SPECTROSCOPY
AND
MULTIVARIATE
SYSTEMS
2.1 Spectroscopic Trends In Glucose Estimation

There are a number of techniques to estimate glucose based on IR radiations. The major accessories in IR studies are to name few like non-contact thermometers, radiometers, spectroradiometers, monochromators, interferometers, power and energy meters, velocity meters, reflectivity meters, transmissivity meters, humidity meters, etc and are classified as first group accessories. The systems from the second group are categorized under the detection, ranging and tracking systems. Similarly other aspects of IR radiation which are used for defense, industry, environment study, geological surveys remote sensing, etc are classified in other groups.

The relationships that relate energy, frequency and wavelength are \( c = \nu \lambda \) & \( E = h\nu \); where \( c \) = the speed of light in vacuum \( (3.00 \times 10^8 \text{ m/s}) \), \( \nu \) = frequency and \( \lambda \) is wavelength, \( h \) is a special constant called Planck's constant \( (6.63 \times 10^{-34} \text{Js}) \). Combining these two, we find that frequency and energy are inversely proportional to wavelength. These relations allow us to convert one form into other. Here in the thesis we will be representing the wavelength in nanometers (as the same is preferred by Engineers and \( \text{cm}^{-1} \) by Physicist). Various units used by Physicists and Engineers for radiation based experiments are given in table 2.1.

2.1.1 IR Spectroscopy

Methods based on IR spectroscopy have the potential to satisfy the demands that chemical analysis does not. IR spectroscopy probes the vibrations of the functional groups of molecules by letting the radiation with the sample under investigation. This is generally a non-destructive, and possibly non-invasive, method which is, in principle, capable of
identifying and quantifying organic molecules. Each organic molecule has an unique IR spectrum, and the

Table 2.1: Energy conversion and respective scales.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Equivalent measurements, comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavenumber (cm(^{-1}))</td>
<td>A wavelength of energy is also called a reciprocal centimeter. Wavenumbers are obtained when frequency is expressed in Hertz and the speed of light is expressed in cm/s. This unit is commonly used in infrared spectroscopy.</td>
</tr>
<tr>
<td>Kilojoules per mole (kJ/mol, kJ mol(^{-1}))</td>
<td>A Joule, J, is the SI unit of energy and is defined as one kg m(^2)/s(^2). As the energies associated with a single molecule or atom are quite small, we often find it easier to discuss the energy found in one mole of the substance, hence &quot;per mole&quot;. To get the energy for one molecule, divide kJ/mol by Avogadro's number, 6.022 x 10(^{23}).</td>
</tr>
<tr>
<td>Kilocalories per mole (kcal/mol, kcal mol(^{-1}))</td>
<td>A calorie was originally defined as the amount of energy required to raise the temperature of one gram of water by one degree Celsius. One calorie = 4.184 J. When we count calories in our food, we are actually referring to kilocalories; e.g. 1 dietary calorie = 1,000 cal = 1 kcal.</td>
</tr>
<tr>
<td>Nanometer (nm)</td>
<td>A nanometer refers to energy with a wavelength that is 1/1,000,000,000 of a meter. Visible light is made of up electromagnetic radiation that has wavelengths ranging from roughly 300 to 700 nm.</td>
</tr>
<tr>
<td>Hertz (s(^{-1}), Hz, /s)</td>
<td>A Hertz is a unit of frequency defined as a reciprocal second (s(^{-1})). For example, AC current cycles with oscillators 60 times per second, so we call this as 60 Hz = 60 s(^{-1}). Human hearing has a frequency range from 20Hz to 20,000 Hz.</td>
</tr>
<tr>
<td>Electron Volt (eV)</td>
<td>The electron volt is the energy emitted by electron when it falls by 1 volt potential difference. 1 eV = 1.602 x 10(^{-19}) J.</td>
</tr>
</tbody>
</table>

strength of this spectrum is proportional to the concentration of the molecule. This means that simultaneous determination of several trace components from one spectrum is feasible,
provided that, the spectra are sufficiently distinct in the spectral regions that are accessible. Unfortunately, the aqueous solutions absorbs strongly in the infrared spectral region because of characteristics water absorption. The signals of interest from the trace components are small in comparison to the water absorption, and must be extracted from this strong background.

Historically, MIR spectroscopy has been used primarily for identification of pure substances in organic chemistry. The arrival of minicomputers and the Fast Fourier Transform techniques in the 1960’s made the Fourier transform infrared (FT-IR) spectrometer practical. It became the instrument of choice because it improves Signal-to-Noise Ratio (SNR) by orders of magnitude compared to conventional instruments. At about the same time multivariate methods for analysis of spectral data with many independent overlapping variations found applications in the chemical field. These methods have made quantitative analysis a reality where simple measurements based on peak heights to determine concentrations are not satisfactory, as the peaks are masked by other variations in the spectra that are comparable or much larger. Since then, improvements in instrumentation, progress in data processing capability (due to advances in electronic computers and progress in the development of algorithms for numerical analysis) had tremendous impact improving the SNR. An application of MIR spectroscopy for quantitative analysis of substances has greatly succeeded in gas analysis, where spectral lines are sharp and isolated, and with the transparent background.

The use of NIR spectroscopy has emerged since the 1970’s as a technique for online monitoring and process control. Within the biomedical field, vibrational spectroscopy is emerging as a potential diagnostic tool with many diverse applications. The increasingly

67
prevalent disease diabetes mellitus has created a demand for continuous non-invasive monitoring of blood glucose concentration. Therefore, methods to do so based on a variety of techniques, including near and mid infrared spectroscopy, has been sought intensively in latter years.\textsuperscript{42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52}

In a simple model of a molecule, one finds that the bond strength in molecules and their mass determines the resonant frequencies at which vibrations are excited and fraction of light is absorbed as shown in eq. 2.1. In the simplest description, two atoms in a bond are regarded as a simple harmonic oscillator with resonant frequency.

\[ \nu = \frac{1}{2\pi c} \sqrt{K/\mu} \]

2.1

Where 'K' is the force constant, \( \mu=m_1m_2/(m_1+m_2) \) is the reduced mass of the two molecules with masses 'm1' and 'm2', respectively. Typical atomic masses are 1, 12, and 16 atomic units (exemplified by the H, C and O atom, respectively) and bond strengths are 8.5, 4.5 and 16 N/cm (exemplified by the CH bond, the OH bond in H\textsubscript{2}O and the CO bond in CO\textsubscript{2}, respectively) Therefore, resonant frequencies is in the NIR part of the electromagnetic spectrum from 2,500 ~ 20,000 nm. Stretching, bending and even more complicated vibration modes, including several atoms in a molecule, may be excited. Stretching vibrations have large force constants and exist in the lower frequency region, whereas bending vibrations have comparatively smaller force constants and exist in the higher frequency regions. In addition, overtone and combination bands may be excited because of non-harmonic nature of the vibrations, such that light with higher frequencies may excite these vibrations. For a vibration to be IR active, it must cause a change in the dipole moment of the molecule. Therefore, vibrations around a center of symmetry are not infrared active. For molecules in the gas state, fine structure of absorption bands arise from
the different rotational states. Molecules in the liquid state interact so frequently by collisions that absorption bands are broadened. Molecules in the solid state are locked such that peaks are more distinct than in the liquid state. The infrared spectrum of a given molecule provides information about the functional groups present in it and may be used to identify the molecule. The concentration of a given molecule may also be determined. Interestingly, the thermal radiation from matter at temperatures and upto 800K has it’s maximum between 4000 – 20,000 nm. The maximum moves towards higher frequencies as the temperature is increased. Therefore, infrared spectroscopy is also used for measurements of temperature, determination of spectral emmissivity, and remote sensing.

2.1.2 FTIR Technique

The Fourier transform infrared spectrometer is basically a Michelson interferometer with a broadband light source, a detector, and an accurate control of the mirror displacement. A schematic drawing is shown in Fig. 2.1. The mirror displacement is controlled by measuring the zero-crossings of the interference signal from a HeNe laser. Several different practical realizations of this instrument are possible and commercially available. The mirrors can be replaced with the corner-cubes instead of flat mirrors as shown in Fig. 2.1. The FT-IR spectrometer measures an interferogram \( \mathbf{I}_n \), which is an array of discretely measured points, containing the AC variation of the intensity at the detector as a function of the displacement of the movable mirror. This interferogram, is converted to an intensity spectrum of the light by Fourier transformation. Because of the high accuracy of the mirror position, determined by the interference signal of the HeNe laser, multiple scans may be co-added such that a linear averaging of the signal may be carried out. In principle, this allows improvement of the signal-to-noise ratio with the square
root of the number of scans. The maximum displacement of the movable mirror determines
the spectral resolution of the spectrometer.

Figure 2.1: Block of the FT-IR spectrophotometer.

2.1.2.1 Advantages Over Grating Instrument

The FT-IR spectrometer has replaced the conventional instrument because it possesses a number of advantages. These advantages have been discussed by Hirschfeld, Griffiths and de Haseth and etal. In a traditional grating instrument, each spectral point is measured sequentially. In contrast, a FT-IR instrument measures all spectral points simultaneously. For measurements on aqueous solutions, the spectral range is usually limited to a narrow region selected by the chosen pathlength of the transmission cell and the spectral resolution is chosen to be low because of the broad absorption found in the liquid state. If we assume a spectral range of 10,000 nm with a spectral resolution of 150 nm, we have an improvement by roughly a factor of eight compared to a grating instrument. A
grating instrument with a detector array will also measure all spectral points simultaneously and an FT-IR spectrometer will possess no Felgett advantage over such an instrument. The absence of a slit in an FT-IR instrument increases it’s light gathering power compared to a grating instrument. The light gathering power is usually expressed as the product of the allowed solid angle of the incoming beam and it’s cross sectional area and is known as the throughput.

2.2 NIR Absorption Of Glucose In Aqueous Media

Water, the major component of biological tissues, has a simple IR spectrum and a rich combination and overtone spectrum that extends into the NIR. The assignment of the NIR absorption bands of water has been established. The intensity of the NIR absorption bands for water is sensitive to solute concentration and temperature. It decreases as solute concentration increases because of change in the molar ratio of water. This is referred to as water displacement. The physical properties of aqueous solutions have been determined from the temperature dependence of the intensity of the 1400 nm band. A similar temperature sensitivity of the NIR absorption of water is observed in tissue.

