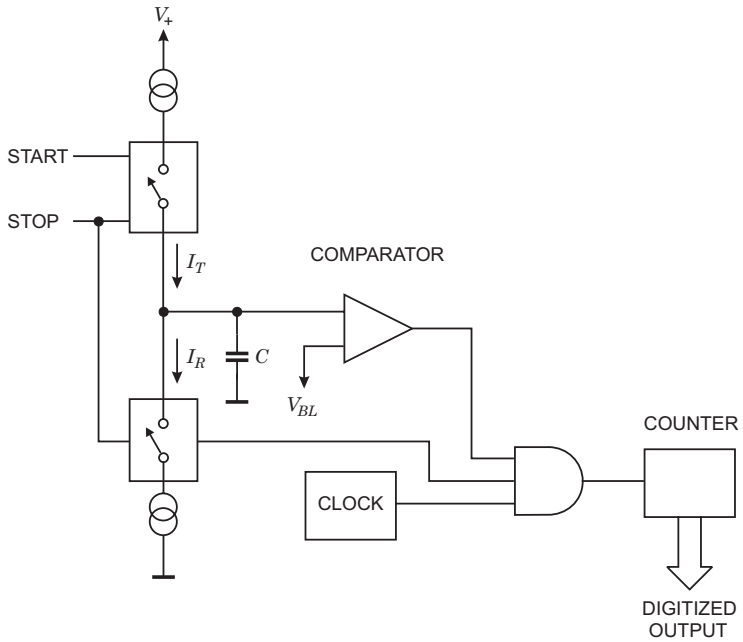


Figure 36: Combining a time-to-amplitude converter with an ADC forms a time digitizer capable of ps resolution. The memory capacitor C is charged by the current I_T for the duration $T_{start} - T_{stop}$ and subsequently discharged by a Wilkinson ADC.

as illustrated in Figure 36. This technique provides high resolution, but at the expense of dead time and multi-hit capability.

14 Signal Transmission

Signals are transmitted from one unit to another through transmission lines, often coaxial cables or ribbon cables. When transmission lines are not terminated with their characteristic impedance, the signals are reflected. As a signal propagates along the cable, the ratio of instantaneous voltage to current equals the cable's characteristic impedance $Z_0 = \sqrt{L/C}$, where L and C are the inductance and capacitance per unit length. Typical impedances are 50 or 75 Ω for coaxial cables and $\sim$ 100 Ω for ribbon cables. If at the receiving end the cable is connected to a resistance different from the cable impedance, a different ratio of voltage to current must be established. This occurs through a reflected signal. If the termination is less than the line impedance, the voltage

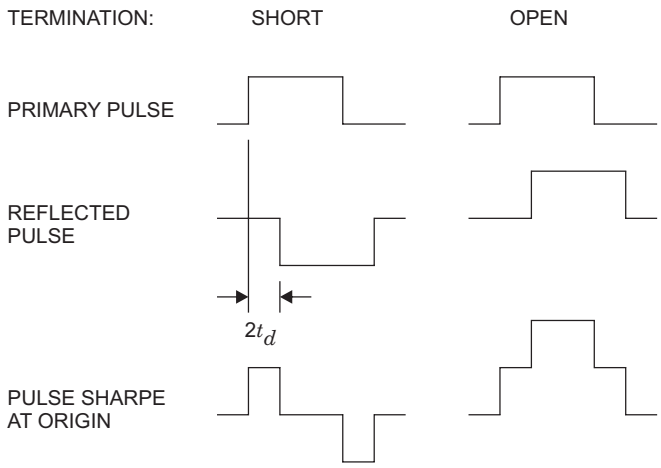


Figure 37: Voltage pulse reflections on a transmission line terminated either with a short (left) or open circuit (right). Measured at the sending end, the reflection from a short at the receiving end appears as a pulse of opposite sign delayed by the round trip delay of the cable. If the total delay is less than the pulse width, the signal appears as a bipolar pulse. Conversely, an open circuit at the receiving end causes a reflection of like polarity.

must be smaller and the reflected voltage wave has the opposite sign. If the termination is greater than the line impedance, the voltage wave is reflected with the same polarity. Conversely, the current in the reflected wave is of like sign when the termination is less than the line impedance and of opposite sign when the termination is greater. Voltage reflections are illustrated in Figure 37. At the sending end the reflected pulse appears after twice the propagation delay of the cable. Since in the presence of a dielectric the velocity of propagation $v = c/\sqrt{\varepsilon}$, in typical coaxial and ribbon cables the delay is 5 ns/m.

Cable drivers often have a low output impedance, so the reflected pulse is reflected again towards the receiver, to be reflected again, *etc.* This is shown in Figure 38, which shows the observed signal when the output of a low-impedance pulse driver is connected to a high-impedance amplifier input through a 4 m long 50 Ω coaxial cable. If feeding a counter, a single pulse will be registered multiple times, depending on the threshold level. When the amplifier input is terminated with 50 Ω, the reflections disappear and only the original 10 ns wide pulse is seen.

There are two methods of terminating cables, which can be applied either

individually or – in applications where pulse fidelity is critical – in combination. As illustrated in Figure 39 the termination can be applied at the receiving or the sending end. Receiving end termination absorbs the signal pulse when it arrives at the receiver. With sending-end termination the pulse is reflected at the receiver, but since the reflected pulse is absorbed at the sender, no additional pulses are visible at the receiver. At the sending end the original pulse is attenuated two-fold by the voltage divider formed by the series resistor and the cable impedance. However, at the receiver the pulse is reflected with the same polarity, so the superposition of the original and the reflected pulses provides the original amplitude.

This example uses voltage amplifiers, which have low output and high input impedances. It is also possible to use current amplifiers, although this is less common. Then, the amplifier has a high output impedance and low input impedance, so shunt termination is applied at the sending end and series termination at the receiving end.

Terminations are never perfect, especially at high frequencies where stray capacitance becomes significant. For example, the reactance of 10 pF at 100 MHz is 160 Ω. Thus, critical applications often use both series and parallel termination, although this does incur a 50% reduction in pulse amplitude. In the μs regime, amplifier inputs are usually designed as high impedance, whereas timing amplifiers tend to be internally terminated, but one should always check if this is the case. As a rule of thumb, whenever the propagation delay of cables (or connections in general) exceeds a few percent of the signal risetime, terminations are required.

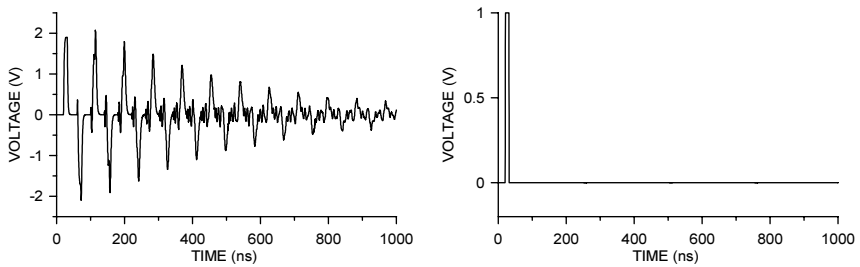


Figure 38: Left: Signal observed in an amplifier when a low-impedance driver is connected to the amplifier through a 4 m long coaxial cable. The cable impedance is 50 Ω and the amplifier input appears as 1 kΩ in parallel with 30 pF. When the receiving end is properly terminated with 50 Ω, the reflections disappear (right).

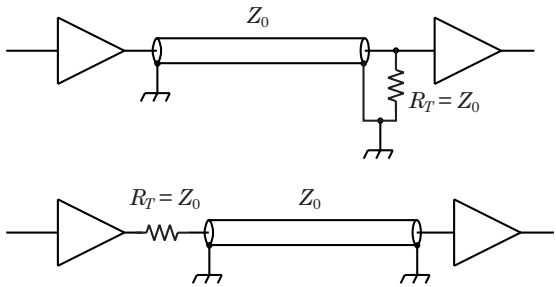


Figure 39: Cables may be terminated at the receiving end (top, shunt termination) or sending end (bottom, series termination).

15 Interference and Pickup

The previous discussion analyzed random noise sources inherent to the sensor and front-end electronics. In practical systems external noise often limits the obtainable detection threshold or energy resolution. As with random noise, external pickup introduces baseline fluctuations. There are many possible sources, radio and television stations, local RF generators, system clocks, transients associated with trigger signals and data readout, *etc.* Furthermore, there are many ways through which these undesired signals can enter the system. Again, a comprehensive review exceeds the allotted space, so only a few key examples of pickup mechanisms will be shown. A more detailed discussion is in refs [1] and [2]. Ott[14] gives a more general treatment and texts by Johnson and Graham[15][16] give useful details on signal transmission and design practices.

15.1 Pickup mechanisms

The most sensitive node in a detector system is the input. Figure 40 shows how very small spurious signals coupled to the sensor backplane can inject substantial charge. Any change in the bias voltage ΔV directly at the sensor backplane will inject a charge $\Delta Q = C_d \Delta V$. Assume a silicon strip sensor with 10 cm strip length. Then the capacitance C_d from the backplane to a single strip is about 1 pF. If the noise level is 1000 electrons ($1.6 \cdot 10^{-16}$ C), ΔV must be much smaller than $Q_n/C_d = 160\mu$V. This can be introduced as noise from the bias supply (some voltage supplies are quite noisy; switching power supplies can be clean, but most aren't) or noise on the ground plane can

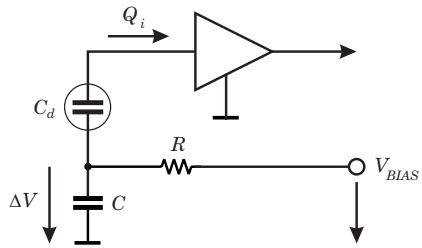


Figure 40: Noise on the detector bias line is coupled through the detector capacitance to the amplifier input.

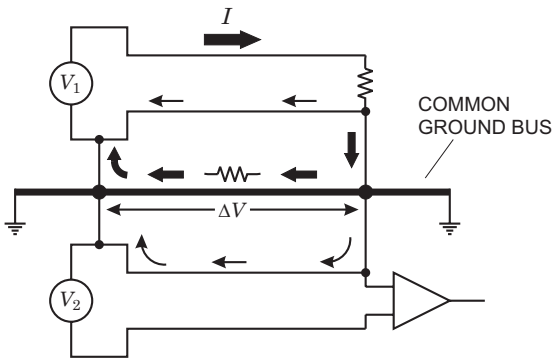


Figure 41: Shared current paths introduce common voltage drops to different circuits.

couple through the capacitor C. Naively, one might assume the ground plane to be "clean", but it can carry significant interference for the following reason.

One of the most common mechanisms for cross-coupling is shared current paths, often referred to as "ground loops". However, this phenomenon is not limited to grounding. Consider two systems. The first is transmitting large currents from a source to a receiver. The second is similar, but is attempting a low-level measurement. Following the prevailing lore, both systems are connected to a massive ground bus, as shown in Figure 41. Current seeks the path of least resistance, so the large current from source V_1 will also flow through the ground bus. Although the ground bus is massive, it does not have zero resistance, so the large current flowing through the ground system causes a voltage drop ΔV.

In system 2 (source V_2) both signal source and receiver are also connected to the ground system. Now the voltage drop ΔV from system 1 is in series with the signal path, so the receiver measures $V_2 + \Delta V$. The cross-coupling has nothing to do with grounding *per se*, but is due to the common return path. However, the common ground caused the problem by establishing the shared path. This mechanism is not limited to large systems with external ground busses, but also occurs on the scale of printed circuit boards and micron-scale integrated circuits. At high frequencies the impedance is increased due to skin effect and inductance. Note that for high-frequency signals the connections can be made capacitively, so even if there is no DC path, the parasitic capacitance due to mounting structures or adjacent conductor planes can be sufficient to close the loop.

The traditional way of dealing with this problem is to reduce the impedance of the shared path, which leads to the "copper braid syndrome". However, changes in the system will often change the current paths, so this "fix" is not very reliable. Furthermore, in many detector systems – tracking detectors, for example – the additional material would be prohibitive. Instead, it is best to avoid the root cause.

15.2 Remedial techniques

Figure 42 shows a sensor connected to a multistage amplifier. Signals are transferred from stage to stage through definite current paths. It is critical to maintain the integrity of the signal paths, but this does not depend on grounding – indeed Figure 42 does not show any ground connection at all. The most

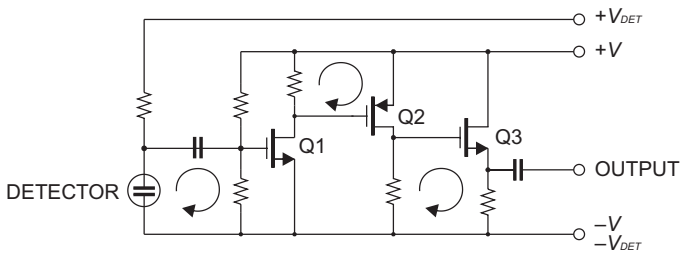


Figure 42: The signal is transferred from the sensor to the input stage and from stage to stage via local current loops.

critical parts of this chain are the input, which is the most sensitive node, and the output driver, which tends to circulate the largest current. Circuit diagrams usually are not drawn like Figure 42; the bottom common line is typically shown as ground. For example, in Figure 40 the sensor signal current flows through capacitor C and reaches the return node of the amplifier through "ground". Clearly, it is critical to control this path and keep deleterious currents from this area.

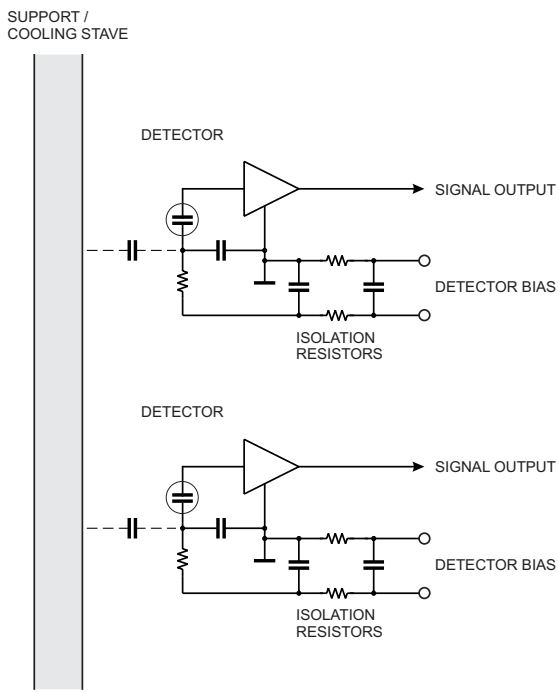


Figure 43: Capacitive coupling between detectors or detector modules and their environment introduces interference when relative potentials and stray capacitance are not controlled.

However superfluous grounding may be, one cannot let circuit elements simply float with respect to their environment. Capacitive coupling is always present and any capacitive coupling between two points of different potential will induce a signal. This is illustrated in Figure 43, which represents individual detector modules mounted on a support/cooling structure. Interference can couple through the parasitic capacitance of the mount, so it is crucial to reduce

this capacitance and control the potential of the support structure relative to the detector module. Attaining this goal in reality is a challenge, which is not always met successfully. Nevertheless, paying attention to signal paths and potential references early on is much easier than attempting to correct a poor design after it's done. Troubleshooting is exacerbated by the fact that current paths interact, so doing the "wrong" thing sometimes brings improvement. Furthermore, only one mistake can ruin system performance, so if this has been designed into the system from the outset, one is left with compromises. Nevertheless, although this area is rife with myths, basic physics still applies.

16 Conclusion

Signal processing is a key part of modern detector systems. Proper design is especially important when signals are small and electronic noise determines detection thresholds or resolution. Optimization of noise is well understood and predicted noise levels can be achieved in practical experiments within a few percent of predicted values. However, systems must be designed very carefully to avoid extraneous pickup.

References

[1] H. Spieler, *Semiconductor Detector Systems*, Oxford University Press, Oxford, 2005. ISBN 0-19-852784-5

[2] http://www-physics.lbl.gov/∼spieler

[3] J. Butler, Triggering and Data Acquisition General Considerations, in *Instrumentation in Elementary Particle Physics, AIP Conf. Proc.* **674** (2003) 101–129

[4] T. Kondo *et al.*, Construction and performance of the ATLAS silicon microstrip barrel modules. *Nucl. Instr. and Meth.* **A485** (2002) 27–42

[5] I. Kipnis, H. Spieler and T. Collins, A Bipolar Analog Front-End Integrated Circuit for the SDC Silicon Tracker, *IEEE Trans. Nucl. Sci.* **NS-41/4** (1994) 1095–1103

[6] S. Ramo, Currents Induced by Electron Motion. *Proc. IRE* **27** (1939) 584–585

[7] F. S. Goulding, Pulse Shaping in Low-Noise Nuclear Amplifiers: A Physical Approach to Noise Analysis, *Nucl. Instr. Meth.* **100** (1972) 493–504

[8] F. S. Goulding and D.A. Landis, Signal Processing for Semiconductor Detectors, *IEEE Trans. Nucl. Sci.* **NS-29/3** (1982) 1125–1141

[9] V. Radeka, Trapezoidal Filtering of Signals from Large Germanium Detectors at High Rates, *Nucl. Instr. Meth.* **99** (1972) 525–539

[10] V. Radeka, Signal, Noise and Resolution in Position-Sensitive Detectors, *IEEE Trans. Nucl. Sci.* **NS-21** (1974) 51–64

[11] H. Spieler, Fast Timing Methods for Semiconductor Detectors, *IEEE Trans. Nucl. Sci.* **NS-29/3** (1982) 1142–1158

[12] D. H. Wilkinson, A Stable Ninety-Nine Channel Pulse Amplitude Analyser for Slow Counting. *Proc. Cambridge Phil. Soc.* **46/3** (1950) 508–518

[13] M. Garcia-Sciveres *et al.*, The SVX3D integrated circuit for dead-timeless silicon strip readout, *Nucl. Instr. and Meth.* **A435** (1999) 58–64

[14] H. W. Ott, *Noise Reduction Techniques in Electronic Systems* (2nd edn), Wiley, New York, 1988, ISBN 0-471-85068-3, TK7867.5.087

[15] H. Johnson and M. Graham, *High-Speed Digital Design*, Prentice Hall PTR, Upper Saddle River, 1993 ISBN, 0-13-395724-1, TK7868.D5J635

[16] H. Johnson and M. Graham, *High-Speed Signal Propagation*, Prentice Hall PTR, Upper Saddle River, 2002, ISBN 0-13-084408-X, TK5103.15.J64

Review Talks

The Electroweak Interactions in the Standard Model and beyond

Guido Altarelli
CERN, Department of Physics, Theory Division
CH–1211 Geneva 23, Switzerland

ABSTRACT

We present a concise review of the status of the Standard Model and of the models of new physics.

