3

SOLID SURFACES SENSING IN THE VISIBLE AND NEAR INFRARED

The visible and near-infrared regions of the electromagnetic spectrum have been the most commonly used in remote sensing of planetary surfaces. This is partially due to the fact that this is the spectral region of maximum illumination by the sun and most widely available detectors (electro-optical and photographic). The sensor detects the electromagnetic waves reflected by the surface and measures their intensity in different parts of the spectrum. By comparing the radiometric and spectral characteristics of the reflected waves to the characteristics of the incident waves, the surface reflectivity is derived. This in turn is analyzed to determine the chemical and physical properties of the surface. The chemical composition and crystalline structure of the surface material has an effect on the reflectivity because of the molecular and electronic processes which govern the interaction of waves with matter. The physical properties of the surface, such as roughness and slope, also affect the reflectivity mainly due to geometric factors related to the source–surface–sensor relative angular configuration.

Figure 3-1. The surface spectral imprint is reflected in the spectrum of the reflected wave.

Thus, information about the surface properties is acquired by measuring the modulation that the surface imprints on the reflected wave (Fig. 3-1) by the process of wave–matter interactions, which will be discussed in this chapter.

3-1 SOURCE SPECTRAL CHARACTERISTICS

By far the most commonly used source of illumination in the visible and near infrared is the sun. In the most simple terms, the sun emits approximately as a hot blackbody at 6000 °K temperature. The solar illumination spectral irradiance at the Earth’s distance is shown in Figure 3-2. The total irradiance is measured to be approximately 1370 W/m² above the Earth’s atmosphere. This irradiance decreases as the square of the distance from the sun because of the spherical geometry. Thus, the total solar irradiance at Venus is about twice the value for Earth while it is half that much at Mars (see Table 3-1).
As the solar waves propagate through a planet’s atmosphere, they interact with the atmospheric constituents leading to absorption in specific spectral regions, which depends on the chemical composition of these constituents. Figure 3-2 shows the sun illumination spectral irradiance at the Earth’s surface. Strong absorption bands exist in the near infrared, particularly around 1.9, 1.4, 1.12, 0.95, and 0.76 μm. These are mainly due to the presence of water vapor (H₂O), carbon dioxide (CO₂), and, to a lesser extent, oxygen (O₂). In addition, scattering and absorption lead to a continuum of attenuation across the spectrum. For comparison, the transmittivity of the Martian atmosphere is shown in Figure 3-3. In the near infrared, the Martian atmosphere is opaque only in very narrow spectral bands near 2.7 and 4.3 μm, which are due to CO₂. The absorption band near 2 μm is due to H₂O.

Figure 3-2. Sun illumination spectral irradiance at the Earth’s surface. (From Chahine, et al. 1983.)

Figure 3-3. Transmittivity of the Martian atmosphere. The model uses band parameters for CO₂, H₂O, and O₃. Surface pressure assumed is 6 mbar. (Courtesy of D. Crisp and D. McCleese.)
TABLE 3-1. Solar Irradiance at the Distance of the Different Planets

<table>
<thead>
<tr>
<th>Planets</th>
<th>Mean Distance to the Sun (AU)</th>
<th>Million Kilometers</th>
<th>Solar Irradiance (W/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mercury</td>
<td>0.39</td>
<td>58</td>
<td>9000</td>
</tr>
<tr>
<td>Venus</td>
<td>0.72</td>
<td>108</td>
<td>2640</td>
</tr>
<tr>
<td>Earth</td>
<td>1.0</td>
<td>150</td>
<td>1370</td>
</tr>
<tr>
<td>Mars</td>
<td>1.52</td>
<td>228</td>
<td>590</td>
</tr>
<tr>
<td>Jupiter</td>
<td>5.19</td>
<td>778</td>
<td>50</td>
</tr>
<tr>
<td>Saturn</td>
<td>9.51</td>
<td>1427</td>
<td>15</td>
</tr>
<tr>
<td>Uranus</td>
<td>19.13</td>
<td>2870</td>
<td>3.7</td>
</tr>
<tr>
<td>Neptune</td>
<td>30</td>
<td>4497</td>
<td>1.5</td>
</tr>
<tr>
<td>Pluto</td>
<td>39.3</td>
<td>5900</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Another important factor in visible and near-infrared remote sensing is the relative configuration of the sun–surface–sensor. Because of the tilt of the Earth’s rotation axis relative to the plane of the ecliptic, the sun’s location in the sky varies as a function of the seasons and the latitude of the illuminated area. In addition the ellipticity of the Earth’s orbit has to be taken into account (Dozier and Strahler, 1983; Wilson, 1980).

With recent and continuing development, high-power lasers are becoming viable sources of illumination even from orbital altitudes. These sources have a number of advantages, including controlled illumination geometry, controlled illumination timing which can be used in metrology, and the possibility of very high powers in very narrow spectral bands. However, these advantages have to be compared to the lack of instantaneous wide-spectral coverage and the need for orbiting the laser source with its associated weight and power requirements.

3-2 WAVE-SURFACE INTERACTION MECHANISMS

When the electromagnetic wave interacts with a solid material, there are a number of mechanisms which affect the properties of the resulting wave. Some of these mechanisms operate over a narrow band of the spectral region, while others are wide band and thus affect the entire spectrum from 0.3 to 2.5 μm. These interaction mechanisms are summarized in Table 3-2 and discussed in detail in this section. The narrow band interactions are usually associated with resonant molecular and electronic processes. These mechanisms are strongly affected by the crystalline structure leading to splitting, displacement, and broadening of the spectral lines into spectral bands. The wide band mechanisms are usually associated with nonresonant electronic processes which affect the material index of refraction (i.e., velocity of light in the material).

3-2-1 Reflection, Transmission, and Scattering

When an electromagnetic wave is incident on an interface between two materials (in the case of remote sensing one of the materials is usually the atmosphere), some of the energy is reflected in the specular direction, some of it is scattered in all directions of the incident medium, and some of it
is transmitted through the interface (Fig. 3–4). The transmitted energy is usually absorbed in the bulk material and either reradiated by electronic or thermal processes or dissipated as heat.

**TABLE 3-2. Interaction Mechanisms**

<table>
<thead>
<tr>
<th>General Physical Mechanisms</th>
<th>Specific Physical Mechanisms</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometrical and physical optics</td>
<td>Dispersive refraction</td>
<td>Chromatic aberration, rainbow</td>
</tr>
<tr>
<td>Scattering</td>
<td>Molecular vibration</td>
<td>Blue sky, rough surfaces</td>
</tr>
<tr>
<td>Reflection, refraction, and interference</td>
<td></td>
<td>Mirror, polished surfaces, oil</td>
</tr>
<tr>
<td>film on water, lens coating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diffraction grating</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vibrational excitation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Molecular vibration</td>
<td></td>
<td>Opal, liquid crystals, gratings</td>
</tr>
<tr>
<td>silicon + oxygen</td>
<td></td>
<td>H₂O, aluminum + oxygen, or</td>
</tr>
<tr>
<td>Ion vibration</td>
<td></td>
<td>Hydroxyl ion (OH)</td>
</tr>
<tr>
<td>Electronic excitation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crystal field effects in transition</td>
<td></td>
<td>Turquoise, most pigments,</td>
</tr>
<tr>
<td>metal compounds</td>
<td></td>
<td>some fluorescent materials</td>
</tr>
<tr>
<td>Crystal field effects in transition</td>
<td></td>
<td>Ruby, emerald, red sandstone</td>
</tr>
<tr>
<td>metals impurities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charge transfer between molecular orbits</td>
<td></td>
<td>Magnetite, blue sapphire</td>
</tr>
<tr>
<td>Conjugated bonds</td>
<td></td>
<td>Organic dyes, most plant colors</td>
</tr>
<tr>
<td>Transitions in materials with energy bands</td>
<td></td>
<td>Metallic: copper, silver, gold;</td>
</tr>
<tr>
<td>diamond, galena</td>
<td>semiconductors: silicon, cinnabar,</td>
<td></td>
</tr>
</tbody>
</table>

When the interface is very smooth relative to the incident wavelength \( \lambda \) (i.e., \( \lambda \gg \) interface roughness), the reflected energy is in the specular direction and the reflectivity is given by Snell’s law. The reflection coefficient is a function of the complex index of refraction \( n \) and the incidence angle. The expression of the reflection coefficient is given by:

\[
| R_h |^2 = \frac{\sin^2 (\theta - \theta_i)}{\sin^2 (\theta + \theta_i)} \quad (3–1)
\]

for horizontally polarized incidence wave, and

\[
| R_v |^2 = \frac{\tan^2 (\theta - \theta_i)}{\tan^2 (\theta + \theta_i)} \quad (3–2)
\]

for vertically polarized incidence wave. \( \theta_i \) is the transmission angle and is given by:

\[
\sin \theta = n \sin \theta_i. \quad (3–3)
\]
The behavior of the reflection coefficient as a function of the incidence angle is illustrated in Figure 3-5. One special aspect of the vertically polarized configuration is the presence of a null (i.e., no reflection) at an angle which is given by

$$\tan \theta = n$$

and which corresponds to $R_\psi = 0$. This is called the Brewster angle.

In the case of normal incidence, the reflection coefficient is:

$$R_n = R_\psi = R = \left( \frac{n-1}{n+1} \right) = \left( \frac{N_r+iN_i-1}{N_r+iN_i+1} \right)$$

(3–5)

and

$$|R|^2 = \frac{(N_r-1)^2 + N_i^2}{(N_r+1)^2 + N_i^2}$$

(3–6)

**Figure 3-5.** Reflection coefficient of a half-space with two indices of refraction ($n = 3$ and $n = 8$) as a function of incidence angle. The continuous curve corresponds to horizontal polarization. The dashed curve corresponds to vertical polarization.
where $N_r$ and $N_i$ are the real and imaginary parts of $n$, respectively. In the spectral regions away from any strong absorption bands, $N_i \ll N_r$ and
\[
|R|^2 = \left( \frac{N_r - 1}{N_r + 1} \right)^2
\]
(3–7)

However, in strong absorption bands, $N_i \gg N_r$ and
\[
|R|^2 = 1
\]
(3–8)

Thus, if wide spectrum light is incident on a polished surface, the reflected light will contain a relatively large portion of spectral energy around the absorption bands of the surface material (Fig. 3-6). This is the restrahlen effect.

\[\text{Figure 3-6.} \quad \text{For a polished surface there is an increase in the reflected energy near an absorption spectral line.}\]

In actuality, most natural surfaces are rough relative to the wavelength and usually consist of particulates. Thus, scattering plays a major role and the particulates’ size distribution has a significant impact on the spectral signature. An adequate description of the scattering mechanism requires a rigorous solution of Maxwell’s equations, including multiple scattering. This is usually very complex, and a number of simplified techniques and empirical relationships are used. For instance, if the particles are small relative to the wavelength, then Rayleigh’s law is applicable, and it describes adequately the scattering mechanism. In this case, the scattering cross section is a fourth power of the ratio $a/\lambda$ (i.e., scattering is $\sim (a/\lambda)^4$) of the particles’ size over the wavelength. The Rayleigh scattering explains the blue color of the sky; molecules and very fine dust scatter blue light about four times more than red light making the sky appear blue. The direct sunlight, on the other hand, is depleted of the blue and therefore appears reddish.

In the case of a particulate surface, the incident wave is multiply scattered and some of the energy which is reflected toward the incident medium penetrates some of the particles (Fig. 3-7). Thus, if the material has an absorption band, the reflected energy is depleted of energy in that band. Usually, as the particles get larger, the absorption features become more pronounced even though the total reflected energy is decreased. This is commonly observed in measured spectra, as illustrated in Figure 3-8. This is the usual case in remote sensing in the visible and infrared.
In the case of a particulate layer the volume scattering and resulting absorption leads to a decrease of the reflected (scattered) energy near an absorption spectral line.

In the general case of natural surfaces, the reflectivity is modeled by empirical expressions. One such expression is the Minnaert law (Minnaert, 1941), which gives:

\[ B \cos \theta_s = B_0 (\cos \theta_i \cos \theta_s)^\kappa \]  (3–9)

where \( B \) is the apparent surface brightness, \( B_0 \) is the brightness of an ideal reflector at the observation wavelength, \( \kappa \) is a parameter that describes darkening at zero phase angle, and \( \theta_i, \theta_s \) are the incidence and scattering (or emergence) angles, respectively. For a Lambertian surface, \( \kappa = 1 \) and

\[ B = B_0 \cos \theta_i \]  (3–10)

which is the commonly used cosine darkening law.

In some special cases, diffraction gratings are encountered in nature. For instance, opal consists of closely packed spheres of silicon dioxide and a little water imbedded in a transparent...
matrix that has similar composition but slightly different index of refraction. The spheres have a
diameter of about 0.25 \( \mu m \). Dispersion of white light by the three-dimensional diffraction grating
gives rise to spectrally pure colors that glint from within an opal.

The reflection of visible and near-infrared waves from natural surfaces occurs within the top
few microns. Thus, surface cover plays an important role. For instance, the iron oxide in desert
varnish can quench or even completely mask the spectrum of the underlying rocks.

3-2-2 Vibrational Processes

Vibrational processes correspond to small displacements of the atoms from their equilibrium
positions. In a molecule composed of \( N \) atoms, there are \( 3N \) possible modes of motion because each
atom has three degrees of freedom. Of these modes of motion, three constitute translation, three
(two in the case of linear molecules) constitute rotation of the molecule as a whole, and \( 3N - 6 \)
(\( 3N - 5 \) in the case of linear molecules) constitute independent type of vibrations (Fig. 3-9). Each
mode of motion leads to vibration at a classical frequency \( \nu_i \). For a molecule with many classical
frequencies \( \nu_i \), the energy levels are given by

\[
E = (n_1 + \frac{1}{2})h\nu_1 + \cdots + (n_{3N-6} + \frac{1}{2})h\nu_{3N-6}
\]

where \( (n + \frac{1}{2})h\nu \) are the energy levels of a linear harmonic oscillator, and \( n_i \) are vibrational
quantum numbers (\( n_i = 0, 1, \ldots \)). The number and values of the energy levels for a material are
thus determined by the molecular structure (i.e., number and type of the constituent atoms, the
molecular geometry, and the strength of the bonds).

