Optical Astronomical Instrumentation

Jaap Tinbergen

Dwingeloo, NL
## 5.3. POLARIMETRIC INSTRUMENTATION IN PRACTICE

- **5.3.1. Telescopes** ......................................................................................................................... 72
- **5.3.2. Optical polarimeters** ............................................................................................................ 74
- **5.3.3. Radio polarimeters** ............................................................................................................... 78
- **5.3.4. Infrared and (sub)mm polarimeters** ....................................................................................... 81

## 6. THE EARTH'S ATMOSPHERE

- **6.1. ATMOSPHERIC DISPERSION** .................................................................................................. 83
- **6.2. SEEING, SCINTILLATION AND EXTINCTION** ......................................................................... 85
- **6.3. FARADAY ROTATION** ............................................................................................................. 86
- **6.4. UNDOING THE EFFECTS OF THE ATMOSPHERE** ..................................................................... 87
  - **6.4.1. Atmospheric dispersion correction** ....................................................................................... 87
  - **6.4.2. Eliminating extinction effects** ............................................................................................... 88
  - **6.4.3. Avoiding scintillation effects** ............................................................................................... 88
  - **6.4.4. Undoing the effects of seeing** ............................................................................................... 89
  - **6.4.5. Derotation of Faraday-degraded data** .................................................................................. 92

## 7. OPTICAL-ASTRONOMICAL INSTRUMENTS

- **7.1. TELESCOPES REVISITED** ....................................................................................................... 93
- **7.2. CORRELATOR INTERFEROMETRY AND APERTURE SYNTHESIS** ........................................ 98
- **7.3. IMAGING** .................................................................................................................................. 102
- **7.4. PHOTOMETRY** ......................................................................................................................... 105
- **7.5. POLARIMETRY** ....................................................................................................................... 107
- **7.6. SPECTRO(POLAR)METRY** ....................................................................................................... 108
- **7.7. ASTROMETRY** .......................................................................................................................... 115
- **7.8. HYBRID INSTRUMENTS** ........................................................................................................... 116
- **7.9. CALIBRATION** .......................................................................................................................... 118

## 8. OPTICAL COMPONENTS IN ENGINEERING

- **8.1. BASIC IMAGING COMPONENTS** ............................................................................................. 121
  - **8.1.1. Transmission optics** ............................................................................................................ 121
  - **8.1.2. Reflective optics** .................................................................................................................. 124
  - **8.1.3. Spherical/aspheric/off-axis optics** ...................................................................................... 125
  - **8.1.4. Cylindrical optics** ................................................................................................................ 129
  - **8.1.5. Non-classical imaging optics** ............................................................................................... 129
  - **8.1.6. Filters** .................................................................................................................................... 132
  - **8.2. THIN FILMS AND OTHER COATINGS** .................................................................................. 133
    - **8.2.1. Anti-Reflection coatings** .................................................................................................. 133
    - **8.2.2. Mirror coatings** ............................................................................................................... 136
    - **8.2.3. Dichroic coatings** .............................................................................................................. 137
  - **8.3. SYSTEMS** ............................................................................................................................... 140
    - **8.3.1. Zoom/telephoto/macro** ..................................................................................................... 140
    - **8.3.2. Telecentric** ...................................................................................................................... 141
    - **8.3.3. Microscope objectives** ..................................................................................................... 141
    - **8.3.4. Eyepieces** .......................................................................................................................... 141
    - **8.3.5. Condensors and diffusers, non-imaging optics** ................................................................ 142
    - **8.3.6. Etc.** ..................................................................................................................................... 142
  - **8.4. MANIPULATION** .................................................................................................................... 142
    - **8.4.1. Filters** .................................................................................................................................. 144
    - **8.4.2. Polarization modifiers** ....................................................................................................... 145
    - **8.4.3. Time-dependent devices/modulators** .................................................................................. 144
  - **8.5. DESIGN** .................................................................................................................................... 150
    - **8.5.1. Optical systems and ray-tracing** ...................................................................................... 150
    - **8.5.2. Other design activities** ...................................................................................................... 151
  - **8.6. MANUFACTURING TECHNIQUES** ......................................................................................... 154
  - **8.7. TEST AND MEASUREMENT** .................................................................................................. 159

