Pulse Oximetry

The principal advantage of optical sensors for medical applications is their intrinsic safety since there is no electrical contact between the patient and the equipment. (An added bonus is that they are also less susceptible to electromagnetic interference). This has given rise to a variety of optical techniques to monitor physiological parameters: for example, the technique of laser doppler velocimetry to measure red blood cell velocity. However, in this lecture course, we will concentrate on the technique of pulse oximetry for the non-invasive measurement of arterial oxygen saturation in the blood.
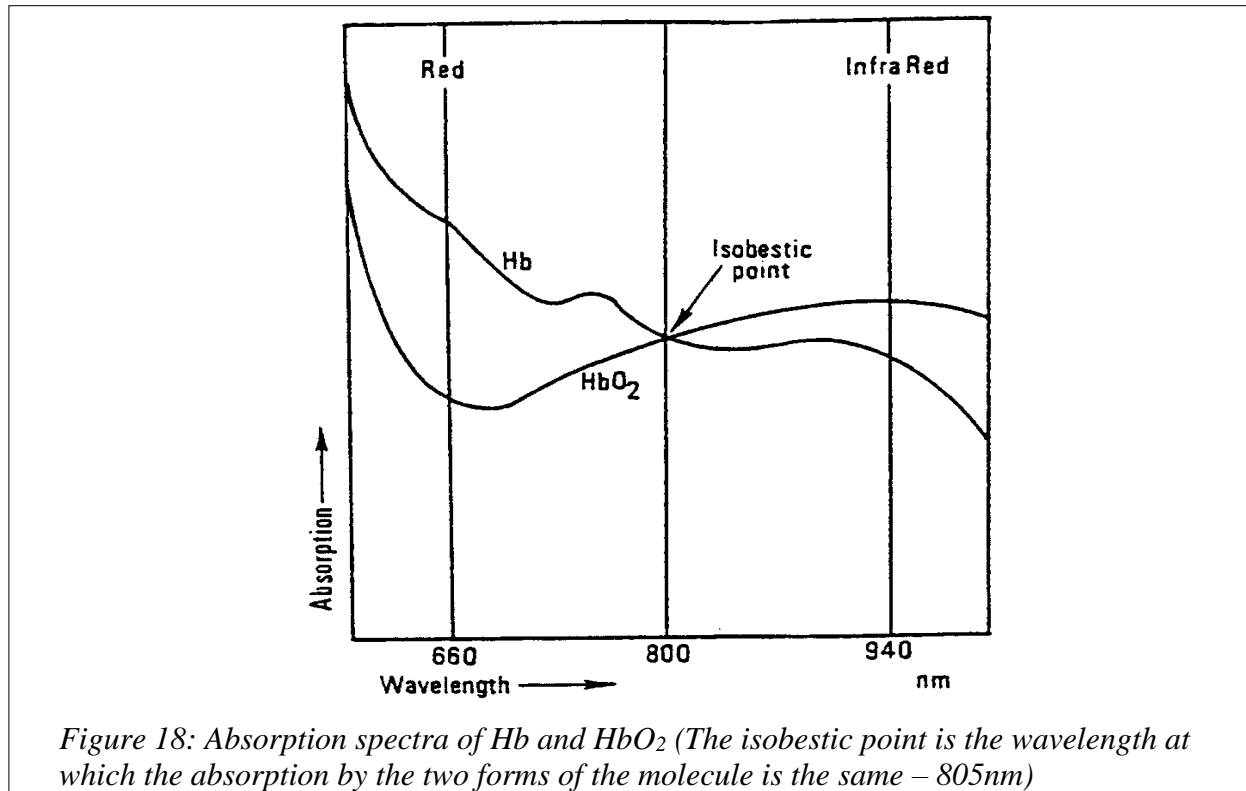
For patients at risk of respiratory failure, it is important to monitor the efficiency of gas exchange in the lungs; i.e., how well the *arterial* blood is oxygenated. Preferably, such information should be available to clinicians on a continuous basis (rather than every few hours). Both of these requirements can be met non-invasively[1] via wearables with the technique of *pulse oximetry*. The technique is now well-established and is in regular clinical use during anaesthesia and intensive care (especially *neonatal* intensive care since many premature infants undergo some form of ventilator therapy). Pulse oximetry is also being used in the monitoring of pulmonary disease in adults and in the investigation of sleep disorders.

Principles of oximetry

It was discovered in the 1860s that the coloured substance in blood, *haemoglobin*, was also its carrier of oxygen. (Haemoglobin is a protein that is bound to the red blood cells.) At the same time, it was noticed that the absorption of visible light by a haemoglobin solution varied with oxygenation. This is because the two common forms of the molecule,

---

1 i.e. using instrumentation which does not require surgery

oxidised haemoglobin (HbO₂) and reduced haemoglobin (Hb), have significantly different optical spectra in the wavelength range from 500nm to 1000nm, as shown in Figure 18.