1 Precision Tests of the Standard Model

The results of the electroweak precision tests as well as of the searches for the Higgs boson and for new particles performed at LEP and SLC are now available in nearly final form. Taken together with the measurements of m_t, m_W and the searches for new physics at the Tevatron, and with some other data from low energy experiments, they form a very stringent set of precise constraints [1] to compare with the Standard Model (SM) or with any of its conceivable extensions. When confronted with these results, on the whole the SM performs rather well, so that it is fair to say that no clear indication for new physics emerges from the data [2].

All electroweak Z pole measurements, combining the results of the 5 experiments, are summarised in Table 1. Information on the Z partial widths are contained in the quantities:

$$\sigma^0_{\rm h} \;=\; \frac{12\pi}{m^2_{\rm Z}}\,\frac{\Gamma_{\rm ee}\Gamma_{\rm had}}{\Gamma^2_{\rm Z}}\,, \qquad R^0_\ell \;=\; \frac{\sigma^0_{\rm h}}{\sigma^0_\ell} \;=\; \frac{\Gamma_{\rm had}}{\Gamma_{\ell\ell}}\,, \qquad R^0_{\rm q} \;=\; \frac{\Gamma_{\rm q\bar{q}}}{\Gamma_{\rm had}}\,. \quad (1)$$

Here $\Gamma_{\ell\ell}$ is the partial decay width for a pair of massless charged leptons. The partial decay width for a given fermion species are related to the effective vector and axial-vector coupling constants of the neutral weak current:

$$\Gamma_{\rm f\bar{f}} \;=\; N^f_C\frac{G_{\rm F}m^3_{\rm Z}}{6\sqrt{2}\pi}\left(g^2_{\rm Af}C_{\rm Af}+g^2_{\rm Vf}C_{\rm Vf}\right)+\Delta_{\rm ew/QCD}\,, \quad (2)$$

where N^f_C is the QCD colour factor, $C_{\{\rm A,V\}f}$ are final-state QCD/QED correction factors also absorbing imaginary contributions to the effective coupling

Observable	Measurement	SM fit
$m_{\rm Z}$ [GeV]	91.1875 ± 0.0021	91.1873
$\Gamma_{\rm Z}$ [GeV]	2.4952 ± 0.0023	2.4965
$\sigma^0_{\rm h}$ [nb]	41.540 ± 0.037	41.481
R^0_ℓ	20.767 ± 0.025	20.739
$A^{0,\ell}_{\rm FB}$	0.0171 ± 0.0010	0.0164
$\mathcal{A}_\ell$ (SLD)	0.1513 ± 0.0021	0.1480
$\mathcal{A}_\ell$ (P_τ)	0.1465 ± 0.0033	0.1480
$R^0_{\rm b}$	0.21644 ± 0.00065	0.21566
$R^0_{\rm c}$	0.1718 ± 0.0031	0.1723
$A^{0,{\rm b}}_{\rm FB}$	0.0995 ± 0.0017	0.1037
$A^{0,{\rm c}}_{\rm FB}$	0.0713 ± 0.0036	0.0742
$\mathcal{A}_{\rm b}$	0.922 ± 0.020	0.935
$\mathcal{A}_{\rm c}$	0.670 ± 0.026	0.668
$\sin^2\theta^{\rm lept}_{\rm eff}$ $(Q^{\rm had}_{\rm FB})$	0.2324 ± 0.0012	0.23140
$m_{\rm W}$ [GeV]	80.425 ± 0.034	80.398
$\Gamma_{\rm W}$ [GeV]	2.133 ± 0.069	2.094
$m_{\rm t}$ [GeV] ($\rm p\overline{p}$ [5])	178.0 ± 4.3	178.1
$\Delta\alpha^{(5)}_{\rm had}(m^2_{\rm Z})$ [6]	0.02761 ± 0.00036	0.02768

Table 1: Summary of electroweak precision measurements at high Q^2 [3]. The first block shows the Z-pole measurements. The second block shows additional results from other experiments: the mass and the width of the W boson measured at the Tevatron and at LEP-2, the mass of the top quark measured at the Tevatron, and the contribution to $\alpha(m^2_{\rm Z})$ of the hadronic vacuum polarisation.

constants, $g_{\rm Af}$ and $g_{\rm Vf}$ are the real parts of the effective couplings, and Δ contains non-factorisable mixed corrections.

Besides total cross sections, various types of asymmetries have been measured. The results of all asymmetry measurements are quoted in terms of the asymmetry parameter $\mathcal{A}_{\rm f}$, defined in terms of the real parts of the effective coupling constants, $g_{\rm Vf}$ and $g_{\rm Af}$, as:

$$\mathcal{A}_{\rm f} \;=\; 2\frac{g_{\rm Vf}g_{\rm Af}}{g_{\rm Vf}^2+g_{\rm Af}^2} \;=\; 2\frac{g_{\rm Vf}/g_{\rm Af}}{1+(g_{\rm Vf}/g_{\rm Af})^2}\,, \qquad A_{\rm FB}^{0,\,{\rm f}} \;=\; \frac{3}{4}\mathcal{A}_{\rm e}\mathcal{A}_{\rm f}\,. \tag{3}$$

The measurements are: the forward-backward asymmetry ($A_{\rm FB}^{0,\,{\rm f}} = (3/4)\mathcal{A}_{\rm e}\mathcal{A}_{\rm f}$), the tau polarisation ($\mathcal{A}_\tau$) and its forward backward asymmetry ($\mathcal{A}_{\rm e}$) measured at LEP, as well as the left-right and left-right forward-backward asymmetry measured at SLC ($\mathcal{A}_{\rm e}$ and $\mathcal{A}_{\rm f}$, respectively). Hence the set of partial width and asymmetry results allows the extraction of the effective coupling constants. In particular, from the measurements at the Z, lepton universality of the neutral weak current was established at the per-mille level.

Using the effective electroweak mixing angle, $\sin^2\theta_{\rm eff}^{\rm f}$, and the ρ parameter, the effective coupling constants are given by:

$$g_{\rm Af} \;=\; \sqrt{\rho}\;T_3^f\,, \qquad \frac{g_{\rm Vf}}{g_{\rm Af}} \;=\; 1-4|q_f|\sin^2\theta_{\rm eff}^{\rm f}\,, \tag{4}$$

where T_3^f is the third component of the weak iso-spin and q_f the electric charge of the fermion. The effective electroweak mixing angle is thus given independently of the ρ parameter by the ratio $g_{\rm Vf}/g_{\rm Af}$ and hence in a one-to-one relation by each asymmetry result.

The various asymmetries determine the effective electroweak mixing angle for leptons with highest sensitivity. The results on $\sin^2\theta_{\rm eff}^{\rm lept}$ are compared in Figure 1. The weighted average of these six results, including small correlations, is:

$$\sin^2\theta_{\rm eff}^{\rm lept} \;=\; 0.23150 \pm 0.00016\,. \tag{5}$$

Note, however, that this average has a χ^2 of 10.5 for 5 degrees of freedom, corresponding to a probability of 6.2%. The χ^2 is pushed up by the two most precise measurements of $\sin^2\theta_{\rm eff}^{\rm lept}$, namely those derived from the measurements of $\mathcal{A}_\ell$ by SLD, dominated by the left-right asymmetry $A_{\rm LR}^0$, and of the forward-backward asymmetry measured in $\rm b\overline{b}$ production at LEP, $A_{\rm FB}^{0,\,{\rm b}}$, which differ by about 2.9 standard deviations. No experimental effect in either measurement has been identified to explain this, thus the difference is presumably either the

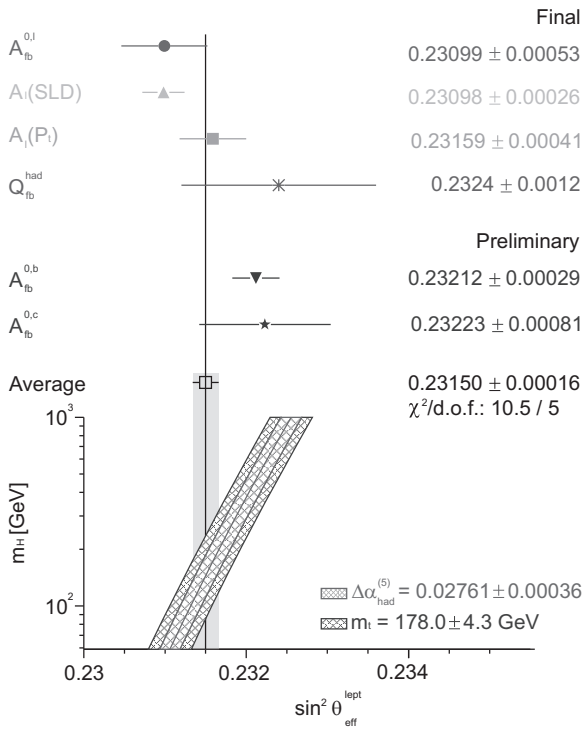


Figure 1: Effective electroweak mixing angle $\sin^2\theta_{\rm eff}^{\rm lept}$ derived from measurement results depending on lepton couplings only (top) and also quark couplings (bottom) [1]. Also shown is the prediction of $\sin^2\theta_{\rm eff}^{\rm lept}$ in the SM as a function of $m_{\rm H}$, including its parametric uncertainty dominated by the uncertainties in $\Delta\alpha^{(5)}_{\rm had}(m_{\rm Z}^2)$ and $m_{\rm t}$, shown as the bands.

effect of statistics or an unidentified systematics or a hint for new physics, as further discussed below.

Also shown in Table 1 are the results on m_W obtained at LEP-2 and at the Tevatron, and the new world average of the top mass.

For the analysis of electroweak data in the SM one starts from the input parameters: as in any renormalisable theory masses and couplings have to be specified from outside. One can trade one parameter for another and this freedom is used to select the best measured ones as input parameters. As a result, some of them, α, G_F and $m_{\rm Z}$, are very precisely known [4], some other ones, $m_{f_{light}}$, $m_{\rm t}$ and $\alpha_s(m_Z)$ are far less well determined while $m_{\rm H}$ is largely unknown. Note that the new combined CDF and DØ value for $m_{\rm t}$ [5], as listed in Table 1, is higher than the previous average by nearly one standard deviation.

Among the light fermions, the quark masses are badly known, but fortunately, for the calculation of radiative corrections, they can be replaced by $\alpha(m_Z)$, the value of the QED running coupling at the Z mass scale. The value of the hadronic contribution to the running, $\Delta\alpha^{(5)}_{\rm had}(m^2_{\rm Z})$, reported in Table 1, is obtained through dispersion relations from the data on $\rm e^+e^-\to$ hadrons at low centre-of-mass energies [6]. From the input parameters one computes the radiative corrections to a sufficient precision to match the experimental accuracy. Then one compares the theoretical predictions and the data for the numerous observables which have been measured, checks the consistency of the theory and derives constraints on $m_{\rm t}$, $\alpha_{\rm S}(m^2_{\rm Z})$ and $m_{\rm H}$.

The computed radiative corrections include the complete set of one-loop diagrams, plus some selected large subsets of two-loop diagrams and some sequences of resummed large terms of all orders (large logarithms and Dyson resummations). In particular large logarithms, e.g., terms of the form $(\alpha/\pi\ \ln\ (m_Z/m_{f_\ell}))^n$ where f_ℓ is a light fermion, are resummed by well-known and consolidated techniques based on the renormalisation group. For example, large logarithms dominate the running of α from m_e, the electron mass, up to $m_{\rm Z}$, which is a 6% effect, much larger than the few per-mille contributions of purely weak loops. Also, large logs from initial state radiation dramatically distort the line shape of the Z resonance observed at LEP-1 and SLC and must be accurately taken into account in the measurement of the Z mass and total width.

Among the one loop EW radiative corrections, a remarkable class of contributions are those terms that increase quadratically with the top mass. The large sensitivity of radiative corrections to m_t arises from the existence of these terms. The quadratic dependence on m_t (and possibly on other widely broken isospin multiplets from new physics) arises because, in spontaneously broken gauge theories, heavy loops do not decouple. On the contrary, in QED or QCD,

the running of α and α_s at a scale Q is not affected by heavy quarks with mass $M \gg Q$. According to an intuitive decoupling theorem [7], diagrams with heavy virtual particles of mass M can be ignored for $Q \ll M$ provided that the couplings do not grow with M and that the theory with no heavy particles is still renormalizable. In the spontaneously broken EW gauge theories both requirements are violated. First, one important difference with respect to unbroken gauge theories is in the longitudinal modes of weak gauge bosons. These modes are generated by the Higgs mechanism, and their couplings grow with masses (as is also the case for the physical Higgs couplings). Second, the theory without the top quark is no more renormalisable because the gauge symmetry is broken if the b quark is left with no partner (while its measured couplings show that the weak isospin is 1/2). Because of non decoupling, precision tests of the electroweak theory may be sensitive to new physics even if the new particles are too heavy for their direct production.

While radiative corrections are quite sensitive to the top mass, they are unfortunately much less dependent on the Higgs mass. If they were sufficiently sensitive, by now we would precisely know the mass of the Higgs. However, the dependence of one loop diagrams on $m_{\rm H}$ is only logarithmic: $\sim G_{\rm F} m_{\rm W}^2 \log(m_{\rm H}^2/m_{\rm W}^2)$. Quadratic terms $\sim G_{\rm F}^2 m_{\rm H}^2$ only appear at two loops and are too small to be important. The difference with the top case is that $m_{\rm t}^2 - m_{\rm b}^2$ is a direct breaking of the gauge symmetry that already affects the relevant one loop diagrams, while the Higgs couplings to gauge bosons are "custodial-SU(2)" symmetric in lowest order.

We now discuss fitting the data in the SM. One can think of different types of fit, depending on which experimental results are included or which answers one wants to obtain. For example, in Table 2 we present in column 1 a fit of all Z pole data plus $m_{\rm W}$ and $\Gamma_{\rm W}$ (this is interesting as it shows the value of m_t obtained indirectly from radiative corrections, to be compared with the value of m_t measured in production experiments), in column 2 a fit of all Z pole data plus $m_{\rm t}$ (here it is $m_{\rm W}$ which is indirectly determined), and, finally, in column 3 a fit of all the data listed in Table 1 (which is the most relevant fit for constraining $m_{\rm H}$). From the fit in column 1 of Table 2 we see that the extracted value of $m_{\rm t}$ is in perfect agreement with the direct measurement (see Table 1). Similarly we see that the experimental measurement of $m_{\rm W}$ in Table 1 is larger by about one standard deviation with respect to the value from the fit in column 2. We have seen that quantum corrections depend only logarithmically on $m_{\rm H}$. In spite of this small sensitivity, the measurements are precise enough that one still obtains a quantitative indication of the mass range. From the fit in column 3 we obtain: $\log_{10} m_{\rm H}({\rm GeV}) = 2.05 \pm 0.20$ (or $m_{\rm H} = 113^{+62}_{-42}$ GeV). This result on the Higgs mass is particularly remarkable. The value of $\log_{10} m_{\rm H}({\rm GeV})$ is right

on top of the small window between ~ 2 and ~ 3 which is allowed, on the one side, by the direct search limit ($m_{\rm H} \gtrsim 114$ GeV from LEP-2 [8]), and, on the other side, by the theoretical upper limit on the Higgs mass in the minimal SM, $m_{\rm H} \lesssim 600 - 800$ GeV [9].

Fit	1	2	3
Measurements	$m_{\rm W}$, $\Gamma_{\rm W}$	$m_{\rm t}$	$m_{\rm t}$, $m_{\rm W}$, $\Gamma_{\rm W}$
$m_{\rm t}$ (GeV)	$178.5^{+11.0}_{-8.5}$	177.2 ± 4.1	178.1 ± 3.9
$m_{\rm H}$ (GeV)	117^{+162}_{-62}	129^{+76}_{-50}	113^{+62}_{-42}
log $[m_{\rm H}({\rm GeV})]$	$2.07^{+0.38}_{-0.33}$	2.11 ± 0.21	2.05 ± 0.20
$\alpha_s(m_{\rm Z})$	0.1187 ± 0.0027	0.1190 ± 0.0027	0.1186 ± 0.0027
χ^2/dof	16.3/12	15.0/11	16.3/13
$m_{\rm W}$ (MeV)		80386 ± 23	

Table 2: Standard Model fits of electroweak data. All fits use the Z pole results and $\Delta\alpha^{(5)}_{\rm had}(m^2_{\rm Z})$ as listed in Table 1, also including constants such as the Fermi constant $G_{\rm F}$. In addition, the measurements listed in each column are included as well. For fit 2, the expected W mass is also shown. For details on the fit procedure see [3].

A different way of looking at the data is to consider the epsilon parameters. As well known these parameters vanish in the limit of tree level SM plus pure QED or pure QCD corrections. So they are a measure of the weak quantum corrections. Their experimental values are given by [1]:

$$\epsilon_1\ 10^3 = 5.4 \pm 1.0 \quad (6)$$

$$\epsilon_2\ 10^3 = -8.9 \pm 1.2 \quad (7)$$

$$\epsilon_3\ 10^3 = 5.25 \pm 0.95 \quad (8)$$

$$\epsilon_b\ 10^3 = -4.7 \pm 1.6 \quad (9)$$

The experimental values are compared to the SM predictions as function of m_t and m_H in Figure 2. We see that ϵ_3 points to a light Higgs, that ϵ_b is a bit too large because of A^b_{FB} and ϵ_2 a bit too small because of m_W.

Thus the whole picture of a perturbative theory with a fundamental Higgs is well supported by the data on radiative corrections. It is important that there is a clear indication for a particularly light Higgs: at 95% c.l. $m_H \lesssim 237$ GeV. This is quite encouraging for the ongoing search for the Higgs particle. More general, if the Higgs couplings are removed from the Lagrangian the resulting

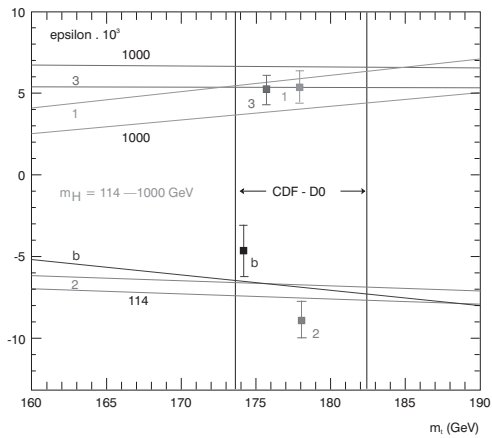


Figure 2: The epsilon variables: comparison of the data with the SM predictions. The data should be horizontal bands but they are shown here near the central value of m_t.

theory is non renormalisable. A cutoff Λ must be introduced. In the quantum corrections $\log m_H$ is then replaced by $\log \Lambda$ plus a constant. The precise determination of the associated finite terms would be lost (that is, the value of the mass in the denominator in the argument of the logarithm). A heavy Higgs would need some unfortunate conspiracy: the finite terms, different in the new theory from those of the SM, should accidentally compensate for the heavy Higgs in a few key parameters of the radiative corrections (mainly ϵ_1 and ϵ_3, see, for example, [10]). Alternatively, additional new physics, for example in the form of effective contact terms added to the minimal SM lagrangian, should accidentally do the compensation, which again needs some sort of conspiracy.