The transitions between the ground state (all \( n_i = 0 \)) to a state where only one \( n_i = 1 \) are
called fundamental tones. The corresponding frequencies are \( \nu_1, \nu_2, \ldots, \nu_i \). These usually occur in
the far-to mid-infrared (> 3 \( \mu m \)). The transitions between the ground state to a state where only one
\( n_i = 2 \) (or some multiple integer) are called overtone tones. The corresponding frequencies are
\( 2\nu_1, 2\nu_2, \ldots \) (or higher order overtones). The other transitions are called combination tones, which
are combinations of fundamental and overtone transitions. The corresponding frequencies are
\( l\nu_i + m\nu_j \), where \( l \) and \( m \) are integers. Features due to overtone and combination tones usually
appear between 1 and 5 \( \mu m \).

As an illustration, let us consider the case of the water molecule (Fig. 3-9a). It consists of
three atoms (e.g., \( N = 3 \)) and has three classical frequencies \( \nu_1, \nu_2, \nu_3 \), which correspond to the
three wavelengths:

\[
\lambda_1 = 3.106 \ \mu m , \text{ which corresponds to the symmetric OH stretch}
\lambda_2 = 6.08 \ \mu m , \text{ which corresponds to the HOH bend}
\lambda_3 = 2.903 \ \mu m , \text{ which corresponds to the asymmetric OH stretch}
\]
These wavelengths are the fundamentals. The lowest order overtones correspond to the frequencies $2\nu_1$, $2\nu_2$, and $2\nu_3$, and to wavelengths $\lambda_1/2$, $\lambda_2/2$, and $\lambda_3/2$. An example of a combination tone can be $\nu = \nu_3 + \nu_2$, which has a wavelength given by

$$\frac{1}{\lambda} = \frac{1}{\lambda_3} + \frac{1}{\lambda_2} \rightarrow \lambda = 1.87 \mu m$$

or $\nu' = 2\nu_1 + \nu_3$, which has a wavelength $\lambda' = 0.962 \mu m$.

Figure 3-9. (a) H$_2$O molecule fundamental vibrational modes. (b) CO$_2$ molecule (linear) fundamental vibrational modes.

In the spectra of minerals and rocks, whenever water is present, two bands appear, one near 1.45 $\mu$m (due to $2\nu_1$) and one near 1.9 $\mu$m (due to $\nu_2 + \nu_3$). The presence of these bands is usually diagnostic of the presence of the water molecule. These bands could be sharp, indicating that the water molecules are located in well-defined, ordered sites, or they may be broad, indicating that they occupy unordered or several unequivalent sites. The exact location and appearance of the spectral bands give quite specific information about the way in which the molecular water is associated with the inorganic material. Figure 3-10 illustrates this effect by showing spectra of various material that contains water. The $2\nu_3$ and $\nu_2 + \nu_3$ tones are clearly displayed and the variation in the exact location and spectral shape is clearly visible.
Figure 3-10. Spectra of water-bearing minerals illustrating the variations in the exact location and shape of the spectral lines associated with two of the tones of the water molecule: $2\nu_3$ near 1.4 $\mu$m and $\nu_2 + \nu_3$ near 1.9 $\mu$m. (From Hunt, 1977).

All the fundamental vibrational modes of silicon, magnesium, and aluminum with oxygen occur near 10 $\mu$m or at longer wavelengths. Because the first overtone near 5 $\mu$m is not observed, one does not expect to detect direct evidence of them at higher order overtones in the near infrared. What is observed in the near infrared are features due to the overtones and combinations involving materials that have very high fundamental frequencies. There are only a few groups that provide such features and by far the most common of these are the near-infrared bands that involve the OH stretching mode.

The hydroxyl ion, OH, very frequently occurs in inorganic solids. It has one stretching fundamental which occurs near 2.77 $\mu$m, but its exact location depends upon the site at which the OH ion is located and the atom attached to it. In some situations this spectral feature is doubled, indicating that the OH is in two slightly different locations or attached to two different minerals. Al-OH and Mg-OH bending have fundamental wavelengths at 2.2 and 2.3 $\mu$m, respectively. The first overtone of OH ($2\nu$ at about 1.4 $\mu$m) is the most common feature present in the near-infrared spectra of terrestrial materials. Figure 3-11 shows examples of spectra displaying the hydroxyl tones.
Carbonate minerals display similar features between 1.6 and 2.5 μm which are combinations and overtones of the few fundamental internal vibrations of the $\text{CO}_3^{2-}$ ion.

### 3-2-3 Electronic Processes

Electronic processes are associated with electronic energy levels. Electrons in an atom can only occupy specific quantized orbits with specific energy levels. The most common elements in rocks, namely silicon, aluminum, and oxygen, do not possess energy levels such that transitions between them can yield spectral features in the visible or near infrared. Consequently, no direct information is available on the bulk composition of geological materials. However, indirect information is available as a consequence of the crystal or molecular structure imposing its effect upon the energy levels of specific ions. The most important effects relevant to remote sensing are crystal field effect, charge transfer, conjugate bonds, and transitions in materials with energy bands.

*Crystal Field Effect*

One consequence of the binding together of atoms is a change in the state of valence electrons. In an isolated atom, the valence electrons are unpaired and are the primary cause of color. In a molecule and in many solids, the valence electrons of adjacent atoms form pairs which constitute the
chemical bonds that hold atoms together. As a result of this pair formation, the absorption bands of the valence electrons are usually displaced to the UV. However, in the case of transition metal elements such as iron, chromium, copper, nickel, cobalt, and manganese, the atoms have inner shells that remain only partially filled. These unfilled inner shells hold unpaired electrons which have excited states that often fall in the visible spectrum. These states are strongly affected by the electrostatic field which surrounds the atom. The field is determined by the surrounding crystalline structure. The different arrangement of the energy levels for different crystal fields leads to the appearance of different spectra for the same ion. However, not all possible transitions may occur equally strongly. The allowed transitions are defined by the selection rules. In this situation, the most pertinent rule is the one related to the electron spins of the state involved. This rule allows transitions between states of identical spins only. Typical examples of such situations are ruby and emerald.

The basic material of ruby is corundum, an oxide of aluminum \((\text{Al}_2\text{O}_3)\). A few percent of the aluminum ions are replaced by chromium ions \((\text{Cr}^{3+})\). Each chromium ion has three unpaired electrons whose lowest possible energy is a ground state designated \(4A_2\) and a complicated spectrum of excited states. The exact position of the excited state is determined by the crystal electric field in which the ion is immersed. The symmetry and strength of the field are determined in turn by the nature of the ions surrounding the chromium and their arrangement. In the case of the ruby, there are three excited states \((2E, 4T_1, 4T_2)\) with energy bands in the visible range.

Selection rules forbid a direct transition from \(4A_2\) to \(2E\), but allow transitions to both \(4T_1\) and \(4T_2\) (see Fig. 3-12). The energies associated with these transitions correspond to wavelengths in the violet and yellow-green regions of the spectrum. Hence, when white light passes through a ruby, it emerges as a deep red (i.e., depleted of violet, yellow, or green).

Because of the selection rules, electrons can return from the \(4T\) states to the \(4A_2\) ground state only through the \(2E\) level. The \(4T\) to \(2E\) transition releases infrared waves, but the \(2E\) to \(4A_2\) transition gives rise to strong red light emission.

In the case of emerald, the impurity is also \(\text{Cr}^{3+}\); however, because of the exact structure of the surrounding crystal, the magnitude of the electric field surrounding a chromium ion is somewhat reduced. This results in lower energy for the \(4T\) states (Fig. 3-12). This shifts the absorption band into the yellow-red, hence the green color of emerald.

Crystal field effects can arise whenever there are ions bearing unpaired electrons in a solid. Aquamarine, jade, and citrine quartz have iron instead of chromium. Blue or green azurite, turquoise, and malachite have copper as a major compound instead of impurity. Similar situations occur with red garnets where iron is a major compound.

One of the most important cases in remote sensing involves ferrous ion \((\text{Fe}^{2+})\). For a ferrous ion in a perfectly octahedral site only one spin-allowed transition is in the near infrared. However, if the octahedral site is distorted, the resulting field may lead to splitting of the energy level, thus allowing additional transitions. If the ferrous ions are in different nonequivalent sites, as is the case in olivine, additional characteristic transitions occur. Thus important information on the bulk
structure of the mineral can be derived indirectly by using the ferrous ion spectral feature as a probe. Figure 3-13 shows the reflection spectra of several minerals that contain ferrous ion, illustrating the spectral changes resulting from the bulk structure.

**Figure 3-12.** Basis for the color characteristics of emerald and ruby.

The crystal field effect is not only limited to electrons in transition metal ions. If an electron is trapped at a structural defect, such as a missing ion or an impurity, similar transition energy level occurs. These anomalies are called color centers or F centers (from the German Farbe, color). The color center effect is responsible for the purple color of fluorite ($\text{CaF}_2$), in which each calcium ion is surrounded by eight fluorine ions. An F center forms when a fluorine ion is missing from its usual position. In order to keep electrical neutrality, an electron is trapped in the fluorine location and is bound to the location by the crystal field of the surrounding ions. Within this field it can occupy a ground state and various excited states.

**Charge Transfer**

In many situations paired electrons are not confined to a particular bond between two atoms but move over longer distances. They even range throughout the molecule or even throughout a macroscopic solid. They are then bound less tightly and the energy needed to create an excited state is reduced. The electrons are said to occupy molecular orbits (in contrast to atomic orbits). One mechanism by which molecular orbits can contribute to the spectral characteristics in the visible and NIR is the transfer of electric charges from one ion to another. An example of such effect occurs in materials where iron is present in both its common values $\text{Fe}^{2+}$ and $\text{Fe}^{3+}$. Charge transfers between these forms give rise to colors ranging from deep blue to black, such as in the black iron ore magnetite. A similar mechanism occurs in blue sapphire. In this case the impurities are $\text{Fe}^{2+}$ and
An excited state is formed when an electron is transferred from the iron to the titanium giving both a $+3$ charge. About 2 eV are needed to drive the charge transfer creating a broad absorption band that extends from the yellow through the red leaving the sapphire with a deep blue color.

**Figure 3-13.** Spectra of minerals that contain ferrous ions in different crystal fields or in different sites. All the features indicated (vertical line is band minimum and shaded area shows bandwidth) in these spectra are due to spin-allowed transitions. (From Hunt, 1977.)

The spectral features which occur as a result of charge transfers typically are very intense, much more intense than the features corresponding to crystal field effects.

**Conjugate Bonds**

Molecular orbital transitions play a major role in the spectral response of biological pigments and many organic substances in which carbon, and sometime nitrogen, atoms are joined by a system of alternating single and double bonds called conjugate bonds. Because each bond represents a pair of shared electrons, moving a pair of electrons from each double bond to the adjacent single bond reverses the entire sequence of bonds leading to an equivalent structure. Actually the best representation of the structure shows all the atoms connected by single bonds with the remaining pairs of bonding electrons distributed over the entire structure in molecular orbitals, which in this instance are called pi-orbitals.
The extended nature of pi-orbitals in a system of conjugate bonds tends to diminish the excitation energy of the electron pairs allowing absorption in the visible spectrum. A number of biological pigments owe their spectral properties to extended systems of pi-orbitals. Among them are the green chlorophyll of plants and the red hemoglobin of blood.

**Materials with Energy Bands**

The spatial extent of electron orbitals reaches its maximum value in metals and semi-conductors. Here the electrons are released entirely from attachment to particular atoms or ions and even move freely throughout a macroscopic volume.

In a metal all the valence electrons are essentially equivalent since they can freely exchange places. The energy levels form a continuum. Thus a metal can absorb radiation at any wavelength. This might lead someone to conclude that metals should be black. However when an electron in a metal absorbs a photon and jumps to an excited state, it can immediately reemit a photon of the same energy and return to its original state. Because of the rapid and efficient reradiation, the surface appears reflective rather than absorbent; it has the characteristic metallic luster. The variations in the color of metallic surfaces result from differences in the number of states available at particular energies above the Fermi level. Because the density of states is not uniform, some wavelengths are absorbed and reemitted more efficiently than others.

In the case of semiconductors there is a splitting of the energy levels into two broad bands with a forbidden gap (Fig. 3-14). The lower energy levels in the valence band are completely occupied. The upper conduction band is usually empty. The spectral response of a pure semiconductor depends on the width of the energy gap. The semiconductor cannot absorb photons with energy less than the gap energy. If the gap is small, all wavelengths in the visible spectrum can be absorbed. Small-gap semiconductors in which reemission is efficient and rapid, such as silicon, have a metal-like luster.

![Figure 3-14](image.png)

**Figure 3-14.** Configurations of energy bands for different types of solid materials.

If the gap is large, no wavelengths in the visible can be absorbed (i.e., photon energy less than the gap energy) and the material is colorless. Diamond is such a material, with an energy gap of 5.4 eV (i.e., $\lambda = 0.23 \ \mu m$). Where the energy gap is intermediate, the semiconductor will have a definite color. Cinnebar (HgS) has a band gap of 2.1 eV (i.e., $\lambda = 0.59 \ \mu m$). All photons with energy higher than this level (i.e., blue or green) are absorbed, and only the longest visible
wavelengths are transmitted; as a result, cinnebar appears red (see Fig. 3-15). Semiconductor materials are characterized with a sharp edge of transition in their spectrum due to the sharp edge of the conduction band. The sharpness of the absorption edge is a function of the purity and crystallinity of the material. For particulate materials, the absorption edge is more sloped.

If the semiconductor is doped with impurity, new intermediate energy levels are available allowing some large-gap semiconductors to have spectral signature in the visible. The technique of doping semiconductors is often used in the manufacture of detector materials. Since the doped layer has a lower transition energy, longer wavelength radiation can typically be detected. An example is the case where silicon is doped with arsenic. The resulting detectors have sensitivity that extend well into the infrared; far beyond the normal cut-off wavelength of pure silicon at about 1.1 µm.

Figure 3-15. Visible and infrared bidirectional reflection spectra of particulate samples of different materials all of which display sharp absorption edge effect. (Siegal and Gillespie © 1980 John Wiley & Sons. Reprinted with permission.)

3-2-4 Fluorescence

As illustrated in the case of the interaction of light with ruby, energy can be absorbed at one wavelength and reemitted at a different wavelength due to the fact that the excited electrons will cascade down in steps to the ground state. This is called fluorescence. This effect can be used to acquire additional information about the composition of the surface. In the case of the sun illumination, it would seem at first glance that it is not possible to detect fluorescence because the emitted fluorescent light could not be separated from the reflected light at the fluorescence
wavelength. However, this can be circumvented due to the fact that the sun spectrum has a number of very narrow dark lines, called Fraunhofer lines, which are due to absorption in the solar atmosphere. These lines have widths ranging from 0.01 to 0.1 µm, and the central intensity of some of them is less than 10% of the surrounding continuum. The fluorescence remote sensing technique consists of measuring to what extent a Fraunhofer “well” is filled up relative to the continuum due to fluorescence energy resulting from excitation by shorter wavelengths (see Fig. 3-16). Thus, by comparing the depth of a Fraunhofer line relative to the immediate continuum in the reflected light and the direct incident light, surface fluorescence can be detected and measured.