Figure 18: Absorption spectra of Hb and HbO₂ (The isobestic point is the wavelength at which the absorption by the two forms of the molecule is the same – 805nm)

The oxygen chemically combined with haemoglobin inside the red blood cells makes up nearly all of the oxygen present in the blood (there is also a very small amount which is dissolved in the plasma). Oxygen saturation, which is usually referred to as SaO_2 in oximetry (SpO_2 for pulse oximetry), is defined as the ratio of oxyhaemoglobin (HbO_2) to the total concentration of haemoglobin present in the blood (i.e. oxyhaemoglobin + reduced haemoglobin):

$$SaO_2 = \frac{[HbO_2]}{[Total\ haemoglobin]}$$

Arterial SaO_2 is the parameter measured with oximetry and is normally expressed as a percentage. Under normal physiological conditions arterial blood is 97% saturated, whilst venous blood is 75% saturated.

It is possible to use the difference in absorption spectra of HbO_2 and Hb for the measurement of arterial oxygen saturation *in vivo* because the wavelength range between 600 nm and 1000 nm is also the range for which there is least attenuation of light by body tissues (tissue and pigmentation absorb blue, green and yellow light and water absorbs the longer infra-red wavelength).
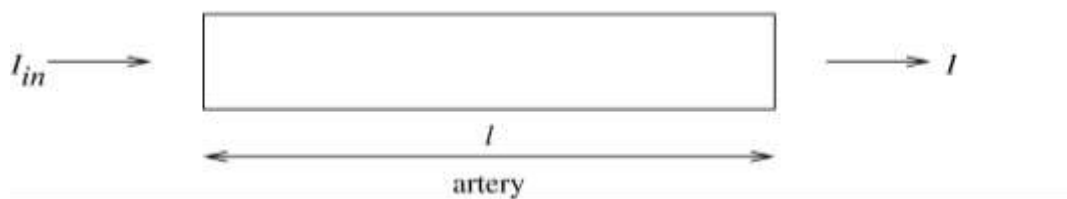

Figure 19: Model artery for Beer-Lambert law

By measuring the light transmitted through the fingertip (or the earlobe) at *two* different wavelengths, one in the red and the other in the near infra-red part of the spectrum, the oxygen saturation of the arterial blood in the finger (or ear) can be determined. If we assume initially that the transmission of light through the arterial bed is influenced only by the relative concentrations of HbO_2 and Hb and their absorption coefficients at the two measurement wavelengths, then the light intensity will decrease logarithmically with path length according to the well-known Beer-Lambert law. If we consider, as shown in Figure 19, an artery of length l through which light, initially of intensity I_{in} is shone, the Beer-Lambert law states that the intensity I is given by:

At wavelength λ_1, $I_1 = I_{in1} 10^{-(\alpha_{01}C_0 + \alpha_{r1}C_r)l}$

At wavelength λ_2, $I_2 = I_{in2} 10^{-(\alpha_{02}C_0 + \alpha_{r2}C_r)l}$

where

- C_o is the concentration of oxyhaemoglobin (HbO$_2$)

- C_r is the concentration of reduced haemoglobin (Hb)

- α_{on} is the absorption coefficient of HbO$_2$ at wavelength λ_n

- α_{rn} is the absorption coefficient of Hb at wavelength λ_n

If we let

$$R = \frac{\log_{10}(I_{in1}/I_1)}{\log_{10}(I_{in2}/I_2)}$$

then it is simple enough to show:

$$SaO_2 = \frac{C_0}{C_0 + C_r} = \frac{\alpha_{r2}R - \alpha_{r1}}{(\alpha_{r2} - \alpha_{02})R - (\alpha_{r1} - \alpha_{01})} \tag{1}$$

Equation (1) simplifies further if λ_2 is chosen to be the isobestic wavelength, in which case $\alpha_{r2} = \alpha_{02}$ (at 805 nm).

A brief history of oximetry

The first devices to measure oxygen saturation in human blood by transilluminating it with coloured light were built in the 1930s. These devices were incapable of distinguishing between arterial and venous (and capillary) blood. An attempt to exclude venous and capillary blood was then made by using one of two methods: zeroing the oximeter by taking a "bloodless" reading from an earlobe compressed between two fingers or arterialising the blood by heating it to 43°C (this method was

developed during the Second World War as part of a project to investigate the problem of the loss of consciousness by R.A.F. pilots during dog-fights). However, the light transmitted through the ear (or finger) is not only attenuated by the arterial, venous and capillary blood but also by the skin (whose pigmentation and hence absorption properties will vary from person to person) and by other tissues such as muscle, bone, etc.