## 9. ONE FOR THE ROAD

- **9.1. THE INTEGRAL AS A SUMMATION** .......................................................................................... 162
Preface

This document is a companion volume to Johan Hamaker's "Instrumental Techniques of Aperture Synthesis" (which will be referred to, often with section numbers added, as "Hamaker"). Its purpose is to provide background to the 'optical' projects undertaken by ASTRON/NFRA, primarily for the benefit of those working in those projects. However, nothing in the course is too technical for university astronomy students, and they may benefit from the mildly engineering way of looking at the instrumentation they use; a course aimed more directly at them is slowly materialising as the 'Dictionary of Astronomical Instrumentation' (DAI), at [http://www.strw.leidenuniv.nl/dai/index.html](http://www.strw.leidenuniv.nl/dai/index.html) (where clicking to "Books, lecture notes" gives access to an annotated list of useful reference books with much more detail than the present course).

The level of assumed previous understanding (whether by formal training or by experience) is loosely defined as 'Dutch HBO' ('polytechnic graduate', equally loosely), but very little is needed in the way of mathematics. Complex-number representation of phasors is treated in Hamaker's course; in addition, I explain and use a simple form of the "integral" concept in order to discuss the Fourier Transform and I use a simple matrix representation of polarization.

I have attempted to emphasize instrumental principles, leaving the details to be dealt with when one needs them. The unity of optical and radio instrumentation is illustrated whenever the opportunity arises.

Now have a quick look at the [final chapter](#).
# 1. Astronomical Instrumentation Overview

## 1.1. Astronomical requirements

Modern astronomers are interested in a very large variety of objects in space. They (have to) use a large variety of instrumental systems to observe these objects and satisfy mankind's curiosity. Examples of this are:

<table>
<thead>
<tr>
<th>&quot;Object&quot;</th>
<th>Instrumental system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radio spectrum of remote galaxies</td>
<td>Westerbork + MFFE</td>
</tr>
<tr>
<td>Hot gas in clusters of galaxies</td>
<td>X-ray imagers</td>
</tr>
<tr>
<td>Gas motions in stellar atmospheres</td>
<td>WHT/VLT + spectrometers (R = 10^5)</td>
</tr>
<tr>
<td>Solar corona</td>
<td>X-ray imagers, optical coronographs, total eclipses</td>
</tr>
<tr>
<td>Molecules in interstellar gas clouds</td>
<td>JCMT + spectral line receivers</td>
</tr>
<tr>
<td>Close binary stars</td>
<td>X-ray photometers, WHT/VLT + fast optical spectrometers</td>
</tr>
<tr>
<td>Details of polarization in pulsar pulses</td>
<td>Westerbork + PUMA, WHT/VLT + fast polarimeter</td>
</tr>
<tr>
<td>γ-ray bursters</td>
<td>(γ + X + optical) photometers, &quot;immediate response&quot;</td>
</tr>
<tr>
<td>&quot;Warm&quot; dust round emerging stars</td>
<td>ISO, VLT + VISIR</td>
</tr>
<tr>
<td>Milliarcsecond structures (including polarization)</td>
<td>VLBI including polarization, VLTI + MIDI (+ polarization option)</td>
</tr>
</tbody>
</table>

The 'instrumental system' one chooses depends on the wavelength at which one wishes to observe and this in general depends on the temperature of the object one wishes to observe. The Planck curve for the particular temperature indicates the wavelengths that contain useful amounts of "black-body" flux. See also the table below.

Radiation mechanisms other than thermal exist and they can yield measurable radiation at other wavelengths than one would expect from the temperature of the source. Familiar examples are: much of the (synchrotron) radio emission from e.g. galaxies and supernova remnants; blue sky (scattered sunlight).