Let $F_s$ be the ratio of solar illumination intensity at the center of a certain Fraunhofer line ($I_0$) to the intensity of the continuum illumination ($I_c$) immediately next to the line (Fig. 3-16)

\[
F_s = \frac{I_0}{I_c} \tag{3-12}
\]

and let $F_r$ be the ratio for the reflected light. If the reflecting material is not fluorescent, then

\[
F_r = \frac{RI_0}{RI_c} = \frac{I_0}{I_c} = F_s \tag{3-13}
\]

where $R$ is the surface reflectivity. If the surface is fluorescent with fluorescence emission near the Fraunhofer line being measured, then an additional intensity $I_f$ is added in the reflected light:

\[
F_r = \frac{RI_0 + I_f}{RI_c + I_f} > F_s \tag{3-14}
\]

and the intensity of the fluorescence illumination $I_f$ can be derived.
3-3 SIGNATURE OF SOLID SURFACE MATERIALS

Solid surface materials can be classified into two major categories: geologic materials and biologic materials. Geologic materials correspond to the rocks and soils. Biologic materials correspond to the vegetation cover (natural and human-grown). For the purpose of this text, snow cover and urban areas are included in the biologic category.

3-3-1 Signature of Geologic Materials

As discussed earlier, the signature of geologic materials in the visible and near infrared is mainly a result of electronic and vibrational transitions. The absorption bands of specific constituents are strongly affected by the crystalline structure surrounding them, their distribution in the host material, and the presence of other constituents. A spectral signature diagram for a variety of geologic materials is shown in Figure 3-17 and is based on the work by Hunt (1977). In the case of vibrational processes, the water molecule (H$_2$O) and hydroxyl (–OH) group play a major role in characterizing the spectral signature of a large number of geologic materials. In the case of electronic processes, a major role is played by the ions of transition metals (e.g., Fe, Ni, Cr, Co), which are of economic importance. The sulfur molecule illustrates the conduction band effect on the spectral signature of a certain class of geologic materials.

Figure 3-17. Spectral signature diagram of a variety of geologic materials. (Based on Hunt, 1977.)
The wide variety of geologic material with their corresponding composition makes it fairly complicated to uniquely identify a certain material based on a few spectral measurements. The measurement of surface reflectivity at a few spectral bands would lead to ambiguities. The most ideal situation would be to acquire the spectral signature of each pixel element in an image all across the whole accessible spectrum from 0.35 to 3 µm and beyond in the infrared. This would allow unique identification of the constituents. However, it requires a tremendous amount of data handling capability, as discussed later. A more feasible approach would be to concentrate on diagnostic regions for specific materials. For instance, a detailed spectrum in the 2.0 to 2.4 µm region will allow the identification of the OH group minerals. Figure 3-18 illustrates the detailed structure in this spectral region for minerals typically associated with hydrothermal alterations.

3-3-2 Signature of Biologic Materials

The presence of chlorophyll in vegetation leads to strong absorption at wavelengths shorter than 0.7 µm. In the 0.7 to 1.3 µm region, the strong reflectance is due to the refractive index discontinuity between air and the leaf cell. In the region from 1.3 to 2.5 µm the spectral reflectance curve of a leaf is essentially the same as that of pure water. Figure 3-19 shows the spectral reflectance of corn, soybean, bare soil, and a variety of foliages.

![Figure 3-18. High-resolution laboratory spectra of common minerals typically associated with hydrothermal alteration. (From Goetz et al., 1983.)](image-url)
One of the major objectives in remotely sensing biologic materials is to study their dynamic behavior through a growing cycle and monitor their health. Thus, the variations in their spectral signature as a function of their health is of particular importance. As illustrated in Figure 3-20, leaf moisture content can be assessed by comparing the reflectances near 0.8, 1.6, and 2.2 µm. Even though there are diagnostic water bands at 1.4 and 1.9 µm, it should be kept in mind that these are also regions of high atmospheric absorption due to atmospheric water vapor.

The amount of green biomass also affects the reflectance signature of biologic materials, as illustrated in Figure 3-21 for alfalfa through its growth cycle. The bare field signature is presented by the zero biomass curve. As the vegetation grows, the spectral signature becomes dominated by the vegetation signature. In principle, the biomass can be measured by comparing the reflectance in the 0.8–1.1 µm region to the reflectance near 0.4 µm.

Figure 3-19. Spectral reflectance of a variety of biological materials. (a) Reflectance of some cultivated vegetation compared to reflectance of bare soil and wet soil. (b) Reflectance of various types of foliage. (From Brooks, 1972.)

Figure 3-22 shows another example of the changes that occur in the spectral signature of a beech leaf through its growing cycle, which in turn reflect changes in chlorophyll concentration. Both the position and slope of the rise (called red edge) near 0.7 µm change as the leaf goes from active photosynthesis to total senescence.

High spectral resolution analysis of the red edge is allowing the detection of geochemical stress resulting from alteration in the availability of soil nutrients. A number of researchers have noted a blue shift, consisting of about 0.01 µm of the red edge or chlorophyll shoulder to slightly shorter wavelengths in plants influenced by geochemical stress (Figs. 3-23 and 3-24). This shift, due to mineral-induced stress, may be related to subtle changes in the cellular environment.
In many situations, geologic surfaces are partially or fully covered by vegetation. Thus, natural spectral signatures will contain a mixture of features which characterize the cover and the underlying material (see Fig. 3-25). The relative contributions depend on the percent of the vegetation cover and the intensity of the indentifying feature (i.e., absorption band or step) being observed.

**Figure 3-20.** Progressive changes in the spectral response of a sycamore leaf with varying moisture content (Short, 1982).

**Figure 3-21.** Variations in spectral reflectance as functions of amounts of green biomass and percent canopy cover (Short, 1982).
Figure 3-22. Reflectance spectra for a healthy beech leaf (1) and beech leaves in progressive phases of senescence (2–4). (From Knipling, 1969.)

Figure 3-23. Blue shift in the spectrum of conifers induced by a sulfide zone. (From Collins et al., 1983. Reproduced from *Economic Geology*, Vol. 78, p. 728.)
The reflectivity of surfaces in the visible and near infrared is fully governed by the reflectivity of the top few microns. Weathered rocks often show a discrete iron-rich surface layer which could be compositionally different from the underlying rock, as in the case of desert varnish. Thus, it is important to determine the depth of penetration of the sensing radiation. This is usually done by conducting laboratory measurements. Buckingham and Sommer (1983) made a series of such measurements using progressively thicker samples. They found that as the sample thickness increases, absorption lines become more apparent (i.e., higher contrast). After a certain critical thickness, an increase in sample thickness does not affect the absorption intensity. This corresponds to the maximum thickness being probed by the radiation.

A typical curve showing the relationship between sample thickness and absorption intensity is shown in Figure 3-26. The penetration depth for 0.9 $\mu$m radiation in ferric materials is at most 30 $\mu$m, and this thickness decreases as the concentration of ferric material increases. In the case of the 2.2 $\mu$m radiation, a 50 $\mu$m penetration depth was measured for kaolinite.

The penetration effect can be easily quantified by including the complex wave vector $k$ in the field expression (Equation 2-10). When the medium is absorbing, the index of refraction is given by
where \( k_0 = \frac{2\pi}{\lambda} \). Thus the field expression becomes:

\[
E = A e^{i(N_i k_0 + iN_i k_0)} e^{-i\omega t} = E_0 e^{-N_i k_0 r} \tag{3–17}
\]

where \( E_0 \) is the field when there is no absorption. Equation 3-17 shows that the field decreases exponentially as its energy is absorbed by the medium. The “skin” depth or penetration depth \( d \) is defined as the thickness at which the field energy is reduced by a factor \( e^{-1} \). Thus

\[
d = \frac{1}{2 N_i k_0} = \frac{\lambda}{4\pi N_i} \tag{3–18}
\]
3-4 PASSIVE IMAGING SENSORS

A large number of visible and infrared imaging sensors have been flown in space to study the Earth and planetary surfaces. These include Gemini, Apollo, and Skylab cameras, the series of Landsat cameras, including Multispectral Scanner (MSS), Landsat Thematic Mapper (TM), and the Enhanced Landsat Thematic Mapper Plus (ETM+), the series of SPOT Satellite Imagers, and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER). A number of planetary imaging cameras have also been flown, including the Galileo Jupiter Multispectral Imager and the Mars Orbiter Camera.

3-4-1 Imaging Basics

Passive imaging systems collect information about the surface by studying the spectral characteristics of the electromagnetic energy reflected by the surface, as shown in Figure 3-1. As the source energy propagates through the atmosphere on its way to the surface, the spectral
characteristics of the atmosphere are imprinted on the source signal, as shown in Figure 3-2 for the case of the Earth’s atmosphere. This incoming energy is then reflected by the surface, and propagates through the atmosphere to the collecting aperture of the imaging sensor.

Let the incoming radiant flux density be $F_{i}$. The spectrum of this incoming radiant flux at the surface is given by

$$S_{i}(\lambda) = S(\lambda, T_{s}) \times \left(\frac{R_{s}}{d}\right)^{2} a(\lambda)$$

(3-22)

where $S(\lambda, T_{s})$ is the energy radiated from the source with temperature $T_{s}$, but modified by the absorption spectrum $a(\lambda)$ of the atmosphere. The second term represents the fact that the energy is radiated spherically from the sun surface with radius $R_{s}$ to the body at distance $d$.

The fraction of this incoming energy that is reflected by the surface is described by the surface reflectance $\rho(\lambda)$, also known as the surface albedo. The most complete description of the surface reflectance is given by the so-called Bidirectional Reflectance Distribution Function (BRDF), which gives the reflectance of the surface as a function of both the illumination and the viewing geometry. In the simpler case of a Lambertian surface, this function is a constant in all directions, which means that the reflected energy is spread uniformly over a hemisphere. If the surface area responsible for the reflection is $dS$, the radiant flux at the sensor aperture is

$$S_{r}(\lambda) = S_{i}(\lambda) \rho(\lambda) dS \frac{1}{2\pi r^{2}} a(\lambda) = S(\lambda, T_{s}) \left(\frac{R_{s}}{d}\right)^{2} a^{2}(\lambda) \frac{\rho(\lambda) dS}{2\pi r^{2}}$$

(3-23)

where $r$ is the distance between the aperture and the surface area reflecting the incoming energy. If the aperture size is denoted by $dA$, the power captured by the aperture, per unit wavelength, will be

$$P(\lambda) = S_{r}(\lambda) dA = S(\lambda, T_{s}) \left(\frac{R_{s}}{d}\right)^{2} a^{2}(\lambda) \frac{\rho(\lambda) dS dA}{2\pi r^{2}}$$

(3-24)

The aperture will typically collect the incoming radiation for a short time, known as the dwell time or aperture time. In addition, the sensor will typically collect radiation over a finite bandwidth with an efficiency described by the sensor transfer function $h(\lambda)$. Denoting the dwell time by $\tau$, we find the total energy collected by the sensor from the surface element to be

$$E_{r} = \int_{\lambda_{1}}^{\lambda_{2}} S(\lambda, T_{s}) \left(\frac{R_{s}}{d}\right)^{2} a^{2}(\lambda) \frac{\rho(\lambda) dS dA}{2\pi r^{2}} h(\lambda) \tau d\lambda$$

(3-25)

If the relative bandwidth is small, this expression can be approximated by

$$E_{r} \approx S(\lambda_{0}, T_{s}) \left(\frac{R_{s}}{d}\right)^{2} a^{2}(\lambda_{0}) \frac{\rho(\lambda_{0}) dS dA}{2\pi r^{2}} h(\lambda_{0}) \tau \Delta\lambda$$

(3-26)
where \( \lambda_0 \) is the wavelength at the center of the measurement bandwidth, and \( \Delta \lambda \) is the bandwidth. This received energy is compared with the intrinsic noise of the sensor system to determine the image signal-to-noise ratio.

### 3-4-2 Sensor Elements

The main elements of an imaging system are sketched in Fig. 3-27. The collecting aperture size defines the maximum wave power that is available to the sensor. The collector could be a lens or a reflecting surface such as a plane or curved mirror.

The focusing optics focuses the collected optical energy onto the detecting element or elements. The focusing optics usually consists of numerous optical elements (lenses and/or reflectors) for focusing, shaping, and correcting the wave beam.

A scanning element is used in some imaging systems to allow a wide coverage in the case where few detecting elements are used. When a large array of detecting elements or film is used, the scanning element is usually not necessary.

![Figure 3-27](image.png)

**Figure 3-27.** Sketch of major elements of an imaging sensor. The elements are not to scale and their order could be different depending on the exact system configuration.

In order to acquire imagery at different spectral bands, the incident wave is split into its different spectral components. This is achieved by the dispersive element, which could consist of a set of beamsplitters/dichroics, a set of filters on a spinning wheel, or dispersive optics such as a prism or grating (see Fig. 3-28).

The wave is finally focused on the detecting element where its energy is transformed into a chemical imprint in the case of films or is transformed into a modulated electrical current in the case of array detectors.

An imaging system is commonly characterized by its response to electromagnetic energy that originates from a fictitious “point source” located infinitely far away from the sensor system.
Electromagnetic waves from such a point source would reach the sensor system as plane waves. Because the sensor system consists of elements that are of finite size, the waves will be diffracted as it propagates through the sensor until it reaches the focal plane. Here, the point source no longer appears as a single point, but the energy will be spread over a finite patch in the focal plane. The descriptive term point spread function is commonly used to describe the response of an imaging system to a point source.