The fundamental IR absorption bands of glucose have been reported in solid pellets and in solution. The strongest bands that can generate intense combinations and overtones are the broad OH stretch at 2,857 nm and the C_H stretch vibrations at 3,377 and 3,393 nm. Possible combination bands are a second OH overtone band at 939 nm (3νOH) and a second harmonic CH overtone band at 1,126 nm (3νCH). A first OH overtone band can be assigned at 1,408 nm (2νOH). The 1,536 nm band can be assigned an OH and CH combination band (νOH + νCH). The 1,688 nm band is assigned as a CH overtone band (2νCH). Other bands at >2,000 nm are possibly combinations of a CH stretch and a CCH, OCH deformation at 2,261 nm (νCH + νCCH, OCH) and 2,326 nm (νCH + νCCH, OCH).
presence of the CCH, OCH ring deformation component confers some glucose specificity on these bands. The calculated NR overtone and combination spectra of glucose overlap with several (more intense) combinations and overtone bands of water and fat and haemoglobin absorption bands. These are the major potential interferences with the NI determination of glucose. NIR (2000–2500 nm) spectrophotometric determination of glucose has been achieved in aqueous media.

2.2.1 Optical Properties Of Tissue In NIR

The near-IR spectral region is commonly proposed or used in all reported technologies. The 600-1100 nm region of the spectrum represents a window between the haemoglobin, visible absorption bands and water IR absorption. Light can penetrate deep enough into the tissue to allow a spectral measurement or a therapeutic procedure. This spectral region is used for oxygen saturation, pulse oximetry, and laser-Doppler flow measurements.

The transport equation (Eq. 3.1), and the diffusion theory approximation of this equation describes the path of photons through human tissue. It expresses light propagation in tissues by a set of spectroscopic properties as shown in Equation 2.2 & 2.3; the absorption coefficient $\mu_a$, the scattering coefficient $\mu_s$, the refractive index of the cells and the interstitial fluid; and the anisotropy factor $g$ (the average cosine of the angle at which a photon is scattered). Another set of properties are the transport properties, such as the reduced scattering coefficient $\mu'_s$, where $\mu'_s = \mu_s [1 - g]$. The absorption coefficient $\mu_a$ equals the absorbance per unit path length, 2.303 $\varepsilon C$ cm$^{-1}$, where $\varepsilon$ is the molar
absorptivity and 'C' is the molar concentration. The scattering coefficient \( \mu_s = \sigma p \) where '\( \sigma \)' is the scattering cross-section and 'p' is the number density of the particle. It has the same unit as \( \mu_a \) (cm\(^{-1}\)) and is equivalent to the product of an absorptivity caused by scattering and the concentration of the scattering centers. Attenuation of light in tissue is described, according to light transport theory, by the effective attenuation coefficient \( \mu_{\text{eff}} \), i.e.:

\[
I = I_0 e^{-\mu_{\text{eff}} l}
\]

Where:

\[
\mu_{\text{eff}} = \sqrt{3\mu_a (\mu_a + \mu_s')} = \sqrt{3\mu_a [\mu_a + \mu_a (1 - g)]}
\]

An exact solution of the light transport equation, in turbid media can be modeled by following the path of each individual photon and calculating the probability of scattering or absorption in a series of steps, using the Monte Carlo simulation. This modeling is used to study the path of photons in tissues and is widely used for optimization of photodynamic therapy, improvement in pulse oximetry, laser-Doppler flowmetry, and optical mammography; all have very important clinical utility. Several recent volumes of the proceedings of the Society of Photo-optical Instrumentation Engineers (SPIE) have covered these topics. Methods that are used for measuring the optical properties of tissues (\( \mu_s \), \( \mu_a \), and \( g \)) include transmission, diffusion, localized reflectance, and frequency domain measurements\(^{84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97} \).

2.2.1.1 Effect Of Glucose On Absorption Properties Of Tissues
Glucose can affect the measured transmitted or reflected signal by absorption of light at the overtone and combination band wavelengths. Light absorption can be expressed as \( I = I_0 e^{-\mu_a l} \), where \( 'l' \) is the effective path length in the medium, and \( \mu_a (2.303eC) \) is the absorption coefficient. Changes in glucose concentration can influence the measured \( \mu_a \) of tissue through changes in absorption corresponding to water displacement or changes in its intrinsic absorption. Changes in \( \mu_a \) because of water displacement are nonspecific, and analytes with higher molecular weights will displace more water than glucose. Changes in the temperature and hydration status of the body may affect water absorption bands and act as noise sources for a glucose measurement. The glucose \( \mu_a \) in the NIR is low and is much smaller than that of water. It is higher at longer wavelengths. However, its magnitude is too small to allow for direct absorption measurements at wavelengths <1400 nm. Attenuation of light (<1400 nm) in a small body part such as an average-sized finger varies in the range 3-4 absorption units, and the expected change in absorption because of a 5 mmol/L change in glucose concentration is ~10-5 absorption units.

Changes in the glucose concentration affect the intensity of light scattered by tissue. The reduced scattering coefficient of a tissue can be expressed in a function form as: \( \mu_s' = F(\rho, a, n_{cells}/n_{medium}) \), where \( \rho \) is the number density of scattering cells in the observation volume, \( 'a' \) is the average diameter of the cells, \( n_{cells} \) is their refractive index, and \( n_{medium} \) is the refractive index of interstitial fluid. Changes in the \( n_{medium} \) are not specific for a particular analyte and are affected by any change in the total concentration of solutes in blood and interstitial fluid. The calculated \( \Delta n_{water} \) as a function of the change in the concentration of several metabolites, as calculated from the slope of tabulated 'n' values at different solute concentrations. During hyperglycemic episodes, the glucose concentration changes rapidly, whereas other analyte concentrations presumably change at a slower rate. It may be possible to relate \( \mu_a \) to changes in glucose concentration over a short time.
span. The measured $n_{\text{water}}$ decreases as temperature increases. This affects $n_{\text{cells}}/n_{\text{medium}}$ in tissue and presents a source of error in scattering measurements. Values of $\mu_s$ are reported to decrease with increasing concentrations of glucose and other sugars in tissue-simulating phantoms because of their effect on $n_{\text{medium}}$. A recent Monte Carlo modeling of the effect of physiologic concentrations of glucose (5–30 mmol/L) on the diffuse reflectance or transmittance of tissue-simulating phantoms predicts very small changes in signal. The modeling is performed at 800 and 960 nm with water as the only absorber. The estimated $\mu_t$ is $<1 \times 10^{-3}$/mmol of glucose, which is much higher than the changes $\mu_a$. Other physiological factors, such as changes in the water content, temperature, and protein concentrations are considered. The value of $\mu_t$ caused by an increase of 1 mmol/L in glucose concentration is equal to a $2 \times 10^{-3}$ increase of water content, a $1 \times 10^{-3}$ increase of protein concentration, or a 0.1 °C decrease in temperature. In addition, the concentrations of pigments in the epidermis strongly influence light penetration in the NIR. The temperatures and body water content of an individual are tightly regulated, but differ between individuals. Although body temperature is regulated within a fraction of a degree, temperature at the extremities may vary by $>1$ °C. A temporary change in water content of 240 mL for a 75-kg individual can lead to an error equivalent to 2 mmol/L glucose. The reference range for albumin spans 20% of the mean concentration. This spread will affect the signals from different subjects with normal albumin concentrations. Haemoglobin has not been considered as an absorber in this simulation. However, changes in its concentration will affect optical signals in the 600–900 nm range. Positioning errors and body interface effects seem to be the largest contributors to measurement errors.
2.3 Dynamics Of Tissue Matrix And Signal Conditioning

In the concepts of signal extraction the interest focuses on the influence of drift and noise as two key factors that determine the success of any quantitative measurement. The advantage of a full spectral measurement, as in FTIR spectroscopy, to distinguish between these two quantities and the possibilities for reducing their impact on quantitative measurements are discussed. In signal extraction, one traditionally measures the absorbance of a substance to determine a concentration. The absorption \( A(v) \), at a given frequency \( v \), is given by

\[
A(v) = -\log_{10} \frac{I(v)}{I_0(v)}
\]

Where, \( I(v) \) is a measurement of the sample intensity and \( I_0(v) \) is a measurement of the reference intensity. The choice of reference depends on the application. In many cases a reference with no sample present is used. In other cases a reference is chosen which resembles the sample. Most grating instruments measure the ratio \( I(v)/I_0(v) \), known as the transmittance, directly, but the Fourier transform instrument measures the sample and reference intensity spectra separately. By taking the ratio of \( I(v) \) and \( I_0(v) \), one creates a dimensionless number which ideally eliminates all the dependency on the instrument, including spectral intensity distribution of the source and sensitivity of the detector. Note that noise is not eliminated and still depends on the source and detector. The noise is increased because the relevant signal is composed of two measurements, each containing noise. The increase in noise pays for elimination of instrumental effects, including drift.

Sample extraction analysis is based on Beer’s law, which states that the absorbance of a substance, at a given wavenumber \( v \), is proportional to the molar concentration \( c \) of the substance and the pathlength \( l \).
\[ A(\nu) = e(\nu)cl; \]

Where \( e(\nu) \), known as the molar absorptivity, it is the wavenumber dependent proportionality constant. Beer's law says that the intensity decays exponentially with path length and with the concentration of the substance.

Traditionally drift and influences from instrumental effects have been considered more problematic than noise. In applications where small signals are to be detected, and with the available methods of data analysis, noise may prove to be more problematic than instrumental effects. If one considers a single point measurement, at a given wavenumber, noise and drift are indistinguishable. If one has a spectral range, on the other hand, drift and influences of instrumental variations may be reduced because they have a spectral structure that may be included in the calibration process. Noise, is impossible to remove from a single wavelength because it is uncorrelated from point to point in the spectrum. Noise may be reduced by averaging a large number of measurements, but the SNR is proportional to the square root of the number of measurements. This strongly limits this procedure because the measurement time becomes prohibitively large. In practice it turns out that many FT-IR instruments fail to signal average well beyond a certain point. The spectrum may be put to low-pass filter, which remove high frequency noise. But such a procedure is based on a prior assumption about the spectrum, which may not be true, and reduces the spectral resolution of the data. In addition, it is doubtful that such a smoothing represents an advantage in a calibration. Smoothing is probably most justified when used as a graphical technique to guide the eye. High-pass filtering, including derivation, may remove baseline variations, but such a procedure will also remove any broad-band variation which is part of the signal of interest. In general, data pre-treatment which is carried out to remove drift will result in an increased noise or degraded spectral resolution. Noise therefore ultimately limits
the calibration. For this reason, single-beam spectra (or logarithm of single-beam spectra) are also used today in quantitative analysis. In this case, variations and drift are included in the modeling of the data instead of being eliminated in the measurement process. The FT-IR instrument has the advantage over the grating instrument in possessing a superior SNR. In contrast, the dual-beam IR grating instrument is made to eliminate drift by measuring the transmittance directly. The dual-beam, optical null, FT-IR spectrometer also eliminate drift in the measurement process, but does so by measuring a difference between two samples instead of a ratio. The variability of the sample population also influences the accuracy of any calibration. With a large variability in a sample population, many calibrations and many independent spectral points will be necessary to maintain stability and accuracy. The variability of the sample population is usually given as an intrinsic part of a job, and reducing a population variability then means to exclude certain classes of samples. This may be necessary, but is seldom desirable.