### 1.1.1. The astronomical spectrum

Astronomy's staple diet is electromagnetic (EM) radiation. One does also glean something from detection of cosmic ray particles and from direct exploration (of the solar system), but these are marginal for our grand view of the Universe and will be ignored from here on.
EM radiation comes in a variety of colours (wavelengths), from the short $\gamma$-rays via X-rays, ultra-violet (UV), visible and infra-red (IR) to (sub)mm, cm, metre (and even longer) radio waves. The correspondence in nomenclature is roughly:

<table>
<thead>
<tr>
<th>Object</th>
<th>Solar corona</th>
<th>Hot stars</th>
<th>Solar photosphere</th>
<th>Interstellar dust/gas</th>
</tr>
</thead>
<tbody>
<tr>
<td>BB Temp (K)</td>
<td>$10^6$</td>
<td>$10^5$</td>
<td>$10^4$</td>
<td>$10^3$</td>
</tr>
<tr>
<td>Wavelength (mm)</td>
<td>$10^{-6}$</td>
<td>$10^{-5}$</td>
<td>$10^{-4}$</td>
<td>$10^{-3}$</td>
</tr>
<tr>
<td>Spectral region</td>
<td>$\leq \gamma$</td>
<td>EUV</td>
<td>UV</td>
<td>visible</td>
</tr>
</tbody>
</table>

Instrumentation contributed to by Dutch astronomical engineering in various spectral regimes is shown in the table below. Note the dominance of the two technical organisations in this age of high technology. This is in a sense the tragedy of modern astronomy, but it is unavoidable; university departments cannot host the very large groups of engineering specialists that are now required. What we can and must do is to ensure cohesion between university departments and engineering labs; the present set of lectures is one contribution towards this end: even within ASTRON a certain effort is needed to ensure that every specialist within a project understands the wider scene that it is embedded in. The other side of the same argument is that university astronomers must acquire a rudimentary understanding of instrumental principles; the website [http://www.strw.leidenuniv.nl/dai/index.html](http://www.strw.leidenuniv.nl/dai/index.html) is a contribution towards that.

<table>
<thead>
<tr>
<th>Spectral region</th>
<th>$\gamma$</th>
<th>X</th>
<th>UV</th>
<th>Visible</th>
<th>IR</th>
<th>(sub)mm</th>
<th>cm/dm/m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument system</td>
<td>COMPTEL, LETG, ANS, SAX, DOT</td>
<td>UES, TAURUS, IRAS, VISIR, ISIS-POL, MIDI</td>
<td>backends</td>
<td>JCMT</td>
<td>D'loo 25-m, W'bork, SKA &amp; LOFAR, Polarimetry, PUMA, Correlators, AIPS++</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL organisation</td>
<td>$&lt;====== SRON ======&gt;$</td>
<td>Utrecht(†), Roden†</td>
<td>SRON, ASTRON</td>
<td>$&lt;====== ASTRON ======&gt;$</td>
<td>Utrecht</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A number of different paradigms (canonized conceptual models) dominate instrumental systems in different parts of the spectrum. They are listed below and their approximate spectral coverage is indicated. Aperture synthesis has been treated in extenso by Hamaker; the present set of lectures will concern mainly the 'telescope/instrument/detector' model, as used at "optical" (i.e. roughly Xray/UV/Visible/IR) wavelengths.

<table>
<thead>
<tr>
<th>Spectral region</th>
<th>γ</th>
<th>X</th>
<th>UV</th>
<th>Visible</th>
<th>IR</th>
<th>(sub)mm</th>
<th>cm/dm/m/dam/hm</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Character&quot; of EM radiation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EM propagation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Character&quot; of instrumentation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detector/receptor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 1.1.2. The optical domain and its instrumentation

Some characteristic aspects of what we loosely call 'the optical domain' are:

1. Very large telescopes and components (in terms of wavelength), therefore mostly geometrical optics.
2. Photon-counting detectors (or nearly so).
3. No thermal background from instrument, very little light from atmosphere at the best sites.
4. Seeing and scintillation are troublesome atmospheric effects that strongly influence the design of earthbound instrumentation.
5. Lots of good optical materials, refractive and reflective.
6. Sophisticated design methods.
7. Fast modulation allowed (even for CCDs).