In Table 3 we collect the results on low energy precision tests of the SM obtained from neutrino and antineutrino deep inelastic scattering (NuTeV [11]), parity violation in Cs atoms (APV [12]) and the recent measurement of the parity-violating asymmetry in Moller scattering [13]. The experimental results are compared with the predictions from the fit in column 3 of Table 2. We see the agreement is good except for the NuTeV result that shows a deviation by three standard deviations. The NuTeV measurement is quoted as a measurement of $\sin^2\theta_{\rm W} = 1 - m_{\rm W}^2/m_{\rm Z}^2$ from the ratio of neutral to charged current deep inelastic cross-sections from ν_μ and $\bar{\nu}_\mu$ using the Fermilab beams. There is growing evidence that the NuTeV anomaly could simply arise from an underestimation

of the theoretical uncertainty in the QCD analysis needed to extract $\sin^2\theta_{\rm W}$. In fact, the lowest order QCD parton formalism on which the analysis has been based is too crude to match the experimental accuracy. In particular a small asymmetry in the momentum carried by the strange and antistrange quarks, $s - \bar{s}$, could have a large effect [14]. A tiny violation of isospin symmetry in parton distributions, too small to be seen elsewhere, can similarly be of some importance. In conclusion we believe the discrepancy has more to teach about the QCD parton densities than about the electroweak theory.

Observable	Measurement	SM fit
$\sin^2\theta_{\rm W}$ (νN [11])	0.2277 ± 0.0016	0.2226
$Q_{\rm W}$(Cs) (APV [12])	-72.83 ± 0.49	-72.91
$\sin^2\theta^{\rm lept}_{\rm eff}$ (e^-e^- [13])	0.2296 ± 0.0023	0.2314

Table 3: Summary of other electroweak precision measurements, namely the measurements of the on-shell electroweak mixing angle in neutrino-nucleon scattering, the weak charge of cesium measured in an atomic parity violation experiment, and the effective weak mixing angle measured in Moller scattering, all performed in processes at low Q^2. The SM predictions are derived from fit 3 of Table 2. Good agreement of the prediction with the measurement is found except for νN.

When confronted with these results, on the whole the SM performs rather well, so that it is fair to say that no clear indication for new physics emerges from the data. However, as already mentioned, one problem is that the two most precise measurements of $\sin^2\theta^{\rm lept}_{\rm eff}$ from $A_{\rm LR}$ and $A^{0,\,\rm b}_{\rm FB}$ differ nearly three standard deviations. In general, there appears to be a discrepancy between $\sin^2\theta^{\rm lept}_{\rm eff}$ measured from leptonic asymmetries ($(\sin^2\theta_{\rm eff})_l$) and from hadronic asymmetries ($(\sin^2\theta_{\rm eff})_h$), see also Figure 1. In fact, the result from $A_{\rm LR}$ is in good agreement with the leptonic asymmetries measured at LEP, while all hadronic asymmetries, though their errors are large, are better compatible with the result of $A^{0,\,\rm b}_{\rm FB}$.

The situation is shown in Figure 3 [15]. The values of $(\sin^2\theta_{\rm eff})_l$, $(\sin^2\theta_{\rm eff})_h$ and their formal combination are shown each at the $m_{\rm H}$ value that would correspond to it given the central value of $m_{\rm t}$. Of course, the value for $m_{\rm H}$ indicated by each $\sin^2\theta^{\rm lept}_{\rm eff}$ has an horizontal ambiguity determined by the measurement error and the width of the $\pm1\sigma$ band for $m_{\rm t}$. Even taking this spread into account it is clear that the implications on $m_{\rm H}$ are sizably different. One might imagine that some new physics effect could be hidden in the $\rm Zb\bar{b}$ vertex. Like for the top quark mass there could be other non decoupling effects from new

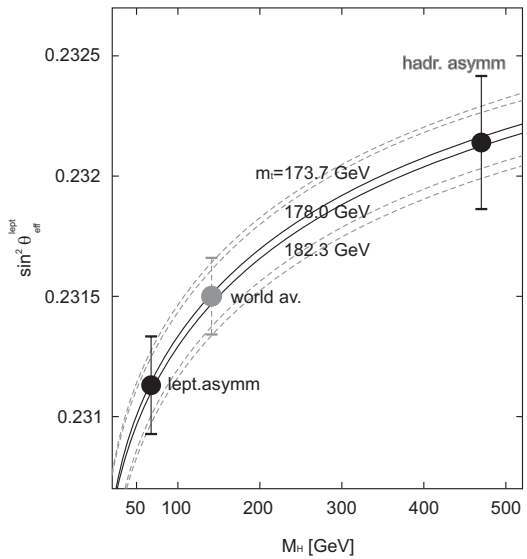


Figure 3: The data for $\sin^2\theta_{\rm eff}^{\rm lept}$ are plotted vs m_H. For presentation purposes the measured points are shown each at the m_H value that would ideally correspond to it given the central value of m_t (updated from [15]).

heavy states or a mixing of the b quark with some other heavy quark. However, it is well known that this discrepancy is not easily explained in terms of some new physics effect in the $Zb\bar{b}$ vertex. In fact, $A_{\rm FB}^{0,\,\rm b}$ is the product of lepton- and b-asymmetry factors: $A_{\rm FB}^{0,\,\rm b} = (3/4)\mathcal{A}_{\rm e}\mathcal{A}_{\rm b}$. The sensitivity of $A_{\rm FB}^{0,\,\rm b}$ to $\mathcal{A}_{\rm b}$ is limited, because the $\mathcal{A}_{\rm e}$ factor is small, so that a rather large change of the b-quark couplings with respect to the SM is needed in order to reproduce the measured discrepancy (precisely a $\sim 30\%$ change in the right-handed coupling, an effect too large to be a loop effect but which could be produced at the tree level, e.g., by mixing of the b quark with a new heavy vectorlike quark [16]). But then this effect should normally also appear in the direct measurement of $\mathcal{A}_{\rm b}$ performed at SLD using the left-right polarized b asymmetry, even within the moderate precision of this result, and it should also be manifest in the accurate measurement of $R_{\rm b} \propto g_{\rm Rb}^2 + g_{\rm Lb}^2$. The measurements of neither $\mathcal{A}_{\rm b}$ nor $R_{\rm b}$ confirm the need of a new effect. Even introducing an ad hoc mixing the overall

fit is not terribly good, but we cannot exclude this possibility completely. Alternatively, the observed discrepancy could be due to a large statistical fluctuation or an unknown experimental problem. The ambiguity in the measured value of $\sin^2\theta_{\rm eff}^{\rm lept}$ could thus be larger than the nominal error, reported in Equation 5, obtained from averaging all the existing determinations.

We have already observed that the experimental value of $m_{\rm W}$ (with good agreement between LEP and the Tevatron) is a bit high compared to the SM prediction (see Figure 4). The value of $m_{\rm H}$ indicated by $m_{\rm W}$ is on the low side, just in the same interval as for $\sin^2\theta_{\rm eff}^{\rm lept}$ measured from leptonic asymmetries. It is interesting that the new value of $m_{\rm t}$ considerably relaxes the previous tension between the experimental values of $m_{\rm W}$ and $\sin^2\theta_{\rm eff}^{\rm lept}$ measured from leptonic asymmetries on one side and the lower limit on $m_{\rm H}$ from direct searches on the other side [17, 18]. This is also apparent from Figure 4.

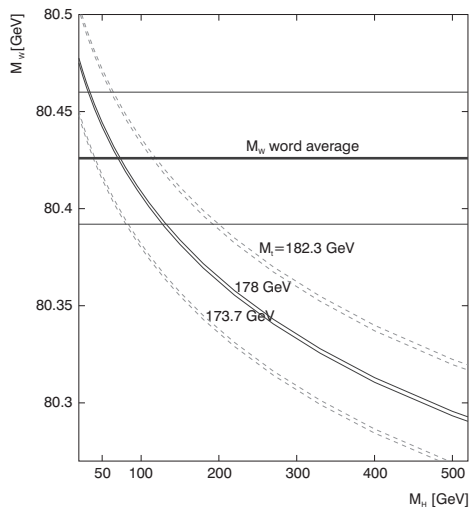


Figure 4: The world average for m_W is compared with the SM prediction as a function of m_H (updated from [15]).

The main lesson of precision tests of the standard electroweak theory can be summarised as follows. The couplings of quark and leptons to the weak gauge bosons $\rm W^{\pm}$ and Z are indeed precisely those prescribed by the gauge symmetry. The accuracy of a few per-mille for these tests implies that, not only the tree level, but also the structure of quantum corrections has been verified. To a lesser

accuracy the triple gauge vertices γW^+W^- and ZW^+W^- have also been found in agreement with the specific prediction of the $SU(2) \bigotimes U(1)$ gauge theory. This means that it has been verified that the gauge symmetry is unbroken in the vertices of the theory: the currents are indeed conserved. Yet there is obvious evidence that the symmetry is otherwise badly broken in the masses. Thus the currents are conserved but the spectrum of particle states is not at all symmetric. This is a clear signal of spontaneous symmetry breaking. The practical implementation of spontaneous symmetry breaking in a gauge theory is via the Higgs mechanism. The Higgs sector of the SM is still very much untested. What has been tested is the relation $m_W^2 = m_Z^2 \cos^2 \theta_W$, modified by computable radiative corrections. This relation means that the effective Higgs (be it fundamental or composite) is indeed a weak isospin doublet. The Higgs particle has not been found but in the SM its mass can well be larger than the present direct lower limit $m_H \gtrsim 114$ GeV obtained from direct searches at LEP-2. The radiative corrections computed in the SM when compared to the data on precision electroweak tests lead to a clear indication for a light Higgs, not too far from the present lower bound. No signal of new physics has been found. However, to make a light Higgs natural in presence of quantum fluctuations new physics should not be too far. This is encouraging for the LHC that should experimentally clarify the problem of the electroweak symmetry breaking sector and search for physics beyond the SM.

2 Outlook on Avenues beyond the Standard Model

Given the success of the SM why are we not satisfied with that theory? Why not just find the Higgs particle, for completeness, and declare that particle physics is closed? The reason is that there are both conceptual problems and phenomenological indications for physics beyond the SM. On the conceptual side the most obvious problems are that quantum gravity is not included in the SM and the related hierarchy problem. Among the main phenomenological hints for new physics we can list coupling unification, dark matter, neutrino masses, baryogenesis and the cosmological vacuum energy.

The computed evolution with energy of the effective SM gauge couplings clearly points towards the unification of the electro-weak and strong forces (Grand Unified Theories: GUT's) at scales of energy $M_{GUT} \sim 10^{15} - 10^{16}\ GeV$ which are close to the scale of quantum gravity, $M_{Pl} \sim 10^{19}\ GeV$. One is led to imagine a unified theory of all interactions also including gravity (at present superstrings provide the best attempt at such a theory). Thus GUT's and the realm of quantum gravity set a very distant energy horizon that modern particle theory cannot ignore. Can the SM without new physics be valid up to such

large energies? This appears unlikely because the structure of the SM could not naturally explain the relative smallness of the weak scale of mass, set by the Higgs mechanism at $\mu \sim 1/\sqrt{G_F} \sim 250\ GeV$ with G_F being the Fermi coupling constant. This so-called hierarchy problem is related to the presence of fundamental scalar fields in the theory with quadratic mass divergences and no protective extra symmetry at $\mu = 0$. For fermion masses, first, the divergences are logarithmic and, second, they are forbidden by the $SU(2) \bigotimes U(1)$ gauge symmetry plus the fact that at $m = 0$ an additional symmetry, i.e. chiral symmetry, is restored. Here, when talking of divergences, we are not worried of actual infinities. The theory is renormalisable and finite once the dependence on the cut off is absorbed in a redefinition of masses and couplings. Rather the hierarchy problem is one of naturalness. We should see the cut off as a parameterization of our ignorance on the new physics that will modify the theory at large energy scales. Then it is relevant to look at the dependence of physical quantities on the cut off and to demand that no unexplained enormously accurate cancellations arise.

The hierarchy problem can be put in very practical terms: loop corrections to the higgs mass squared are quadratic in Λ. The most pressing problem is from the top loop. With $m_h^2 = m_{bare}^2 + \delta m_h^2$ the top loop gives

$$\delta m_{h|top}^2 \sim \frac{3G_F}{\sqrt{2}\pi^2} m_t^2 \Lambda^2 \sim (0.3\Lambda)^2 \tag{10}$$

If we demand that the correction does not exceed the light Higgs mass indicated by the precision tests, Λ must be close, $\Lambda \sim o(1\ TeV)$. Similar constraints arise from the quadratic Λ dependence of loops with gauge bosons and scalars, which, however, lead to less pressing bounds. So the hierarchy problem demands new physics to be very close (in particular the mechanism that quenches the top loop). Actually, this new physics must be rather special, because it must be very close, yet its effects are not clearly visible (the "LEP Paradox" [19]). Examples of proposed classes of solutions for the hierarchy problem are:

Supersymmetry. In the limit of exact boson-fermion symmetry the quadratic divergences of bosons cancel so that only log divergences remain. However, exact SUSY is clearly unrealistic. For approximate SUSY (with soft breaking terms), which is the basis for all practical models, Λ is replaced by the splitting of SUSY multiplets, $\Lambda \sim m_{SUSY} - m_{ord}$. In particular, the top loop is quenched by partial cancellation with s-top exchange.

Technicolor. The Higgs system is a condensate of new fermions. There are no fundamental scalar Higgs sector, hence no quadratic devergences associated to the μ^2 mass in the scalar potential. This mechanism needs a very

strong binding force, $\Lambda_{TC} \sim 10^3\ \Lambda_{QCD}$. It is difficult to arrange that such nearby strong force is not showing up in precision tests. Hence this class of models has been disfavoured by LEP, although some special class of models have been devised aposteriori, like walking TC, top-color assisted TC etc (for recent reviews, see, for example, [20]).

Large compactified extra dimensions. The idea is that M_{PL} appears very large, that is gravity seems very weak because we are fooled by hidden extra dimensions so that the real gravity scale is reduced down to $o(1\ TeV)$. This possibility is very exciting in itself and it is really remarkable that it is compatible with experiment.

"**Little Higgs**"**models**. In these models extra symmetries allow $m_h \neq 0$ only at two-loop level, so that Λ can be as large as $o(10\ TeV)$ with the Higgs within present bounds (the top loop is quenched by exchange of heavy vectorlike new charge-2/3 quarks).

We now briefly comment in turn on these possibilities.

SUSY models are the most developed and most widely accepted. Many theorists consider SUSY as established at the Planck scale M_{Pl}. So why not to use it also at low energy to fix the hierarchy problem, if at all possible? It is interesting that viable models exist. The necessary SUSY breaking can be introduced through soft terms that do not spoil the good convergence properties of the theory. Precisely those terms arise from supergravity when it is spontaneoulsly broken in a hidden sector. This is the case of the MSSM [21]. Of course, minimality is only a simplicity assumption that could possibly be relaxed. The MSSM is a completely specified, consistent and computable theory which is compatible with all precision electroweak tests. In this most traditional approach SUSY is broken in a hidden sector and the scale of SUSY breaking is very large of order $\Lambda \sim \sqrt{G_F^{-1/2} M_{Pl}}$. But since the hidden sector only communicates with the visible sector through gravitational interactions the splitting of the SUSY multiplets is much smaller, in the TeV energy domain, and the Goldstino is practically decoupled. But alternative mechanisms of SUSY breaking are also being considered. In one alternative scenario [22] the (not so much) hidden sector is connected to the visible one by ordinary gauge interactions. As these are much stronger than the gravitational interactions, Λ can be much smaller, as low as 10-100 TeV. It follows that the Goldstino is very light in these models (with mass of order or below 1 eV typically) and is the lightest, stable SUSY particle, but its couplings are observably large. The radiative decay of the lightest neutralino into the Goldstino leads to detectable photons. The signature of photons comes out naturally in this SUSY breaking pattern: with respect to the MSSM, in the gauge mediated model there are typically more photons and less missing energy. The main appeal of gauge mediated models

is a better protection against flavour changing neutral currents but naturality problems tend to increase. As another possibility it has been pointed out that there are pure gravity contributions to soft masses that arise from gravity theory anomalies[23]. In the assumption that these terms are dominant the associated spectrum and phenomenology have been studied. In this case gaugino masses are proportional to gauge coupling beta functions, so that the gluino is much heavier than the electroweak gauginos, and the wino is most often the lightest SUSY particle.

What is really unique to SUSY with respect to all other extensions of the SM listed above is that the MSSM or similar models are well defined and computable up to M_{Pl} and, moreover, are not only compatible but actually quantitatively supported by coupling unification and GUT's. At present the most direct phenomenological evidence in favour of supersymmetry is obtained from the unification of couplings in GUTs. Precise LEP data on $\alpha_s(m_Z)$ and $\sin^2\theta_W$ show that standard one-scale GUTs fail in predicting $\sin^2\theta_W$ given $\alpha_s(m_Z)$ (and $\alpha(m_Z)$) while SUSY GUTs are in agreement with the present, very precise, experimental results. If one starts from the known values of $\sin^2\theta_W$ and $\alpha(m_Z)$, one finds [24] for $\alpha_s(m_Z)$ the results: $\alpha_s(m_Z) = 0.073 \pm 0.002$ for Standard GUTs and $\alpha_s(m_Z) = 0.129 \pm 0.010$ for SUSY GUTs to be compared with the world average experimental value $\alpha_s(m_Z) = 0.119 \pm 0.003$. Another great asset of SUSY GUT's is that proton decay is much slowed down with respect to the non SUSY case. First, the unification mass $M_{GUT} \sim$ few 10^{16} GeV, in typical SUSY GUT's, is about 20-30 times larger than for ordinary GUT's. This makes p decay via gauge boson exchange negligible and the main decay amplitude arises from dim-5 operators with higgsino exchange, leading to a rate close but still compatible with existing bounds (see, for example,[25]). It is also important that SUSY provides an excellent dark matter candidate, the neutralino. We finally recall that the range of neutrino masses as indicated by oscillation experiments, when interpreted in the see-saw mechanism, point to M_{GUT} and give additional support to GUTs [26].