In the early 1970s, the Hewlett-Packard company developed an instrument which attempted to circumvent these problems by measuring the transmission of light across the earlobe at more than two wavelengths. A multi-component model of the ear was set up, composed of m light-absorbing substances (skin, tissues, Hb and HbO_2, etc.). The model further assumed that each light absorber acted independently of the others. A set of simultaneous equations were then written for the total absorbance of light by the m substances at each of the measurement wavelengths. Empirical calibration coefficients were derived from a number of studies on a sample of volunteers. However, the high cost of the instrumentation together with the need for measurements at *eight* different wavelengths meant that it never found regular clinical use.

Principles of pulse oximetry

It is the recent development of *pulse* oximetry which has led to oximetry being accepted as a useful non-invasive technique for the measurement of arterial SaO_2. With pulse oximetry, only that part of the signal directly related to the inflow of arterial blood into the body segment is used for the calculation of oxygen saturation. The intensity of light transmitted across the fingertip, for example, varies as shown in Figure 20. A pulsatile signal, which varies in time with the heart beat, is

superimposed on a d.c. level. (The amplitude of this cardiac-synchronous pulsatile signal is approximately 1% of the d.c. level.)
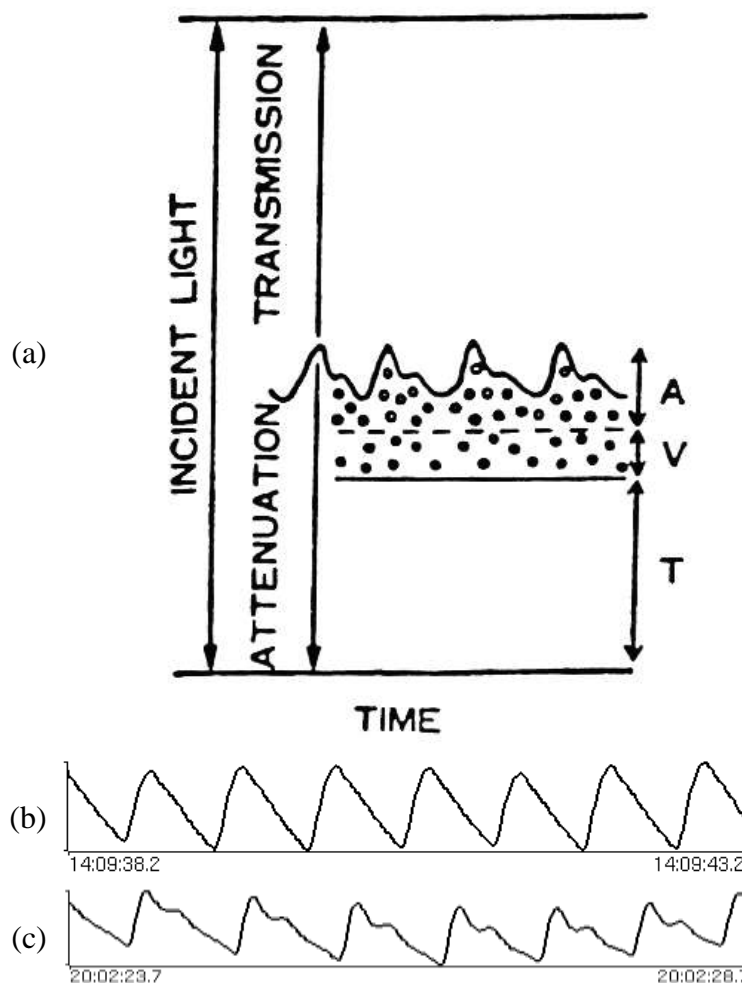
(a)

Figure 20: (a) Transmission of light through the finger when the attenuation of light is caused by arterial blood (A), venous blood (V) and tissues (T); (b) and (c) show typical pulsatile signals recorded with a finger probe when light is shone through the finger.

Pulse oximetry assumes that the attenuation of light by the body segment can be split into the three independent components shown in Figure 20(a): arterial blood, venous blood and tissues.

If we assume that the *increase* in attenuation of light is caused only by the inflow of arterial blood into the fingertip, we can calculate the oxygen saturation of the arterial blood by subtracting the d.c.

component of the attenuation from the total attenuation, leaving only the cardiac-synchronous pulsatile component for the dual-wavelength determination of oxygen saturation.

It is a simple matter to show that the general oximetry equation derived earlier is equally valid for pulse oximetry if R is now given by:

$$R = \frac{\log_{10}\left(\left(I_{dc+ac}\right)/\left(I_{dc}\right)\right)_{\lambda 1}}{\log_{10}\left(\left(I_{dc+ac}\right)/\left(I_{dc}\right)\right)_{\lambda 2}} \tag{2}$$

Figures 20(b) and 20(c) show typical cardiac-synchronous pulsatile waveforms recorded when red or near infra-red (NIR) light is shone through a finger. Note that the d.c. content (or baseline) has been removed from these traces. It is clear from these plots that there is a wide variation in the shape of the waveforms between subjects - note the secondary peak for each heart beat in the plot shown in Figure 20(c). This peak is relatively common and is known as the *dicrotic notch*.