We shall come across these aspects in later chapters. Note that these criteria are approximate and that not all of them may apply in a particular case. Let us agree that the patient shows the 'optical instrumentation' syndrome and 'optical treatment' should be tried, if 4 or more of the above symptoms are exhibited.
2. The Telescope/Instrument/Detector Model

In this preliminary chapter, the ground will be cleared for the remainder of the course, the main concern of which will be: the principles of optical astronomical instrumentation. Since most of the course will concern the instruments that analyse the EM radiation, the telescope (which concentrates it) and the detector (which converts it into a signal suitable for permanent filing) will be discussed separately here.

2.1. Introduction

In the central part of the astronomical spectrum, an observing system generally consists of the following, classical, 'optical' subsystems:

1. Telescope: concentrates the radiation and images the sky in the 'focal plane'.
2. Instrument: isolates and/or manipulates the radiation.
3. Detector: converts power of the radiation into an (electrical) signal which can be recorded.

N.B.: At the shortest astronomical wavelengths (X, γ) reflective or refractive components do not exist and a conventional 'telescope' can't be built, so opaque masks must be used to define the direction(s) from which radiation is accepted.

N.B.: At long wavelengths, optical components (telescopes, but also prisms, gratings and waveplates) would have to be very large, too large to be at all feasible, so that autocorrelation/FT techniques are used (aperture synthesis, autocorrelation spectrometers).

Simple telescopes (i.e. excluding correlation interferometers and aperture synthesis) can for a large part be discussed in terms of their geometric-optical properties (since the optical components are very much larger than the wavelength of the radiation). Classical design uses ray optical theory (including aberration theory based on ray optics). However, for extreme and emphatically 'modern' aspects (very high spatial resolution, very long wavelengths) one does have to take the wave character of EM radiation into account. Ray optics can be considered to be just a limiting case of technical astronomy's central discipline of wave optics (the sequence of linked concepts being: waves --> wavefronts --> rays --> beams, images).

In optical detectors, the particle character of EM radiation comes to the fore. The radiation particles ('photons') are often detected individually ('photon counting'), or the granularity of the energy flux has to be taken into account (photon-noise-limited performance). This distinguishes the detector from the telescope and optical instrument, where the EM radiation can be considered to be an absolutely 'smooth' wave phenomenon, and wave amplitude and phase are more relevant concepts than energy flux. Another distinction is that, at astronomical signal levels, telescope and instrument always have linear response, whereas detectors always deviate from response linearity to some extent.
2.1.1. Alternative concepts

Although the *telescope/instrument/detector* model is generally useful, at times one may wish to use other models (other ways of splitting up the 'system' into 'components'). Be on the lookout for such opportunities; they are often enlightening. A case in point is the classical telescope/slit-spectrometer combination, below.

![Diagram of telescope and spectrometer](image)

**Figure 2-1: Telescope and spectrometer rearranged as beamsize converter and objective-prism telescope**

Traditionally, the 'telescope' produces an 'image', part of which is selected by a diaphragm or slit. The light from the selected object(s) passes into the 'slit-spectrometer', in which the 'collimator' renders the rays (from each point within the slit) parallel. These parallel beams (bundles of rays) then strike a prism or other dispersing element and are deviated by an angle which depends on wavelength, without affecting the parallelism of the beam (the beam that originated from any one point in the slit, for any one wavelength). The 'camera' then converts the 'angular spectrum' back into an image of the slit (the position of which depends on the wavelength). This is what optical astronomers call 'the spectrum' of the selected object, but it is something quite different from what an audio engineer would call a spectrum; a single-frequency component in an audio spectrum has identifiable amplitude and phase, while this optical spectrum has only intensity: it is in fact a spatial representation of the power spectrum (its intensity is the square of...
the amplitude; phase is irrelevant); more descriptively, one could call it a 'wavelength-dispersed power image' of the source.