The exact shape of the point spread function depends on the design of the optical system of the sensor, and on the physical shapes of the apertures inside the optical system. For example, light from a point source that propagated through a circular aperture with uniform transmission would display an intensity pattern that is described by

$$ I(u,v) = I_0 \frac{\pi a^2}{\lambda^2} \left[ \frac{2J_1(ka\sqrt{u^2 + v^2})}{ka\sqrt{u^2 + v^2}} \right]^2 $$

where $a$ is the radius of the circular aperture, $\lambda$ is the wavelength of the incident radiation, $k = 2\pi/\lambda$ is the wavenumber, and $u$ and $v$ are two orthogonal angular coordinates in the focal plane. This diffraction pattern, shown in Figure 3-29, is commonly known as the Airy pattern, after Airy, who derived (3-27) in 1835. The first minimum in the Airy pattern occurs at the first root of the Bessel function $J_1$ at

$$ w = \sqrt{u^2 + v^2} = 0.610 \frac{\lambda}{a} = 1.22 \frac{\lambda}{D} $$

Figure 3-28. Multispectral wave dispersion techniques. (a) Beamsplitter used on the Landsat TM. (b) Dispersive optics. A third technique is the use of a spinning wheel with a bank of filters.
where \( D = 2a \) is the diameter of the aperture. The area inside the first minimum of the Airy function is commonly referred to as the size of the Airy disk. The light from the point source is not focused to a point; instead it is spread over the neighborhood of the point in the shape of the Airy function – a process known as diffraction. Looking again at (3-27) or (3-28), we observe that the size of the patch of light that represents the original point source in the instrument focal plane is a function of the electrical size of the aperture as measured in wavelengths. The larger the aperture, the smaller the Airy disk for a given wavelength. On the other hand, if the aperture size is fixed, the Airy disk will be larger at longer wavelengths than at shorter wavelengths. This is the case in any telescope system that operates over a number of wavelengths. The Airy disk at 0.7 \( \mu \text{m} \) would be nearly twice as wide as that at 0.4 \( \mu \text{m} \) for a fixed aperture size.

**Figure 3-29.** Diffraction pattern of a circular aperture with uniform illumination. The image on the left shows the logarithm of the intensity and is scaled to show the first three diffraction rings in addition to the central peak - a total dynamic range of 3 orders of magnitude. The diffraction pattern is also shown as a three-dimensional figure on the right, displayed using the same logarithmic scale.

**Figure 3-30.** These graphs show cuts through the composite diffraction patterns of two point sources. Three cases are shown: two point sources in the left graph are completely resolved, the two points in the middle graph are just resolved, exactly satisfying the Rayleigh criterion, and the two point sources in the right-hand graph are not resolved. Note that here the intensities are displayed using a linear scale, so the diffraction rings are only barely visible.
The resolution, or resolving power, of an imaging system describes its ability to separate the images of two closely spaced point sources. The most commonly used definition of resolution is that defined by Lord Rayleigh according to which two images are regarded as resolved when they are of equal magnitude and the principal maximum of the one image coincides with the first minimum of the second image. Therefore, (3-28) describes the smallest angular separation between two point sources of equal brightness if a telescope with a circular aperture with radius $a$ would be used to observe them. Figure 3-30 shows three cases where two point sources are completely resolved (left graph), the two points are just resolved (middle graph), exactly satisfying the Rayleigh criterion, and the two point sources are not resolved (right graph).

It should be mentioned that not all imaging systems employ circular apertures of the type that would result in a point source being displayed as an Airy function in the image plane of the sensor. Figure 3-31 shows the point spread function of a square aperture with equal intensity across the aperture, and also that of a circular aperture in which the intensity is tapered as a Gaussian function from the center with a 50% reduction at the edges of the circular aperture. Notice the reduction in the brightness of the sidelobes of the circular aperture with a Gaussian taper when compared to Figure 3-29. This technique, known as *apodization*, is commonly used to reduce the sidelobes of imaging.
systems, but comes at the expense of energy throughput coupled with a slight broadening of the central disk of the point spread function.

Figure 3-32. Imaging geometry showing the instantaneous field-of-view of a single detector.

The resolution definition described by (3-27) assumes that the diffraction pattern of the system is adequately sampled in the focal plane. In modern remote sensing systems, the focal plane of the sensor is commonly populated with electronic detectors. Due to the finite size of these detectors, energy from a small solid angle, corresponding to the instantaneous field-of-view (IFOV) of an individual detector, will be integrated and reported as a single value (Figure 3-32). Therefore, if the IFOV of the detector is larger than the size of the Airy disk, it is possible that two point sources can be further apart than the resolution limit of the telescope optics, but the light from these two sources can still end up on a single detector, and hence they will not be distinguishable. In this case, the resolution of the sensor is driven by the size of detector, and not by the resolving power of the optics. In any case, the intersection of the detector IFOV and the surface being imaged gives the size of the surface element that is imaged by that detector. This area is commonly known as the size of a “pixel” on the ground, referring to the fact that this area will be reported as one picture element in the final image. This area represents the reflecting surface area \(dS\) in (3-23) – (3-26).

3-4-3 Detectors

The detector transforms the incoming wave into a form of recordable information. Optical films are one type of detector. Another type is the electro-optic detector, which transforms the wave energy into electrical energy that is usually transmitted to a digital recording medium. Electro-optic detectors are generally classified on the basis of the physical processes by which the conversion
from radiation input to electrical output is made. The two most common ones are thermal detectors and quantum detectors.

Thermal detectors rely on the increase of temperature in heat sensitive material due to absorption of the incident radiation. The change in temperature leads to change in resistance (bolometers usually using Wheatstone bridge) or voltage (thermocouplers usually using thermoelectric junctions), which can be measured. Typically, thermistor bolometers use carbon or germanium resistors with resistance change of about 4% per degree. Thermal detectors are usually slow, have low sensitivity, and their response is independent of wavelength. They are not commonly used in modern visible and near-infrared remote sensors.

Quantum detectors use the direct interaction of the incident photon with the detector material, which produces free charge carriers. They usually have high sensitivity and fast response, but they have limited spectral region of response (see Fig. 3-33). The detectivity of a quantum detector is defined as

\[ D = \frac{1}{\text{NEP}} \]  

(3-29)

where \( \text{NEP} \) is the noise equivalent power, which is defined as the incident power on the detector that would generate a detector output equal to the r.m.s. noise output. In other words, the noise equivalent power is that incident power that would generate a signal-to-noise ratio that is equal to 1.

For many detectors, both the detectivity and the \( \text{NEP} \) are functions of the detector area and the signal bandwidth. To take this into account, the performance of quantum detectors are usually characterized by the normalized detectivity, \( D^* \), which is equal to:

\[ D^* = D \sqrt{A \Delta f} = \frac{\sqrt{A \Delta f}}{\text{NEP}} = \frac{(S/N) \sqrt{A \Delta f}}{W} \]  

(3–30)

where \( A \) is the detector area, \( \Delta f \) is the circuit bandwidth, \( S/N \) is the signal-to-noise ratio, and \( W \) is the radiation incident on the detector.

Quantum detectors are typically classified into three major categories: photoemissive, photoconductive, and photovoltaic.

In photoemissive detectors, the incident radiation leads to electron emission from the photosensitive intercepting surface. The emitted electrons are accelerated and amplified. The resulting anode current is directly proportional to the incident photon flux. These detectors are usually operable at wavelengths shorter than 1.2 \( \mu \text{m} \) because the incident photon must have sufficient energy to overcome the binding energy of the electron in the atom of the photosensitive surface. This is expressed by the Einstein photoelectric equation:

\[ E = \frac{1}{2} m \nu^2 = h\nu - \phi \]  

(3–31)

where \( \phi \) = work function = energy to liberate an electron from the surface, \( m \) is the electron mass, and \( \nu \) is the velocity of the ejected electron. Thus the photon energy \( h\nu \) must be greater than \( \phi \) in order to liberate an electron. The critical wavelength of the incident wave is that wavelength for which the photon energy is equal to the work function of the detector material, and is given by:

32
The lowest \( \phi \) for photoemissive surfaces is for alkali metals. Cesium has the lowest \( \phi = 1.9 \text{ eV} \), which gives \( \lambda_c = 0.64 \text{ \( \mu \)m} \). Lower values of \( \phi \) can be achieved with composite surfaces. For instance, for a silver–oxygen–cesium composite, \( \phi_c = 0.98 \text{ eV} \) and \( \lambda_c = 1.25 \text{ \( \mu \)m} \).

In a photoconductive detector, incident photons with energy greater than the energy gap in the semiconducting material produce free-charge carriers, which cause the resistance of the photosensitive material to vary in inverse proportion to the number of incident photons. This requires substantially less energy than electron emission, and, consequently, such detectors can operate at long wavelengths in the thermal infrared. The energy gap for silicon, for example, is 1.12 eV, allowing operation of silicon detectors to about 1.1 \( \mu \)m. Indium antimonide has an energy gap of 0.23 eV, giving a cut-off wavelength of 5.9 \( \mu \)m.

In the case of photovoltaic detectors, the light is incident on a p-n junction modifying its electrical properties, such as the backward bias current.

Quantum detectors can also be built in arrays. This feature allows the acquisition of imaging data without the need of scanning mechanisms which are inefficient and provide short integration time per pixel. For this reason, most of the advanced imaging sensors under development will use detector arrays such as charge coupled devices (CCD) arrays (Fig. 3-34).

![Figure 3-33](image_url)  

**Figure 3-33.** Comparison of the \( D^* \) of various infrared detectors when operated at the indicated temperature (Hudson, 1969).
Silicon CCD detectors are among the most commonly used detectors in the visible and NIR part of the spectrum. These detectors typically have excellent sensitivity in the wavelength region 0.4 – 1.1 microns. Fundamentally, a CCD array is made up of a one- or two-dimensional array of Metal Oxide Silicon (MOS) capacitors that collect the charge generated by the free-charge carriers. Each capacitor accumulates the charge created by the incident radiation from a small area, known as a pixel, in the total array. To register the image acquired by the sensor, the charges accumulated on these capacitors must be read from the array. This is typically done using a series of registers into which the charges are transferred sequentially, and the contents of these registers are then sent to a common output structure where the charges are converted to a voltage. The voltage is then digitized and registered by the read-out electronics. The speed at which the charges can be read determines the frame rate at which a CCD detector can be used to acquire images. Various different read-out schemes are used to increase this frame rate. The fastest frame rates typically result when two CCD arrays are implemented side by side. One device is illuminated by the incoming radiation and is used to integrated charges, while the second device is covered with a metal layer to prevent light from entering the device. Once the charges are accumulated on the imaging CCD, the charges are quickly transferred to the covered device. These charges can then be read from the second device, while the first is again used to integrate charges from the scene being imaged.

Traditional CCD detectors have reduced sensitivity to the shorter wavelengths in the blue portion of the visible spectrum. The polysilicon gate electrodes, used to clock out the charge from the capacitors in the imaging part of the array, strongly absorb blue wavelength light. As pixel geometries get smaller, this problem is exacerbated. The blue response of the sensor decreases rapidly with pixel size. The short wavelength sensitivity can be improved using a technique known as thinning and back-illumination. Essentially, the CCD is mounted upside down on a substrate,
and charges are collected through the back surface of the device, away from the gate electrodes. On the longer wavelength side of the spectrum, the sensitivity of silicon CCDs can be improved by doping the silicon with impurities, such as arsenic. These devices, known as extrinsic CCDs, have sensitivities that extend well into the infrared part of the spectrum.

Recently, CMOS detectors have generated much interest. The basic difference between CMOS and CCD detectors is that in the case of the CMOS detectors, the charge to voltage conversion takes place in each pixel. This makes it easy to integrate most functions on one chip using standard CMOS processes. This integration into a single structure usually means that signal traces can be shorter, leading to shorter propagation delays and increased readout speed. CMOS arrays also support random access, making it easy to read only a portion of the array, or even one pixel. The drawback for CMOS detectors is that the charge-to-voltage conversion transistors must be placed in the pixel, and therefore takes up area that would otherwise be part of the optically sensitive part of the detector. This leads to a lower fill factor for CMOS devices as compared to CCDs, and generally lower sensitivity.

Indium Antimonide (InSb) detectors are commonly used to cover the near infrared part of the spectrum, as is the case for bands 5 and 7 of the Landsat ETM+ sensor.

Figure 3-35. Different types of imaging sensor implementations.

In the case of films, black and white, true color, and false color films can be used. The true color film is sensitive to the visible part of the spectrum. The false color film records a small part of the highly reflective portion of the infrared spectrum (0.75 – 0.9 μm) characteristic of vegetation. With this type of film, a yellow filter is used to remove the blue part so that the blue, green, and red
sensitive layers of emulsion in the film are available to record the green, red, and infrared parts of the radiation, respectively. Since healthy vegetation is strongly reflective in the infrared, it appears bright red in false color infrared images, while unhealthy vegetation appears blue to gray.

3-5 TYPES OF IMAGING SYSTEMS

Depending on how the sensor acquires and records the incoming signal, imaging systems can be divided into three general categories: framing cameras, scanning systems, and pushbroom imagers (see Fig. 3-35 and Table 3-3).

TABLE 3-3. Comparison of Different Imaging Systems

<table>
<thead>
<tr>
<th>Type</th>
<th>Advantage</th>
<th>Disadvantage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Film framing camera</td>
<td>Large image format</td>
<td>Transmission of film</td>
</tr>
<tr>
<td></td>
<td>High information density</td>
<td>Potential image smearing</td>
</tr>
<tr>
<td></td>
<td>Cartographic accuracy</td>
<td>Wide field of view optics</td>
</tr>
<tr>
<td>Electronic framing camera</td>
<td>Broad spectral range</td>
<td>Difficulty in getting large arrays or sensitive surface</td>
</tr>
<tr>
<td></td>
<td>Data in digital format</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Simultaneous sampling of image, good geometric fidelity</td>
<td></td>
</tr>
<tr>
<td>Scanning systems</td>
<td>Simple detector</td>
<td>Low detector dwell time</td>
</tr>
<tr>
<td></td>
<td>Narrow field-of-view optics</td>
<td>Moving parts</td>
</tr>
<tr>
<td></td>
<td>Wide sweep capability</td>
<td>Difficult to achieve good image, geometric fidelity</td>
</tr>
<tr>
<td></td>
<td>Easy to use with multiple wavelengths</td>
<td></td>
</tr>
<tr>
<td>Pushbroom imagers</td>
<td>Long dwell time for each detector</td>
<td>Wide field of view optics</td>
</tr>
<tr>
<td></td>
<td>Across track geometric fidelity</td>
<td></td>
</tr>
</tbody>
</table>

A framing camera takes a snapshot of an area of the surface, which is then projected by the camera optics on a film or a two-dimensional array of detectors located in the camera focal plane. As detector arrays with more pixels become available, framing cameras will become more common.

An example of a framing camera is the panchromatic camera used on the two Mars Exploration Rovers that landed on Mars in January 2004. Framing cameras have the major advantage that excellent geometric fidelity can be achieved because the entire image is acquired at once. This comes at the price that the optics system must typically have excellent response over a wide field of view.