In spite of all these virtues it is true that the lack of SUSY signals at LEP and the lower limit on m_H pose problems for the MSSM. The lightest Higgs particle is predicted in the MSSM to be below $m_h \lesssim 135\ GeV$ (the recent increase of m_t helps in this respect). The limit on the SM Higgs $m_H \gtrsim 114\ GeV$ considerably restricts the available parameter space of the MSSM requiring relatively large $\tan\beta$ ($\tan\beta \gtrsim 2-3$: at tree level $m^2_{h} = m^2_Z \cos^2 2\beta$) and rather heavy stop (the loop corrections increase with $\log m^2_t$). Stringent naturality constraints also follow from imposing that the electroweak symmetry breaking occurs at the right place: in SUSY models the breaking is induced by the running of the H_u mass starting from a common scalar mass m_0 at M_{GUT}. The squared Z mass

m_Z^2 can be expressed as a linear combination of the SUSY parameters m_0^2, $m_{1/2}^2$, A_t^2, μ^2,... with known coefficients. Barring cancellations that need fine tuning, the SUSY parameters, hence the SUSY s-partners cannot be too heavy. The LEP limits, in particular the chargino lower bound $m_{\chi+} \gtrsim 100\ GeV$, are sufficient to eliminate an important region of the parameter space, depending on the amount of allowed fine tuning. For example, models based on gaugino universality at the GUT scale are discarded unless a fine tuning by at least a factor of 20 is not allowed. Without gaugino universality [27] the strongest limit remains on the gluino mass: $m_Z^2 \sim 0.7\ m_{gluino}^2 + \dots$ which is still compatible with the present limit $m_{gluino} \gtrsim 200\ GeV$.

The non discovery of SUSY at LEP has given further impulse to the quest for new ideas on physics beyond the SM. Large extra dimensions [28] and "little Higgs" [29] models are the most interesting new directions in model building. Large extra dimension models propose to solve the hierarchy problem by bringing gravity down from M_{Pl} to $m \sim o(1\ TeV)$ where m is the string scale. Inspired by string theory one assumes that some compactified extra dimensions are sufficiently large and that the SM fields are confined to a 4-dimensional brane immersed in a d-dimensional bulk while gravity, which feels the whole geometry, propagates in the bulk. We know that the Planck mass is large because gravity is weak: in fact $G_N \sim 1/M_{Pl}^2$, where G_N is Newton constant. The idea is that gravity appears so weak because a lot of lines of force escape in extra dimensions. Assume you have $n = d - 4$ extra dimensions with compactification radius R. For large distances, $r >> R$, the ordinary Newton law applies for gravity: in natural units $F \sim G_N/r^2 \sim 1/(M_{Pl}^2 r^2)$. At short distances, $r \lesssim R$, the flow of lines of force in extra dimensions modifies Gauss law and $F^{-1} \sim m^2(mr)^{d-4}r^2$. By matching the two formulas at $r = R$ one obtains $(M_{Pl}/m)^2 = (Rm)^{d-4}$. For $m \sim 1\ TeV$ and $n = d - 4$ one finds that $n = 1$ is excluded ($R \sim 10^{15} cm$), for $n = 2$ R is at the edge of present bounds $R \sim 1\ mm$, while for $n = 4, 6$, $R \sim 10^{-9}, 10^{-12}\ cm$. In all these models a generic feature is the occurrence of Kaluza-Klein (KK) modes. Compactified dimensions with periodic boundary conditions, as for quantization in a box, imply a discrete spectrum with momentum $p = n/R$ and mass squared $m^2 = n^2/R^2$. There are many versions of these models. The SM brane can itself have a thickness r with $r <\sim 10^{-17}\ cm$ or $1/r >\sim 1\ TeV$, because we know that quarks and leptons are pointlike down to these distances, while for gravity there is no experimental counter-evidence down to $R <\sim 0.1\ mm$ or $1/R >\sim 10^{-3}\ eV$. In case of a thickness for the SM brane there would be KK recurrences for SM fields, like W_n, Z_n and so on in the TeV region and above. There are models with factorized metric ($ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu + h_{ij}(y)dy^i dy^j$, where y (i,j) denotes the extra dimension coordinates (and indices), or models with warped metric

($ds^2 = e - 2kR|\phi|\eta_{\mu\nu}dx^\mu dx^\nu - R^2\phi^2$ [30]. In any case there are the towers of KK recurrences of the graviton. They are gravitationally coupled but there are a lot of them that sizably couple, so that the net result is a modification of cross-sections and the presence of missing energy.

Large extra dimensions provide a very exciting scenario [31]. Already it is remarkable that this possibility is compatible with experiment. However, there are a number of criticisms that can be brought up. First, the hierarchy problem is more translated in new terms rather than solved. In fact, the basic relation $Rm = (M_{Pl}/m)^{2/n}$ shows that Rm, which one would apriori expect to be 0(1), is instead ad hoc related to the large ratio M_{Pl}/m. In this respect the Randall-Sundrum variety is more appealing because the hierarchy suppression m_W/M_{Pl} could arise from the warping factor $e^{-2kR|\phi|}$, with not too large values of kR. Also it is not clear how extra dimensions can by themselves solve the LEP paradox (the large top loop corrections should be controlled by the opening of the new dimensions and the onset of gravity): since m_H is light $\Lambda \sim 1/R$ must be relatively close. But precision tests put very strong limits on Λ. In fact, in typical models of this class, there is no mechanism to sufficiently quench the corrections. No simple, realistic model has yet emerged as a benchmark. But it is attractive to imagine that large extra dimensions could be a part of the truth, perhaps coupled with some additional symmetry or even SUSY.

In the extra dimension general context an interesting direction of development is the study of symmetry breaking by orbifolding and/or boundary conditions. These are models where a larger gauge symmetry (with or without SUSY) holds in the bulk. The symmetry is reduced in the 4 dimensional brane, where the physics that we observe is located, as an effect of symmetry breaking induced geometrically by suitable boundary conditions. There are models where SUSY, valid in $n > 4$ dimensions is broken by boundary conditions [32], in particular the model of Ref.[33], where the mass of the Higgs is computable and can be extimated with good accuracy. Then there are "Higgsless models" where it is the SM electroweak gauge symmetry which is broken at the boundaries [34]. Or models where the Higgs is the 5th component of a gauge boson of an extended symmetry valid in $n > 4$ [35]. In general all these alternative models for the Higgs mechanism face severe problems and constraints from electroweak precision tests [36]. At the GUT scale, symmetry breaking by orbifolding can be applied to obtain a reformulation of SUSY GUT's where many problematic features of ordinary GUT's (e.g. a baroque Higgs sector, the doublet-triplet splitting problem, fast proton decay etc) are improved [31], [37].

In "little Higgs" models, the symmetry of the SM is extended to a suitable global group G that also contains some gauge enlargement of $SU(2) \bigotimes U(1)$, for example $G \supset [SU(2) \bigotimes U(1)]^2 \supset SU(2) \bigotimes U(1)$. The Higgs particle is a

pseudo-Goldstone boson of G that only takes mass at 2-loop level, because two distinct symmetries must be simultaneously broken for it to take mass, which requires the action of two different couplings in the same diagram. Then in the relation between δm_h^2 and Λ^2 there is an additional coupling and an additional loop factor that allow for a bigger separation between the Higgs mass and the cut-off. Typically, in these models one has one or more Higgs doublets at $m_h \sim\ 0.2\ TeV$, and a cut-off at $\Lambda \sim\ 10\ TeV$. The top loop quadratic cut-off dependence is partially canceled, in a natural way guaranteed by the symmetries of the model, by a new coloured, charge-2/3, vectorial quark χ of mass around $1\ TeV$ (a fermion not a scalar like the s-top of SUSY models). Certainly these models involve a remarkable level of group theoretic virtuosity. However, in the simplest versions one is faced with problems with precision tests of the SM [38]. Even with vectorlike new fermions large corrections to the epsilon parameters arise from exchanges of the new gauge bosons W' and Z' (due to lack of custodial $SU(2)$ symmetry). In order to comply with these constraints the cut-off must be pushed towards large energy and the amount of fine tuning needed to keep the Higgs light is still quite large. Probably these bad features can be fixed by some suitable complication of the model (see for example, [39]). But, in my opinion, the real limit of this approach is that it only offers a postponement of the main problem by a few TeV, paid by a complete loss of predictivity at higher energies. In particular all connections to GUT's are lost.

Finally, we stress the importance of the cosmological constant or vacuum energy problem [40]. The exciting recent results on cosmological parameters, culminating with the precise WMAP measurements [41], have shown that vacuum energy accounts for about 2/3 of the critical density: $\Omega_\Lambda \sim 0.65$, Translated into familiar units this means for the energy density $\rho_\Lambda \sim (2\ 10^{-3}\ eV)^4$ or $(0.1\ mm)^{-4}$. It is really interesting (and not at all understood) that $\rho_\Lambda^{1/4} \sim \Lambda_{EW}^2/M_{Pl}$ (close to the range of neutrino masses). It is well known that in field theory we expect $\rho_\Lambda \sim \Lambda_{cutoff}^4$. If the cut off is set at M_{Pl} or even at $0(1\ TeV)$ there would an enormous mismatch. In exact SUSY $\rho_\Lambda = 0$, but SUSY is broken and in presence of breaking $\rho_\Lambda^{1/4}$ is in general not smaller than the typical SUSY multiplet splitting. Another closely related problem is "why now?": the time evolution of the matter or radiation density is quite rapid, while the density for a cosmological constant term would be flat. If so, them how comes that precisely now the two density sources are comparable? This suggests that the vacuum energy is not a cosmological constant term, buth rather the vacuum expectation value of some field (quintessence) and that the "why now?" problem is solved by some dynamical mechanism.

Clearly the cosmological constant problem poses a big question mark on the relevance of naturalness as a relevant criterion also for the hierarchy prob-

lem: how we can trust that we need new physics close to the weak scale out of naturalness if we have no idea on the solution of the cosmological constant huge naturalness problem? The common answer is that the hierarchy problem is formulated within a well defined field theory context while the cosmological constant problem makes only sense within a theory of quantum gravity, that there could be modification of gravity at the sub-eV scale, that the vacuum energy could flow in extra dimensions or in different Universes and so on. At the other extreme is the possibility that naturalness is misleading. Weinberg [42] has pointed out that the observed order of magnitude of Λ can be successfully reproduced as the one necessary to allow galaxy formation in the Universe. In a scenario where new Universes are continuously produced we might be living in a very special one (largely fine-tuned) but the only one to allow the development of an observer. One might then argue that the same could in principle be true also for the Higgs sector. Recently it was suggested [43] to abandon the no-fine-tuning assumption for the electro-weak theory, but require correct coupling unification, presence of dark matter with weak couplings and a single scale of evolution from the EW to the GUT scale. A "split SUSY" model arises as a solution with a fine-tuned light Higgs and all SUSY particles heavy except for gauginos, higgsinos and neutralinos, protected by chiral symmetry. Or we can have a two-scale non-SUSY GUT with axions as dark matter. In conclusion, it is clear that naturalness can be a good heuristic principle but you cannot prove its necessity.

3 Summary and Conclusion

Supersymmetry remains the standard way beyond the SM. What is unique to SUSY, beyond leading to a set of consistent and completely formulated models, as, for example, the MSSM, is that this theory can potentially work up to the GUT energy scale. In this respect it is the most ambitious model because it describes a computable framework that could be valid all the way up to the vicinity of the Planck mass. The SUSY models are perfectly compatible with GUT's and are actually quantitatively supported by coupling unification and also by what we have recently learned on neutrino masses. All other main ideas for going beyond the SM do not share this synthesis with GUT's. The SUSY way is testable, for example at the LHC, and the issue of its validity will be decided by experiment. It is true that we could have expected the first signals of SUSY already at LEP, based on naturality arguments applied to the most minimal models (for example, those with gaugino universality at asymptotic scales). The absence of signals has stimulated the development of new ideas like those of large extra dimensions and "little Higgs" models. These ideas are

very interesting and provide an important reference for the preparation of LHC experiments. Models along these new ideas are not so completely formulated and studied as for SUSY and no well defined and realistic baseline has so far emerged. But it is well possible that they might represent at least a part of the truth and it is very important to continue the exploration of new ways beyond the SM.

I would like to express my gratitude to the Organisers of the ICFA '03 School for their invitation and their magnificent hospitality in Itacuruca. In particular I would like to thank Bernard Marechal.

References

[1] The LEP EW Working Group, hep-ex/0212036.

[2] G. Altarelli and M. Grunewald, hep-ph/0404165.

[3] The ALEPH, DELPHI, L3, OPAL, SLD Collaborations and the LEP Electroweak Working Group, *A Combination of Preliminary Electroweak Measurements and Constraints on the Standard Model*, hep-ex/0312023, and references therein.

[4] The Particle Data Group, *Phys. Rev.* **D66** (2002) 1.

[5] The CDF Collaboration, the DØ Collaboration, and the Tevatron Electroweak Working Group, *Combination of CDF and DØ Results on the Top-Quark Mass*, hep-ex/0404010.

[6] H. Burkhardt, B. Pietrzyk, *Update of the Hadronic Contribution to the QED Vacuum Polarization*, Phys. Lett. B 513 (2001) 46.

[7] Th. Appelquist, J. Carazzone, *Infrared Singularities and Massive Fields*, *Phys. Rev.* **D11** (1975) 2856.

[8] The ALEPH, DELPHI, L3 and OPAL Collaborations, and the LEP Working Group for Higgs Boson Searches, *Search for the Standard Model Higgs Boson at LEP*, hep-ex/0306033, *Phys. Lett.* **B565** (2003) 61–75.

[9] Th. Hambye, K. Riesselmann, *Matching Conditions and Higgs Mass Upper Bounds Revisited*, hep-ph/9610272, *Phys. Rev.* **D55** (1997) 7255–7262.

[10] G. Altarelli, R. Barbieri, F. Caravaglios, *Electroweak Precision Tests: A Concise Review*, hep-ph/9712368, *Int. Jour. Mod. Phys.***A13** (1998) 1031–1058.

[11] The NuTeV Collaboration, G.P. Zeller *et al.*, *Phys. Rev. Lett.* **88** (2002) 091802.

[12] M. Yu. Kuchiev and V. V. Flambaum, *Radiative Corrections to Parity Non-Conservation in Atoms*, hep-ph/0305053.

[13] The SLAC E158 Collaboration, P.L. Anthony *et al.*, *Observation of Parity Non-Conservation in Moller scattering*, hep-ex/0312035, *A New Measurement of the Weak Mixing Angle*, hep-ex/0403010. We have added 0.0003 to the value of $\sin^2\theta(m_Z)$ quoted by E158 in order to convert from the MSbar scheme to the effective electroweak mixing angle [4].

[14] S. Davidson et al, hep-ph/0112302.

[15] P. Gambino, *The Top Priority: Precision Electroweak Physics from Low-Energy to High-Energy*, hep-ph/0311257.

[16] D. Choudhury, T.M.P. Tait, C.E.M. Wagner, *Beautiful Mirrors and Precision Electroweak Data*, hep-ph/0109097, *Phys. Rev.* **D65** (2002) 053002.

[17] M. S. Chanowitz, *Electroweak Data and the Higgs Boson Mass: A Case for New Physics*, hep-ph/0207123, *Phys. Rev.* **D66** (2002) 073002.

[18] G. Altarelli, F. Caravaglios, G.F. Giudice, P. Gambino, G. Ridolfi, *Indication for Light Sneutrinos and Gauginos from Precision Electroweak Data*, hep-ph 0106029, *JHEP* 0106:018, 2001.

[19] R. Barbieri and A. Strumia, hep-ph/0007265.

[20] K. Lane, hep-ph/0202255,
R. S. Chivukula, hep-ph/0011264.

[21] For a recent introduction see, for example, S. P. Martin, hep-ph/9709356.

[22] M. Dine and A. E. Nelson, *Phys. Rev.* **D48** (1993) 1277;
M. Dine, A. E. Nelson and Y. Shirman, *Phys. Rev.* **D51** (1995) 1362;
G. F. Giudice and R. Rattazzi, *Phys. Rept.* **322** (1999) 419.

[23] L. Randall and R. Sundrum, *Nucl. Phys.* **B557** (1999) 79;
G.F. Giudice et al, *JHEP* **9812** (1998) 027.

[24] P. Langacker and N. Polonsky, *Phys. Rev.* **D52** (1995) 3081.

[25] G. Altarelli, F. Feruglio and I. Masina, *JHEP* 0011:040,2000.

[26] For a review see, for example, G. Altarelli and F. Feruglio, hep-ph/0405048.

[27] G. Kane et al, *Phys.Lett.* **B551** (2003) 146.

[28] For a review and a list of refs., see, for example, J. Hewett and M. Spiropulu, hep-ph/0205196.

[29] For a review and a list of refs., see, for example, M. Schmaltz, hep-ph/0210415.

[30] L. Randall and R. Sundrum, *Phys. Rev. Lett.* **83** (1999) 3370, **83** (1999) 4690.

[31] For a review, see F. Feruglio, hep-ph/0401033.

[32] I. Antoniadis, C. Munoz and M. Quiros, *Nucl. Phys.* **B397** (1993) 515; A. Pomarol and M. Quiros, *Phys. Lett.* **B438** (1998) 255.

[33] R. Barbieri, L. Hall and Y. Nomura, *Nucl. Phys.* **B624** (2002) 63; R.Barbieri, G. Marandella and M. Papucci, hep-ph/0205280, hep-ph/0305044, and refs therein.

[34] see for example, C. Csaki et al, hep-ph/0305237 , hep-ph/0308038, hep-ph/0310355; S. Gabriel, S. Nandi and G. Seidl, hep-ph/0406020 and refs therein.

[35] see for example, C. A. Scrucca, M. Serone and L. Silvestrini, hep-ph/0304220 and refs therein.

[36] R. Barbieri, A. Pomarol and R. Rattazzi, hep-ph/0310285.

[37] Y. Kawamura, *Progr. Theor. Phys.* **105** (2001) 999.

[38] J. L. Hewett, F. J. Petriello, T. G. Rizzo, hep-ph/0211218; C. Csaki et al, hep-ph/0211124, hep-ph/0303236. H-C. Cheng and I. Low, hep-ph/0405243.

[39] H-C. Cheng and I. Low, hep-ph/0405243.

[40] For orientation, see, for example, M. Turner, astro-ph/0207297.

[41] The WMAP Collaboration, D. N. Spergel et al, astro-ph/0302209.

[42] S. Weinberg, *Phys. Rev. Lett.* **59** (1987) 2607.

[43] N.Arkani-Hamed and S. Dimopoulos, hep-th/0405159; G. Giudice and A. Romanino, hep-ph/0406088.

Near beam detectors*

Hélio da Motta
Centro Brasileiro de Pesquisas Físicas – CBPF/CLAFEX
Rua Dr. Xavier Sigaud 150
22280-190 Rio de Janeiro – RJ -Brazil
E-mail: helio@fnal.gov

ABSTRACT

There are instances of high energy particle interactions where one, or both, of the interacting particles survives the proccess experiencing only a slight scattering that deviates it from the original beam. The identification of the particles scattered at small angles is of uppermost importance in the study of such events. Typical 4π detectors usually miss these particles. We describe here a series of detectors specially built to operate very close to the beam enlarging the capability of the experiment to study the physics of this kind of event.