One may (as a thought experiment) combine the optical elements differently. The 'camera' may be regarded as a small kind of telescope, with a so-called objective prism in front of it. This assembly looks at parallel rays of white light, apparently coming from a source at infinity. The 'telescope' and 'collimator' then function purely as a 'beamsize reducer' (or, equivalently, as a 'field-angle magnifier'). This way of looking at the system is useful in showing that, in general, larger telescopes require larger instruments, as follows:

• For a certain acceptable smearing of the spectrum, the above apparent source at infinity should be no larger than a given angular size (at the input to the disperser); the actual value of this will depend on detector pixel size and the focal length of the 'camera' lens.

• For a given angular sky-field input to the (large) telescope, the field-angle magnifier should therefore have no more than a certain maximum magnification factor. Since this factor is equal to the ratio of the focal lengths of telescope and collimator, a larger input telescope (at the same F-ratio) implies an increased collimator focal length and hence a larger collimator diameter (to catch all of the beam as concentrated by the telescope), hence also a larger disperser and camera. Q.E.D.

• One could of course escape from this dilemma by making the accepted angular field in the sky smaller, but the atmosphere smears to about 1 arcsecond, so at that point a narrower slit starts to reduce light from a 'point source' and the gain from the increased telescope size is disappointing.

2.2. Telescopes

The term 'telescope' once denoted a complete instrument consisting of an objective and an eyepiece; it only needed the human eye to complete the optical system. In optical textbooks, one will still find the term used in this way (one may also find 'instrument' used for what we now call a 'telescope'). However, in astronomy, the objective became enormous and the eyepiece evolved into a much more extensive instrument as other detectors took over from the human eye. So, within astronomy, 'telescope' is now synonymous with what others call an 'objective' and astronomical telescope engineering is mainly concerned with achieving as large an objective as possible, with as wide an angular field as possible and with images as sharply defined as possible (needless to say, these requirements are partly contradictory, so the art of successful compromise is at the heart of such engineering; this tends to drive astronomers to distraction, they want everything, of course - rightly so, but also impossibly so). Since the objectives are so huge, their mechanical support is non-trivial. Hence mechanical heavy engineering of extreme precision is the other aspect of telescope design that is more or less unique.

Astronomical telescopes were once constructed from lenses, but size was limited to about 1 metre (deformation under own weight: stiffness proportional to $D^2$, weight to $D^3$), so all telescopes now are reflective (mirrors can be supported from the rear and kept in their proper shape, by passive or active means; lenses can only be supported at their edge).
Telescopes may consist of single optical elements, or they may be compound. Some of the designs in use at the present time are:

<table>
<thead>
<tr>
<th>Class</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Paraboloid only</td>
<td>Dwingeloo 25-metre dish</td>
</tr>
<tr>
<td>b) Paraboloid + secondary</td>
<td>Cassegrain, Gregorian</td>
</tr>
<tr>
<td>c) More complex primary and secondary</td>
<td>Ritchey-Chrétien: large field</td>
</tr>
<tr>
<td>d) Spherical primary + corrector</td>
<td>Schmidt: very large field (chromatic)</td>
</tr>
<tr>
<td>e) Nasmyth and broken Cassegrain</td>
<td>instrument orientation constant wrt gravity</td>
</tr>
<tr>
<td>f) Coudé</td>
<td>instrument stationary</td>
</tr>
<tr>
<td>g) Grazing incidence (X-rays)</td>
<td>Wolter types, Chandra multiple shell</td>
</tr>
</tbody>
</table>

**Question:** Why are MFFE, UES, VISIR and MIDI located at their respective foci and not at some other focus?

### 2.3. Detectors

A detector is any component or system that yields an output signal (usually voltage) proportional to input **power**. In the radio domain, the detector is situated late in the system, after amplification (and frequency conversion) of a **voltage** which is proportional to the EM radiation **field** exciting the feed antenna. At shorter wavelengths, detection is normally direct, i.e. the EM radiation itself strikes the detector (after filtering, wavelength dispersion or other manipulation, at **signal frequency**).