2.3.1 The Properties Of Human Tissue And Blood

Optical imaging and non-invasive diagnosis of the human body depends strongly on the optical and physical properties of skin and blood. The composition and morphology of the skin is very complicated. Therefore, to build a reasonable optical model of the skin, its composition and structure must be studied.

2.3.1.1 Skin Tissue Composition And Structure

The structure and properties of skin vary considerably in different parts of the body. A typical structure of skin is shown in fig. 2.8, while table 2.7 and table 2.8 list the average elemental composition and the biochemical composition, respectively. The skin is usually divided into three layers, namely, the epidermis, dermis, and subcutaneous fat, each with
their own sub-layers. The outermost layer of the epidermis is composed of a relatively thin, but rough, protective top layer of dead and dry skin cells, known as the stratum corneum or horny layer. The remainder of the epidermis, including the stratum lucidum, stratum granulosum and stratum spinosum, is made up of cells called keratinocytes as well as melanocytes, which are pigment cells responsible for skin pigmentation. The thickness of the epidermis varies from 0.1mm in the eyelids to nearly 1mm on the palms and soles. The dermis consists of a variety of cells, fibres, amorphous ground substance, nerves, oil glands, sweat glands, blood vessels and hair roots. Its upper layer is called the papillary dermis and contains the vascular network and sensory nerve endings, whereas the deeper layer, referred to as reticular dermis, consists mainly of a loose connective structure and epithelial-derived structures such as glands and follicles. The thickness of the dermis varies from 0.3mm in the eyelids to about 3mm in the palm and soles. Subcutaneous fat is composed of fat cells, which form a cushioning layer between the skin and the deeper muscles. It also has abundant blood content.

Fig. 2.3: A structure of human skin.
Table 2.2: Average elemental composition of the skin, percentage by mass.98

<table>
<thead>
<tr>
<th></th>
<th>O</th>
<th>C</th>
<th>H</th>
<th>N</th>
<th>Na</th>
<th>Mg</th>
<th>P</th>
<th>Cl</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>59.4~69.5</td>
<td>25~15.8</td>
<td>10~10.1</td>
<td>4.6~3.7</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Table 2.3: Percentage constituents of adult human skin.99

<table>
<thead>
<tr>
<th></th>
<th>Water</th>
<th>Protein</th>
<th>Lipid</th>
<th>Ash</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>58.6~72.1</td>
<td>22~27.2</td>
<td>5.2~13.5</td>
<td>0.7</td>
</tr>
</tbody>
</table>

2.3.1.2 Optical Absorption

Being composed of water as well as proteins and lipids, the chemical make-up of the skin influences its optical absorption properties. Water absorbs photons at wavelengths longer than the middle IR range, while proteins are strongly absorbing in the UV and Violet region (fig. 2.9). Fortunately, the optical absorption capacity of water, proteins and lipids is small in the red and NIR region. This region, known as the “tissue optical window”, having range from 600 nm to 2,300 nm and the light can penetrate to a depth of few hundreds of micrometers to a few mm into the skin tissue. As a result, it can be exploited for a variety of purposes, including diagnosis, imaging or therapy.

Fig. 2.4: The absorption spectrum of tissues.100
At the shorter wavelengths of the tissue optical window, from 600 nm to 1,100 nm, the most important photon absorbing chromophores are blood and melanin. Water becomes dominant at incident wavelengths longer than 1,150 nm. The epidermis does not contain any blood and its water content is also much lower that of the dermis. However, the stratum granulosum and stratum spinosum comprise some melanocytes, including melanin, which is involved in skin pigmentation. Because the absorption capacity of melanin is stronger than that of blood and water, it is the dominant source of absorption in the epidermis at shorter NIR wavelengths. The volume fraction of melanosomes in the epidermis can vary from 1.3 ~ 6.3% for light-skinned adults, 11 ~ 16% for well-tanned adults and 18 ~ 43% for darkly pigmented Africans. The blood content of the dermis is about 0.2 ~ 5%, representing the main source of absorption at wavelengths shorter than 1100 nm. If the optical wavelength exceeds the NIR range, water content becomes an important consideration in terms of optical absorption. It is a well-known fact that the measured values of absorption coefficient of a tissue are different \textit{in vitro} and \textit{in vivo} measurements. This can be explained on a number of grounds. First, soaking the tissue sample in saline prior to an \textit{in vitro} measurement may alter its optical properties, and increase the amount of reflectance. Also other kinds of tissue treatments, including freezing, drying, heating or deforming, may change the optical properties of the sample. Second, measuring and calibration procedures may introduce an error into the determined values for diffuse reflectance and total transmittance. Third, the use of simplified calculation methods may result in an incorrect interpretation of the measured data, as in the case of internal reflectance at tissue boundaries.
2.3.2 A Model For Human Tissue Calibration In NIR For Glucose

NIR spectroscopy is widely used for analytical measurements in complex matrixes. A key feature of this methodology is the ability to extract quantitative information for numerous substances from a single NIR spectrum. In addition, NIR spectroscopic analyses are nondestructive, simple, fast, and require no sample pretreatment, which makes this technology ideally suited for in-situ process monitoring and quality control. Biological and clinical applications of NIR spectroscopy frequently involves measurements in samples composed primarily of water. An important special case is the measurement of glucose in complex biological matrixes. Even in matrixes as complex as whole blood, water is the single most important component that defines the fundamental analytical performance. Measurements in aqueous samples are complicated by the strong NIR absorption properties of water, which accentuate the impact of sample thickness, instrumental SNR, and spectral range.

2.3.2.1 Light Transport In Biological Tissue

Reticular dermis (about 1500 mm thick) also includes small arteries and veins (20-60 μm inner diameter) which are orientated almost perpendicularly to skin surface. These small vessels constitute routes for blood supply and drainage to veins and arteries (50-100 μm inner diameter) from deep blood plexus (220 μm thick) and subcutaneous fat (Fig. 2.10). Behavior of blood in skin capillaries of less than 100mm in diameter is non-Newtonian\textsuperscript{37, 38}, which present difficulties for a phantom modeling and computer simulation.

To understand the technique of NIR tissue spectroscopy in more detail, it is important to establish an understanding of the manner in which light propagates through biological tissue. Description of the mechanism that governs light transport through physical media and identifying the components of biological tissue are described here. In
first part, the constituents of biological tissue are responsible for the absorption of NIR light; thus highlighting the important features of absorption spectra. In second part, the basic theory behind light scattering and the parameters used to define it are introduced. The discussion is limited to elastic scattering, in which there is no loss in energy of the scattered light, on the grounds that non-linear scattering effects are generally not significant in the NIR region. The biological structures that give rise to scattering in tissue are also identified.

Fig. 2.5: Dimension of the typical skin layers model.

Models commonly used for transport of light through turbid media, i.e. those in which multiple scattering occurs are described. In particular the discussion concentrates on the assumptions and approximations made for each model and hence their validity in 'real' situations.
2.3.2.2 Light Absorption Mechanism

When light radiation is incident on matter composed of discrete electrical charges; the charges are forced to oscillate at the frequency of the incident electric field. The range of frequencies covered by radiation in the IR region of the electromagnetic spectrum (300 GHz -300 THz) is comparable to the natural frequencies at which atoms or molecules will vibrate in the absence of an applied field. Thus when infrared radiation is incident on a biological molecule, the resonance will occur around the natural frequencies of various bonds, whereby energy is transferred from the incident field to the system and its amplitude of vibration is greatly increased. Although the lifetime of the excited state is around 8-10 ns, the atoms or molecules will usually lose their energy by colliding with one another within 10–12 ns, thereby raising the kinetic energy of the other particles involved in the collisions. Hence, the energy associated with the incident field is most often dissipated as heat within the medium. This process is known as absorption. The overall effect of absorption is a reduction in the intensity of the light beam traversing the medium. A relationship between the absorption of light in a purely absorbing medium and the thickness of the medium was first determined in 1729 by Bouguer. Some years later Lambert (1760) derived the following mathematical expression for the relationship, known as the Lambert-Bouguer law (eq. 2.5),

$$\frac{dI}{I} = \mu_a dl$$

which describes how each successive layer ‘dl’ of the medium absorbs the same fraction ‘dI/I’ of the incident intensity ‘I’ for a constant ‘\(\mu_a\)’, the latter known as the absorption coefficient with units of inverse length (usually mm\(^{-1}\)). For incident intensity ‘\(Jo\)’, therefore, the transmitted intensity ‘\(I\)’ through a distance ‘\(l\)’ will be
The absorption coefficient \( \mu_a \) can thus be interpreted as the probability that a photon will be absorbed by the medium per unit length. The reciprocal of the absorption coefficient, known as the absorption length, is the distance required for the intensity of the beam to fall to \( e^{-1} \) of the initial intensity. When Equation (2.6) is expressed in base 10 logarithms then the constant \( K \) is known as the extinction coefficient. The absorbance of the medium is defined as given below (eq. 2.8)

\[
I = I_0 e^{-\mu al}
\]

2.6

\[
A = \log_{10} \left( \frac{I_0}{I} \right) = K l
\]

2.8

where 'A' is 'optical density' (OD). Hence, the units of \( K \) are 'OD' per unit length (usually ODcm\(^{-1}\)). The extinction coefficient and the absorption coefficient are conceptually the same, differing only by the base of the logarithm used in the Lambert-Bouguer expression. For the same unit length, therefore, the extinction coefficient is related quantitatively to the absorption coefficient by a factor of 0.434. The absorption coefficient of a compound is linearly related to its concentration \( 'c' \) diluted in a non-absorbing medium (Beer, 1852)

\[
\mu_a = \alpha c
\]

2.9

where \( \alpha \) is known as the specific absorption coefficient. Substituting for \( \mu_a \) in the Lambert-Bouguer law gives what is known as the Beer-Lambert law

\[
I = I_0 e^{-\alpha cl}
\]

2.10
Expressing the Beer-Lambert law in log10 gives

\[ I = I_0 10^{-\varepsilon cl} \]

where \( \varepsilon \) is the specific extinction coefficient, usually expressed in units of ODcm\(^{-1}\)mM\(^{-1}\). In a solution containing a mixture of \( n \) absorbing compounds, the total absorption is the sum of the individual extinction coefficients multiplied by the distance \( l \).