Scanning systems use a scanning mirror that projects the image of one surface resolution element on a single detector. To make an image, across-track scanning is used to cover the imaged swath across the track. In some cases a limited number of detectors are used so that each scan covers a set of across-track lines instead of a single one. In this case, the imaging system is known as a whiskbroom scanner. The platform motion carries the imaged swath along the track. The major
disadvantage of such a system is the presence of moving parts and the low detection or dwell time for each pixel. In addition, images acquired with scanning systems typically have poorer geometric fidelity than those acquired with framing cameras. Examples of scanning systems are the Landsat instruments such as the Multispectral Scanner (MSS) and Thematic Mapper (TM) and the Enhanced Thematic Mapper Plus (ETM+).

**Figure 3-36.** Conceptual sketch of an imaging spectrometer. A narrow strip AB is imaged at a specific instant in time. The light from this strip is dispersed by the dispersive element such that each line on the array of detectors will correspond to the image of the strip in a very narrow spectral band.

Pushbroom imagers delete the scanning mechanism and use a linear array of detectors to cover all the pixels in the across-track dimension at the same time. This allows a much longer detector dwell time on each surface pixel, thus allowing much higher sensitivity and a narrower bandwidth of observation. Examples of such systems are the SPOT and the ASTER cameras. A pushbroom system can be thought of as a framing camera with an image frame that is long in the across-track direction, and much narrower in the along-track direction. Pushbroom sensors do not require a moving scan mirror in order to acquire an image. As a result, these sensors can be expected to exhibit longer operating life than a scanner. In addition, the fixed geometry afforded by the detector arrays results in high geometric accuracies in the line direction, which will simplify the image reconstruction and processing tasks.

The imaging spectrometer goes one step further. It utilizes a spectrometer section to separate the spectral channels and an area array detector to acquire images simultaneously in a large number of spectral channels. A narrow strip of the surface, one resolution element wide and a swath width long, is imaged through a slit followed by a dispersive element that disperses the energy in the line image into a series of line images of different spectral bands. This dispersed line image is projected onto a two-dimensional detector array. Each array line will detect the image of one spectral image line. Figures 3-36 and 3-37 illustrate one possible concept and design of an imaging spectrometer.

As the remote sensing instrument is carried along the satellite orbit, it typically images a strip on the surface along the orbit track. The width of this strip is known as the *swath* imaged by the camera. In the case of a framing camera, the swath may simply be the width of each image frame in the direction orthogonal to the satellite movement. Successive frames are combined to
form an image of which the length is limited by on-board storage or data downlink capability. In the case of a scanning system, the width of the swath is determined by the angle over which the scanning is done. The width of the swath and the altitude of the camera above the surface determine the field of view (FOV) of the imaging instrument, usually expressed as the total angular extent over which an image is acquired. Note that this definition of the FOV means that the instrument FOV may be different than the FOV of the optics of the telescope. This will be the case when either the instrument uses a scanning technique for imaging, or when the detector array in the focal plane only covers part of the telescope field of view.

![Figure 3-37](image_url)

**Figure 3-37.** One possible design for the optical system of the imaging spectrometer.

### 3-6 DESCRIPTION OF SOME VISIBLE/INFRARED IMAGING SENSORS

The complexity of an imaging sensor is usually directly related to the number of spectral channels, the number of detector elements, and the imaging resolution. Figure 3-38 gives a comparison of some of the most advanced imaging systems.

To illustrate the complexity associated with some of the advanced sensors, let us consider the data rate associated with a 100-channel imager with a 25-m resolution and a 100-km image swath. As a satellite in earth orbit moves at about 7.5 km/sec, the number of imaged pixels per second is given by:

\[
n = \frac{100,000 \times 7500}{25} = 1.2 \times 10^6
\]

and the total number of pixels is:

\[
N = 100 \text{ channels} \times n = 1.2 \times 10^8
\]

Assuming that each pixel brightness is digitized to 8 bits, the total bit rate is then \(8 \times 1.2 \times 10^8\), which is approximately 1 gigabit per second. This is well beyond the capability of present data transmission systems. In fact, it is often the case that the capability of spaceborne imaging systems is limited more by the data downlink transmission rate than by the instrument technology. This
problem is particularly bad in the case of deep space missions. Thus, some intelligent data reduction system is required.

**Figure 3-38.** Comparison of different imaging systems.

### 3-6-1 Landsat Enhanced Thematic Mapper Plus (ETM+)

The Landsat Enhanced Thematic Mapper Plus is a multispectral imaging system of the scanning type. It is flown on the Landsat-7 (1999 launch) spacecraft, and its characteristics are summarized in Table 3-4. The sensor is a derivative of the Thematic Mapper (TM) instruments flown on earlier Landsat satellites. The primary changes of the ETM+ over the TM's are the addition of a panchromatic band and improved spatial resolution for the thermal band.

The Landsat-7 spacecraft was put into a 705 km altitude orbit with an inclination of 98.2 degrees. A 185-km swath is imaged (Fig. 3-39). This allows complete coverage utilizing the 233 orbits in 16 days, which is the repeat cycle of the orbit. The satellite orbit is sun synchronous at approximately 10:00 a.m. local time. The data is transmitted to the ground as two streams of 150 Mbps each for a total rate of 300 Mbps using an X-band communications link. The primary receiving station is the US Geological Survey's (USGS) EROS Data Center (EDC) in Sioux Falls, South Dakota. Images can be acquired by EDC through real-time downlink, or playback from an on-board, 380 gigabit (100 scenes) solid state recorder. Since data is split in two streams, one can contain real-time data, while the other is playing back data from the recorder, or both streams can be used to play back data from the recorder simultaneously. Images can also be received by a worldwide network of receiving stations either in real-time, or direct downlink at X-band.
<table>
<thead>
<tr>
<th>Band Number</th>
<th>Spectral Range (µm)</th>
<th>Spatial Resolution (m)</th>
<th>Quantization Levels (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.45–0.52</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>0.53–0.61</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>0.63–0.69</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>0.78–0.9</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>1.55–1.75</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>10.4–12.5</td>
<td>60</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>2.09–2.35</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>0.52–0.9</td>
<td>15</td>
<td>8</td>
</tr>
</tbody>
</table>

**TABLE 3-5. Significant ETM+ Parameters**

<table>
<thead>
<tr>
<th>Orbit</th>
<th>Sun synchronous</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>705.3 km altitude</td>
</tr>
<tr>
<td></td>
<td>98.9 min period</td>
</tr>
<tr>
<td></td>
<td>98.2° inclination</td>
</tr>
<tr>
<td></td>
<td>16-day repeat cycle</td>
</tr>
<tr>
<td>Scan</td>
<td>185-km swath</td>
</tr>
<tr>
<td></td>
<td>7.0-Hz rate</td>
</tr>
<tr>
<td></td>
<td>85% efficiency</td>
</tr>
<tr>
<td>Optics</td>
<td>40.6-cm aperture</td>
</tr>
<tr>
<td></td>
<td>f/6 at prime focus</td>
</tr>
<tr>
<td></td>
<td>42.5 µrad, IFOV, bands 1–4</td>
</tr>
<tr>
<td></td>
<td>f/3 at relay focus</td>
</tr>
<tr>
<td></td>
<td>43.8 µrad IFOV, bands 5, 7</td>
</tr>
<tr>
<td></td>
<td>170 µrad IFOV, band 6</td>
</tr>
<tr>
<td>Signal</td>
<td>52 kHz, 3 dB, bands 1–5, 7</td>
</tr>
<tr>
<td></td>
<td>13 kHz, 3 dB, Band 6</td>
</tr>
<tr>
<td></td>
<td>1 sample/IFOV</td>
</tr>
<tr>
<td></td>
<td>8 bits/sample</td>
</tr>
<tr>
<td></td>
<td>84.9 Mbps multiplexed output</td>
</tr>
</tbody>
</table>

The ETM+ optics are similar to that of the older Thematic Mapper, shown in Figure 3-40, and its key parameters are given in Table 3-5. The bi-directional scan mirror moves the view of the telescope back and forth across the ground track (Fig. 3-35). The ground track is subdivided into 16 raster lines (32 for the panchromatic channel), which correspond to an array of 16 along-track detectors per spectral channel, and 32 for the panchromatic channel. There are six arrays of 16 detectors, each with an optical filter to define the corresponding spectral band for the visible and near IR. The thermal IR channel has a eight-element array. Thus, for each scan, 16 across-track
lines are mapped in the visible and near IR, 32 are mapped in the panchromatic band, and eight lines are mapped in the thermal IR. This corresponds to a strip that has a width of

\[ 16 \times 30 = 480 \text{ m} \]

The satellite moves at a speed of about 7 km/sec. Thus we require about \( \frac{7000}{480} = 14.6 \) scans or a 7.3 Hz back and forth scans per second. The scan mirror assembly consists of a flat mirror supported by flex pivots on each side, a torquer, a scan angle monitor, 2 leaf spring bumpers and scan mirror electronics. The motion of the mirror in each direction is stopped by the bumper, and is boosted by precision torque pulses during the turnaround period. The scan mirror is constructed of beryllium with an egg crate internal structure for maximum stiffness and minimum inertia. It has an elliptical shape and a size of 52 cm \( \times \) 41 cm.

![Diagram](image)

**Figure 3-39.** Landsat-D mapping geometry. (From Blanchard and Weinstein, © 1980 IEEE.)

The telescope is a Ritchey–Chretién design with primary and secondary mirror surfaces of hyperbolic figure to provide for simultaneous correction of spherical aberration and coma. An \( f/6 \) design allows a reasonable size detector with a 40.64 cm aperture, the focal length is 2.438 m. The mirrors are made from ultra-low expansion (ULE) glass, and have enhanced silver coatings. A detector of 0.1 mm across gives an IFOV (instantaneous field of view) of 42.5 \( \mu \text{rad} \), which corresponds to a 30 m spot from 705 km altitude.

A scan line corrector preceding the detectors compensates for the tilt of the array swath due to the spacecraft motion, so the scan lines will be straight and perpendicular to the ground track.
The prime focal plane is located at the focus of the primary telescope and contains the filters and silicon detectors of the first four spectral bands. Two mirrors, a folding mirror and a spherical mirror, are used to also relay the image from the primary focal plane to a second focal plane where the remaining three filters and detectors are located. These detectors require cooling to achieve good performance. Bands 5 and 7 use indium antimonide (InSb) detectors, and band 6 uses a mercury cadmium telluride (HgCdTe) detector. These detectors are kept at temperatures around 91°K using radiative cooling.

Figure 3-40. Thematic Mapper Optical System. (From Blanchard and Weinstein, © 1980 IEEE.)

3-6-2 Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)

ASTER, an advanced multispectral imager operating in the pushbroom configuration, is one of five instruments that were launched on board NASA’s Terra spacecraft in December 1999. ASTER is a cooperative effort between NASA and Japan's Ministry of Economy Trade and Industry (METI), formerly known as Ministry of International Trade and Industry (MITI).

The Terra satellite is in a sun-synchronous orbit at an altitude of 705 km at the equator, with an descending equator crossing time of 10:30 a.m. local time. Terra flies in a loose formation with the Landsat 7 satellite, crossing the equator approximately 30 minutes after the Landsat satellite. The orbit repeat period is 16 days.

The complete ASTER system actually consists of three telescopes covering a wide spectral region with 14 bands from the visible to the thermal infrared. The visible and near-infrared system has three bands with a spatial resolution of 15 m, and an additional backward-looking telescope that
is used for stereo imaging. Each ASTER scene covers an area of 60 x 60 km. The visible and near-infrared telescopes can be rotated to 24° on either side of the nadir line, providing extensive cross-track pointing capability. Table 3.6 shows the spectral passbands of the ASTER system. Here we shall describe the visible and near-infrared system; the others will be discussed in more detail in subsequent Chapters.

**TABLE 3-6. ASTER Characteristics**

<table>
<thead>
<tr>
<th>Subsystem</th>
<th>Band No.</th>
<th>Spectral Range (µm)</th>
<th>Spatial Resolution (m)</th>
<th>Quantization Levels (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VNIR</td>
<td>1</td>
<td>0.52-0.60</td>
<td>15</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.63-0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3N</td>
<td>0.78-0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3B</td>
<td>0.78-0.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWIR</td>
<td>4</td>
<td>1.60-1.70</td>
<td>30</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.145-2.185</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>2.185-2.225</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>2.235-2.285</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>2.295-2.365</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>2.360-2.430</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TIR</td>
<td>10</td>
<td>8.125-8.475</td>
<td>90</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>8.475-8.825</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>8.925-9.275</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>10.25-10.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>10.95-11.65</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The ASTER visible and near-infrared subsystem consists of two independent telescope assemblies, used for stereo imaging, to minimize image distortion in the backward and nadir looking telescopes. The detectors for each of the bands consist of 5000 element silicon charge-coupled detectors (CCDs). Only 4000 of these detectors are used at any one time for the total image swath width of 60 km. The reason for using only 80% of the available detectors has to do with the stereo imaging geometry. In this mode, the first image is acquired with the nadir-looking telescope, and the second with the back-looking telescope. This means that a time lag occurs between the acquisition of the backward image and the nadir image. During this time, earth rotation displaces the relative image centers. Based on orbit position information supplied by the Terra platform, the ASTER visible and near-infrared subsystem automatically extracts the appropriate 4000 pixels from each image to ensure that the same ground area is covered in each stereo image.

The ASTER visible and near-infrared optical system is a reflecting-refracting improved Schmidt design. The backward looking telescope focal plane contains only a single detector array and uses an interference filter for wavelength discrimination. The focal plane of the nadir telescope contains 3 line arrays and uses a dichroic prism and interference filters for spectral separation allowing all three bands to view the same area simultaneously. The telescope and detectors are maintained at 296 ± 3 K using thermal control and cooling from a cold plate. On-board calibration of the two telescopes is accomplished using a halogen lamp as a radiation source. These measures ensure that the absolute radiometric accuracy is ± 4% or better.
Figure 3-41. The picture shown here was taken by the Mars Orbiter Camera narrow angle (high resolution) camera and "colorized" by applying the colors of Mars obtained by the MOC wide angle cameras. The picture shows gullies in a crater at 42.4°S, 158.2°W, which exhibits patches of wintertime frost on the crater wall, and dark-toned sand dunes on the floor.

The visible and near-infrared subsystem produces the highest data rate of the three ASTER imaging subsystems. With all four bands operating (3 nadir and 1 backward) the data rate including image data, supplemental information and subsystem engineering data is 62 Mbps.