Calibration of pulse oximeters

The first pulse oximeters, which were manufactured in the early 1980s, used equation (1) to compute the values of SpO_2. However, the Beer-Lambert law, on which this equation is based, does not take into account the multiple scattering of light by the red blood cells. Although oximetry is a differential technique, the effect of scattering is only partially compensated for since scattering is wavelength dependent.

Equation (1) is therefore an over-simplification: Figure 21 shows the relationship between oxygen saturation and the ratio R for two cases: when it is calculated using the Beer-Lambert law and when it is based on empirical data. Consequently, instruments based on the Beer-

Lambert law tended to give erroneous estimates of the true value of oxygen saturation, especially at low values. There have been a few attempts to modify the theory in order to take scattering into account, but most pulse oximeters use look-up tables derived from calibration studies on large numbers of healthy volunteers whose oxygen saturation is also measured invasively.
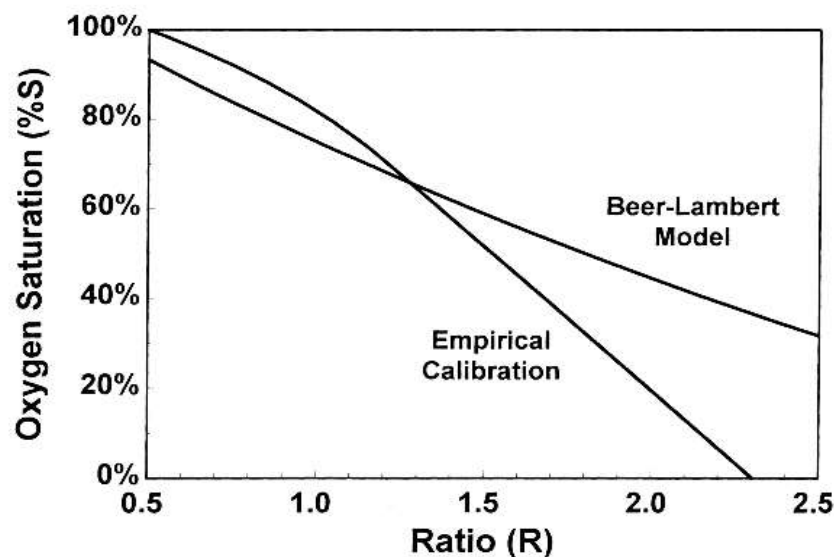

Figure 21: Two relationships between the ratio R and the Oxygen Saturation of the patient

Design of pulse oximetry instrumentation

A block diagram of the circuit for a pulse oximeter is shown in Figure 22. The main sections of this block diagram are now described.
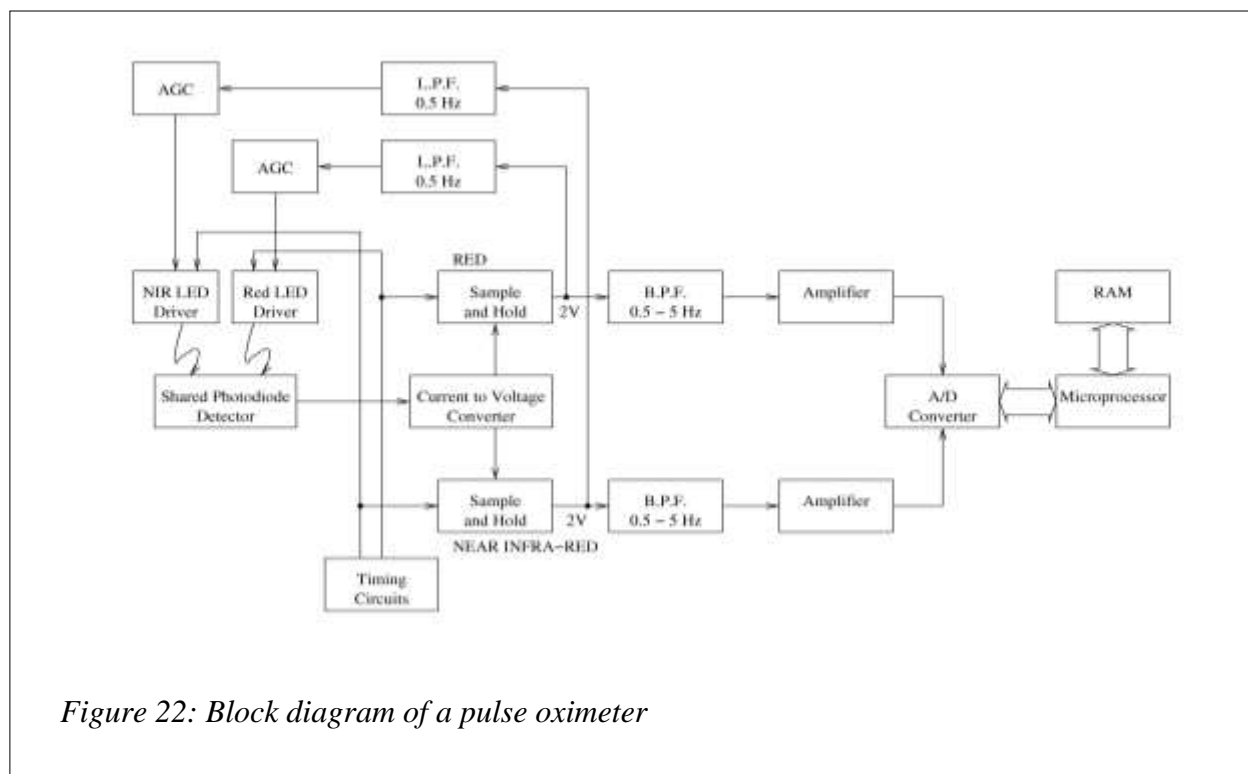
Figure 22: Block diagram of a pulse oximeter