\[ A = (K_1 + K_2 + \cdots + K_n)l \]
\[ = (\varepsilon_1 c_1 + \varepsilon_2 c_2 + \cdots + \varepsilon_n c_n)l \]

2.12

The Beer-Lambert law is only valid under certain limited conditions: the light entering the medium must be monochromatic and perfectly collimated, and the medium itself must be purely and uniformly absorbing. Therefore, certain errors will arise when applying the law to practical spectroscopic measurements since, for example, even lasers are not perfectly monochromatic. The consequences on experimental measurements of the limitations imposed by the Beer-Lambert law has been discussed elsewhere (Cope, 1991). There are many compounds in biological tissue which absorb light radiation, collectively known as tissue chromophores, each of which has its own unique spectrum. As expressed in Equation (2.12), the total extinction coefficient ‘K’ of a mixture of compounds is equal to the sum of their individual extinction coefficients, weighted by their relative concentrations. Therefore, approximating a tissue as a homogeneous mixture of compounds, the overall light absorption in tissue at a given wavelength depends on the type and concentration of chromophores present. The following sections will discuss the chromophores present in biological tissues and their absorption spectra in the NIR.
2.3.3 Major Varients In Human Tissue

2.3.3.1 Water

Water is the most abundant chemical substance in the human body, accounting for 60% to 80% of total body mass (Marieb, 1995). The water content varies with tissue type and it is age and gender-dependent. For example, the new-born brain comprises 90% water by mass, whereas the water content in adult skeletal muscle is around 74% (White et al, 1991). Because of its high concentration in most biological tissue, water is considered to be one of the most important chromophores in tissue spectroscopy measurements. The absorption spectrum of water is shown in Figure 2.11 over the wavelength range 200–10,000 nm (Hale and Querry, 1973) and on an expanded scale from 650 to 1050 nm (Cooper et al, 1996). Between 200 and 900 nm there exists a region of relatively low absorption. Above 900 nm the absorption coefficient increases rapidly to a peak at about 970 nm, and following a minor trough continues to increase at longer wavelengths into the midinfrared. The region of low absorption acts as a ‘window’ of transparency, allowing NIR spectroscopic measurements through several centimeters of tissue to.

Water is a critical matrix component for near infrared spectra of aqueous based clinical samples, such as the human body. The high concentration of water in clinical samples coupled with the relatively strong absorptivity of OH groups result in large water absorption bands. The strong absorption of water dictates using the regions between these water bands where sufficient amounts of light is transmitted. The following three regions are generally accessible: 1) the combination region: 2000 – 2500 nm; 2) the first overtone region: 1,540 – 1,820nm; and 3) the short-wavelength near infrared (SW-NIR) region: 700 – 1,330 nm. Glucose has three absorption bands in both the combination region (centered at 2,100, 2,270, and 2,320 nm) and the first overtone region (centered at 1,730, 1,690 and
1,610nm). Although glucose absorption bands are difficult to measure in the SW-NIR owing to their extremely low absorptivities, bands centered at 760, 920, and 1,000nm are reported.

![Absorption spectrum of pure water.](image)

**Figure 2.6:** Absorption spectrum of pure water. a) Plotted on a log10 scale from 200-10,000nm (Hale and Querry, 1973), b) NIR region from 650-1050 nm (Cooper et al., 1996),

### 2.3.3.2 Haemoglobin

Within the window of transparency for water the most dominant absorption of NIR light is by haemoglobin in its various forms. Haemoglobin is carried in red blood cells, or erythrocytes, and constitutes approximately 40–45% of whole blood. It is responsible for delivering oxygen from the lungs to the body tissues and returning waste gases, such as carbon dioxide, to the lungs to be exhaled. Haemoglobin consists of the protein globin bound to four ‘haem’ groups. Each ‘haem’ group contains an iron atom at the centre of a ring-like structure. An iron atom in the ferrous (Fe²⁺) form will bind physically to an oxygen molecule to become oxygenated, as opposed to oxidised which would involve a
chemical bond. Thus, one haemoglobin molecule with its four iron centres can carry a total of four molecules of oxygen, the state is said to be 100% saturated. In the oxygenated state haemoglobin is known as oxyhaemoglobin (HbO₂). The de-oxygenated form, with no oxygen molecules attached, is known as deoxyhaemoglobin (Hb).

The specific absorption spectra of oxy- and deoxyhaemoglobin, shown in Figure 2.12 (Cope, 1991), differ significantly, particularly in the red region of the visible and the NIR. This difference in absorption explains the visible colour difference between venous and arterial blood. Arterial blood, which in adults is usually about 98% oxygen saturated,

![Specific absorption spectra of Hb and HbO2](image)

Figure 2.7: Specific absorption spectra of Hb and HbO2 in the NIR from 650-1,050nm (Cope, 1991)

And has bright red in colour, whereas venous blood, which is approximately 75% saturated, appears dark red to purple in colour. Although the NIR absorption is reduced in amplitude relative to those in the visible, the spectra are still sufficiently different to distinguish between the two forms of haemoglobin. A typical value for haemoglobin concentration in the adult brain has been calculated to be 84 μM (Cope, 1991). Other forms of haemoglobin that absorb in the NIR are carboxyhaemoglobin (HbCO), methaemoglobin (metHb) and sulfhaemoglobin (SHb). However, these derivatives are generally ignored in spectroscopic
measurements, either because they have a low percentage concentration in blood or a low specific absorption, or both (Cope, 1991).

### 2.3.3.3 Lipids

Most of the lipid in the body exists in the form of triglycerides (neutral fats) and is found in subcutaneous tissues and around internal organs. Phospholipids, another group of lipids, are the main component of cell membranes and are thus found in every organ in the body. The lipid content in the brain, which also contains steroidal lipids, varies with age from 2.6% in the new-born to 11.6% in the adult (White et al, 1991). In adipose tissue, found in the subdermis, the lipid concentration is again age- and gender-dependent, in the range 23–47% for new-born infants and 68–87% for adults (White et al, 1991). The importance of lipid as an absorber in NIR spectroscopy depends upon the tissue in question. Since the water content is much greater than the lipid content in the brain, absorption due to lipid may be insignificant. In the forearm, however, the lipid content is seen to be highly variable, depending on the fat-to-muscle tissue ratio (Matcher et al, 1994), in which case lipid absorption may well be significant in spectroscopic measurements.

### 2.3.3.4 Other Chromophores

There are two other tissue chromophores that are worth mentioning: melanin and myoglobin. Melanin, the pigment found in the epidermal layer of human skin, has a large scattering coefficient in the ultraviolet region, which protects the skin damaging due to UV radiation from the sun, and a significant absorption coefficient in the NIR region (Cope, 1991).
2.4 Noninvasive Determination Of Glucose

2.4.1 Near-IR Transmission And Reflectance

Robinson et al. 101 Haaland et al. 102 and Ward et al. 103 have attempted NI glucose detection in two overlapping regions, 750-1050 nm and 850-1300 nm, which encompass the 3υOH and 3υCH glucose overtone bands. Hiese and co-workers 104,105,106,107,108,109 and Marbach et al. 110 have reported a series of studies on the determination of glucose in oral mucosa in the 1,111-1,835 nm spectral range. This range encompasses bands corresponding to the 3υOH, 2υOH, υOH + υCH, and 2υCH glucose vibrational combinations. Jagemann et al. 111 Muller et al. 112 and Fischbacher et al; 113 have measured reflected light in the wavelength range 900-1200 nm through a fiber bundle touching the skin. Khalil and Malin 114 have proposed the use of reflected signals from five distinct near-IR spectral ranges 1320-1340, 1440-1460, 1670-1690, 1940-1960 and 2120-2280 nm (As shown in Fig 2.2 ;only two regions are shown). The 1320-1340 nm range is used for correcting optical coupling and sample positioning errors. The 1440-1460 nm or 1940-1960 nm ranges are where the highly absorbing water bands lie. The 1670-1690 nm range (2υCH glucose overtone band) and the 2120-2280 nm range, are where glucose combination bands υCH + υOH, CH are located. The correction method involves normalizing the signal to the shortest wavelength range, subtracting the normalized signal, and correlating the subtracted signals with glucose concentration. Bruneister etal. 115, have used the 1400-2000 nm range for NI determination of glucose with the tongue as the body site.

Rosenthal et al. 116 at Futrex Inc. have proposed using the spectral band in 600-1100 nm, NIR transmitted or reflected through a body part and analyzing the light by a set of
filters and detectors. This spectral range encompasses the water spectrum and the 3\(\nu\)OH and 3\(\nu\)CH glucose overtone absorption bands.

Purdy and co-workers\textsuperscript{117} and Barnes et al.\textsuperscript{118} at Biocontrol Inc. have used light at wavelengths >1100 nm falling on the forearm of a patient from light-emitting diodes. The reflected light is directed into a spectrometer and analyzed. The spectral range encompasses 3\(\nu\)CH, 2\(\nu\)OH, \(\nu\)CH + \(\nu\)OH, and the water bands.

![Glucose absorption spectrum](image)

Fig 2.8: Absorption spectrum of Glucose in IR region (Source: THz-bridge).

2.4.2 Mechanical Manipulation Of Tissues

In a method developed by VivaScan Inc., measurements are made by altering the blood volume by changing pressure on a body part in a controlled way, and performing a NIR transmission measurement in the 1300–1600 nm range. The spectral range encompasses 2\(\nu\)OH, \(\nu\)CH + \(\nu\)OH, and water absorption bands. Difference data is obtained by comparing spectra collected with different compressions within the exposed tissue samples\textsuperscript{119}. Wavelength pairs in the 1300–1600 nm range are selected by use of an acousto-optic modulator, and measurements are taken at different compressions of a thin body part.
such as a web or an earlobe. This is considered as a hardware compensation for tissue contributions to the signal.

2.4.3 NIR Kromoscopic© Measurements

Kromoscopy, developed by Optix Corp., is reported extensively in the patent and commercial literature\textsuperscript{120,121,122,123,124,125}. It is claimed to have superior sensitivity compared with spectroscopic methods. Kromoscopy presumably has the potential for NI determination of glucose by using broad band light illumination, broad band overlapping filters, and multiple detectors; however, there is limited theoretical basis for the sensitivity improvement over photometric methods. The method reportedly is based on the ability of the eye to determine slight changes in colour; however, the effect of light scattering on Kromoscopic measurements, are not available in the literature.