1 Introduction

In high energy physics, diffraction is a proccess in which no quantum number is exchanged between interacting particles which remain intact. Diffraction encompasses events in which one or both incoming particles undergo diffractive dissociation with any surviving particle having a small angle with respect to the beam axis. It can account for 40% of the inclusive cross section of a process.

The interest in diffractive scattering has been growing since the observation of diffractive production of jets by the UA8 collaboration[1, 2]. The experimental difficulty in observing these events is that the scattered proton tends to remain in the beam pipe and can not be detected using typical collider central detector as DØ and CDF. It is necessary to add special forward particle detectors to the central assembly, close to the beam and at large distances from the collision point.

Although rapidity gap (absence of particles in a region of the detector) signatures can also be used to tag diffractive events, only the addition of a forward particle detector allow access to the full kinematics of the scattered particle.

*ICFA SCHOOL – Itacuruçá – Rio de Janeiro, Brazil – 8 – 20 December 2003

By detecting the scattered p or $\bar{p}$ one can measure its momentum $|\vec{p}|$ and derive the variables: $x_p = |\vec{p}|/|\vec{p}_{beam}|$; $t = (p_{beam} - p)^2$; $\xi = 1 - x_p$ where t is the four-momentum transfer of the p (or $\bar{p}$)and x_p is the fractional longitudinal momentum of the scattered particle.

2 Near the beam

The detection of the particles scattered at small angles requires the positioning of special detectors very close to the beam, usually inside the accelerator beampipe. That must be done with consideration to the very critical conditions existing in such environment, specially the ultra high vacuum (UHV) that may be present in the area. The detectors may have to be moved close to beam or removed to clear the area during accelerator insertion periods. Its location must be known within some precision and accuracy and a high degree of repeatability must be achieved.

Accelerator vacuum may reach up to 10^{-10} Torr, or even more, and any apparatus built to operate in such environment must not disturd this vacuum. We can use a simple description to help understanding the meaning of such vacuum.

Consider a cube whose sides are 10.0 cm long (a 1 L volume). Imagine the inside of this cube completely void of particles. That would mean a 0.0 Torr vacuum. If we cover the cube surface with a single layer of N_2 molecules we will need 6.0×10^{17} N_2 molecules. Throwing all these molecules into the cube volume the vacuum will correspond to 1.7×10^{-2} Torr. This is 8 orders of magnitude bigger than 10^{-10} Torr. To achieve UHV, we must remove N_2 molecules from inside the cube, leaving only 1 out of every $100,000,000$ molecules in there !

This is not simple to do. The tinyiest impurity on the material surface would make our efforts vain. Even when completely clean of impurities, the material would degass (liberating particles that are embeded in the material, like water, nitrogen). This degassing will also ruin the attempts to achieve UHV.

A UHV chamber is built following a series of steps, from which we shall mention:

- Use of 360L stainless steel.
- TIG (Tungsten Inert Gas) welding (a proccess where a cloud of inert gas is set around the welding point). Welding is made in the interior of the chamber, so there are no air pockets left inside.
- Cleaning of all parts with demineralized water and neutral detergent.

- Baking the parts (this speeds up degassing, cleaning up the parts)
- Once the chamber is finished, it is sealed, and a ion pump is attached to it.
- Vacuum is done by means of a conventional pump. Then, the chamber is baked and the ion pump turned on. During the baking stage, that may last a few days, the vacuum increases until UHV. Baking is then interruped but the ion pump is left on.

Another issue to be considered when putting a detector close to the beam is the need to have a precise knowledge of its position and the ability to move it precisely and accuretely with a high level of repeatibility. All those are required if one does not want to risk the beam and also to have a reliable measurement of the quantities described above. These mechanical conditions demand high precision and reliable mechanical devices.

The area close to the beam may also present high levels of radiation, in which case all detectors and components must be radiation hard.

All these conditions make it a very complex task the building of any system designed to operate close to the beam. We describe now a set of detectors that have been built to such use.

3 Roman Pots

Roman pots are stainless containers that allow the position detector to function outside of the machine ultra high vacuum, but close to the beam. The scattered particle traverses a thin steel window at the entrance of the pot, goes through the detector that rests in the pot and exit the pot through another thin window. The pots are remotely controlled and can be moved close to the beam during stable conditions.

The first ever built Roman pot[3] was used in 1970-1972 at the ISR by the CERN-Rome group. It did not have a thin window and housed a small hodoscope of scintillating counters. Figure1 shows the first Roman pot and a later Roman pot that was used by UA4 at SPS. The UA4 Roman pot section facing the beam is concave in shape, allowing a closer aproach to the beam. A thin window is used to reduce the material between the beam and the detector inside the pot.

More recently, CDF installed a set of three Roman pots with a scintillating fiber detector, as viewed in Figure2. When hit by the scattered particle, the fiber scintillates and the light is guided to a multianode photomultiplier that collects the light. Figure2 also shows details of the CDF scintillating fiber detector.

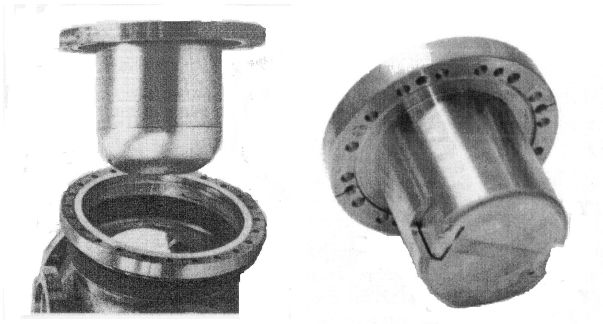


Figure 1: First Roman pot (left). UA4 Roman pot (right). The concave bottom of the UA4 Roman pot allows a closer aproach to the beam and the thin window reduces the material the particle has to traverses before hiting the detector that rests inside the pot.

4 The Forward Proton Detector

The DØ collaboration has built a very comprehensive system of Roman pot detectors that is know as Forward Proton Detector (FPD)[4]. It consists of a series of momentum spectrometers that make use of accelerator magnets in conjunction with position detectors along the beam line in order to determine the kinematic variables (t and ξ) of the scattered p and $\bar{p}$.

The detector itself consists of an array of square scintillating fibers organized in three different directions (x,u and v). Clear fibers transport the light signal to multi anode photomultipliers (MAPMT), whose outputs feed an electronic shaping and amplifying system and are read out by electronics. The detectors are placed in Roman Pot structures.

The FPD has 18 Roman pots arranged in 6 stainless steel chambers called castles located at different distances with respect to the DØ interaction point and in locations that do not interfere with any accelerator element. The experimental arrangement of the FPD is shown in Figure3. Four castles are located downstream of the low beta quadrupole magnets on each side of the colliding point: two on the p side (P1 and P2)and two on the $\bar{p}$ side (A1 and A2). Each of these quadrupole castles contains 4 Roman pots arranged to cover most of the area around the beam. Two castles (D2 and D1) are located on the outgoing $\bar{p}$ side after the dipole magnet. Each of these dipole castles contain only one Roman pot.

Figures 4,5 and 6 show the driving system that makes it possible for the pot

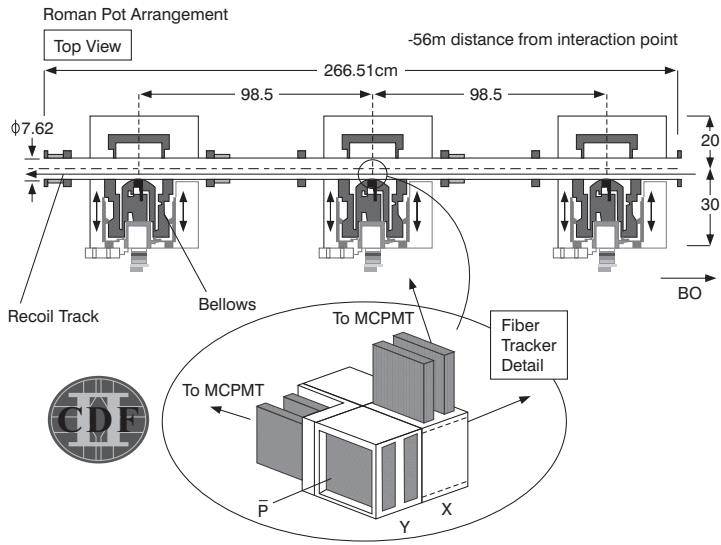


Figure 2: CDF Roman pot system. Schematics of the detector is shown. It consists of an array of scintillating fibers oriented in the X and Y direction. The scattered particle leaves a signal in a pair of XY channel that is used to determine its position.

to move perpendicularly to the beam in a very precise, safe and accurate way. The system is operated by a step motor and a set of reduction gears allows pot motion with a precision of aproximately 5 μm. A set of cylindrical and conical bearings allows adjustment of the pot alignment and a linear variable differential transducer (LVDT) monitors the pot position. A steel bellows guaranties the movement of the pot whithout affecting the vacuum.

The FPD driving assembly is basically a cylindrical tube threaded on its outer surface that is actuated by a worm-gear system. The system has a reduction of 120 times which accounts both for a precision of about $5\mu m$ movement of the detector and for compensation for the inward force of about $2,000$ N resulting from the vacuum inside the castle. This arrangement makes possible the use of small low torque motors and makes unecessary the use of any vacuum compensation system. An exploded view of the driving system is shown in Figure5. The compacteness of the whole piece makes it easy to employ in as many units as needed. Details of the driving assembly structure is presented in Figure6. The moving parts get a thin layer of a molibidenium lubricant.

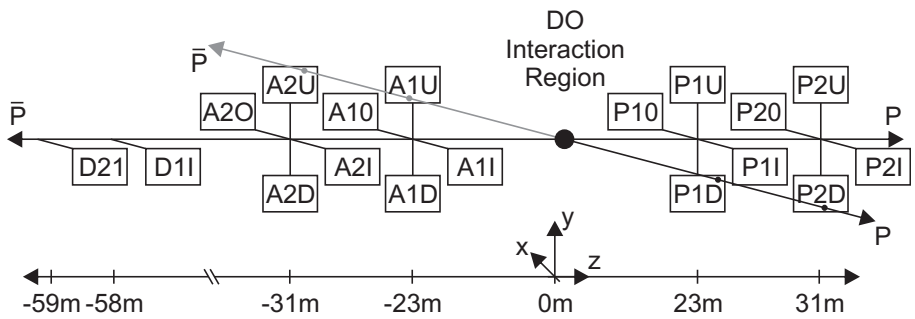


Figure 3: FPD layout at DØ experimental area. Quadrupole castles are named P or A when placed on the p side or the $\bar{p}$, respectively.

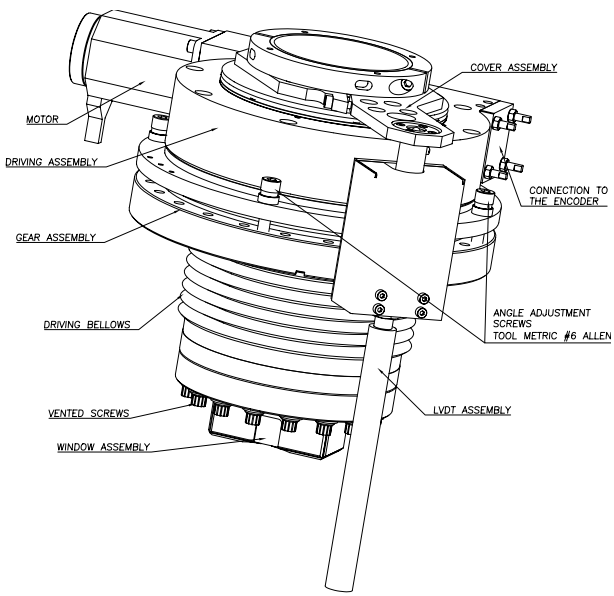


Figure 4: FPD driving assembly.

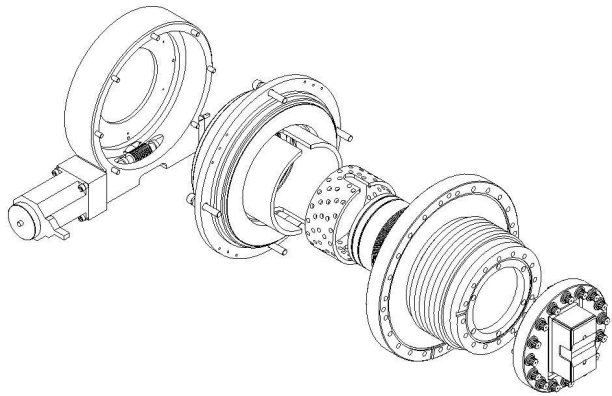


Figure 5: FPD driving assembly (exploded view).

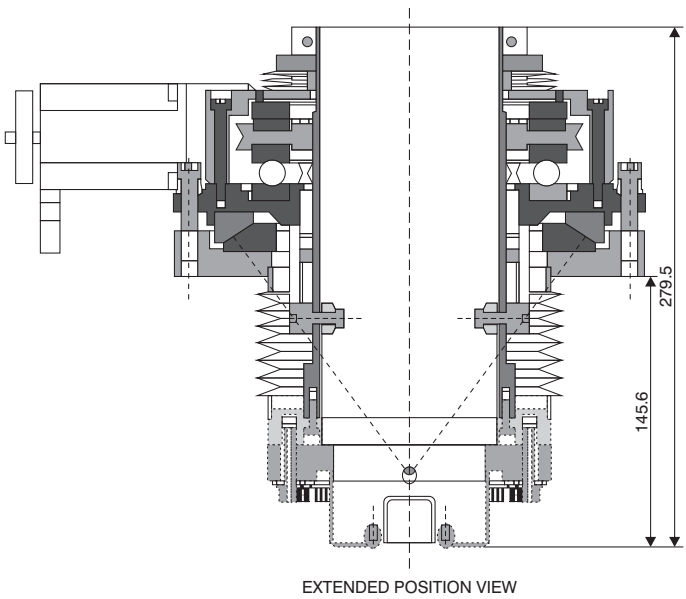


Figure 6: FPD driving assemby (cut view).

Each position detector is made of 0.8 mm thick double-clad square scintillating fibers (Bycron BCF10) bundled in groups of four forming a scintillating structure measuring 0.8 $mm \times 3.2$ mm. One end of the detector element is aluminized (aproximately 3 μm thick layer) to increase the light yield and the other end of each scintillating fiber is spliced to a double-clad clear fiber of square cross section (Bycron BCF98) with the same dimensions[†]. The scattered p or $\bar{p}$ goes through 3.2 mm of scintillating material yielding aproximately 10 photoelectrons. The 4 clear fibers then take the light of one element to a single channel of a Hamamatsu H6568 16-channel multi anode photomultiplier (MAPMT).

The use of clear fibers, spliced to the scintillating fibers that constitute the active part of the detector, is conceived as a way to reduce the effect of halo background in the fibers, the optical cross talk in the fibers, and to minimize light attenuation since clear fibers have an attenuation length longer than scintillating fibers.

Each detector consists of six planes in three views (U, V and X) in order to minimize ghost hit problems and to reduce the reconstruction ambiguities. Each view is made of two planes ($U-U'$,$V-V'$ and $X-X'$), the unprimed layer being offset by 2/3 of a fiber with respect to primed ones. This arrangement yields a theoretical detector point resolution of 80 μm. U and V planes are oriented at $\pm 45^{\circ}$ with respect to the horizontal bottom of the detector, while the X plane is at 90°. There are 20 channels in each layer of the U and V planes and 16 channels in each of the X layers. There are 112 channels (each with four fibers) per detector, giving a total of 2016 channels in the 18 Roman pots. Each detector needs 7 MAPMT and also includes a trigger scintillator read out by a fast photomultiplier (Phillips XP2282). Figure7 shows the fiber arrangement and the way they are connected to the MAPMT. Figure8 shows the FPD position detector.

[†]The use of square fibers gives an increase of about 20 % in light output compared to round fibers.

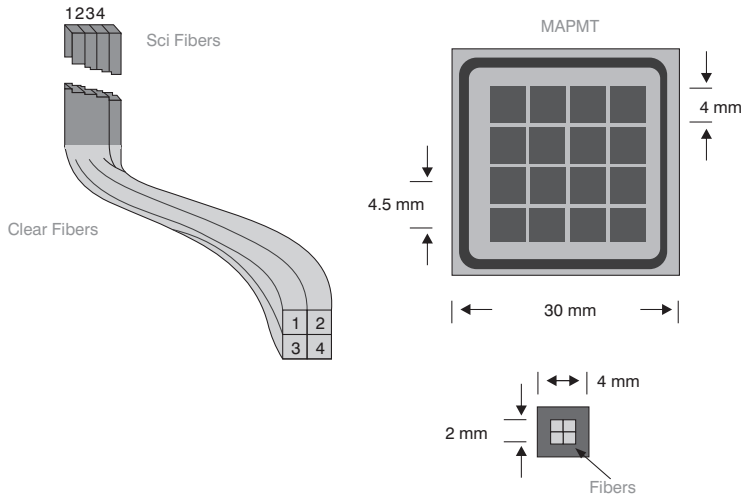


Figure 7: FPD detector fibers and the MAPMT.

5 Near the beam without a pot

Fermilab experiments E710 and E811 used a detector made of a bundle of scintillating fibers placed directly into the accelerator vacuum. No pot or any other structure was used and so there was no residual dead space between the detector edge and the beam. The bundle of scintillating fibers was placed just above the beam, with the fibers aligned along the beamline. A scattered particle hits one of the fibers originating a light signal that os guided to a CCD camera attached to the other end of the fibers bundle (a photomultiplier could also be used). This would give information on the scattered particle position. Figure9 shows a scheme of this detector.

A potless device is now under development possibly to be used at the LHC. This device, known as microstation and whose schematic is shown in Figure10 will use a silicon detector arranged in two moving structures placed around the beam. Small actuators will move the detectors to place them in the desired position around the beam. The interior of the chamber is kept under vacuum.

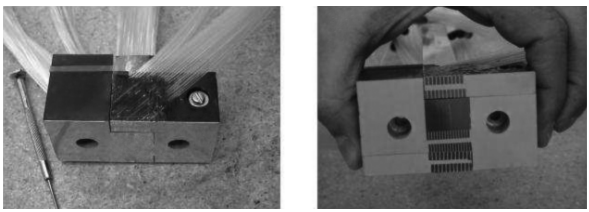


Figure 8: FPD position detector.