The large majority of optical detectors now in use in astronomy is photo-electric, i.e. radiation particles ('photons') each liberate one or more electrons (into a vacuum or into a 'conduction band' in a semiconductor) and these electrons are in one way or another moved out of the component as a current, which can be passed through a resistor to yield a detectable voltage; that voltage is then amplified before further processing (such as demodulation, integration, counting pulses). An obsolescent detector that has served optical astronomy very well is the photographic emulsion, in which individual grains of silver halide are each activated by a few photons and are treated chemically later (to be 'developed' -- reduced to metallic silver). Photographic detection is now used only for very-large-scale surveys, when something like 50 x 50 cm detectors are needed.

A 'bolometer' is any type of detector of EM radiation, in which the energy of the radiation is used to heat the detector and this heating is detected in some way or another. Whenever no other type of detector is available, bolometers may be pressed into service; historically, bolometers played a role at IR, submm and X-ray wavelengths (I think it was Fraunhofer who held a blackened thermometer in the solar spectrum, beyond the reddest red he could see). A sensitive bolometer
must have very small heat capacity and it should be possible to measure with great precision a property sensitive to its temperature.

Historically, optical astronomy used (after the aeons of the human retina) photography whenever an extended object or field of point sources had to be imaged, or a spectrum had to be recorded. Though the 'quantum efficiency' (the fraction of incident photons actually used) of the photographic emulsion was (and is) only of order 1%, the parallel-processing advantage of the large photographic plates actually offset this disadvantage by a large factor. The photographic emulsion is not very linear in its response and the techniques for obtaining calibrated photometry from photographic registrations were many, specialised and complicated.

Whenever strict linearity of response was required, or high speed of response (e.g. modulation methods), the photomultiplier tube (PMT) was used (early on and in special cases, the photodiode). Since they were used in analogue mode, considerable effort was necessary to make amplifiers perform well enough to match these detectors, effort of a kind similar to that in radio-astronomy. Feedback (servo) methods were mostly used to make the response of a passive component like resistor or condensor determine the amplification and linearity of the amplifier. With PMTs, however, another option was to detect the output pulses resulting from individual photons and to count these; this is possible only because the cascade amplification process in a PMT dynode chain is almost noise-free: one uses internal gains of order 1 million and thus obtains output pulses easily large enough to be detected and counted by simple electronics. Obtaining linearity then reduces to keeping the intervals between pulses large compared to their length (so that virtually no double pulses are counted as single). This photon-counting mode was used for many years to obtain fundamental linear photometry of faint sources for calibration of photographic data; until CCDs took over, 'photo-electric photometry' was almost synonymous with 'PMT photon-counting'.

The modern successor to the PMT for specialised applications (such as detecting optical interference fringes by rapidly sweeping them back and forth over the detector within a fraction of the effective scintillation cycle) is the avalanche photodiode (APD). This semiconductor component has the advantage of an efficiency several times larger than that of even the best PMT photocathodes, and it can be made to operate in photon-counting mode; it is also small and does not need excessive supply voltages.

For mainstream optical astronomy, the detector of choice is now the CCD (although active-pixel detectors of various kinds are in development and may supersede CCDs in the future). A good introduction to CCDs for non-specialists is given at the commercial site of Apogee Instruments [http://www.apogee-ccd.com/ccdu.html](http://www.apogee-ccd.com/ccdu.html); a taste of the laboratory work needed just to verify performance is given by a [1995 paper by Tim Abbott](http://www.not.iac.es/CCD-world/). Proper design of CCDs and their associated electronics is an engineering discipline in itself and involves other aspects of instrument design. To take just one example:

To obtain high Q.E., anti-reflection coatings (Figure 8-13) are a must (since silicon has a very high refractive index and bare silicon therefore has very substantial reflection losses). Complicated dielectric coatings designed for normal incidence do not always perform well at oblique angles. This usually limits the acceptance angle to something less than would be desirable in a system on a large telescope (where final beams are inevitably very-wide-angle in order to force the image onto the small CCD detector). Large (assemblies of) CCDs are the (expensive) solution, and overall cost optimization is a complex matter, involving also optics, mechanics, cryogenics, backend electronics and computers.