2.4.4 Spatially-Resolved Diffuse Reflectance R(r) Measurements

In this technique, a narrow beam of light illuminates a restricted area on the surface of a sample or a body part, and the diffuse reflectance is measured at several distances from the illumination point. This method is denoted as R(r) measurement. The value of \( \mu_{\text{eff}} \) can be calculated from the data, and both \( \mu_a \) and \( \mu_s' \) values for tissue can be deduced. Changes in the values of \( \mu_a \) and \( \mu_s'' \) can then are used to calculate the change in concentration of an analyte affecting the tissue optical properties\textsuperscript{126}.

2.4.5 NIR Frequency Domain Reflectance Measurements

Frequency-domain reflectance measurements use an optical system similar to that used for spatially resolved diffuse reflectance R(r), except that the light source and the detector are modulated at a high frequency. The difference in phase angle and modulation between injected and reflected beam is used to calculate \( \mu_a \) and \( \mu_s'' \) of the tissue\textsuperscript{127}.
2.4.6 Optical Activity And Polarimetry

Polarimetry has been used for quantitative analysis of solutions of optically active (chiral) compounds such as glucose. When a plane polarized light beam is transmitted through a solution, its plane of polarization is rotated by an angle $\alpha$ which is related to the concentration of the optically active solute$^{128}$. A 5.5 mmol/L change in glucose concentration yields the highest calculated short-term effect on $\alpha$. It may be possible to detect hyperglycemic swings, assuming that changes in the concentrations of other optically active compounds occur over a longer time frame than that for a change in glucose concentration. Scattering is bound to depolarize the light and decrease the measured value of $\alpha$.

2.4.7 NIR Photo-Acoustic (PA) Spectroscopy

Photo-Acoustic is alternative detection technology for near-IR light interaction with tissues. PA is used to detect weak absorbance in liquids and gases. The tissue is excited at a wavelength that is absorbed by glucose molecules, by pulses of a 1000–1800 nm near-IR laser light. Subsequent optical absorption causes microscopic localized heating. The increase in temperature causes rapid thermal expansion, which generates an ultrasound pressure wave detectable by a hydrophone or a piezoelectric device in the cuvette or located at the skin surface$^{129}$. The magnitude of a pulsed PA signal, ‘P’ (eq. 2.4), is related to the absorption coefficient of solution by:

$$P = K(\mu_a, \beta, \sqrt{\nu})/C_p$$

2.4
where $\mu_a$ is the optical absorption coefficient, $\beta$ is the thermal expansion coefficient, $\nu$ is the sound velocity, $C_p$ is the specific heat of the solution, and $K$ is a proportionality constant that is related to the bulk modulus of the medium. The $C_p$ of a solution decreases, whereas the acoustic velocity increases with increasing glucose concentration. At a glucose absorption wavelength, change in the PA signal is the result of changes in $\mu_a$, $\nu$, and $C_p$. This multiplicative effect increases the PA signal as a function of concentration. The speed of sound and the $C_p$ values change as the total solute concentration changes. When excited in the NIR, PA detects the absorption caused by the overtones of O-H and C-H bond vibrations of glucose and other analytes; this absorption is subsequently converted into an acoustic pulse. PA measurements have some sensitivity advantages over other NIR detection methods, because a PA detector collects all generated signals in a volume of the tissue and detects signals generated at wavelengths longer than the range of silicon or gallium arsenide detectors.

2.5 Research Methodology For NI Glucose Measurements

Two research methodologies are used for NI glucose measurements. An empirical approach involves the following steps: (a) collect NI signals from non-diabetic individuals and diabetic patients while performing an oral glucose tolerance test (OGTT) or a glucose challenge test; (b) simultaneously measure blood glucose concentrations by an invasive method; and (c) compute models based on the correlation between measured glucose values and NI signals. This approach does not measure the effect of other metabolites and interferences, biological noise, or variability in instrument-body interface, but attempts to compute-out these noise contributions. The number of variables and the complexity of data analysis necessitates the use of multivariate chemometric techniques such as principal component analysis, partial least squares (PLS), or artificial neural network (ANN).
methods. Special criteria are needed to avoid overfitting data and to eliminate overshoots without affecting the fidelity of data presentation.

A second approach is the physical model approach. This method involves the following steps: (a) measurement of a glucose optical signal in a simple matrix, (b) progressive increases in the complexity of the matrix to mimic human tissues; (c) demonstration of accuracy and precision at each step (d) correlation of the data with a mathematical model for light propagation in tissue. Finally, the detection system and the measurement method are applied to body parts. The in vivo signals are again correlated with the invasive data by use of chemometric techniques. This stepwise approach allows the identification of noise components so that strategies may be derived to minimize their contribution to the signal before the use of chemometric techniques.

2.5.1 Calibration Of NI Glucose Measurements

2.5.1.1 Analytical Calibration

Analytical calibration, defined as the determination of the concentration of the analyte from a calibration curve generated by calibrators and standards, is difficult for NI glucose measurements. There are numerous sources of error that can affect the measurement, which need to be either eliminated or compensated by a calibration method. Some of these error sources that can affect the measurement, but cannot be easily incorporated into the calibration are detector positioning error, temperature and cardiac pulse effect, motion, mechanical pressure of the test device, hydration state, sweat, blood volume, and haematocrit change.

Several tissue-simulating phantoms have been proposed as calibration systems. Some of these phantoms are suspensions of lipids or polystyrene particles in solutions having different concentrations of glucose. A phantom containing fat and glucose solution
has been used to mimic tissue glucose absorption in the 2000 – 2500 nm range\textsuperscript{132}. Glucose concentrations used in these studies have generally been higher than the physiological range. Because of the non-specificity of the signal measured by several technologies, developing tissue-simulating phantoms as analytical standards for NI glucose determination in tissue is a challenging goal.

2.5.1.2 Clinical Calibration

In vitro and in vivo measurements are performed on a fasting subject at time intervals during an OGTT, meal-tolerance test, or a glucose clamp procedure. These methods offer a concentration range over which the glucose signal can be monitored. Data from an OGTT can be used to establish an NI instrument response as glucose concentration for an individual is changing. Data that are generated during the test period are used to predict glucose concentrations in subsequent NI measurements. Because the response of an NI instrument may embody non-glucose-related physiological effects, relying on clinical calibration based on the correlation of OGTT data with NI instrument response leads to a calibration curve that is unique to the individual tested. This calibration curve may need to be periodically updated, by use of an invasive test.

Use of an OGTT or a meal-tolerance test for calibration leads to a series of measurements that are sequential in time. Spurious drift and time-dependent artifacts can influence the results from multivariate calibrations when randomized sampling cannot be performed\textsuperscript{133}. Thus, the temporal distribution of signal and noise may lead to erroneous glucose correlation.
2.6 Sources, Detector And Accessories For IR Spectroscopy

At present, only terminology related to quantities of IR radiations and detectors for these radiations have been relatively well standardized in the International Lighting Vocabulary published by the International Commission on Illumination (CIE) and the International Electrotechnical Commission (IEC) in 1987. However, there are vast areas of the IR technology where terminology is not standardized, mostly due to the fact that scientists and engineers of completely different background work nowadays in IR technology. (see table 2.2).

Table 2.4 Recommendation of the CIE for spectral bands.

<table>
<thead>
<tr>
<th>Spectral Bands</th>
<th>Wavelength</th>
</tr>
</thead>
<tbody>
<tr>
<td>UV-C</td>
<td>100 – 280 nm</td>
</tr>
<tr>
<td>UV-B</td>
<td>280- 315 nm</td>
</tr>
<tr>
<td>UV-A</td>
<td>315- 400nm</td>
</tr>
<tr>
<td>VIS</td>
<td>360 – 400nm to 760 – 800nm</td>
</tr>
<tr>
<td>IR-A</td>
<td>780 – 1400 nm</td>
</tr>
<tr>
<td>IR-B</td>
<td>1400 – 3000nm</td>
</tr>
<tr>
<td>IR-C</td>
<td>3000 – 10,00,000</td>
</tr>
</tbody>
</table>

2.6.1 Sources

The most commonly used light source used in the MIR region is a ‘SiC’ globar. This is a rod, heated to a temperature of about 800 °C by a current passing through it, which emits thermal radiation with a maximum intensity at approximately 4,800 nm. In the near infrared region, quartz halogen tungsten filament lamps are usually employed (Newport supply several types of filaments rated at various operating voltages and watts as shown in Table 2.3). They have a temperature of about 2,500 °C, providing maximum intensity at 1,818 nm. The absorption of quartz in the MIR region prevents application of this light
source in that spectral region. The power generated can be quantified using different physical unit as shown in Table 2.3.

Table 2.5 (a): Various Lamps and their parameters (Source: Newport).

<table>
<thead>
<tr>
<th>Lamp Wattage</th>
<th>DC Voltage</th>
<th>Approximate Flux (V)</th>
<th>Filament Size (WxH)</th>
<th>Filament Type</th>
<th>Colour Temperature</th>
<th>Permissible Burning Position</th>
<th>Average Life (Hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 W QTH</td>
<td>6</td>
<td>200</td>
<td>1.7 x 0.65</td>
<td>TC</td>
<td>3200</td>
<td>0 or 90°</td>
<td>100</td>
</tr>
<tr>
<td>20 W QTH</td>
<td>6</td>
<td>450</td>
<td>2.3 x 0.8</td>
<td>RD</td>
<td>3200</td>
<td>0 or 90°</td>
<td>100</td>
</tr>
<tr>
<td>45 W Calibrated QTH</td>
<td>6.8</td>
<td>705</td>
<td>4.75 H x 1.5 Dia.</td>
<td>CC</td>
<td>****</td>
<td>0 or 90°</td>
<td>****</td>
</tr>
<tr>
<td>45 W QTH</td>
<td>6.8</td>
<td>705</td>
<td>4.75 H x 1.5 Dia.</td>
<td>CC</td>
<td>****</td>
<td>0 or 90°</td>
<td>****</td>
</tr>
<tr>
<td>50 W, Short Filament</td>
<td>12</td>
<td>1600</td>
<td>3.3 x 1.6</td>
<td>RD</td>
<td>3300</td>
<td>0 or 90°</td>
<td>50</td>
</tr>
<tr>
<td>50 W Long Filament*</td>
<td>12</td>
<td>850</td>
<td>5.2 x 1.5</td>
<td>TC</td>
<td>3000</td>
<td>0 or 90°</td>
<td>3000</td>
</tr>
<tr>
<td>100 W QTH</td>
<td>12</td>
<td>3600</td>
<td>4.2 x 2.3</td>
<td>RD</td>
<td>3300</td>
<td>0 or 90°</td>
<td>50</td>
</tr>
<tr>
<td>200 W Calibrated QTH</td>
<td>30</td>
<td>5000</td>
<td>7.0 H x 3.6 Dia.</td>
<td>CC</td>
<td>****</td>
<td>0 or 90°</td>
<td>****</td>
</tr>
<tr>
<td>200 W QTH</td>
<td>30</td>
<td>5000</td>
<td>7.0 H x 3.6 Dia.</td>
<td>CC</td>
<td>****</td>
<td>0 or 90°</td>
<td>****</td>
</tr>
<tr>
<td>250 W QTH</td>
<td>24</td>
<td>10000</td>
<td>7.0 x 3.5</td>
<td>RD</td>
<td>3400</td>
<td>0 or 90°</td>
<td>50</td>
</tr>
<tr>
<td>600 W QTH</td>
<td>120</td>
<td>16500</td>
<td>4.0 x 13.5</td>
<td>CC</td>
<td>3200</td>
<td>Any</td>
<td>75</td>
</tr>
<tr>
<td>1000 W QTH</td>
<td>120</td>
<td>28000</td>
<td>5.0 x 18.0</td>
<td>CC</td>
<td>3200</td>
<td>Any</td>
<td>150</td>
</tr>
<tr>
<td>1000 W FEL QTH</td>
<td>120</td>
<td>27500</td>
<td>6.0 x 16.0</td>
<td>CC</td>
<td>3200</td>
<td>0 or 90°</td>
<td>300</td>
</tr>
<tr>
<td>1000 W Calibrated FEL QTH</td>
<td>120</td>
<td>27500</td>
<td>6.0 x 16.0</td>
<td>CC</td>
<td>3200</td>
<td>0 or 90°</td>
<td>****</td>
</tr>
<tr>
<td>1000 W FEL QTH</td>
<td>120</td>
<td>27500</td>
<td>6.0 x 16.0</td>
<td>CC</td>
<td>3200</td>
<td>0 or 90°</td>
<td>****</td>
</tr>
</tbody>
</table>