In order to build finger (or earlobe) probes which are small and unobtrusive, we need miniature light sources and detectors. Light-emitting diodes (LEDs) which work in the red and Near-Infra Red (NIR) part of the spectrum are readily available. However, the average power which can be obtained from *standard* LEDs is limited and a very sensitive detector (such as a photomultiplier tube) would be required to detect the small amount of light transmitted through the finger.

This problem can be overcome by using special-purpose LEDs; for example, red LEDs are available with internal lensing systems to give high intensity outputs. Similarly, high current NIR LEDs are designed to be *pulsed* so that the *peak* power available from them can be increased without increasing the average power. This makes it possible to detect the light transmitted through the finger with a simple, compact, solid-state photodetector such as a photodiode.

If we pulse *both* light sources, we can then use a single photodetector in the finger probe, since silicon devices are responsive to light in the visible and NIR parts of the spectrum. We could, for example, use

timing circuits to supply, say, 50 μs pulses to the red and NIR LED drivers at a repetition rate of 1 kHz, as shown in Figure 23 (a frequency of 1 kHz is suitable because such a frequency is well above the maximum frequency present[2] in the arterial pulse).
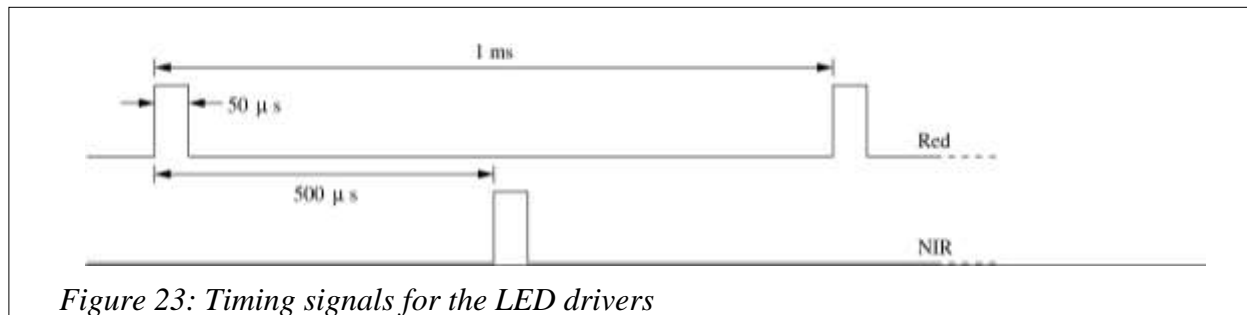

Figure 23: Timing signals for the LED drivers

In this mode of operation, high-intensity light outputs can be obtained with the NIR LED with currents of up to 1A over a low duty cycle. The transmitted light detected by the photodiode is amplified and converted to a voltage using an op-amp configured as a current-to-voltage converter. At this point in the circuit the signal is fed to two identical sections, one for each of the transmitted wavelengths.

Since the light is pulsed, we need to use a sample-and-hold circuit to reconstitute the waveforms at each of the two wavelengths. The same timing circuits which were used to control the red and NIR LED drivers are also used to provide the control pulses for the corresponding sample-and-hold circuits. The outputs from these circuits are then filtered with a band-pass filter (with 0.5 Hz and 5 Hz cut-off frequencies) in order to remove primarily the d.c. component but also high-frequency noise. The resulting signals thus represent the cardiac-synchronous information in the waveforms and these are further amplified before they are converted to digital format for subsequent analysis by the microprocessor.