Do you see why optical-detector science has its own specialists? To get an impression, see the archive of the email list "CCD-world" [www.not.iac.es/CCD-world/](http://www.not.iac.es/CCD-world/).
A final aspect of modern optical detectors that should be mentioned is that, whenever new types of detector are invented, great efforts are invariably made to implement them as arrays; such is the pressure for enormous parallel-processing advantage to push astronomy to observing ever greater numbers of ever fainter sources, by other means than ever larger telescopes. In the radio domain, SCUBA is a manifestation of this drive, even though full replication of the entire electronic chain is required. In the CCD detectors for optical astronomy, only the detector pixels are replicated N x M times (N, M of order 1000), the readout electronics usually only N (or N + M), while the processing chain is present in at most a few parallel copies per chip (many chips may go into a modern detector system, though, and full parallel replication generally holds at that level). The essential element in avoiding full replication within the chip is a pixel-level memory, normally a capacitor (one per pixel) in which the electric charge generated by the incident photons can be stored. CIDs, active-pixel and infra-red detectors such as the BIB (blocked-impurity-band, whatever that means) types used in VISIR are also available in array form. Even PMTs are now made in 10 x 10 units within a single glass envelope and APDs also come in modest arrays. In the early 70's a DIY array of photodiodes existed at Sacramento Peak solar observatory within a spectro-polarimeter; it had several hundred individual diodes and measured several metres!

There is little point in discussing technical details of optical detectors in a course for 'outsiders' (i.e. those who will probably never design detector subsystems themselves). It is more appropriate to list the properties that determine where and how certain detectors can be used in astronomical instrumentation.

### 2.3.1. Efficiency

The sole purpose in life of a detector is to convert radiation (power) into a current, voltage or other signal that can be processed and recorded. Its efficiency can be expressed in such units as amperes per watt, but such measures are detector-specific, readout-specific, wavelength-specific, etc. A more fundamental measure is the so-called quantum efficiency (Q.E.), which is the fraction of the incident radiation power that actually leads to an output signal; it is like an internal transmission coefficient for the detector. For photo-electric detectors and the shorter optical wavelengths, it is the ratio of the number of separately detectable pulses in the output to the number of photons present in the incident radiation. Since the fractional noise in a measurement of an optical signal of N photons is $\frac{1}{\sqrt{N \times Q.E.}}$ (exactly analogous to the case of electron shot noise in an electrical current, see [Hamaker](15.2)), it is important that Q.E. is as high as possible (it can never exceed 1, obviously). For modern CCDs, Q.E. can be as high as 90% in the 0.3. to 0.9 μm region and also in the X-ray region. Detectors incorporating photocathodes (PMTs, ICCDs) often do no better than 10-50%, while the photographic emulsion has a Q.E. of order 1%; if for some reason one is forced to use photographic plates rather than a CCD, one must increase optical efficiency, telescope collecting area, exposure time, or a combination of these, 100-fold (!) to obtain a signal of equivalent Signal-to-Noise ratio (for the moment forgetting about other important quality aspects such as linearity, homogeneity and zero-offset).
2.3.2. Noise

Noise is not just a detector property; rather, it is a property of the detector system, i.e. including the readout process. As described in [Hamaker]5.5, the noise at the input of the amplifier chain is the critical parameter. In optical detectors, almost-noise-free amplification can be part of the detector (PMT, ICCD, APD); in such a case the noise in the detector dark signal is likely to be dominant (and cooling the detector will often help). If the detector has no internal gain (CCD), readout noise is often the dominant component of the detector system noise (in the case of the CCD, persistent effort has led to extremely low-noise amplifiers, approaching the ideal of less than 1 electron per readout cycle, which will allow the equivalent of photon-counting; in the case of the CID, non-destructive readout allows reduction of the effective noise by repeated readout and averaging, at the expense of much slower readout).

In addition to noise originating in the detector system, there is shot noise in the optical signal itself. In designing a system, one generally aims to use a signal power large enough for signal shot noise to exceed the noise of the detector system; this can involve tailoring the spatial image scale or the spectral bandwidth, hence it may involve a re-imager selection wheel or a filter selection wheel, considerable complications in some instruments.