RD= Rectangular Dense, CC= Coiled Coil, TC Tight Coil
Table 2.5 (b): Physical quantity for quantifying the light.

<table>
<thead>
<tr>
<th>Radiometric Quantity</th>
<th>Usual Symbol</th>
<th>Units</th>
<th>Photometric Quantity</th>
<th>Usual Symbol</th>
<th>Units</th>
<th>Photon Quantity</th>
<th>Usual Symbol</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radiant Energy</td>
<td>Q_e</td>
<td>J</td>
<td>Luminous Energy</td>
<td>Q_e</td>
<td>lm s</td>
<td>Photon Energy</td>
<td>N_p</td>
<td>s^{-1}</td>
</tr>
<tr>
<td>Radiant Power or Flux</td>
<td>ϕ_e</td>
<td>W</td>
<td>Luminous Flux</td>
<td>ϕ_e</td>
<td>lm</td>
<td>Photon Flux</td>
<td>ϕ_p</td>
<td>s^{-1}</td>
</tr>
<tr>
<td>Radiant Exitance or Emittance</td>
<td>M_e</td>
<td>W m^{-2}</td>
<td>Luminous Exitance or Emittance</td>
<td>E_v</td>
<td>lx</td>
<td>Photo Irradiance</td>
<td>E_p</td>
<td>s^{-1}</td>
</tr>
<tr>
<td>Irradiance</td>
<td>E_0</td>
<td>W m^{-2}</td>
<td>Luminance</td>
<td>I_v</td>
<td>cd</td>
<td>Photo Intensity</td>
<td>I_p</td>
<td>s^{-1} sr^{-1}</td>
</tr>
<tr>
<td>Radiant Intensity</td>
<td>I_e</td>
<td>W sr^{-1}</td>
<td>Luminous Intensity</td>
<td>I_0</td>
<td>cd m^{-2}</td>
<td>Photo Radiance</td>
<td>I_p</td>
<td>s^{-1} sr^{-1}</td>
</tr>
</tbody>
</table>

2.6.2 Detectors

The table 2.4 gives spectral bands in IR regions with respect to the sensitivity of the detectors. Wavelength 1,100 nm is a sensitivity limit of popular ‘Si’ detectors. Similarly, wavelength of 3,000 nm is a long-wave sensitivity limit of ‘PbS’ and ‘InGaAs’ detectors; wavelength 6,000 nm is a sensitivity limit of ‘InSb’, ‘PbSe’, ‘PtSi’ detectors and ‘HgCdTe’ detectors optimized for 3,000-5,000 nm atmospheric window; and finally wavelength 15,000 nm is a long-wave sensitivity limit of ‘HgCdTe’ detectors optimized for 8,000-12,000 nm atmospheric window. It may be noted that ‘Ge’ detectors devices are not popular due to their high noise.

Table 2.6: Division of Infrared radiations as per the detectors type.

<table>
<thead>
<tr>
<th>Name</th>
<th>Wavelength range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near infrared NIR</td>
<td>780 – 1,000 nm</td>
</tr>
<tr>
<td>Short wave infrared SWIR</td>
<td>1,000 – 3,000 nm</td>
</tr>
<tr>
<td>Mid-wave infrared MWIR</td>
<td>3,000-6,000 nm</td>
</tr>
<tr>
<td>Very long wave infrared VLWIR</td>
<td>15,000 – 10,00,000nm</td>
</tr>
</tbody>
</table>
Table 2.7 Values of the energy Gap between the valence and conduction bands in semiconductors at room temperature.

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Eg(ev)</th>
<th>Crystal</th>
<th>Eg(ev)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diamond</td>
<td>5.33</td>
<td>PbS</td>
<td>0.34 - 0.37</td>
</tr>
<tr>
<td>Si</td>
<td>1.14</td>
<td>PbSe</td>
<td>0.27</td>
</tr>
<tr>
<td>Ge</td>
<td>0.67</td>
<td>PbTe</td>
<td>0.30</td>
</tr>
<tr>
<td>InAs</td>
<td>0.33</td>
<td>CdS</td>
<td>2.24</td>
</tr>
<tr>
<td>InAs</td>
<td>0.33</td>
<td>CdSe</td>
<td>1.74</td>
</tr>
<tr>
<td>InP</td>
<td>1.25</td>
<td>CdTe</td>
<td>1.45</td>
</tr>
<tr>
<td>GaAs</td>
<td>1.4</td>
<td>ZnO</td>
<td>3.2</td>
</tr>
<tr>
<td>AlSb</td>
<td>1.6-1.7</td>
<td>ZnS</td>
<td>3.6</td>
</tr>
<tr>
<td>GaP</td>
<td>2.25</td>
<td>ZnSe</td>
<td>2.60</td>
</tr>
<tr>
<td>SiC</td>
<td>3</td>
<td>AgCl</td>
<td>3.2</td>
</tr>
<tr>
<td>Te</td>
<td>0.33</td>
<td>AgI</td>
<td>2.8</td>
</tr>
<tr>
<td>ZnSb</td>
<td>0.56</td>
<td>Cu2O</td>
<td>2.1</td>
</tr>
<tr>
<td>GaSb</td>
<td>0.78</td>
<td>TiO2</td>
<td>3</td>
</tr>
</tbody>
</table>

The above discussed detectors are quantum detectors and have non linear detection mechanism and therefore if they are to be used for the instrumentation, a high level program is required to calibrate them for various wavelength of interest. Therefore most of the spectrophotometer employ thermal detectors instead of quantum detectors despite of low efficiency as thermal detector have flat responsivity in entire spectrum of interest. Some of the popular thermal detectors are described below.

Deuterated triglycine sulphate (DTGS) detector is the standard detector in most FT-IR instruments. This is a thermal detector, a so-called pyroelectric bolometer, which consists of a ferroelectric crystal which has a Curie point close to room temperature. The
crystal therefore exhibits a large change in electrical polarizability when exposed to modulated radiation. By placing electrodes on the crystal faces, the crystal acts as a capacitor across which an AC voltage may be measured. This detector is very linear, stable, and has a wide spectral range of operation.

Semiconductor based quantum detectors are used when increased sensitivity and low noise is required. In the MIR spectral region, the mercury cadmium telluride (HgCdTe) which use reference as ‘MCT’ detector is almost exclusively used. This detector usually requires cooling by liquid ‘N₂’. In the near infrared spectral region ‘InSb’ and ‘InAs’ detectors are employed. These detectors may be liquid ‘N₂’ or Peltier Cooled. Compared with the ‘DTGS’ detector, these detectors have lower noise, higher sensitivity, but a narrower spectral range of operation. The levels of intensity they may be exposed is far lower than the ‘DTGS’ accept. The ‘MCT’ detector in particular has a non-linear behavior when exposed to high levels of intensity. The difference in linearity between the MIR ‘MCT’ detector and the NIR ‘InSb’ and ‘InAs’ detectors may be understood from the structure of these detectors. In a semi-conductor structure with a conduction band separated from a valence band by a band-gap, a current can be measured, when free electrons, are created in the conduction band by absorption of photons with energies greater than the band-gap. A mid infrared detector has a small band-gap as it is required to detect radiation with low energies and thermal excitation creates a considerable number of electrons in the conduction band, even when the detector is cooled. The reservoir of electrons that may be excited by the incoming IR radiation is therefore small. This is illustrated in figure 2.3. With too much light intensity reaching the detector, the reservoir is dried out and the detector saturates.

In comparison to MIR detectors NIR detectors are required to detect much higher energies and, consequently, it has a much larger band-gap. For this reason, thermal
excitation does not create as much electrons in the conduction band as in MIR detectors. A higher incident flux of radiation is therefore permissible before saturation sets in. The use of dual-beam techniques where the intensity reaching the detector is twice as high as in the single-beam case is therefore much more limited by the ‘MCT’ detector in the MIR region than by the ‘InAs’ detector in the NIR region.

![Energy band gap diagram for IR quantum detectors.](image)

**Figure 2.9: Energy band gap diagram for IR quantum detectors.**

### 2.6.3 InGaAs Detector

Standard ‘Si’ detectors have a spectral range between 350nm to 1,100nm, and a UV enhanced type extend the range down to 200 nm in the ultraviolet. The Photomultiplier detector is the most sensitive detectors for UV & Visible radiation. In fact they provide relatively noise free signal directly proportional to the incident light, with the sensitivity several orders of magnitude greater than that of any system available. Operating on the principle of solid-state physics, semiconductor devices cover a much wider spectral range than PMT’s. ‘Ge’, ‘InGaAs’ and ‘InAs’ are also quantum devices, with greater sensitivity in the infrared. The fig. 2.4 illustrates the ‘InGaAs’ detector spectral response which indicates the responsivity variations over the wavelength.
Fig: 2.10: Spectral response of InGaAs detector. [1.7µm and InGaAs 2.2µm at +20°C. (InGaAs peak sensitivity moves approx. 1nm towards shorter wavelengths with 1°C cooling temperature.)]