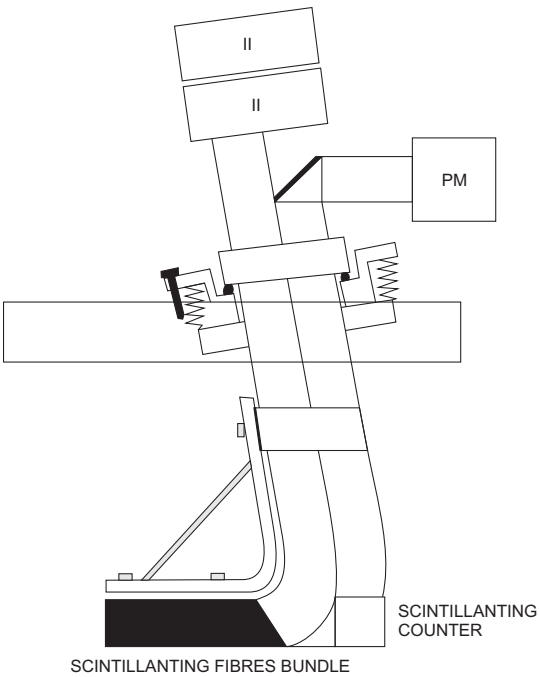


Figure 9: E710/E811 near beam detector.

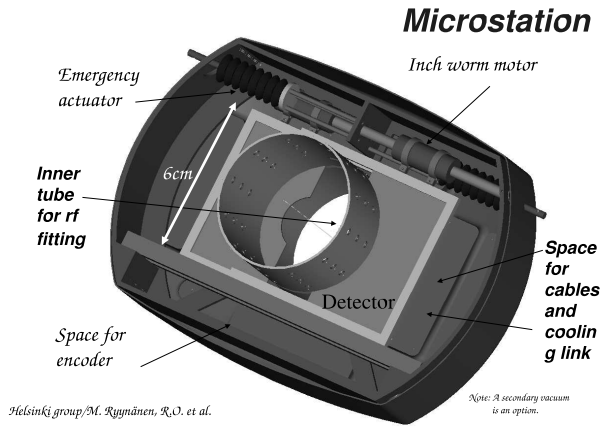


Figure 10: Microstation.

6 Conclusions

Near beam detectors have been used for over 30 years in different accelerators with different beams. They allow the measurement of several proccesses (like elastic scattering and diffraction dissociation at high energy) that can not be as well studied without these detectors.

They are as more important as higher the energy. Consequently, they shall be of great interest in the new LHC era that is to start in a few years. The technology to design and built such detectors is already of common knowledge and has been tested successfully. They are safe to operate and can become an integrated part of the accelerator structure.

7 Acknowledgments

The author thanks the Organizers of the ICFA School for the invitation to present this talk in Itacuruçá, and thanks Marcelo Juni (LNLS) and Regis Neuenschwander (LNLS) for their very useful information on vacuum and mechanics. Special thanks goes to Alberto Santoro (UERJ) whose support and determination has made the FPD possible.

References

[1] R. Bonino et al.(UA8 Collaboration), Phys. Lett. B 211 (1988) 239.

[2] A. Brandt et al.(UA8 Collaboration), Phys. Lett. B 297 (1992) 417.

[3] U. Amaldi et al., Phys. Lett. B 43 (1973) 231.

[4] A. Brandt et al., FERMILAB-Pub-97/377

[5] J. Barreto and J. Montanha, DØ Note 3790 (2000)

[6] G. Alves et al., DØ Note 4054 (2002)

High Energy Physics GRID

Alberto Santoro
DFNAE-IF-UERJ
Rua São Francisco Xavier,524
20550-013 Rio de Janeiro – RJ – Brazil
E-mail: Alberto.Santoro@cern.ch

ABSTRACT

Computing is one of the technologies associated to High Energy Physics (HEP). The needs of the next generation CERN HEP experiments push the development of new computing architectures. The scale of data will be orders of magnitude larger than the one we currently have. In this talk we will present the current status of GRID computing development.

1 Introduction

The Grid will permeate all science, allowing groups around the world to collaborate objectively and giving different regions the oportunity to take part in the science frontier. This technology will be present in our professional life for a long time. New chips and computer memmories, new hard disks, the Raid technology, are developments that help to advance the proposed computing architecture for HEP even faster.

1.1 Origin

HEP physicists have always been involved with technology and, since the begining, computing has been present as one of the main tools. Computing is present in the accelerator, in the proposal, during the development, in each detector project, in the data acquisition, in the analysis. Putting it simple, computing is in all parts of a high energy physics experiment. Each experiment builds its own instrumentation and new tools according to its needs. Grid is a natural step forward. Frontiers of science imply frontier of technologies. The World Wide Web was developed by Tim Bernes Lee team at CERN as a need of HEP because collaborations were becoming more participative and comunication was a priority.

The arrival of **New Technologies** (CPU, storage, networks, new languages) will be very useful in the very high energy collisions (14 TeV) environment with a much larger amount of events (tens of PetaBytes).

The four experiments (ATLAS, ALICE, CMS, LHCb) at LHC (Large Hadron Collider) at CERN will collect, in the first year, an amount about **20 Petabytes** of data. Each experiment will involve an average of 1000 physicists. This is an important parameter since all collaborators must have access to the data.

New fast eletronics and/or photonics are being developed for use in the triggers that will have to deal with higher rates.

A micro-society must be created and organized for the sake of efficiency. This means **Virtual Labs** and **Virtual Organizations** in addition to the real HEP.

1.2 Definitions

Before going ahead, I would like to give a few useful definitions acronyms frequently used in the field.

Petabyte: 1 Petabyte = 10^3 Terabyte = 10^6 Gigabyte = 10^9 Megabyte = 10^{12} = Kilobyte = 10^{15} Bytes. It is the unity generally used in the LHC environment.

GRID: is the best computing combination of distributed and shared CPUs and storage, added by higher bandwidths. No bottleneck is accepted in the network.

EGEE[1]: Enabling Grids for E-science in Europe. This is the European organization for Grid in Science. It is a strong collaboration with high energy physics basic software.

OSG[2]: Open Science Grid. This is the simmilar organization for United States Sciences.

LCG[3]: LHC Computing Grid middleware. This is the organization that takes care of the Grid for the four LHC experiments.

GRID3: Is the organization that coordinates the HEPGRID in United States.

All these organizations are consequence of the development of the GRID in HEP and the interest for this technology by other sciences. Many computing professionals and industries are now involved with Grid Technology meaning that many improvements will soon appear.

1.3 New World

HEP has one of the most advanced and suitable structures for GRID development. A GRID has to take into acount each aspect of one HEP experiment as, (i) Project for Particle Accelerators and Detectors; (ii) Data Acquisition Systems and Data Storage; (iii) Development of software and languages; (iv) Data reconstruction using parallel processing; (v)Monitoring, control, simmulation, security, networking; (vi) Data Analysis (one of the most important activities in GRID computing for LHC era); and (vii) Video Conferencing (already used at global level). All this defines a new world or a new way to work. The development of GRID and applications will certainly represent a new revolution in the internet. Several branches of society are being re-organized in view of these new possibilities.

2 Projects

HEP GRID has a number of projects based on the planning of new collidder experiments. These experiments and related Grid projects are summarized here.

2.1 iVDGL + GryPhyN + PPDG = Trilling

Trilling (the name of a beautiful flower) is a coordination of the three main USA HEP Grid projects:**iVDGL**[4]; **GriPhyN**[5] and **PPDG**[6]. It was noticed that there was a considerable overlap of software developments, people and experiments. This coordination has brought benefits to all of them and has started to produce software like VDT (Virtual Data Toolkit) + PACKMAN (package management and distribution tools). Each project is associated to a set of physics experiments:

iVDGL, GriPhyN is associated to LIGO (an experiment for detecting Einstein's Gravitational Waves) + NVO (National Virtual Observatory) + SDSS (Sloan Digital Sky Survey) nnd **PPDG** is associated to BaBar and Nuclear Physics.

We can name these projects as General Grid Projects for HEP. The only common specific purpose is to be dedicated to one or more HEP experiment. The choice of HEP is clear: it is one of the best "laboratories" to develop an idea like GRID. We encourage the reader to go through the web pages listed at the end of this text to have a complete view of each one of these projects.

2.2 GRID for Alice, Atlas, LHCb, CMS

We would like to summarize the LHC experiments and direct the reader to the web pages for more details. They can be used by anyone who would like to be associate to one of these experiments and, eventually, by referees. I use practically the same web pages and information from GRID groups for each LHC experiment that I used in my talk in Lima (SILAFAE).[7] Figure 1 shows the four detectors to be installed at LHC.

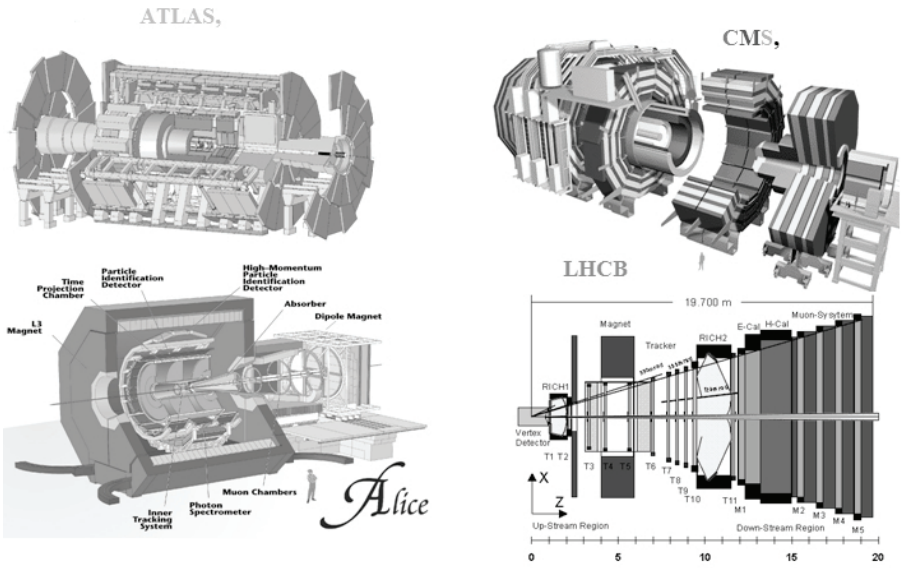


Figure 1: The four detectors under construction for the LHC.

ALICE[8] is dedicated mainly to plasma physics, heavy ions colliding in the center of mass with very high energies. The collaboration count about 1000 physicists from almost one hundred institutions all over the world. The ALIEN (ALIce ENvironment) is a system of softwares created as a grid-like system for job submission and data management. This system intends to help ALICE to build a computing mode to be defined based on functionatility, interoperability, performance, scalability and standards. These properties do not differ much from the ones of the other LHC experiments.

ATLAS[9] is a collaboration of approximately 2000 physicists and engineers from about 150 Institutes from 34 Countries. The Atlas detector weights 7000 Tons, is 25 m in diameter and 46 m long. For more details the reader

should go to the Atlas[9] web page. Many technologies have been developed by ATLAS collaboration to build one of the most interesting general purpose detector. The goals of the experiment are: Detect the Standard Model Higgs Boson, detect Supersymmetric states, study Standard Model QCD (Chromodinamics), EW (Electroweek), HQ (Heavy Quark) Physics, and new physics (to be defined by each collaboration but, in general, we can say that new physics is everything beyond the Standard Model). The Collaboration has been in intensive collaborative development of GRID software. Some of the tools developed in collaboration with LHCb are: GridView (Simple tool to monitor status of testbed), Gripe (unified user accounts), Magda (Manager for Grid Data), Pacman (package management and distribution tool), Grappa (web portal using active notebook technology), GRAT (Grid Application Toolkit), Grdsearcher (browser), GridExpert (Knowledge Database), VOToolkit (Site Authentication,Authorization).

LHCb[10], is a detector dedicated mainly to b physics. The LHCb (Large Hadron Collider Beauty Experiment) experiment has 563 physicists from 50 Institutes from 12 Countries. The experiment expects to get $10^{12} b\bar{b}$ pairs per year, a much higher statistics than the current B factories. Another set of numbers expected for LHCb experiments is: (i) 200,000 reconstructed $B^0 \to J/\psi K_s$ events per year; (ii) 26,000 reconstructed $B^0 \to \pi^+\pi^-$; (iii) all B Mesons and Barions. The LHCb collaboration has produced many useful software for analysis in the near future in cooperation with ATLAS. Two examples, in addition to those pointed out above on ATLAS subsection are GANGA (Gaudi ANd Grid), an user interface for Grid, and DIRAC (Distributed Infrastructure with Remote Agent Control) for Monte Carlo event production. Details of this experiment can be found on the LHCb[10] web page.

CMS[11] collaboration has approximately 2000 physicists from 160 Institutions from about 40 countries. It weights 12,500 Tons, is 15 m in diameter and 22 m in length. It will use a Magnet field of 4 Tesla.

The Detector is composed by Silicon Microstrips as a central Tracker; Electromagnetic and Hadronic Calorimeters; a Superconducting Coil; Iron Yoke; Muon Barrel and Muon Endcaps. CMS will explore a big number of physics topics. In the case of Higgs, the collaboration intends to explore the full range of 100 - 1000 GeV as the allowed region for Higgs mass. Topics as QCD, Heavy Flavor Physics, SUSY, New Phenomena are in the list of future analysis. Diffractive Physics will be explored by the Totem Group (for elastic and total cross section only) and by a large number of physicists interested in exploring all innelastic diffraction and respective topologies. Hard diffraction will be the central part of these studies.

2.3 Other GRID Projects

The idea of the Grid computing architecture can be generalized in science as an useful tool for all data intensive Sciences.

Each area is developing its own framework, with similar characteristics of that used in HEP, many times using almost the same software, as it is the case of the Mamogrid project. Security Services are also common developments using several generic mechanisms (maintaining the data *Integrity, confientiality, authentication, availability*, and several types of appropriated applications.

The interested reader can easily find examples in several fields like medicine (Mammogrid[12]), astronomy (SLOAN-Digital Sky Survey), Gravitation (LIGO, the gravity wave experiment), Biology (the Genoma project), Fusion Physics (project ITER and Gloria D). It is important to keep in mind that all areas with a data intensive system is a good field for the development of a computing Grid.

3 Digital Divide and GRID

Networking is one important component of the GRID systems. In a Grid architecture, all available CPUs are useful only if they are connected by a good bandwidth. The existence of a good bandwidth is the fundamental condition for the stablishing of a good Grid. ICFA (International Committee for Future Accelerators) has set a subcommittee known as SCIC (Standing Committee on Interregional Connectivity) to investigate this question in the HEP field. This committee issued a report[13] that evaluates the world situation of networks and points many regions with strong connectivity problems. The bottleneck may be present in different parts of a network. There are problems like poor last mile connection, long mile connection, poor connectivity in the local institutions network. The idea of Grid for the next generation of experiments will not achieve its purposes if the bottlenecks are not eliminated. See details in the reports of ICFA/SCIC/DD[13].

4 News and Conclusions

After the School HEPGRID has experienced considerable progress in several different projects. HEPGRID BRAZIL[14], that was inaugurated in November at UERJ/BRAZIL with 200 CPUs, is a good example. Progress was also done in the Networks[15] (see the ICFA/SCIC report). Under the leadership of CALTECH, a consortium of institutions (in which Brazilian institutions - UNESP, ANSP, UERJ, RNP - have actively participated) has set a demonstra-

tion that consisted in transmitting a record of 101 Gbps. This was one of the activities of the Bandwidth Challenge of the Supercomputing 2005 conference. We expect that the T2-HEPGRID BRAZIL will soon be connected to the full circuit of CMS Grid. Basic information about Grid systems can be find on the web[16, 17, 18, 19]

5 Acknowledgments

The author thanks the Organizers of the ICFA School for the invitation to present this talk in Itacuruçá, where we found a very good ambient created by the participants. I would like to thank FAPERJ and CNPq by partial financial support. Finally, I thank my colleagues P. Avery, H. Newman, D. Barberis, J. Bunn, R. Gardness, R. Mount, S. Bgnaco, P. Cerello, R. Barbera, P. Buncic, F. Caminati, P. Satz, G. Pulard, N. Brook, C.Eck, J. Marco, F. Gagliardi, T. Wenaus and F. Harri for all the information about their projects. Finally I would like to thank my colleague Helio da Mota for interesting discussions during the preparation of this talk.

References

[1] http://public.eu-egee.org/

[2] http://www.opensciencegrid.org/

[3] http://lcg.web.cern.ch/LCG/

[4] http://www.ivdgl.org/

[5] http://www.griphyn.org/

[6] http://www.ppdg.net/

[7] *Future Experiments - GRID and LHC* Proceedings of the Simpósio Latino Americano de Física de Altas Energias -Lima, Peru, July 12-17,2004.

[8] http://www.alice.cern.ch/

[9] http://www.atlas.cern.ch/

[10] http://www.lhcb.cern.ch/

[11] http://www.cms.cern.ch/

[12] See talk "Mamogrid: applying Grid Technology to Health Care (in breast cancer diagnose)" by Salvator Roberto Amendolia (CERN) in http://www.lishep.uerj.br/

[13] http://icfa-scic.web.cern.ch/ICFA-SCIC/; http://www.weforum.org/site/homepublic.nsf/Content/Global+Digital+Divide+Initiative; and the book "The Digital Divide", Edited by Benjamin M. Campaine, 2003.

[14] http://www.hepgridbrasil.uerj.br/

[15] http://ultralight.caltech.edu/gaeweb/

[16] http://www.cs.wisc.edu/condor/

[17] http://www.globus.org

[18] (i) The Grid Blueprint for a New computing Infrastructure, Edited by Ian Foster and Carl Kesselman, and the second volume too; (ii) Grid Computing Making the Global Infrastructure a Reality - Fram Berman , Anthony J.G. Hey and Geoffrey C. Fox

[19] The Digital Divide - Facing a Crisis or Creating a Myth? - Edited by Benjamin M. Compaine

Laboratory Courses

Laboratory Course on Silicon Sensors

Elisabetta Crescio, Marek Idzik
INFN, Torino, Italy

Danielle Moraes and *Alan Rudge*
CERN, CH-1211, Geneva, Switzerland

ABSTRACT

This laboratory course consists of five different mini sessions, in order to give the student some hands-on experience on various aspects of silicon sensors and related integrated electronics. The five experiments are:

- Characterization of silicon detectors (VCI measurements).
- Double light spot (measurement of the collection time of electrons and holes separately versus voltage in a silicon diode).
- Understanding of the VA read-out chip operation, and measurement of its noise versus detector capacitance characteristics.
- Measurement of the position resolution of a microstrip detector.
- Observation of spectra in a silicon diode.

1 Characterization of silicon diodes for particle detection

1.1 Introduction

Near intrinsic n-type silicon with a metallised p-doped region is the most frequently used semiconductor structure for detecting charged tracks in high-energy physics experiments. A polarisation voltage is applied across the diode structure, which depletes the silicon from charge carriers. Charged particles or photons interacting with the silicon will create electron-hole pairs that drift along the electric field lines to the contacts located on the silicon surface. A schematic picture of a silicon sensor diode is shown in Figure 1.