---

2 The maximum cardiac frequency is never going to be more than a few Hz.

It can be seen from the block diagram in Figure 22 that the output from each sample-and-hold is also passed to a low-pass filter. This is the first stage of an automatic gain control (AGC) circuit which adjusts the light intensity from the corresponding LED so that the d.c. level always remains at the same value (say 2V) whatever the thickness or skin characteristics of the patient's finger.

There are two equally important reasons for deciding to use an AGC circuit: firstly, it means that the amplitude of the a.c. signal (which may vary between 0.1% and 2% of the total signal) is also within a pre-defined range and this makes the amplifier which follows the band-pass filter easier to design. Secondly, the d.c. component of the transmitted red and NIR signals can be set at the same value (2 V) in each case. Hence it can be eliminated from the formula used by the microprocessor to calculate the oxygen saturation. A new index, $R' = \log(I_{ac})_{\lambda 1} / \log(I_{ac})_{\lambda 2}$ is defined (compare with Equation (2)).

In practice, it is not even necessary to convert the a.c. signal amplitudes at the two wavelengths to their logarithmic equivalents: instead a look-up table can be loaded into memory and this will contain the values of oxygen saturation corresponding to each value of the (Red pulse amplitude)/(NIR pulse amplitude) ratio.

We will now consider in turn each of the main circuits shown in the block diagram.


Timing circuit

In this application, the accuracy of the timing is not of paramount importance, hence the timing circuit can be built around the 555 timer integrated circuit. From the data sheet for this i.c., it can easily be worked out that the circuit given in Figure 26 can be configured (for

example by setting $C = 22$ nF, $R_a = 56k\Omega$ and $R_b = 3.3k\Omega$) to give a 50 μs pulse approximately every millisecond, as intended. (N.B., this course does not expect you to remember how to configure a 555 timer.)

$$T_1 = 0.7(R_a + R_b)C$$

$$T_2 = 0.7R_bC$$

Figure 26: Generating the timing pulses for pulse oximetry
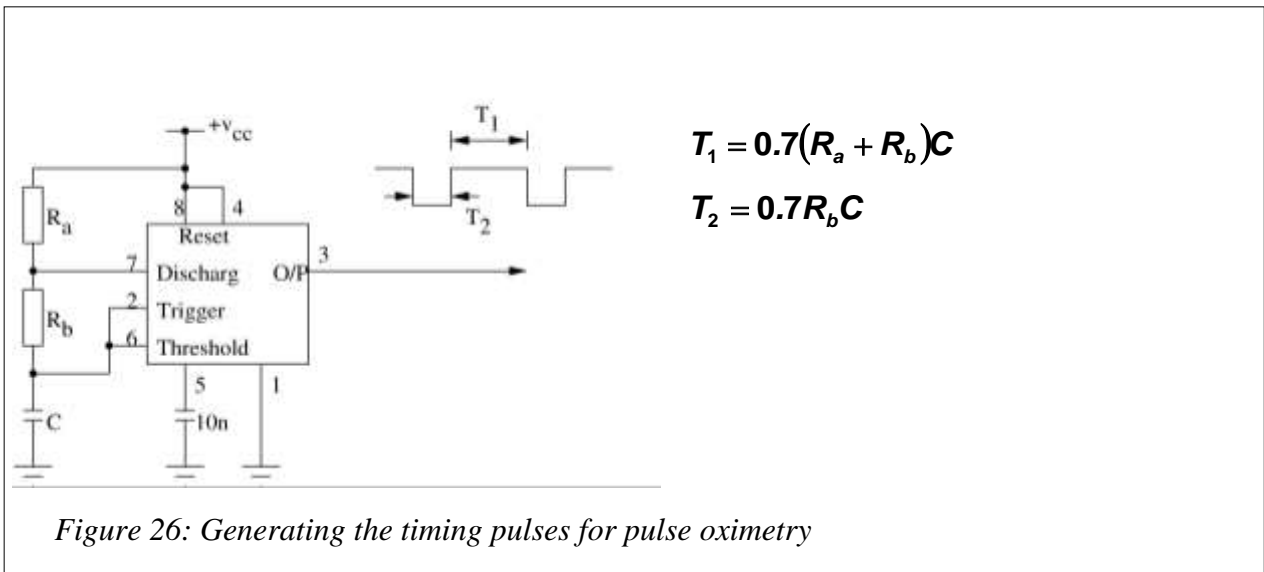
Pulsing the light output from the LEDs

The output from the LED can be pulsed by connecting an n-channel enhancement-mode MOSFET across it as shown in Figure 27. The pulses from the output pin of the 555 timer (pin 3) are taken to the gate of the transistor. The FET needs to be an enhancement-mode MOSFET for it to be turned fully off and on by the gate pulses. The MOSFET chosen for this task should also be capable of handling the maximum current flowing through the LED.