2.6.4 Monochromator Control Module

Fourier Transform (FT) spectroradiometers differ from the Variable Filter (VF) and monochromator based spectroradiometers not only due to different spectral band selector but there are also significant differences in the role of the optics (fig. 2.5). In case of the variable filter spectroradiometers (a) and dispersive spectroradiometers (b) selection of the desired spectral band is done using convergent beams; while the interferometer is used in FT spectroradiometer (c) which works with parallel beams. Next, the variable filter and the monochromator are self-contained blocks in the sense that major spectral characteristics do not depend very much how you irradiate the input slit of the monochromator (the variable filter) and how you collect the radiation from the exit slit of the monochromator (the filter output). Changing the external optics increases or decreases your sensitivity, and also adds or reduces.
Fig. 2.11: Typical optical layout of external optics of three types of spectroradiometers a) VF spectroradiometer, b) dispersive spectroradiometer c) FT spectroradiometer.

Aberrations. However, in case of the FT spectroradiometers the spectral characteristics depends, also on the external optics playing major role. FT spectroradiometers are characterized by very good spectral resolution and very good sensitivity, better than that offered by other types of spectroradiometer. Very good spectral resolution can be obtained with use of the interferometer as a spectral selector. Better sensitivity originates from the fact that the detector is irradiated not only by the radiation from a desired narrow spectral band (the case of the variable filter spectroradiometer and the dispersive spectroradiometers) but by a full spectrum of radiation coming to the interferometer input. This feature enables design of high-speed, high spectral resolution FT spectroradiometers using standard or thermoelectrically cooled detectors (typically HgCdTe)
detectors) instead of bulky liquid nitrogen cooled detectors needed in the variable filter or dispersive spectroradiometers. However performance of the FT spectroradiometers can be severely reduced even by a very small non-alignment of the optical system which makes this type of spectroradiometers inherently sensitive to shocks and vibrations. Therefore FT spectroradiometers were for the last few decades considered as rather laboratory type equipment that cannot be used in field applications. However, at present this opinion is changing as there are on the market fully mobile FT spectroradiometers. Great majority of the commercially available spectroradiometers are systems enabling measurement of the spectral distribution of radiation emitted or reflected by a single spot and these systems can be termed the spot radiometers. There exists also another group termed the imaging spectroradiometers because these systems offer some imaging capabilities.

2.6.5 Material For Sample Holders

Various window materials and sample accessories (shown in table 2.6) used as in FT-IR spectrometers and must be transparent to IR radiation. The most common materials are salts that are hygroscopic and therefore ill suited for use in connection with aqueous solutions. Most beamsplitters consists of a base of such a material with a proper coating and FT-IR spectrometers are therefore either sealed, with a desiccant material within, or purged with dry air, free of water and carbon dioxide. These two gasses have intense infrared absorption bands and the purge also has the purpose of reducing the concentration of these two gases. The most commonly used window material for measurements on aqueous solutions is ‘CaF2’ which is sparingly dissolved by water and has an index of refraction which is close to that of water. Another commonly used material is ‘ZnSe’ which has a higher index of refraction. In the near infrared spectral range, quartz and sapphire windows may be employed. They have the advantage of being hard and chemically inert.
Table 2.8: Infrared materials for windows.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Material</th>
<th>Spectral range in nm</th>
<th>Refractive index at 10,000nm</th>
<th>Water solubility g/100g</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>KBr</td>
<td>204 -28985</td>
<td>1.52</td>
<td>53.5</td>
</tr>
<tr>
<td>2</td>
<td>NaCl</td>
<td>190-21,881</td>
<td>1.49</td>
<td>35.7</td>
</tr>
<tr>
<td>3</td>
<td>CaF2</td>
<td>125-9,000</td>
<td>1.39</td>
<td>0.0016</td>
</tr>
<tr>
<td>4</td>
<td>ZnSe</td>
<td>666 -21691</td>
<td>2.4</td>
<td>Insol.</td>
</tr>
<tr>
<td>5</td>
<td>Sapphire</td>
<td>250-6218</td>
<td>2.6</td>
<td>Insol.</td>
</tr>
<tr>
<td>6</td>
<td>Suprasil 300</td>
<td>175-3500</td>
<td>2.5</td>
<td>Insol.</td>
</tr>
<tr>
<td>7</td>
<td>Diamond</td>
<td>333-333,333</td>
<td>2.4</td>
<td>Insol.</td>
</tr>
</tbody>
</table>

2.6.6 Signal Conditioning Techniques

The measurement of an absorption spectrum of an aqueous solution is mainly carried out using two different signal capturing techniques, namely using a transmission cell (fig. 3.5) or an attenuated total reflection (ATR) cell (fig. 2.6). The transmission cell consists of two IR transparent windows between which the aqueous solution is placed. The light ray of the interferometer then passes through the IR windows and the aqueous solution before reaching the detector.

Transmission cells are well suited for near infrared spectroscopy of aqueous solutions. In this spectral region optimal pathlengths are in the range 0.5 -10mm and consequently much larger than the wavelength of the light. In the mid infrared region, the pathlength is in the range 7 -50 mm which is of the same order of magnitude as the wavelength of the light. This means that multiple reflections inside the transmission cell...
may cause fringe effects in the detector signal in the MIR region, but not in the NIR region. Attenuated total reflection is a capturing technique where light passes through a crystal or an IR optical fiber and is totally internally reflected. The evanescent wave reaches into the sample and is altered by changes in the absorption and index of refraction of the sample. This is illustrated in figure 2.7.

![Transmission window for IR spectroscopy](image)

**Figure 2.12:** Transmission window for IR spectroscopy.

![Attenuated total reflection (ATR) cell and Evanescent field](image)

**Figure 2.13:** Attenuated total reflection (ATR) cell and Evanescent field.

This allows absorption spectroscopy with penetration depths comparable to the wavelength of the light.
2.7 Signal Processing

In comparison, the execution speed of most DSP algorithms is limited almost completely by the number of multiplications and additions required. The MAC capability of the DSP can be very well used to exploit the operators involved in eq. 2.12 (As it resemble with the FIR filter). For example, the implementation of an FIR digital filter, the most common DSP technique, uses the standard notation. A FIR filter having output ‘\( y[n] \)’ performs this calculation by multiplying appropriate samples from the input ‘\( x[n] \)’ signal by a group of coefficients, ‘\( a_i \)’ and then adding the products (Fig 2.13). In equation form,

\[
y[n] = a_0 x[n] + a_1 x[n-1] + a_2 x[n-2] + a_3 x[n-3] + a_4 x[n-4] + \ldots
\]

2.13

This is simply saying that the input signal has been *convolved* with a filter kernel (i.e., an impulse response) consisting of ‘\( a_i \)’. Depending on \( a_0, a_1, a_2, a_3 \), there may only be a few coefficients in the filter kernel, or many thousands. While there is some data transfer and inequality evaluation in this algorithm, such as to keep track of the intermediate results and control the loops, the math operations dominate the execution time.

Also in comparison, most DSPs are used in applications where the processing is continuous, not having a defined start or end. For instance, consider an engineer designing a DSP system for an audio signal, such as a hearing aid. If the digital signal is being received at 20,000 samples per second, the DSP must be able to maintain a sustained throughput of 20,000 samples per second. However, there are important reasons not to make it any faster than necessary. As the speed increases, so does the cost, the power consumption, the design difficulty, and so on. This makes an accurate knowledge of the execution time critical for selecting the proper device, as well as the algorithms that can be applied.
2.7.1 Circular Buffering

Digital Signal Processors are designed to quickly carry out FIR filters and similar techniques. To understand the hardware, we must first understand the algorithms. In this section we will make a detailed list of the steps needed to implement an FIR filter. In the next section we will see how DSPs are designed to perform these steps as efficiently as possible. To start, we need to distinguish between off-line processing and real-time processing. In off-line processing, the entire input signal resides in the computer at the same time. For example, a geophysicist might use a seismometer to record the ground movement during an earthquake. After the shaking is over, the information may be read into a computer and analyzed in some way. Another example of off-line processing is medical imaging, such as computed tomography and MRI. The data set is acquired while the patient is inside the machine, but the image reconstruction may be delayed until a later time. The key point is that all of the information is simultaneously available to the processing program. This is common in scientific research and engineering, but not in consumer products. Off-line processing is the realm of personal computers and mainframes. In real-time processing, the output signal is produced at the same time that the input signal is being acquired. For example, this is needed in telephone communication, hearing aids, and RADAR & LIDAR. These applications must have the information immediately available, although it can be delayed by a short amount. For instance, a 10 millisecond delay in a telephone call cannot be detected by the speaker or listener. Likewise, it makes no difference if a radar signal is delayed by a few seconds before being displayed to the operator. Real-time applications input a sample, perform the algorithm, and output a sample, over-and-over. Alternatively, they may input a group of consecutive or similar data.
Now look at Fig. 2.14 and imagine that this is an FIR filter being implemented in real-time. To calculate the output sample, we must have access to a certain number of the most recent samples from the input. For example, suppose we use eight coefficients in this filter. This means we $a_0$, $a_1$, to $a_7$ must know the value of the eight most recent samples from the input signal. These eight samples must be stored in memory and $x[n]$, $x[n-1]$, to $x[n-7]$ continually updated related to sampling rate as new samples are acquired. The best way to manage these stored samples is by circular buffering as illustrated in Fig. 2.14.

**2.7.2 Architecture Of The Digital Signal Processor**

One of the biggest bottlenecks in executing DSP algorithms is transferring information to and from memory. This includes data, such as samples from the input signal and the filter coefficients, as well as program instructions, the binary codes that go into the program sequencer. For example, suppose we need to multiply two numbers that reside somewhere in memory. To do this, we must fetch three binary values from memory, the numbers to be multiplied, plus the program instruction describing what to do. Figure 2.15(a), shows how this seemingly simple task is done in a traditional microprocessor. This
is often called Von Neumann architecture. We only need other architectures when very fast processing is required, and we are willing to pay the price of increased complexity. This leads us to the Harvard architecture shown in 2.15(b). Since the buses operate independently, program instructions and data can be fetched at the same time, improving the speed over the single bus design. Most present day DSPs use this dual bus architecture.