A first step in constructing a particle detector based upon a silicon sensor is to characterise the sensor without readout electronics attached. The static characteristics of a sensor are usually adequate to determine if the sensor can be used for particle detection. The leakage current behaviour as a function of

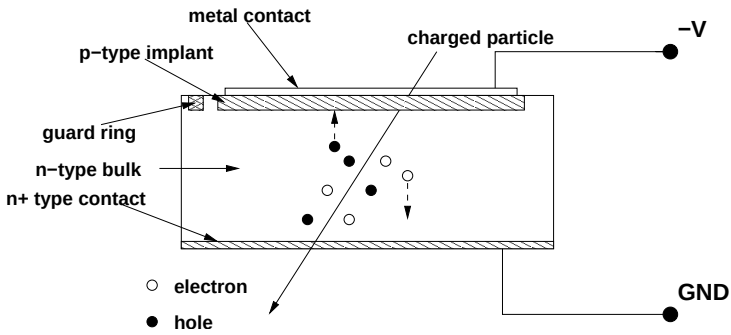


Figure 1: A schematic picture of a silicon sensor diode.

voltage and the voltage needed to fully deplete the sensor are two important parameters. The voltage needed to fully deplete the sensor can be determined by measuring the capacitance between the diode implant and the backplane of the sensor. In the final particle detector system, both the capacitance and the leakage current will influence the performance of the readout electronics. The capacitance and leakage current depend on the geometry of the sensor and the quality of the material and manufactoring process. In a well controlled and uniform process sensors with the same geometrical layout, processed on the same substrate, should have the same behaviour. In reality, there may be variation both in the process and in material and therefore there may be sensors which differ largely from what we naively would expect. When constructing an experiment consisting of many sensors we have to measure them in the laboratory to find the good sensors that can be assembled into the experiment.

This session requires some knowledge of the basic principles of diodes and will give experience handling unprotected diodes, operating the microscope and probe manipulators. Because of the short time available we restrict ourselves to simple DC-coupled diodes.

1.2 Laboratory setup

In this experiment we use the following equipment:

- probe station;
- digital multimeter;
- capacitance meter and power supply;

- vacuum pump;
- microscope;

In Figure 2 a schematic picture of the setup is shown.

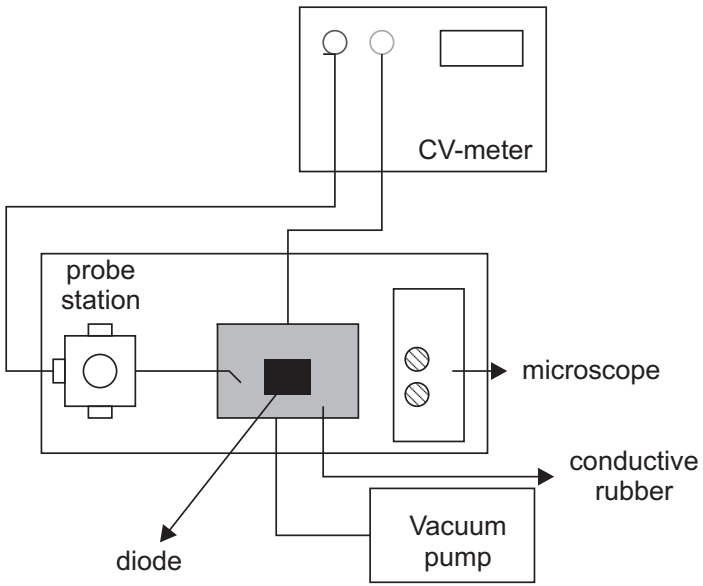


Figure 2: A schematic picture of the setup of CV and IV measurement.

1.3 Measurement of IV-properties of silicon diodes at room temperature

The main sources of leakage current in silicon sensors are:

1. Diffusion of charge carriers from undepleted regions of the detector to the depleted region.
2. Thermal generation of electron-hole pairs in the depleted region.
3. Surface currents depending on contamination, surface defects from processing and edge effect from dicing, etc.

Contribution (1) is generally well controlled and small giving a few nA/cm^2. The contribution from (2) depends largely on the purity of the material since recombination centres and trapping centres increase the creation of electron-hole pairs. The magnitude is higher than that from (1), giving a few μA/cm^2. The leakage current originating from thermal generation is of course temperature dependent. By lowering the temperature of the sensor we may reduce the contribution. By decreasing the temperature by 10^{o}C the leakage current will typically be reduced to a third. In some cases the contribution from (3) may be the dominant source of leakage current. The surface current may be caused by effects on the non-depleted edge region or by a bad processing environment. The leakage current originating from the surface may vary extensively from sensor to sensor. To reduce the effects from surface current a guard ring structure is processed on the silicon. The guard ring can be anything, from a single implant around the diode to a complex structure of alternating implants and floating metal rings around the silicon diode. We will now study the IV-characteristic of a silicon diode sensor with 300 $\mu\,m$ thickness. Execute the steps below:

1. Place the silicon diode on the vacuum chuck with conductive rubber under it.

2. Connect the diode with the probe needle to the negative pole of the battery pack.

3. Connect the chuck (and thus, the silicon backplane) to the positive pole of the battery.

4. Cover the probe station to prevent light from generating current in the diodes, note down the voltage and current at 0 V.

5. Ramp up the voltage and note down the voltage and the corresponding current.

1.4 Measurement of CV-properties of silicon diodes at room temperature

At low reverse bias voltage the capacitance will fall such that $1/C^2$ is proportional to V. When the sensor has reached full depletion the capacitance will not change anymore. This is clearly visible when plotting $1/C^2$ vs. V. The point where the curve shows a kink and does not reduce further gives the point for full depletion. For large single diodes the capacitance to the backplane is the dominant contribution to the total capacitance. For small and segmented sensors such as pixel and microstrip sensors the inter pixel/strip capacitance will dominate over the backplane capacitance when the sensor is fully depleted. The

inter pixel/strip capacitance do not change in the same way as the backplane capacitance when applying reverse bias to the sensor. In this experiment, the capacitance of the silicon diode is measured between the backplane and the p-implant. Execute the steps below:

1. Find the capacitance C_0 of the setup by leaving the circuit open.
2. Follow the procedure outlined for IV-measurements. Measure the full capacitance (capacitance of the diode and capacitance of the setup), calculate the capacitance of the diode by subtracting the value C_0 from the measured one, and note down the measurements.

1.5 Results from the measurements

The results obtained from the measurements are shown in Table 1.5. The corresponding curves are shown in Figure 3. The silicon diode shows a strong increase in leakage current after 40V. This may indicate a breakdown in the structure. The capacitance decreases with higher depletion voltage, and the silicon diode shows the expected behaviour of capacitance vs. depletion voltage. At 60V the silicon diode shows full depletion.

Table 1: IV and CV measurements.

Voltage[V]	Current[pA]	Capacitance[pF]	C_0
0	-12	51	38 pF
10	27	28	
20	56	18	
30	6	16	
40	86	13	
50	114	11	
60	132	10	
70	215	10	
80	360		

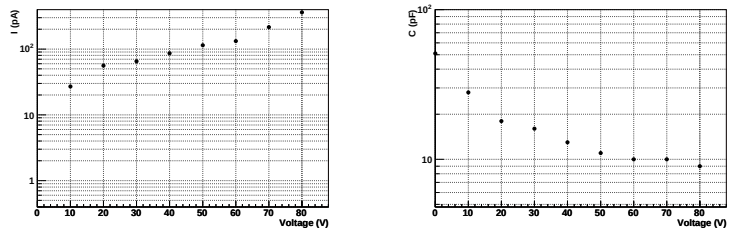


Figure 3: The current of the diode as a function of the applied voltage (left) and the logarithm of the corrected capacitance vs. depletion voltage in V (right).

2 Study of the Viking architecture and noise performance

2.1 The Viking architecture

A number of readout methodologies exist. Among these, the principal ones are the MX series which use double correlated sampling, successfully used in the DELPHI vertex detector, and the Viking "time continuous" shaping. The Amplex/Viking "time continuous" shaping is shown in Figure 4.

A voltage corresponding to the input charge is stored on the sample and hold capacitor and read out sequentially via the output multiplexer. It is important to understand the shift in/out concept to daisy chain many (up to 20) chips. The signals "clockb" and "shift-in" activate the start and stop unit in the chip, which creates an internal clock ("clocki"), a "start" signal for the shift registers and activate signal ("ero") for the output buffer. "Ero" is necessary to allow daisy chaining of chips. The "start" goes into the output multiplexer and "clocki" shifts it through the 128 channels so that each channel is for one clock cycle connected to the output buffer to read the channels out. After 128 clock cycles the outcoming "shift-in" from channel 128 stops the internal clock, disables the output buffer and creates a shift-out which can be used as a shift-in for the next readout chip. Figure 5 shows a timing diagram for the readout sequence of a Viking chip. Corresponding screen shots from a digital oscilloscope are shown in Figure 7(a) and Figure 7(b). Corresponding screen shots from a digital oscilloscope are shown in Figure **??**(a) and **??**(b).

An important mode of operation is the single channel mode, where the output from one amplifier is seen. This is shown in Figure **??**(b) for a signal from an Americium source.

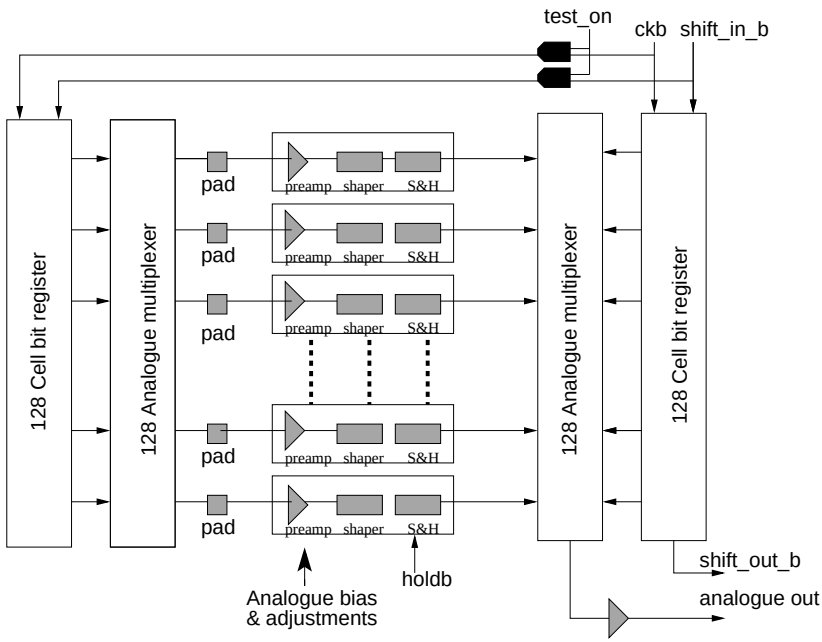


Figure 4: The Amplex/Viking architecture.

2.2 Study of noise performance of the Viking readout circuit

An essential part in operating silicon sensors is a low noise amplifier circuit. The energy needed to create an electron-hole pair at room temperature in silicon is around 3.6 eV. A minimum ionising particle traversing 300 $\mu\, m$ of silicon will on average create 25000 electron-hole pairs. For low energy X-ray applications the requirements for low noise is even more demanding. A 10 KeV X-ray will only produce 2800 electron-hole pairs in silicon. The main contribution of the silicon sensor to the total noise of the assembly comes from:

- the load capacitance of the silicon sensor;
- the leakage current in the silicon sensor;
- possible resistance between the active element of the sensor and ground, or the bias supply.

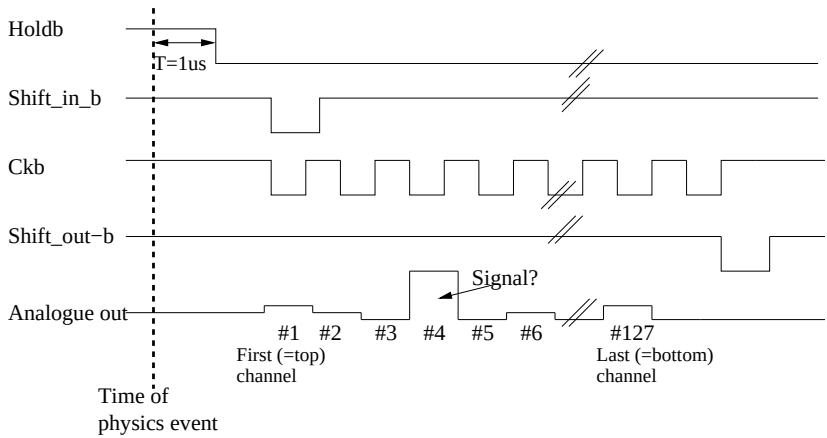
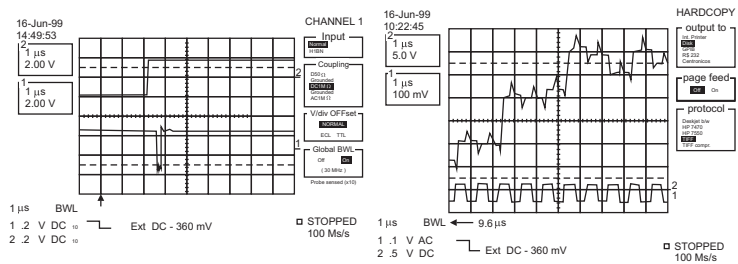


Figure 5: Timing diagram for the readout sequence of a Viking chip.

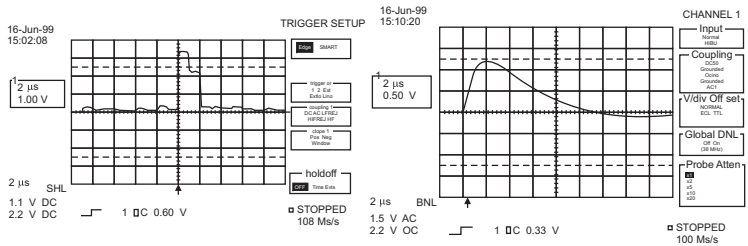


(a) (b)

Figure 6: (a)Screen shots showing hold and shift signals. (b)*Screen shots showing clocking and analogue output pulse.*

The main sources of noise for silicon detector assemblies are: The Equivalent Noise Charge (ENC) of the input FET of the pre-amplifiers proportional to the load capacitance:

$$ENC_{FET} = \frac{Ce}{q}\sqrt{\frac{4kT}{3g_m\tau}} \tag{1}$$



(a) (b)

Figure 7: (a)Screen shots showing output waveform showing hits in two adjacent channels. (b)Output from a single channel.

The detector (and input FET) leakage noise:

$$ENC_l = \frac{e}{q}\sqrt{\frac{qI_l\tau}{4}} \tag{2}$$

The bias and feedback resistor noise:

$$ENC_R = \frac{e}{q}\sqrt{\frac{\tau kT}{2R}} \tag{3}$$

where C is the total load capacitance, I_l is the leakage current, τ is the peaking time of the amplifier, k the Boltzmann constant, T the absolute temperature in Kelvin and R is parallel of the bias resistor and feedback resistor. The parameter g_m is the trasconductance of the MOS device. The total noise contribution is the squared sum of the components listed above:

$$ENC_{total} = \sqrt{ENC_{FET}^2 + ENC_l^2 + ENC_R^2} \tag{4}$$

For the Viking chip the noise is typically $70\,e^- + 12\,e^- C_{load}[pF]$. It is obvious that we want to keep the leakage current small by choosing a sensor that has low leakage current at a voltage that fully depletes the sensor. The contribution to the noise from the detector capacitance and series resistance will be studied in this experiment.

2.3 Laboratory setup

In this session we use the following equipment:

1. Timing unit for readout circuit, VIKING TIMING;
2. Pulse generator for trigger;
3. Current limited power supply for readout circuit;
4. NIM crate;
5. Digital oscilloscope capable of measuring RMS.

We have an assembly with a Viking chip connected to capacitors of various values.

2.4 Measurement

In order to measure the noise, the setup has to be calibrated. This is done by applying a known charge pulse Q_{cal} on the input of the Viking chip and measuring the output from the chip V_{outl}. The input charge is generated by applying a known voltage pulse V_{cal} over a known calibration capacitor C_{cal}:

$$Q_{cal} = \frac{C_{cal} V cal}{e} \tag{5}$$

where e is here the magnitude of the electron charge ($e = 1.6 \times 10^{-19} C$. The calibration capacitance C_{cal} is on our case 1.8 pF. The RMS of the noise V_{noise} from the Viking chip without test pulse in mV can be calculated using a modern digital oscilloscope. Since the calibration is known, the measured value in mV can be converted to ENC by the relation:

$$ENC = \frac{V_{noise} Q_{cal}}{V_{out}} \quad RMS\, e \tag{6}$$

Six channels have been pre-wired with different load capacitances to the input pads of the Viking chip. Determine the calibration and measure the corresponding ENC noise for the channels, and note down the values.

2.5 Results from the measurement

The measured values are reported in Table 2.5.

The plot of noise vs. capacitance is shown in Figure 8.

Table 2: Values of the measured noise.

Capacitance[pF]	V_{cal}[mV]	V_{out}[mV]	V_{noise}[RMS mV]	ENC[RMS e^-]
0	2	2500	10	90
3	2	2500	13	115
13	2	2500	28	250
25	2	2300	43	420
36	2	2200	60	600
51	2	2000	78	880

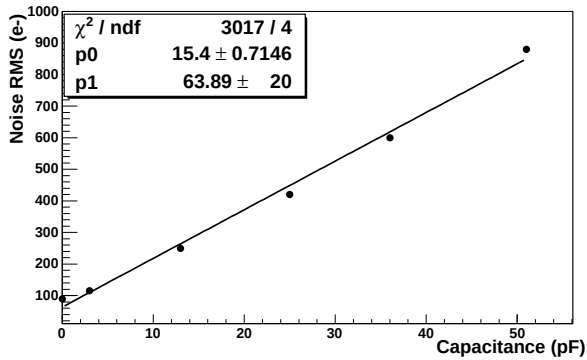


Figure 8: VA2 noise vs. detector capacitance.

3 Study of the position resolution of a silicon microstrip detector

3.1 Introduction

Silicon microstrip sensors are the most commonly used device for high resolution tracking in particle physics. The strip design allows a large sensitive area with relatively few readout channels. The basic strip detector is read out on one side giving information of the track position only in one dimension. Various solutions to measure track position in two dimensions exist. The simplest solution is to glue two single-sided sensors back-to-back, but a more demanding design is to process strips on both sides of the sensor. In this experiment we

will study a single sided sensor which is illuminated by a pulsed laser source on the strip side (because the back of the sensor is fully aluminised).

3.2 The laboratory setup

In this experiment we use the following equipment:

1. Timing unit for readout circuit, VIKING TIMING;
2. Pulse generator for trigger and laser;
3. Current limited power supply for readout circuit;
4. NIM crate;
5. Oscilloscope;
6. Laser diode driver;
7. Battery pack for biasing the sensor;

The silicon microstrip sensor is wire bonded to the Viking readout circuit, which has been placed on a readout PCB (Printed Circuit Board). The assembly has been mounted on a slide which can be precisely moved such that the translation direction is orthogonal to the strips. An optical fibre has been mounted about 100 $\mu\, m$ away from the silicon surface. Figure 9 shows a photograph of the PCB card and the detector mounted on the slide.