Figure 27: Pulsing the LED
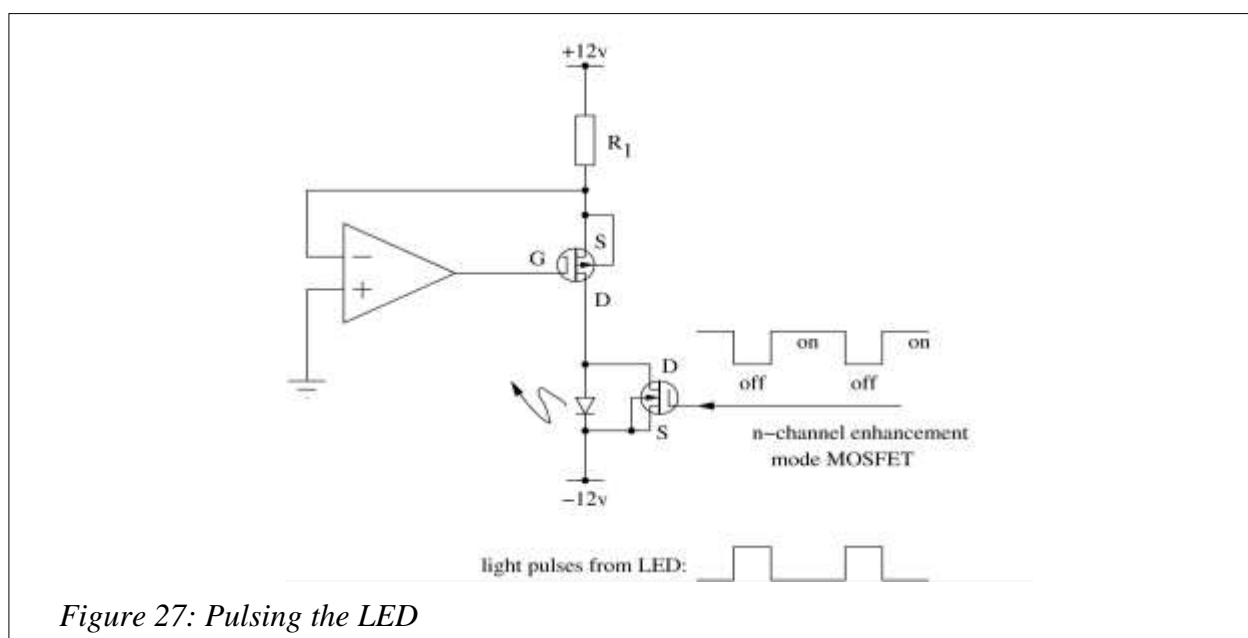
Receiver circuit

The simplest solid-state optical detector is the photodiode. Photodiode detectors normally operate with reverse bias applied to the p-n junction (photoconductive mode). When light falls on the junction region of the photodiode, an electron-hole pair is created; under the influence of the junction (or built-in) field, the hole is swept towards the p-material and the electron towards the n-material. The resulting light current is seen as a large increase in the reverse current.
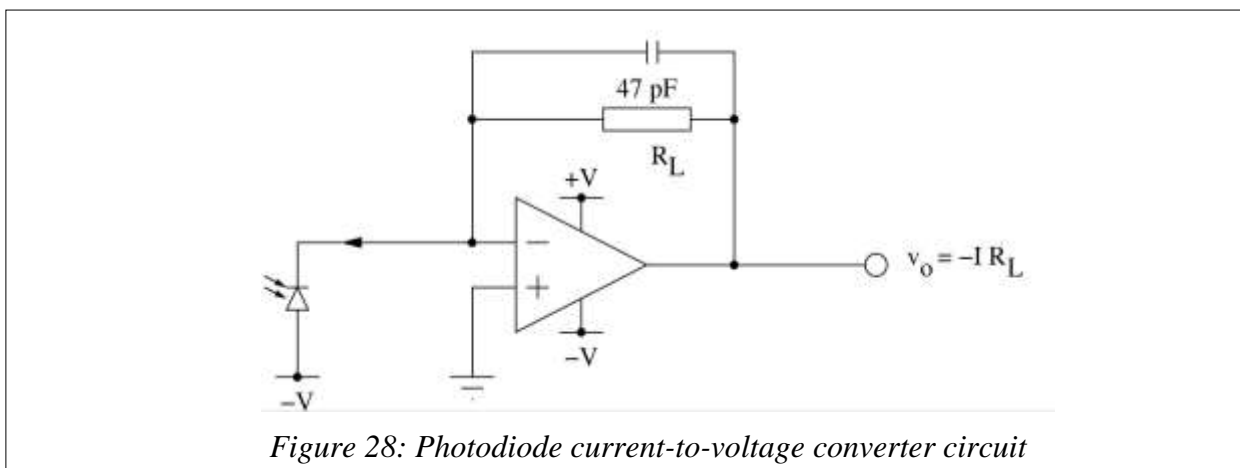