Figure 2.15(c) illustrates the next level of sophistication, the Super Harvard Architecture. This term was coined by Analog Devices to describe the internal operation of their ADSP-2106x and new ADSP-211xx families of Digital Signal Processors. These are called SHARC® DSPs, a contraction of the longer term, Super Harvard ARCHitecture. The idea is to build upon the Harvard architecture by adding features to improve the throughput. While the SHARC DSPs are optimized in dozens of ways, two areas are important enough to be included in Fig. 2.14C: an instruction cache, and an I/O controller. First, let's look at how the instruction cache improves the performance of the Harvard architecture. A handicap of the basic Harvard design is that the data memory bus is busier than the program memory bus. When two numbers are multiplied, two binary values (the numbers) must be passed over the data memory bus, while only one binary value (the program instruction) is passed over the program memory bus. To improve upon this situation, we start by relocating part of the "data" to program memory. For instance, we might place the filter coefficients in program memory, while keeping the input signal in data memory. At first glance, this doesn't seem to help the situation; now we must transfer one value over the data memory bus (the input signal sample), but two values over the program memory bus (the program instruction and the coefficient). In fact, if we were executing random instructions, this situation would be no better at all. However, DSP algorithms generally spend most of their execution time in loops. This means that the same set of program instructions will
continually pass from program memory to the CPU. The Super Harvard architecture takes advantage of this situation by including an **instruction cache** in the CPU.

The math processing is broken into three sections, a **multiplier**, an **arithmetic logic unit** (ALU), and a **barrel shifter**. The multiplier takes the values from two registers, multiplies them, and places the result into another register. The ALU performs addition, subtraction, absolute value, logical operations (AND, OR, XOR, NOT), conversion between fixed and floating point formats, and similar functions. Elementary binary operations are carried out by the barrel shifter, such as shifting (division/multiplication), rotating, extracting and depositing segments, and so on. A powerful feature of the SHARC family is that the multiplier and the ALU can be accessed in parallel. In a single clock cycle, data from registers 0-7 can be passed to the multiplier, data from registers 8-15 can be passed to the ALU, and the two results returned to any of the 16 registers.

**2.7.2.1 Fixed versus Floating Point**

Digital Signal Processing can be divided into two categories, **fixed point** and **floating point**. These refer to the format used to store and manipulate numbers within the devices. Fixed point DSPs usually represent each number with a minimum of 16 bits, although a different length can be used. For instance, Motorola manufactures a family of fixed point DSPs that use 24 bits. There are four common ways that these possible bit patterns can represent a number. In **unsigned integer**, the stored number can take on any integer value from 0 to 65,535. Similarly, **signed integer** uses two's complement to make the range include negative numbers, from -32,768 to 32,767. With **unsigned fraction** notation, the 65,536 levels are spread uniformly between 0 and 1. Lastly, the **signed fraction** format allows negative numbers, equally spaced between -1 and 1.
In comparison, floating point DSPs typically use a minimum of 32 bits to store each value. This results in many more bit patterns than for fixed point, to be exact. In the most common format (ANSI/IEEE Std. 754-1985), the largest and smallest numbers, values are unequally $\pm 3.4 \times 10^{-38}$ $\pm 1.2 \times 10^{38}$ spaced between these two extremes, such that the gap between any two numbers is about ten-million times smaller than the value of the numbers. This is important because it places large gaps between large numbers, but small gaps between small numbers.
2.7.3 How Fast Are DSPs?

To handle these high-power tasks, several DSPs can be combined into a single system. This is called **multiprocessing** or **parallel processing**. The SHARC DSPs were designed with this type of multiprocessing in mind, and include special features to make it as easy as possible. For instance, no external hardware logic is required to connect the external busses of multiple SHARC DSPs together; all of the bus arbitration logic is already contained within each device. As an alternative, the link ports (4 bit, parallel) can be used to connect multiple processors in various configurations. Figure 2.17 shows typical ways that the SHARC DSPs can be arranged in multiprocessing systems. In Fig. 2.17(a) the algorithm is broken into sequential steps, with each processor performing one of the steps in an "assembly line".

![Diagram](image)

Fig. 2.16: The speed of DSP controller for various applications.
a. Data flow multiprocessing

Fig. 2.17: Multiprocessing configuration of DSP processor entity. a) Data flow type b) Cluster type.

b. Cluster multiprocessing

strategy. In fig. 2.17(b), the processors interact through a single shared global memory, accessed over a parallel bus (i.e., the external port). Figure 2.18 shows another way that a large number of processors can be combined into a single system, a 2D or 3D "mesh." Each of these configurations will have relative advantages and disadvantages for a particular task.

To make the programmer's life easier, the SHARC family uses a unified address space. This means that the 4 Gigaword address space, accessed by the 32 bit address bus, is divided among the various processors that are working together. To transfer data from one processor to another, simply read from or write to the appropriate memory locations. The SHARC internal logic takes care of the rest, transferring the data between processors at a rate as high as 240 Mbytes/sec (at 40 MHz).
2.7.4 FPGAs And DSP Processing

FPGAs have gained rapid acceptance and growth over the past decade because they can be applied to a very wide range of applications. A list of typical applications includes: random logic, integrating multiple SPLDs, device controllers, communication encoding and filtering, small to medium sized systems with SRAM blocks, and many more. Other interesting applications of FPGAs are prototyping of designs later to be implemented in gate arrays, and also emulation of entire large hardware systems. The former of these applications might be possible using only a single large FPGA (which corresponds to a small Gate Array in terms of capacity), and the latter would entail many FPGAs connected by some sort of interconnect; for emulation of hardware.
2.7.4.1 Potential Advantages of implementing a DSP function within an FPGA

Algorithm performance improvement in an FPGA-based implementation over the performance in a conventional DSP processor is usually based on a combination of factors. The most common are, increased data path width and/or increased operational speed resulting in a higher overall performance.

Another performance improvement is the ability to separate the data stream into multiple parallel blocks of data which have limited interdependence. Each data block can then be operated independently, resulting in higher relative performance. Taking advantage of any architectural opportunity for maximizing the number or speed of operations is essential to maximizing the performance achievable within an FPGA.

The critical architectural transformation necessary to maximize an algorithm's performance within an FPGA is the process of translating every serial operation or group of operations into the most parallel implementation possible, up to the limits imposed by the resources available within the target FPGA device for implementing a specific function.

A further performance advantage can be gained if the FPGA can perform operations on multiple channels or streams of data. Example applications include Time Division Multiple Access (TDMA) multiplexing, multiple channel communication protocols, and I/Q math based algorithms. Since each channel can be processed in parallel the performance advantage associated with the system can be multiplied by the number of times the channels implemented.

Designs which require signal pre-processing can also benefit, since filtering and signal conditioning algorithms are generally straight-forward to implement within an FPGA architecture. When an algorithm is implemented in a structure which takes advantage of the flexibility of target FPGA architecture, the benefits can be tremendous. Algorithms can be
customized to adjust to system requirements on the fly. Filter coefficients, implementations, and architectures can be updated to reflect changing system conditions and user requirements.

The implementation of an algorithm within an FPGA also provides a range of implementation options. The design team must determine and prioritize their design objectives. It is possible to implement an algorithm as a maximally parallelized architecture, or in a highly serial architecture using a single structure which is fed with sequential data elements. Hybrid architecture can also be implemented, which is a parallel implementation of serial structures or a serial chain of parallel architectures. Each of these design options will have its own set of characteristics including the number of devices required to implement a function, resource requirements within a device, maximum speed, and cost of implementation. The design team has the flexibility to optimize for size, speed, cost, or a target combination of these factors.

An FPGA device also provides a platform for integrating multiple design functions into a single package or a group of packages. Integration of functionality can result in higher performance, reduced real-estate requirements, and reduced power requirements. Resources integrated into the I/O circuitry of FPGAs can further improve system performance by allowing control of drive strength, signal slew rate, and implementation of on-board matching, resulting in fewer required, system-level components.

Further design integration can be implemented by incorporating hard or soft processor cores within an FPGA to implement required control and processing functionality. The availability of pre-verified design functionality through Intellectual Property (IP) availability can also be used to implement and incorporate common functionality. The ability to incorporate multiple system-level components and design
functionality within a smaller quantity of components can potentially reduce risk, cost, and schedule.

### 2.7.4.2 Numerical Representation

There are multiple numeric representation styles which can be used to represent data as it passes through an algorithm within an FPGA. The two largest classes of numeric representation are fixed-point and floating-point. Informed selection of the numeric representation style can maximize the utilization of available FPGA resources. Traditionally fixed-point implementations are considered first for DSP algorithm implementation within FPGAs. This is due to the perceived ability to operate at higher operational data throughput rates and more efficient utilization of FPGA resources. For more complex algorithms floating-point may also be considered. The advantage of a higher dynamic range and elimination of data path scaling, are the primary advantages for floating-point design implementation. The implementation of floating-point numeric representation can come with the penalties of higher resource utilization and less intuitive design implementations.

Fixed-point numbers can be represented as unsigned integers or signed magnitude values. The two most popular signed representations are two's complement and one's complement. Two's complement is the more popular of the two formats, due to its simplified implementation when considering arithmetic overflow conditions. One's complement has the characteristic that negative and positive numbers have identical bit patterns, with the exception of the leading sign bit.

There are also more advanced numeric representations and modified implementations of existing standards. Where possible, it is desirable to implement designs based on existing defined standards. This supports simplified design comprehension for engineers, new to a project and simplified modification for future modifications or enhancements.
2.7.4.3 DSP-Oriented Architectural Features

The architecture of FPGA fabric is inherently suited to implementation of parallel structures. The capability to support very wide buses and implement multiple instantiations of complex structures is a key feature of FPGA technology. There are generally multiple options for implementing individual DSP-related operations within an FPGA. The structures which manufacturers continue to optimize for DSP performance include:

- Clock management and distribution
- Distributed and block memory within the FPGA
- Access to memory external to the FPGA
- Implementation of low-overhead shift registers
- Embedded wide hardware multiplier blocks
- Advanced hardware multiplier blocks with associated accumulator functionality

For an example consider the implementation of the primary DSP algorithm structure; the Multiply Accumulate (MAC) function. A MAC structure can be implemented in one of several different configurations: Both the Multiplier and the Accumulator can be implemented within the logic fabric of the FPGA taking advantage of dedicated structures such as dedicated high-speed carry chains. The Multiplier can be implemented in an optimized multiplier block which does not require the use of FPGA logic, and the Accumulator implemented within the logic fabric of the FPGA. Both the Multiplier and Accumulator can be implemented within an advanced Multiplier block requiring the use of no FPGA logic. Each of these approaches has its own characteristics. Depending on the architecture there may be a limit on the number of available optimized hard multiplier blocks. There can be speed advantages for logic level or hard multiplier block implementations, in addition to using more or less of the available FPGA logic matrix respectively.