3-6-3 Mars Orbiter Camera (MOC)

The Mars Orbiter Camera was launched on the Mars Global Surveyor (MGS) spacecraft in November 1996, and arrived at Mars after a 300-day journey in September 1997. The initial orbit insertion of the MGS spacecraft around Mars left the spacecraft (by design) in a highly elliptical orbit with an orbital period of 44.993 hours and altitudes at periapsis and apoapsis of 262.9 km and 54,025.9 km, respectively. The periapsis altitude, however, was low enough to put the spacecraft
well within the Martian atmosphere. To place the MGS spacecraft in the appropriate nearly circular and nearly polar orbit for science operations, a technique known as aerobraking was used. Aerobraking essentially uses the drag in the Martian atmosphere to slow the spacecraft down near periapsis, which in turn lowers the apoapsis, slowly circularizing the orbit. The final mapping orbit with altitudes at periapsis and apoapsis of approximately 370 km and 435 km, respectively, and a period of 117 minutes was reached in March 1999.

The MOC is a pushbroom system that incorporates both wide angle (140 degree) and narrow angle (0.4 degree) optics for producing global coverage (7.5 km/pixel), selective moderate resolution images (280 m/pixel) and very selective high resolution (1.4 m/pixel) images. The narrow-angle camera optics is a 35 cm aperture, f/10 Ritchie-Cretién telescope, with a 2048 element CCD detector array (13 micron detector size) operating with a passband of 0.5 to 0.9 microns with a maximum resolution of 1.4 meters per pixel on the surface. At a spacecraft ground track velocity of ~3 km/s the narrow-angle camera exposure time is approximately 0.44 milliseconds. A 12 MB buffer is used to store images between acquisition and transmission to earth. This camera system has returned spectacular images of the Martian surface at data rates that vary between 700 bps to 29260 (real-time) bps. The camera system has a mass of 23.6 kg and consumes 6.7 W of power in the standby mode, and 18.7 W when acquiring narrow-angle data. Figure 3.41 shows a narrow-angle MOC image of gullies and sand dunes in a crater on Mars.

3-6-4 Mars Exploration Rover Panchromatic Camera (Pancam)

The stereo panchromatic cameras used on the two Mars Exploration Rovers, Spirit and Opportunity, are examples of framing cameras. The two rovers arrived at Mars in early 2004 at Gusev Crater and Meridiani Planum, respectively, to answer the fundamental question of whether there ever was water present on the surface of Mars.

The Pancam systems are two cameras that combine to form a stereo system. The camera optics for each “eye” consists of identical 3-element symmetrical lenses with an effective focal length of 38 mm and a focal ratio of f/20, yielding an IFOV of 0.28 mrad/pixel and a square FOV of 16.8°×16.8° per eye. The optics and filters are protected from direct exposure to the Martian environment by a sapphire window. A filter wheel is used to capture multi-spectral images. Each filter wheel has eight positions with narrowband interference filters covering the 400-1100 nm wavelength region. Two filter passbands are common between the left- and right cameras, and the left camera has one clear filter. The remaining 13 filter positions (7 on the right camera, and 6 on the left) have different center wavelengths, allowing spectral measurements at a total of 15 different wavelengths between the two cameras.

The images are captured using a 1024 x 2048 pixel CCD array detector for each “eye” of the stereoscopic system. The arrays are operated in frame transfer mode, with one 1024 x 1024-pixel region constituting the active imaging area and the adjacent 1024 x 1024 region serving as a frame transfer buffer. The individual detectors are 12 µm in both directions. The arrays are capable of exposure times from 0 msec (to characterize the “readout smear” signal acquired during the ~5 msec required to transfer the image to the frame transfer buffer) to 30 sec. Analog to digital converters provide a digital output with 12-bit encoding, and SNR > 200 at all signal levels above 20% of full scale.
Radiometric calibration of both Pancam cameras is performed using a combination of preflight calibration data and inflight images of a Pancam calibration target carried by each rover. The Pancam calibration target is placed within unobstructed view of both camera heads and is illuminated by the Sun between 10:00 AM and 2:00 PM local solar time for nominal rover orientations. The calibration target has three gray regions of variable reflectivity (20%, 40%, and 60%) and four colored regions with peak reflectance in the blue, green, red, and near-IR for color calibration.

Figure 3-42 shows a panchromatic image of the site (later named Eagle Crater) where the second Mars Exploration Rover Opportunity landed on January 23, 2004. The layered rocks visible in this panoramic image measure only 10 centimeters tall. Further detailed investigation of these rocks by Opportunity’s suite of scientific instruments located on its robotic arm showed fine layers that are truncated, discordant and at angles to each other, indicating that the sediments that formed the rocks were laid down in flowing water.

**Figure 3-42.** Data from the Mars Exploration Rover Opportunity's panoramic camera's near-infrared, blue and green filters were combined to create this approximate, true-color image of the rock outcrop near the rover's landing site.

### 3-7 ACTIVE SENSORS

With the advances of laser power and efficiency, laser sources could be used to illuminate the surface and remotely sense its properties. Laser sources have two characteristics which give them unique aspects: (1) the transmitted energy can be pulsed, and (2) the transmitted energy has a narrow well-known spectral bandwidth.

A pulsed laser sensor can be used to measure surface topography from orbital altitude. An example of such a system is the Mars Orbiter Laser Altimeter (MOLA) instrument, launched on the Mars Global Surveyor satellite in November 1996. The transmitter is a Q-switched Nd:YAG laser operating at a wavelength of 1064 nm. Pulses with energy of 48 mJ/pulse are transmitted at a 10 Hz rate, and illuminates a spot of approximately 130 m on the surface of Mars. The receiver has a 50 cm mirror and a silicon-avalanche photodiode is used as the detector. The system has a range resolution of 37.5 cm, and a vertical accuracy (shot-to-shot) of 37.5 cm. The absolute vertical accuracy is better than 10 m, limited primarily by the accuracy with which the Mars Global Surveyor orbit can be reconstructed. Elevation measurements are made at intervals of 330 m along the spacecraft track. The instrument mass is 25.85 kg, the power consumption is 34.2 W, and it produces a continuous data rate of 618 bps.

The MOLA instrument only measures profiles of elevation directly underneath the Mars Global surveyor satellite. An extension of this system would scan the laser beam across track to acquire surface topography over a wide swath (see Fig. 3-43). This technique is now used routinely from aircraft by several commercial firms worldwide to perform high accuracy three-dimensional
mapping. Depending on the laser altimeter used and the altitude of operation of the aircraft, ground spot sizes as small as of 20 cm, with elevation accuracies of a few centimeters can be achieved. Typical swath widths range from as small as 50 m to about 10 km, depending on the spot sizes used in the mapping. If the laser altimeter is able to measure the intensity of the returning laser pulse in addition to its round-trip time-of-flight, an image of the surface reflectance at the laser wavelength may be generated. In addition, the narrow spectral spread of a laser source allows the use of the fluorescence technique to identify certain surface materials.

![Figure 3-43. Sketch illustrating the principle of a scanning laser altimeter.](image)

### 3.8 SURFACE SENSING AT VERY SHORT WAVELENGTHS

All radiation at wavelengths shorter than 0.35 \( \mu m \) is strongly absorbed in planetary atmospheres. Thus, this spectral region can be used for surface sensing only in the case of planets without atmosphere (Moon, Mercury, asteroids) or, in the case of Earth, from low flying aircraft.

One of the most useful sensing techniques is \( \gamma \)-ray spectroscopy, which provides a spectral measurement of the particles emitted by a number of radioactive geologic materials. The spectral or energy unit commonly used in this region is the electron volt \( (1 \text{ eV} = 1.610^{-19} \text{ J}) \). Spectral lines of \( \gamma \)-ray emission are usually at many MeV, which correspond to wavelengths of a hundred to a thousand angstroms.

#### 3-8-1 Radiation Sources

In the case of the Earth, the main natural sources of \( \gamma \) radiation are uranium (U), thorium (Th), and potassium-40 \((^{40}\text{K})\).

The radioisotopes of uranium found in nature \((^{238}\text{U} \text{ which constitutes } 99.3\% \text{ of natural uranium and } ^{235}\text{U} \text{ which constitutes } 0.7\% \text{ of natural uranium})\) are long-lived \( \alpha \) emitters and parents of radioactive decay chains. Some of the daughter products with their corresponding energies are given in Table 3-7.
Other major sources of $\gamma$-rays in the Earth surface are (1) thorium ($^{232}$Th), which is a long-lived ($1.38 \times 10^{10}$ years half-life) $\alpha$ emitter and is parent of a radioactive decay chain which contains, among others, $^{228}$Ra, $^{216}$Po, $^{216}$Pb, and $^{208}$Tl, and (2) potassium-40 ($^{40}$K), which is widespread.

### 3-8-2 Detection

Gamma-rays are absorbed or scattered by matter with partial or total loss of energy by photoelectric effect, Compton effect, and pair production (see Fig. 3-44). The ejected electrons dissipate their energy by ionization of surrounding atoms. Two methods are generally used to measure the amount of ionization: (1) measurement of the light emitted, and (2) collection of the created charges.

If the detector absorbing material is optically transparent to the light released by ionization, then a burst of light will accompany each $\gamma$-ray interaction. This property is exhibited by a number of plastics, halides, organic liquids, and phosphors. Photodetectors are then used to detect the emitted light. One phosphor commonly used is the inorganic crystal of thallium-activated sodium iodide NaI (Tl). It has the important property of being able to absorb all of the incident $\gamma$-radiation, even the highest energy rays, due to its high density of $3.67 \text{ g/cm}^3$. The spectral resolution of this type of detector is limited by the number of light photons released by the ionizing electrons created by $\gamma$-ray interactions. The resolution is proportional to the square root of the number of photons released.

### TABLE 3-7. Daughter Products of $^{238}$U

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Half-life</th>
<th>Radiation Type</th>
<th>Energy (MeV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$^{238}$U</td>
<td>$4.5 \times 10^9 \text{ yr}$</td>
<td>$\alpha$</td>
<td>4.18, 4.13</td>
</tr>
<tr>
<td>$^{234}$Th</td>
<td>24.5 days</td>
<td>$\beta$</td>
<td>0.19, 0.1</td>
</tr>
<tr>
<td>$^{234}$Pa</td>
<td>1.1 min</td>
<td>$\gamma$</td>
<td>0.09, 0.06, 0.03</td>
</tr>
<tr>
<td>$^{234}$U</td>
<td>$2.5 \times 10^{15} \text{ yr}$</td>
<td>$\alpha$</td>
<td>4.77, 4.72</td>
</tr>
<tr>
<td>$^{230}$Th</td>
<td>$8.3 \times 10^4 \text{ yr}$</td>
<td>$\alpha$</td>
<td>4.68, 4.62</td>
</tr>
<tr>
<td>$^{226}$Ra</td>
<td>1620 yr</td>
<td>$\alpha$</td>
<td>4.78, 4.59</td>
</tr>
<tr>
<td>$^{214}$Bi</td>
<td>19.7 min</td>
<td>$\beta$</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\gamma$</td>
<td>3.18, and many others</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.43, 2.2, and many others</td>
</tr>
</tbody>
</table>

*Note: $\alpha$ (alpha) particles are positively charged, $\beta$ (beta) particles are negatively charged, $\gamma$ (gamma) particles have no charge.*
Semiconductor detectors have a higher spectral resolution capability. They measure the charge created by ionization within the detector. The resolution in this type of detector is mainly limited by the electronic noise and is somewhat independent of the photon energy.

3-9. IMAGE DATA ANALYSIS

Multispectral remote sensing imagers provide a series of spatial images recorded at a number of different spectral pass-bands. Each of these images consists of a two-dimensional array of picture elements, or pixels; the brightness of each pixel corresponds to the average surface reflectivity for the corresponding surface element. The images from such a multispectral imager can be thought of as a data cube in which horizontal slices represent spatial images for a particular pass-band, and vertical columns represent the spectral signature of a particular surface picture element.

The spectral signature is the most diagnostic tool in remotely identifying the composition of a surface unit. A trade-off generally exists between the spectral resolution, spectral coverage, radiometric accuracy, and identification accuracy. In the case of hyperspectral imagers, the spectral signature corresponds to high resolution (spectrally) radiometric measurements over a fairly broad region of the spectrum. In this case, surface units can be separated, classified, and identified based upon some unique characteristic in their reflectivity spectrum, such as a diagnostic absorption band or combination of absorption bands, a diagnostic reflectivity change at a certain wavelength, or ratio of reflectivities in two separate spectral regions. Because hyperspectral imagers record a large number of spectral measurements for every pixel in the image, these instruments produce large data...
volumes. As an example, a 128 channel image of 15 meter pixels and an image size of 1024 x 1024 pixels (covering an area of 15.36 km x 15.36 km), would require storage of 1 Gbits, assuming 8 bits per pixel are recorded. At a nominal speed of 7.5 km/sec, a satellite will have to record this data every approximately 2 seconds, for an approximate data rate of 500 Mbits/sec! For this reason, most visible and near-infrared imagers record significantly fewer channels than this. The challenge then is to extract the useful information from relatively few spectral measurements.

Figure 3-45. Individual spectral channel images for the 9 visible and near-infrared channels of the ASTER instrument. The channels progress from left to right and from top to bottom in increasing wavelength, with channel 1 in the upper left, and channel 9 in the lower right. The images cover an area of roughly 15 km x 15 km of an area near Cuprite, Nevada.

Various different analysis techniques are used to extract qualitative and quantitative information about the surface from the recorded images. The analysis of radiometric and spectral signatures in surface studies can generally be divided into three steps of increasing complexity: (1) detection and delineation, (2) classification, and (3) identification. We shall describe these in more detail in the next few sections using data acquired by the ASTER instrument (see Section 3-6-2 for a description) over the Cuprite Mining District located in Nevada. This area has been used
extensively as a testing and verification site for remote sensing instruments. Hydrothermal alteration of Cambrian sedimentary rocks and Cenozoic volcanic rocks by acid-sulfate solutions took place during the Miocene. This occurred at shallow depths, producing silicified rocks containing quartz and minor amounts of alunite and kaolinite, opalized rocks containing opal, alunite and kaolinite, and argillized rocks containing kaolinite and hematite. These units are readily mappable using remote sensing instruments, as we shall show later.

Figure 3-45 shows the 9 visible and near infrared channels of an ASTER image of the Cuprite area. The area shown is a subset of a larger image, and covers approximately 15 km x 15 km. The bright feature near the middle of the image is Stonewall Playa. The road that cuts through the image from the upper left to middle bottom is highway 95. Note how similar the individual images are. We shall now analyze this image in more detail in the next few sections, illustrating the most commonly used data analysis techniques.