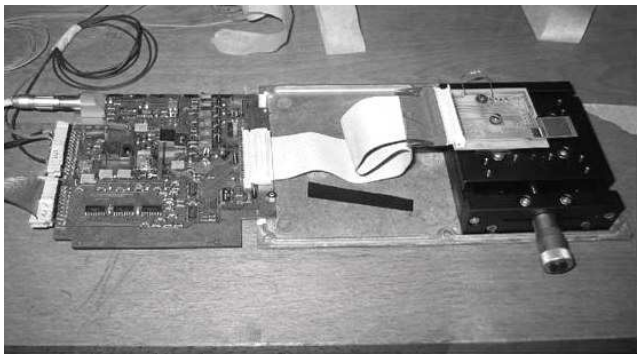


Figure 9: Picture of the PCB board and the detector mounted on the slide.

3.3 Measurement

We will now try to determine the position resolution of the silicon microstrip sensor. The peak of the light source is a few strips wide. We require the information from a number of strips in order to accurately determine the peak position. By moving the fibre closer to the silicon sensor the laser spot size will reduce, but on the other hand we risk mechanically damaging the sensor. Proceed as folows:

1. move the slide by turning the micrometer screw on the right hand side of the box (seen from the repeater electronics board). You will now see the signal from the light moving from one strip to another. You may also notice that this translation is not very smooth. The reason for this is the aluminium on top of the implanted strips which reflect a fraction of the light and reduces the signal locally. This effect would be larger if we had a better focused light spot.

2. Place the sensor in a region with a nicely distributed signal.

3. Determine and write down the amplitudes of the channels in the peak by moving the cursor on the oscilloscope.

4. Four complete turns of the screw correspond to 1 mm. Therefore one major division equals 10 $\mu\,m$. Move the slide by 10 $\mu\,m$ and repeat 3.

5. Repeat 4. and 3.

3.4 Results from the measurement

Table 3.4 shows the measured amplitudes of the signal for different positions of the sensor. Figure 10 shows the oscilloscope output without a laser spot (a), and with the laser spot in two different positions (b and c).

One horizontal division corresponds to about two readout channels. Now perform a gaussian fit for each position and calculate the peak positions. Table 3.4 shows the results from a previous measurement.

Since we are averaging the signals from the Viking in order to minimize the electronics noise, and taking 10 points to determine the peak in the gaussian fit, it can be seen that position accuracy down to a few $\mu\,m$ can be achieved. What are the main error sources?

Table 3: Raw data from the setup.

Strip n	0 $\mu\,m$	10 $\mu\,m$	20 $\mu\,m$	30 $\mu\,m$	40 $\mu\,m$	50 $\mu\,m$
1	304	292	268	240	232	212
2	424	400	368	340	320	288
3	612	584	544	504	472	436
4	888	844	808	760	716	680
5	1030	1000	984	948	912	884
6	1150	1140	1130	1100	1080	1050
7	1140	1160	1170	1150	1130	1120
8	1010	1070	1100	1130	1160	1200
9	736	772	820	864	916	968

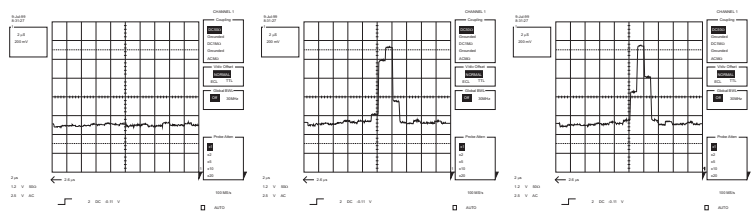


Figure 10: Oscilloscope output without a laser spot (a), and with the laser spot in two different positions (b and c).

Table 4: Results from the position resolution measurement (in $\mu\,m$).

	0 $\mu\,m$	10 $\mu\,m$	20 $\mu\,m$	30 $\mu\,m$	40 $\mu\,m$	50 $\mu\,m$
Calculated peak	264	274	282	292	302	312
Difference		10	8	10	10	10

4 Study of charge transport in silicon with a fast amplifier

4.1 Introduction

In gaseous detectors the mobility for electrons is several orders of magnitude higher than for positive ions. In semiconductors the mobility for holes is only slightly lower than for electrons. In general the signal propagation in semiconductors is a few nanoseconds while the signal propagation in gaseous detectors typically varies from microseconds to milliseconds. In order to study the drifting of electrons and holes in silicon, very fast electronics is required. The drift velocity for electrons and holes in silicon at low electric field strength is given by:

$$v_e = \mu_e E \tag{7}$$

$$v_h = \mu_h E \tag{8}$$

where μ_e and μ_h are the mobilities for electrons and holes respectively, and E is the electric field. The mobility in silicon at room temperature is 1350 cm^2/Vs for electrons, and 480 cm^2/Vs for holes. At high field, the velocity saturates with velocities of the order of $10^7\, cm/s$.
When the sensor is fully depleted, the signals from holes and electrons will arrive almost at the same time. We can try to study the slower transit of holes by shining a short laser pulse on the back side of a non-depleted n-type silicon diode. By choosing a wavelength which does not penetrate far in the silicon the charge can be generated close to the surface of the silicon. The holes have now to drift to the other side of the sensor, while the electrons are formed at the interface. If the electric field is low and the detector has a reasonably large depleted region we will now see a difference in the total time between the signal arising from electrons and holes. It is important to realise that the signal itself arises *immediately* as the charge carriers start moving, and lasts until the last charge carriers are collected.
By starting with a high depletion voltage the signal from the setup will have the same shape as the light pulse from the laser which is shown in Figure 11.

The current signal which is induced on the diode is due to the movement of both charge carriers in the electric field. The transit time for the carriers can be calculated by integrating Equations 7 and , remembering that the field varies as a function of position:

$$E(x) = \frac{qN_D}{\epsilon}x + E_{min} \tag{9}$$

where ϵ is the dielectric constant, E_{min} is a function of the bias voltage, and is

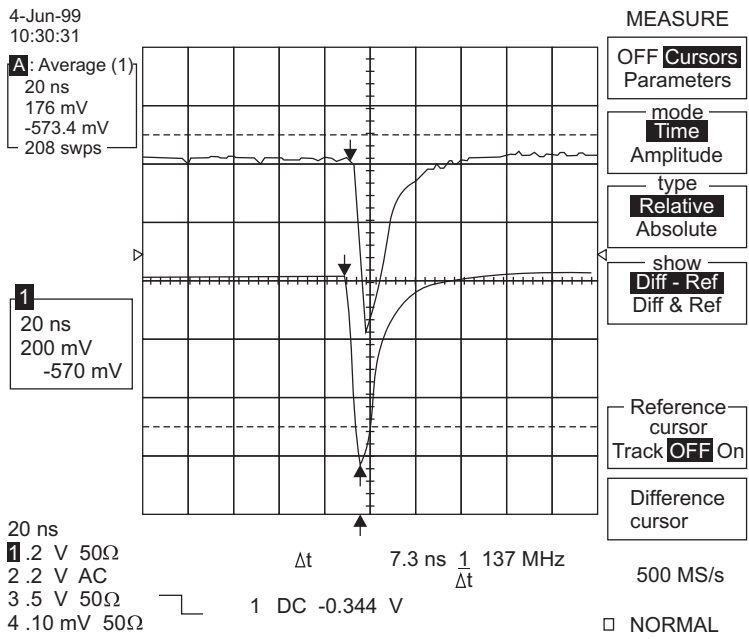


Figure 11: The amplifier output from the laser pulse with a fully depleted silicon diode.

a minimum of zero for a barely depleted detector, then the transit times are:

$$t_h = \frac{\epsilon}{\mu_h N_D} ln\Big(\frac{w + (\epsilon/qN_D)E_{min}}{x_0 + (\epsilon/qN_D)E_{min}}\Big) \tag{10}$$

$$t_e = \frac{\epsilon}{\mu_h N_D} ln\Big(\frac{x_0 + (\epsilon/qN_D)E_{min}}{(\epsilon/qN_D)E_{min}}\Big) \tag{11}$$

where x_0 is the depth of production, measured from the ohmic contact side. The induced current is given by Ramos theorem, which states that the current on the electrode of interest is equal to the charge value multiplied by the dot product of the "weighting field" and the charge velocity. The weighting field is an hypothetical field calculated by putting unit potential on the electrode of interest, zero on all other electrodes, and ignoring any static charges (N_D in our case). For a simple diode, this reduces to a constant $1/w$, and the velocities

are aligned with the electric field, so calculating the scalar velocities gives the induced current:

$$i_h = q\frac{1}{w}\mu_h\Big(E_{min} + \frac{x_0 q N_D)}{\epsilon}\Big)\Big(exp\Big[\mu_h q\frac{N_D}{\epsilon}t\Big]\Big) \tag{12}$$

$$i_e = q\frac{1}{w}\mu_e\Big(E_{min} + \frac{x_0 q N_D)}{\epsilon}\Big)\Big(exp\Big[-\mu_e q\frac{N_D}{\epsilon}t\Big]\Big) \tag{13}$$

In fact these pulses become smeared by the amplifier bandwidth, and the resulting pulse is what is seen on the oscilloscope.

4.2 The setup

In this experiment we use a fast amplifier chain connected to a diode of n-type silicon with the backplane not covered by aluminium. Two light fibres lead the laser light to the two opposite sides of the diodes. A picture of the sensor setup, with a zoom of the sensor region, is shown in Figure 12.

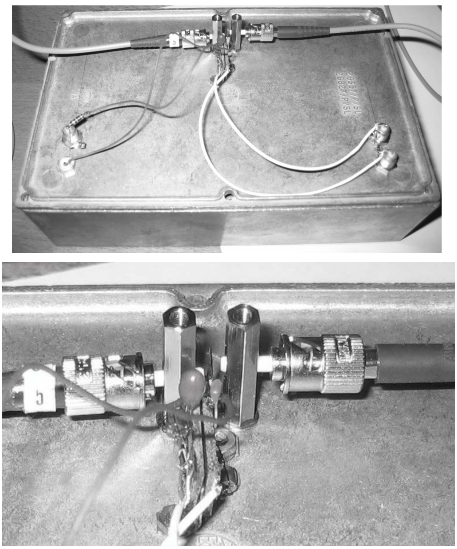


Figure 12: Picture of the sensor setup.Top:box containing the detector with the two optic fibers and cables for the detector polarization and read-out. Bottom: zoom on the detector and electronics.

Other equipment used are:

1. Oscilloscope;
2. Laser diode driver;
3. Battery pack for biasing the sensor;
4. Pulse generator for trigger and laser;
5. Low voltage power supply;
6. NIM crate.

4.3 Signals when shining laser on the junction side (p-side)

With the laser pulse on the p-side, you get the picture on oscilloscope shown in Figure 13.

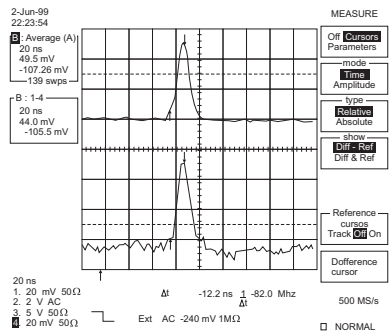


Figure 13: The signal from an unbiased sensor (read out from the p-side) when laser is shined on the p-side.

Turn on the sensor bias and ramp slowly up the voltage. The amplitude of the signal gets larger but the shape stays approximately unchanged. The sensor is depleted from the p-side and the n-side is conducting transporting the electrons to the amplifier.

4.4 Signals when shining laser on the backplane (n-side)

When the pulse is incident on the n-side and the detector is off, there is a very tiny signal. The holes are trapped close to the backplane. As the bias is increased, one sees an increasing signal (see Figure 14), with a double peak

structure. In the double peak structure are visible the contribution coming from fast arriving electrons and the contribution from holes with a lower mobility which have to travel through the depleted region of the sensor.

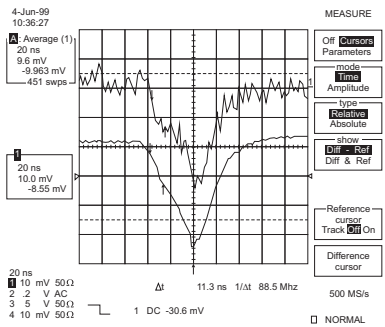


Figure 14: The signal amplitude at low bias voltage when laser is shined on n-side of the silicon sensor (The signal is read out from the n-side).

The double peak structure disappears when the bias voltage gets higher. At full depletion the amplitude of the signal is comparable to that obtained when laser is shined on p and n sides (see Figure 15).

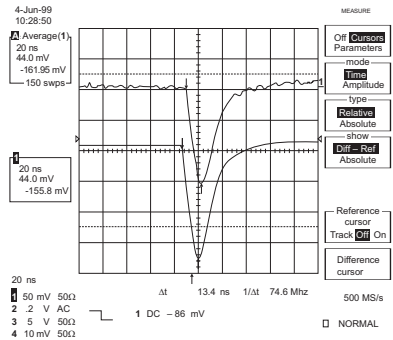


Figure 15: The signal amplitude at full depletion when laser is shined on n-side (read out from n-side).

The optical fibre cables are not labeled; can you discover, by

using one or the other cable, which cable points on the n-side, and which on the p-side?

5 Spectroscopy with the Viking chip and pad detector

5.1 Introduction

During this experiment we will use a silicon sensor to see the spectrum form a γ source and study its energy resolution.

5.2 The setup

The VIKING chip is used in single channel mode and the inputs bonded to a small silicon detector with 36 pads of 1 mm square. Channel 12 is bonded to 1 pad, channel 36 to 2 pads, channel 58 to 4 pads, channel 78 to 8 pads and channel 102 to 21 pads. The other equipment used in this experiment is the following:

1. Oscilloscope;
2. ^{241}Am source;
3. Battery pack for biasing the sensor;
4. Viking Timing unit;
5. Low voltage power supply;
6. NIM crate.
7. Multi Channel Analyzer and PC.

A picture of the sensor setup is shown in Figure 16.

5.3 The measurement

A spectrum of ^{241}Am is taken with channel 12 (only one pad) with the Am source placed under the printed circuit board hosting the detector. The spectrum is shown in Figure 17. The peak corresponding to the 59.95 X-ray line is clearly visible at channel 54. Since the detector is supported on the board with copper and the contact to the backplane of the sensor is done with silver loaded glue, other peaks at lower energies are observed, corresponding to florescence lines excited by the 59.95 keV line from copper and silver. The K_α

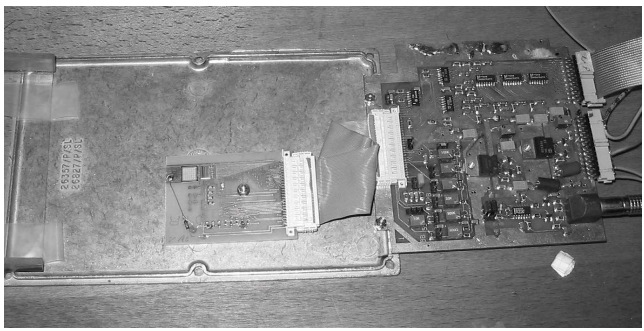


Figure 16: Photograph of the setup. The sensor and the readout chip are mounted on a printed circuit board which is placed into a aluminium box.

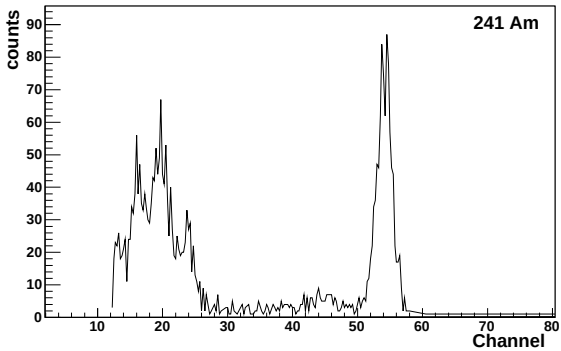
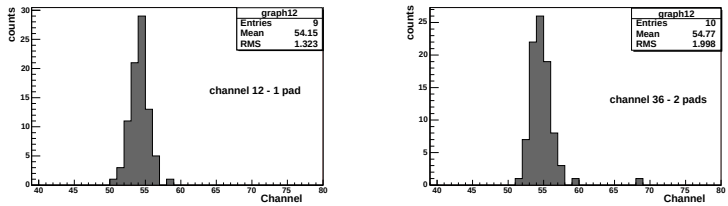


Figure 17: Spectrum from a ^{241}Am source. The source is placed under the printed circuit board hosting the detector.

and K_β lines for copper and silver are: Cu (K_α) 8.03 keV, Cu (K_β) 8.90 keV, Ag (K_α) 21.99 keV and Ag (K_β) 24.94 keV.

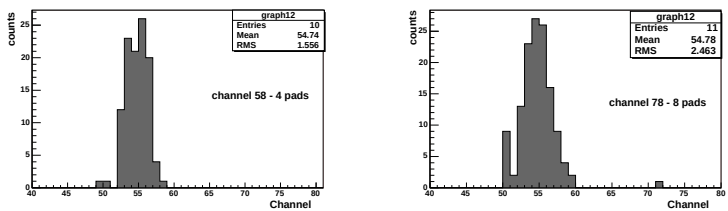
Successively, other measurement are taken with the source placed into the aluminium box on a "bridge" above the sensor. In this case only the peak corresponding to the 59.95 keV line is visible. In the following, the peaks obtained reading out channels 12,36,58,78,102 are shown.

From Figures 18,19 and 20 it can be seen that the energy resolution deteriorates increasing the number of the readout pads, due to the increase of the sensor capacitance. If it assumed that the energy resolution is limited by



(a) (b)

Figure 18: Spectrum from a ^{241}Am source obtained from channel 12, with only one pad connected (a), and from channel 36 with two pads connected (b).



(a) (b)

Figure 19: Spectrum from a ^{241}Am source obtained from channel 58, with 4 pads connected (a), and from channel 78 with 8 pads connected (b).

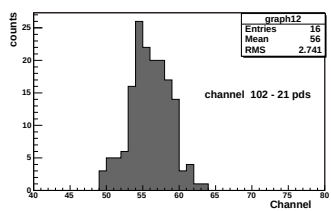


Figure 20: Spectrum from a ^{241}Am source obtained from channel 102, with 21 pads connected.

the electronic noise, then the noise can be calculated knowing the RMS of the energy distribution:

$$Noise\ (RMS\, e-) = \frac{RMS(eV)}{3.6eV} \tag{14}$$

where 3.6 eV represent the energy required to create an electron-hole pair from and incident particle.