Figure 28: Photodiode current-to-voltage converter circuit

For the purposes of signal amplification, the photocurrent must be transformed into a voltage with moderate output impedance; this is achieved with the circuit shown in Figure 28, the op-amp being configured as a current-to-voltage converter. Because of the high junction resistance of the reverse-biased photodiode, the op-amp should be an FET type with a very high input impedance. Since the negative input of the op-amp acts like a virtual earth, the output voltage of the circuit is $v_o = -I\,R_L$. A very large feedback resistance may be used, values as high as several tens of MΩ being typical in practice.

Sample-and-hold circuit

In the sample mode, the output of an ideal sample-and-hold circuit is equal to the input signal at that particular instant. When switched to the hold mode, the output should remain constant at that value of the input

signal which existed at the instant of switching. A simple sample-and-hold circuit is shown in Figure 29.

This circuit uses an FET switch which passes the signal through during the sample period and disconnects it during the hold period. Whatever signal was present at the time the FET is turned off is then held on the capacitor *C*. The choice of a value for *C* is a compromise between two conflicting requirements:

- Leakage currents in the FET and the op-amp cause the capacitor voltage to droop during the hold period according to the equation

$$\frac{dV}{dt} = \frac{I_l}{C}$$

  where I_l is the leakage current. Thus *C* should be as large as possible in order to minimise droop.

- The resistance of the FET when turned on (typically tens of ohms) forms a low-pass filter in combination with *C* and so *C* should be small if high speed signals are to be followed accurately.

Ready-built sample-and-hold circuits are also available as monolithic integrated circuits which simply require the connection of an external hold capacitor.

Figure 29: Sample-and-hold circuit
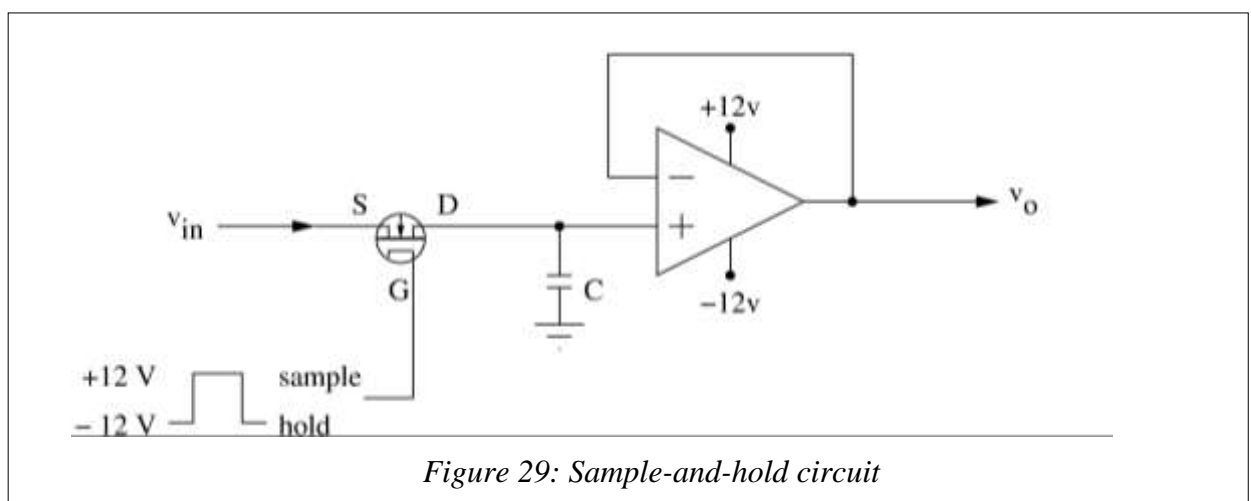
Automatic gain control circuit

The output from the sample-and-hold circuit, as indicated in the general description of the block diagram, is fed to a band-pass filter which extracts the pulsatile signal prior to its further amplification and analysis. The same output is also taken to a low-pass filter with a cut-off frequency of, say, 0.1 Hz, which extracts the d.c. value of the transmitted signal.

There are then several ways of implementing the AGC function. One of the simplest ways is to feed the d.c. signal to one input of a differential amplifier whose other input is a constant, reference voltage (from a zener diode, for example). The difference between these two voltages is then used to generate the voltage v_{in} in Figure 24 which sets the value of the LED current.