3-4-1 Detection and Delineation

The first step in the analysis of surface polychromatic or multispectral images is to recognize and delineate areas with different reflectivity characteristics. This can be done manually or with computers by simply delineating areas with image brightness within a certain range of values. Conceptually, this means that the data space is divided into a number of sub-spaces, and all pixels in the image that have brightness values that fall in a particular subspace are assumed to have similar characteristics. In general, change in brightness is associated with changes in surface chemical composition, biological cover, or physical properties (roughness, slope, etc.). Unfortunately, change in brightness can also result from changes in the illumination geometry or atmospheric conditions. If this is not taken into consideration during the data analysis, areas may be misclassified. An obvious example would be where areas in the shadow of a cloud would be misclassified as low reflectivity areas. Slopes facing the sun will in general appear brighter than those facing away from the sun, even if the two areas have in fact identical characteristics.

The simplest form of image analysis is a simple photo-interpretation of an image. Usually color images are used, displaying three spectral images as the red, green and blue components of the color image, respectively. Surfaces with similar colors and brightness are delineated. The main difficulty is to decide which spectral images to use in the color composite, and which spectral band to assign to each color. Even in the case of the Landsat ETM+ images, where six channels are available (counting all the channels with 30 m resolution), this leads to 120 possible combinations. Experience has shown, however, that not all these combinations are that useful, and therefore, only a few combinations are commonly used. For example, geologic interpretation in semi-arid areas is typically done using bands 1-4-7 of the TM (or ETM+) instrument as blue, green and red displays. Sometimes TM bands 2-4-7 (ASTER bands 1-3-6) are used to reduce the effects of atmospheric scattering, which is more severe in the shorter wavelength TM band 1. Atmospheric scattering is particularly troublesome in humid, tropical areas, where the usefulness of the shorter wavelength bands is reduced dramatically. In these situations, the optimum combination for geologic interpretation is to display bands 4-5-7 as blue, green and red, respectively.

Figure 3-46 shows two ASTER color combinations of the Cuprite scene, bands 1-2-3 (the three visible channels displayed as blue, green and red) on the left, and bands 1-3-6 (roughly equivalent to TM bands 2-4-7) on the right. Because of the high correlation between the
reflectances in the individual bands, especially in the visible part of the spectrum, there is little color information in the image on the left. The image on the right shows more detail in the colors, and differentiates well between the opalized rocks with kaolinite and/or alunite (shown in cyan) and the unaltered volcanics, which are shown in the reddish colors.

**Figure 3-46.** Two color combination displays for the Cuprite scene shown in Figure 3-45. On the left are ASTER channels 1, 2, 3 displayed as blue, green and red, respectively, giving a pseudo natural color image. On the right are channels 1, 3 and 6 displayed as blue, green and red.

Several techniques are used to increase the separation of surfaces in correlated images. One such technique transforms the image from the red, green and blue (RGB) representation of color to the intensity, hue and saturation (IHS) representation. (Most image analysis software packages include this capability.) The color information is contained in the hue of the transform, and the purity of the color is contained in the saturation. After transformation, the saturation image layer is then stretched to increase the color separation. The modified image layers are then transformed back to the red, green, and blue representation for display. Figure 3-47 shows the same two images as Figure 3-46 after enhancing the colors using the IHS transform. Note the dramatic increase in color separation between the two sets of images. Note in particular the area to the left of the road that runs from the top left to the center bottom. This area, shown in gold tones on the left image, and light green tones in the right image, is not easily distinguished in either of the original images in Figure 3-46. We shall show below that this area contains higher concentrations of alunite and kaolinite.

In addition to increasing color separation, this color transformation is often used to enhance images in the following way. One would start with a medium resolution multispectral image and transform three image layers the RGB representation to the IHS representation. After stretching the hue and possibly the saturation channels, the intensity channel is replaced with a higher resolution image, such as the panchromatic band of a SPOT image that has been registered with the original data. The new IHS layers are then transformed back to the RGB representation. The resulting color image appears to have much higher resolution than the original one because of the enhanced
resolution of the intensity layer. Sometimes an image from a completely different type of sensor, such as an imaging radar, may be used as the new intensity layer. Since radar images typically provide excellent separation of surfaces with different geometrical properties such as roughness, this combination better delineates surfaces based on both spectral information (contained in the color of the image) and geometrical properties contained in the intensity of the image.

![Figure 3-47](image1.png) The same images shown in Figure 3-46 after performing a color stretch using the IHS transform. Note the dramatic increase in color separation evident in both images.

The RGB/IHS transformation uses three channels of data as input. As the number of data channels increase, the potential number of color combinations increase rapidly. As an example, a 6 channel system has 120 potential color combinations, while a 200 channel system would have 7880400 such three-band combinations! A more powerful method of data analysis, known as principal components analysis, offers one solution to the problem of which channels to use in the color display. Mathematically, principal component analysis simply transforms an n-dimensional dataset into its eigenvectors. If the original data are real numbers, all eigenvectors are orthogonal. The eigenvector image corresponding to the largest eigenvalue, usually called the first principal component (PC1), is the image with the largest variation in brightness, while the PCN image, corresponding to the smallest eigenvalue, contains the least variation. For most scenes, the first three principal components will account for more than 95% of the variation in the data. Figure 3-48 shows the 9 principal component images for the Cuprite scene. The first principal component image contains 91% of the variation in the data. The next three principal components contain another 8.6% of the variation, with the remaining 0.4% of the variation contained in the last five principal components. The last three principal component images are quite noisy. All the images have been stretched to show the same dynamic range for display.

In practice, it is found that the first eigenvector spectrum typically is closely related to the average scene radiance, which is the incoming solar radiance convolved with the average scene reflectance and the atmospheric attenuation. The PC1 image is typically dominated by topographic effects, strongly highlighting slopes and shadows as seen in the upper left image of Figure 3-48. For this reason, the PC1 layer is not commonly used in the further analysis of the data. It is more
common to use the next three principal component images, corresponding to PC2, PC3 and PC4 in a color analysis of the image. Figure 3-49 shows the color image displaying these three principal components as the blue, green and red channels, respectively. This image shows very little variation in brightness, showing that little information about the absolute albedo remains. Spectral differences, as manifested in the optimum stretch of the colors, are highlighted excellently. Notice how well the alteration zone to the left of the road is identified in this image as compared to the original visible color composite shown on the left in Figure 3-46.

Figure 3-48. Principal component images for the 9 visible and near-infrared channels of Cuprite scene. The principal components progress from left to right and from top to bottom, with PC1 in the upper left, and PC9 in the lower right.

In some cases, principal component analysis is performed separately on subsets of the total number of bands. For example, one might elect to perform the analysis separately on the visible channels of an ASTER image, and then again separately on the near infrared channels. In this way, it may be possible to extract optimally the difference in reflectance properties of each subset of the spectrum. One drawback of principal component-based display techniques is that spectral features
that occur in only a small number of pixels will be lost in the averaging process, and will therefore not be visible in the final product.

**Figure 3-49.** The principal components PC2, PC3, and PC4 are displayed as blue, green and red, respectively, in this color combination of the Cuprite scene. Notice the excellent color separation compared to Figures 3-46 and 3-47.

In the case of multispectral images, the delineation process should take albedo variation in any one of the spectral channels into consideration. In many situations, more accurate delineation of surface units can be achieved by using ratios of reflectivity in two different spectral bands. This would allow the minimization of non-relevant effects such as slope changes. To illustrate, let us assume that the reflectivity of a certain surface unit as a function of wavelength $\lambda$ and incidence angle $\theta$ is given by

$$ R(\lambda, \theta) = g(\lambda) f(\theta) $$

(3–32)

where $g(\lambda)$ denotes the “pure” spectral response of the surface, and $f(\theta)$ denotes the modification of the response by the imaging geometry. If we consider two neighboring areas $A$ and $B$ of identical composition but having different slope aspects, then their reflectivity will be different at each and every spectral band. However, if we consider the ratio $r$ at two separate bands,

$$ r = \frac{R(\lambda_1, \theta)}{R(\lambda_2, \theta)} = \frac{g(\lambda_1)}{g(\lambda_2)} $$

(3–33)
the effect of the slope change is eliminated and only the change in the composition is delineated.

Ratio images are also used to highlight specific spectral differences. For example, if a particular surface has a spectral signature that shows high reflectance at $\lambda_1$, and low reflectance at $\lambda_2$, the ratio image will enhance this difference and delineate these surfaces more clearly. Clay minerals, for example show relatively strong absorption in band 7 of TM images (due to the absorption by the hydroxyl group at 2.2 - 2.3 microns – see Figure 3-11), and little absorption in band 5. The ratio image of bands 5/7 is therefore commonly used to delineate clays that are typically associated with hydrothermal alteration. Similarly, TM band ratios 3/1 or 5/4 are commonly used to delineate the presence of iron oxide. Another example would be the strong reflectance shown by vegetation in the near infrared, compared to the low reflectance in the red part of the visible spectrum, which would lead to large values in a TM band 4/2.

![Figure 3-50](image.png)

**Figure 3-50.** Spectra of some minerals commonly associated with hydrothermal alteration, resampled to show the expected spectra for the ASTER visible and near-infrared passbands. Also shown are the ASTER passbands in the bottom of the figure.

Figure 3-50 shows the spectra of some minerals commonly associated with hydrothermal alteration, resampled to show the expected spectra for the ASTER visible and near-infrared passbands. Also shown are the ASTER passbands in the bottom of the figure. Most of the clays show strong absorption in band 6 at the fundamental wavelength of the Al-OH bending mode. These same minerals show little absorption in ASTER band 4, however. An ASTER band 4/7 ratio image would therefore be expected to highlight the presence of these minerals. The iron oxide minerals in Figure 3-50 show strong spectral ratios when comparing ASTER bands 4/3 or 3/1. Figure 3-51 shows a color image of these three ratios displayed as red (4/7), green (3/1) and blue...
The alteration zones, especially the one to the right of highway 95, are highlighted well in this image.

![Image](image_url)

**Figure 3-51.** Spectral ratio image of the Cuprite scene. The ratios are 4/7 (red), 3/1 (green) and 4/3 (blue).

Ratio images have the advantage that slope and illumination effects are reduced significantly. On the other hand, they have the disadvantage that noise and instrument artifacts in the images are typically amplified. In addition, those surfaces that have spectral features that are similar in the different bands, *i.e.* for which the spectral features are correlated, will show little variation in the ratio images.

### 3-4-2 Classification

The next step after delineation is to classify units based on a set of criteria. Classifications extend not only to individual images, but also to a number of images taken at different times of the same area or of different areas.

Classification of images involves using a set of rules to decide whether different pixels in an image have similar characteristics. These rules in effect divide the total data space into subsets separated by so-called decision boundaries. All pixels that fall within a volume surrounded by such decision boundaries are then labeled as belonging to a single class. The classification criteria range
from the most simple, such as all areas with identical reflectivity in a certain spectral band being put into the same class, to more sophisticated criteria, such as comparing the measured spectra over a large wavelength range. Some intermediate criteria include albedo (simple and composite), specific spectral absorption bands, spectral response slope in specific spectral regions, or the presence of specific spectral features.

Two major approaches are used in classifying images; supervised and unsupervised classifications. In the case of supervised classification, a user will specify so-called feature vectors to be used in the comparison process. These vectors can be thought of as defining the centroids of the decision volumes that are separated by the decision boundaries. These feature vectors can be extracted from the image to be classified, or could come from a library of spectral signatures either measured in the laboratory or in the field. In the case of unsupervised classification, the computer is allowed to find the feature vectors without help from an image analyst. In the simplest form, known as the K-means algorithm, K feature vectors are typically selected at random from the data space.

![Figure 3-52](image)

Figure 3-52. Results of an unsupervised classification of the Cuprite scene. The classification was initialized using randomly chosen features, and then iterated. The number of classes was arbitrarily set to 6.
Once the feature vectors are identified, classification rules are used to assign pixels in the image to one of the feature vectors. Many different classification rules are used ranging from the simple nearest neighbor distance classifier, to neural network schemes, to sophisticated schemes that take into account the expected statistical distributions of the data. To apply these rules during the classification process, a so-called *distance measure* is typically defined. The nearest neighbor scheme, for example, simply calculates the Euclidean distance between a pixel and each of the feature vectors as if each spectral measurement represents an orthogonal axis in the data space. In other cases, the distance definition includes some measure of the probability that a pixel may be similar to a particular feature vector. During the classification process, the distance between a pixel and each of the feature vectors is computed. The pixel is then labeled the same as the feature vector for which this distance is the smallest. If this smallest distance is larger than some threshold specified by the analyst, the pixel will not be classified.

![Figure 3-53](image)

*Figure 3-53.* Results of a supervised classification of the Cuprite scene. The classification was initialized using features selected from a combination of the principal component results shown in Figure 3-49, and the unsupervised classification results in Figure 3-52. The number of classes was limited to 4.
Figure 3-52 shows the results of an unsupervised classification of the Cuprite scene. The classification was arbitrarily asked to produce 6 classes, and the initial feature vectors were selected randomly from the data space. The classification was then performed iteratively, updating the feature centroids after each iteration, until fewer than 0.01% of the pixels were changing between iterations. We note that the two classes colored dark blue and green are found mostly in the altered zones, especially the one to the right of highway 95. All we know at this point is that we found 6 stable “classes” of terrain from the data itself. To attach any significance to the classes, we need to compare the spectra of the feature centroids to some library of spectra. These could be either laboratory spectra, or spectra measured in the field. Since it is unlikely that such large areas will be covered with a homogeneous layer of a single mineral type, field spectra may be more useful at this stage. Nevertheless, the spectra of the classes colored dark blue and Figure 3-52 show large band 4/5 ratios, consistent with the spectra of the alteration minerals shown in Figure 3-50.

The results of any classification depend strongly on the selection of the feature vectors. In selecting these in the case of supervised classification, the analyst may be guided by the results of previous analysis such as those described in the previous section. For example, feature vectors could be selected based on unusual spectral ratios observed in a ratio analysis. Or they may be selected based on specific colors found in a principal component analysis. Finally, the analyst may have some field experience with the scene, and may be picking feature vectors based on known characteristics of certain areas in the scene.

We used the principal component image in Figure 3-49 as a guide to select areas to use as the classification feature vectors in a supervised classification. We selected three areas from the image corresponding to the pink and dark blue tones in the alteration zones, and the green alluvial areas near the top of the image, which represent the unaltered terrain. We added to these areas a feature selected from one of the areas from the green class in the alteration zone to the right of highway 95 from the unsupervised classification. The classification result is shown in Figure 3-53. The alteration zones are clearly highlighted in this result.

In the classification schemes discussed above, it was assumed that each pixel would be assigned to only one feature vector class. In reality, pixels represent areas on the ground that are rarely homogenously covered with only one type of surface. Instead, these surfaces are typically a mixture of different surface covers, each with a different spectral response. The spectral signature measured by the remote sensing instrument is made up of the weighted sum of the individual surface element spectra, where the weights in the summation depend on the relative abundance of that type of cover in the area represented by the pixel. Sub-pixel classification schemes attempt to identify these relative abundances for each pixel. The fundamental assumption for the sub-pixel classification scheme is that the measured spectrum $S_{total}(\lambda)$ is a linear mixture of individual spectra:

$$S_{total}(\lambda) = \sum_{i=1}^{N} a_i S_{ei}(\lambda) + n$$  \hspace{1cm} (3-34)$$

where $a_i$ represents the relative fraction of the measured spectrum contributed by the endmember spectrum $S_{ei}(\lambda)$, and $n$ denotes additive noise. Under the assumption that we have identified an
exhaustive set of endmember spectra to choose from, and that all fractions must be positive, (3-34) is constrained by

$$\sum_{i=1}^{N} a_i = 1; \quad a_i \geq 0$$  \hspace{1cm} (3-35)

If we describe the measured and endmember spectra as $M$-dimensional vectors, where $M$ is the number of spectral channels, we can rewrite (3-34) as

$$\tilde{S} = E\tilde{A} + \tilde{N}$$  \hspace{1cm} (3-36)

Figure 3-54. This image shows the relative abundances of different materials after linear unmixing. The spectra used in the supervised classification were used as endmembers.

The endmember matrix is an $M \times N$ matrix, with each of the $N$ columns representing one endmember spectrum, and is the same for the entire image. Usually, the number of spectral channels is larger than the number of endmember spectra. In that case, the unmixing solution is found as
\[ \tilde{A} = \left( E^T E \right)^{-1} E^T \tilde{S} \]  

(3-37)

Linear unmixing for large dimensional datasets is computationally expensive. The matrix multiplications shown in (3-37) must be performed for each pixel. On the other hand, linear unmixing provides more quantitative information than a simple classification.

**Figure 3-55.** (a) Io’s spectral reflectance showing the step drop near 0.45 µm (from Fanale et al., 1974). (b) Io’s spectral reflectance showing the deep absorption at 4.08 µm associated with SO\(_2\) frost compared to laboratory reflectance spectrum of SO\(_2\) frost (courtesy of D. Nash, JPL).

Identifying the endmember spectra to use in the analysis is a major challenge, since the results are strongly influenced by the choice of the endmember spectra. As in the case of simple classification, several techniques are used to select these. One could use laboratory or field spectra. In that case, the remote sensing dataset must be calibrated. If the analyst has some prior knowledge about the scene, endmember spectra could be selected from the scene itself. Alternatively, multi-dimensional scatter plots could be used to identify the extreme values in the multi-dimensional histogram of the spectra in the image. Since all the pixels in the image fall within the volume enclosed by these extreme endpoints, they form an exhaustive set of endmembers.
As an illustration we used the four features used in the supervised classification shown in Figure 3-53 as our “endmembers” and performed an unmixing of the Cuprite scene. There is no guarantee that these spectra are indeed endmembers in this image, so care should be taken when interpreting the results, as negative abundances may result. The results are shown in Figure 3-54, where we display only the abundances associated with the three spectra selected from the alteration zone. The color assignment is the same as that used in Figure 3-53. As expected, the features that were originally selected from the alteration zone shows high abundances in these areas, and relatively low amounts in the rest of the image. When comparing the results to published mineral maps of the Cuprite area, it is found that the reddish areas in Figure 3-54 show high abundance of hematite and goethite.

3-4-3 Identification

The last step in the spectral analysis of imaging data is the unique identification of the classified elements. This requires a detailed knowledge of the spectral signatures of the materials being sought, as well as of all the other materials in the scene, and the development of a spectral signature library of all expected natural materials. In the ideal case, if a certain material, or family of materials, is the only one which has a certain spectral feature, such as an absorption line at a certain wavelength, the identification becomes simple. The identification feature could be a single absorption line or an association of lines.


**Figure 3-56.** This graph shows a spectrum, taken by the Mars Exploration Rover Opportunity's Mössbauer spectrometer. The Mössbauer spectrometer graph shows the presence of an iron-bearing mineral called jarosite.
If spectral signatures from a reference library were used in the classification to begin with, the identification is automatic. However, if the classification was done using areas selected from the image, or data-driven endmembers, the final identification step involves comparing the spectra of the endmembers or feature vectors to those in a reference library. If a reference spectral library is not available, field checking in one area of each class will allow identification of the constituents in the whole scene.

One example was the case of identifying the presence of sodium on the surface of Io, the first satellite of Jupiter, by ground spectral observation. Figure 3-55 shows the reflectance of Io in the visible and near infrared and compares it to the reflectance spectra of a number of materials including sulfur dioxide frost. Two specific features led to the identification of Io surface material as consisting mainly of sulfur dioxide frost: (1) The sharp drop of the reflectance for wavelengths lower than 0.45 µm. This wavelength corresponds to the energy of the forbidden gap in sulfur (see Fig. 3-15). (2) The deep absorption band at 4.08 µm, which corresponds to the absorption of SO₂ frost.

Another, more recent, example is the identification of the mineral jarosite on the Meridiani plain on Mars by the Mössbauer spectrometer on the Mars Exploration Rover Opportunity. Figure 3-56 shows the measured spectrum of one of the rocks dubbed “EL Capitan” within the rock outcrop inside the crater where the rover Opportunity landed (see Fig. 3-42). The Mössbauer spectrometer graph shows the presence of an iron-bearing mineral called jarosite. The pair of yellow peaks in the graph indicates a jarosite phase, which contains water in the form of hydroxyl as a part of its structure. These data give evidence of water-driven processes that have existed on Mars. Three other phases are also identified in this spectrum: a magnetic phase (blue), attributed to an iron-oxide mineral; a silicate phase (green), indicative of minerals containing double-ionized iron (Fe₂⁺); and a third phase (red) of minerals with triple-ionized iron (Fe³⁺).

**EXERCISES**

3-1. Consider a planet with radius $R$ with its rotation axis inclined at an angle $\psi$ relative to the normal to the plane of its orbit around the sun. Assume that the solar rays reaching the planet are all parallel to the orbital plane.

(a) Calculate the sun angle (angle between the local vertical and the sun direction) at the equator as a function of longitude $\delta$ and the location of the planet in its orbit (i.e. the angle $\alpha$ in Figure 3-A). Consider $\delta = 0$ to be the point closest to the sun. Plot the sun angle for $\psi = 0^\circ$, $27^\circ$ and for $\delta = 0$.

(b) Calculate the sun angle at the poles. Plot for $\psi = 0^\circ$, $27^\circ$.

(c) Calculate the sun angle as a function of latitude $\gamma$ along zero longitude as a function of $\alpha$. Plot for $\psi = 0^\circ$, $27^\circ$ and for $\gamma = 45^\circ$.

(d) Calculate the sun angle as a function of longitude $\delta$ for $\gamma = 45^\circ$, $\psi = 27^\circ$ and $\alpha = 0^\circ$, $90^\circ$ and $180^\circ$. 
3-2. A narrow band filter with a variable center frequency is used to observe a spectral region where the emitted intensity is given by

\[ W(\nu) = 1 - \alpha e^{-\left(\frac{\nu - \nu_0}{\nu_c}\right)^2} \]

The filter transmission function is given by

\[ F(\nu) = e^{-\left(\frac{\nu - \nu_f}{\nu_c}\right)^2} \]

The intensity measured by a detector behind the filter is given by

\[ I = \int T(\nu) \, d\nu = \int W(\nu) F(\nu) \, d\nu \]

(a) Derive the expression for \( I \) as a function of \( \nu_s, \nu_f, \nu_c \) and \( \alpha \).
(b) Plot \( I \) as a function of \( \nu_c/\nu_s \) for \( \nu_f/\nu_s = 0.1, 1, \) and 10, and for \( \alpha = 0.1 \) and 1.
(c) Repeat the calculations for

\[ F(\nu) = \begin{cases} 
1 & \text{for } -1 \leq \frac{\nu - \nu_c}{\nu_f} \leq 1 \\
0 & \text{otherwise}
\end{cases} \]
3-3. Figure 3-B shows the energy levels for three different materials.

(a) Silicon has a bandgap of 1.12 eV. Calculate the cut-off wavelength for this material and sketch the expected absorption spectrum.
(b) Now suppose silicon is doped with arsenic, which places a layer of electrons at 1.07 eV. Calculate the cut-off wavelength for this material and sketch the absorption spectrum.
(c) This material has three discrete energy levels at 1 eV, 1.5 eV and 3 eV. Calculate the positions of the primary absorption lines, and sketch the absorption spectrum of this material.

![Figure 3-B](image)

**Figure 3-B.** Energy levels of three different materials.

3-4. A class of hypothetical granular materials have energy levels which depend on the crystalline field $E$ as shown in Figure 3-C. Describe the behavior of the spectral reflectance as a function of $E$. Only the transitions shown in Figure 3-C are allowed.

![Figure 3-C](image)

**Figure 3-C.** Energy levels and allowable transitions for a hypothetical material
3-5. Calculate and plot the reflection coefficient as a function of the angle of incidence for a half space with index of refraction equal to \( n = 1.7 \) and for \( n = 9 \). In both cases, the upper half space has \( n = 1 \). Consider both horizontal and vertical polarizations.

3-6. Consider a half space with index of refraction \( n = n_r + in_i \) and

\[
 n_i = \frac{\alpha}{\nu_0^2} \left( 1 - \nu^2 / \nu_0^2 \right)
\]

Assume that \( \alpha \ll n_r \). Calculate the reflection coefficient for normal incidence and plot its behavior as a function of \( \nu / \nu_0 \) for \( n_r = 3, \alpha = 0.1 \) and \( \nu / \nu_0 = 0.05 \).

3-7. A telescope is orbiting the earth at 705 km altitude. The telescope lens diameter is 40 cm, the focal length is 120 cm, and the square focal plane is 20.916 cm on a side.
(a) Calculate the power density at the lens if the earth albedo is 50%. (Assume that the solar energy density at the surface of the earth is 1.37 kW/m²).
(b) Calculate the total energy intercepted from a 30m x 30m ground pixel if the shutter stays open for 0.1 millisecond.
(c) Calculate the field-of-view and the swath width of the telescope.
(d) If the instrument is a framing camera, calculate the number of detectors required to acquire a square image with 30m x 30m pixels on the ground. Also, calculate the dwell time if we assume that we use a scanning mirror to stare at the same area with a 75% duty cycle.
(e) In the pushbroom configuration with a single line of detectors, calculate the number of detectors required for a pixel size of 30m x 30m on the ground. Also calculate the dwell time and total energy per pixel assuming that smearing is to be kept less than 1/3 of a pixel.
(f) For a scanning mirror system, calculate the mirror scan rate, the dwell time per pixel, and the total energy intercepted from each pixel assuming that we use an array of 16 detectors arranged in the along-track direction.
(g) Now assume that we place a filter in the focal plane with a 1% transmission bandwidth centered at 0.5 microns. Calculate the number of photons received from a pixel for each of the three cases above.
(h) Calculate the data rate for each of the three cases, assuming we digitize the data from each pixel to 8 bits.

3-8. A telescope is orbiting the earth at 705 km altitude. The telescope lens diameter is 40 cm, the focal length is 120 cm, and the square focal plane is 10 cm on a side. The pixels in the detector are 10 microns on a side. Calculate the swath width, size of the pixels on the ground, and the number of pixels across the swath assuming a pushbroom design. Also, calculate the maximum dwell time per pixel, and the resulting data rate.

The instrument is now changed into a spectrometer. The incoming light is dispersed using a prism such that the different colors are separated spatially in a direction orthogonal to the
pushbroom line array. To measure the different spectral components, the prism mechanism is scanned so that different colors sweep over the line array. If we are to measure 64 spectral channels, calculate the scan rate of the dispersion mechanism. What is the dwell time per spectral channel now? What is the data rate?

Assume that the wavelength region 0.4 microns to 2.4 microns is dispersed over 640 microns in the focal plane. Now assume that we stack 64 line arrays next to each other to cover the dispersed spectrum. Calculate the bandpass of each channel, the dwell time per spectral channel and the resulting data rate.

3-9. A scanning laser altimeter is in a 250 km altitude orbit. The laser sends a series of 20 nanosecond pulses with a peak power of 10 kW and a beam divergence of 0.5 milliradian. What is the beam footprint on the surface and the height measurement accuracy? Let us assume that the surface has a 10% albedo and scatters in an isotropic fashion in the upper hemisphere. What is the power collected by a receiving collector of 50 cm diameter? What number of photons does this energy correspond to assuming that \( \lambda = 0.5 \mu m \) ? Assuming that the detector has an area of \( 10^{-5} \text{ cm}^2 \) and a bandwidth \( \Delta f = 1/\tau \) with \( \tau = 20 \) nanoseconds, calculate the detector minimum \( D^* \). In order to eliminate the signal due to the solar background, a 1% (\( i.e. .005 \mu m \) bandwidth) filter is used before the detector.

Assuming that the solar energy density at the surface is 2 kW/m²/\( \mu m \), what is the signal to background ratio? How can this ratio be increased? Do you think this system can be operated during the day? Assuming that a mapping swath of 40 km is required and the surface has to be sampled every 200 m, how many pulses does the laser need to transmit per second? Assume that the satellite velocity is 7 km/sec.

3-10. Consider a planet with radius 1900 km, located at a distance of \( 1.4 \times 10^9 \) km from the Sun and \( 1.25 \times 10^9 \) km from the Earth. The average surface temperature of the planet is 70 K. The surface reflectivity is 0.3 across the visible and infrared spectrum. Plot the received energy by a 1 m² aperture telescope in Earth orbit as a function of wavelength. Assume that the Sun is a blackbody with a temperature of 6000 K and a radius of \( 7 \times 10^5 \) km. Now let us assume an area of size 100 km x 100 km at the equator consists of an active volcanic caldera with a temperature of 700 K. Can the earth orbiting telescope uniquely detect the presence of the caldera? Can it measure its size and temperature? Explain.

REFERENCES AND FURTHER READING


Hubbard, B.E., J. K. Crowley, and D. R. Zimbelman, Comparative alteration mineral mapping using visible to shortwave infrared (0.4-2.4 \( \mu m \)) Hyperion, ALI, and ASTER imagery. *IEEE Transactions on Geoscience and Remote Sensing*, **GE-41**, 1401 - 1410, 2003.