The term "noise-equivalent power" (NEP) is the lowest signal power that can be detected (at a signal-to-noise ratio of 1), i.e. it is the system noise, expressed in units of signal power (or signal charge for a CCD, hence occasionally NEQ: noise-equivalent charge). The term is used mainly for IR detectors; the system noise is detector background noise plus readout amplifier noise.

The term "noise floor" refers to the lowest attainable system noise; it is often synonymous with "readout noise".

So-called 'fixed-pattern noise' (in array detectors) is in fact variation of sensitivity (or efficiency) from pixel to pixel; such sensitivity variation masquerades as (multiplying) noise in the output. In linear (i.e. 1-D) arrays it often takes the form of a different gain for even- and odd-numbered pixels, hence the name. Needless to say, noise of this kind can be calibrated out by 'flat-fielding'.

2.3.3. Linearity

By "linearity" optical astronomers denote the extent to which output signal (voltage, count rate) of a detector system is strictly proportional to input power. A "linear detector" in optics is a "square-law device" in electronics (output voltage proportional to square of input voltage). One should be aware of this difference in local jargon.

N.B. Linearity is not a desirable property per se. The human visual system is hard to beat and is approximately logarithmic; this is reflected in the astronomical magnitude system. In general, applications of extreme dynamic range (such as a SKA antenna close to Smilde TV tower) may benefit from such a response (of the detector system or, more rarely, of the detector itself), but for most of astronomy, linear response is preferred.
2.3.3.1. Extending the linear range

If a detector (or registration device such as a DVM) is approximately linear, one may use the (assumed) linearity of a passive attenuator (such as a neutral-density filter or a mask at a conjugate optical position, or a pair of resistors in the case of the DVM) to estimate the deviations and improve the system linearity, by examining how the apparent attenuation factor varies with output level. Standard servo theory proves that such linearisation is possible; the following example demonstrates a simple case without making the 'servo' explicit. The situation is shown in the figure. For a detector for which output \( y \) and input \( x \) are related by

\[
y = a + bx + cx^2 + d(x^3 + \ldots) \tag{a,b,c,d constants}
\]

one may take any input pair \( x_1 \) and \( x_2 \) (where \( x_1 \) is the attenuated version of \( x_2 \)) and write, for small deviations from linearity (\( bx_1 \) dominant; this determines the range of validity):

\[
\frac{y_2}{y_1} \approx \frac{a + bx_2 + cx_2^2 + d(x_2^3 + \ldots)}{bx_1} = \frac{x_2}{x_1} + \left(\frac{ax_2}{bx_1}\right) + \left(\frac{cx_2^2}{bx_1}\right) + \left(\frac{dx_2^3}{bx_1}\right) + \ldots = A + B \frac{1}{x_2} + C \cdot x_2 + D \cdot (x_2^2 + \ldots)
\]

since \( \frac{x_2}{x_1} \) is constant by assumption.

The behaviour is shown in [Figure 2-2](#). The second term shows at small \( x_2 \), the third as the slope of the linear range and the fourth at large \( x_2 \); we can determine \( A, B \) and \( C \) and use them to make the system response linear to a much better degree of approximation.

Actual algorithms will be more complex and tailored to the detector in use, but the principle is often the same: use a trusted passive component to calibrate an active one.
2.3.4. Dark signal

Detectors generally show a "dark signal", i.e. an output signal in the absence of any incident radiation. This dark signal adds to any radiation-induced signal and must be eliminated from astronomical observations by suitable extra ('calibration') measurements and subtraction.

In photo-electric detectors, the dark signal is due to that fraction of the bound electrons that have thermal energies in excess of the 'band gap' (the energy needed to free them from the reservoir of bound electrons into the regime of free conduction). As one raises the temperature of the detector material, the average energy of the bound electrons increases and so does the fraction that can jump the barrier into conduction. Conversely, cooling the detector reduces the dark signal.

The more red-sensitive a photo-electric detector has to be, the smaller the bandgap should be in its active material and the more deeply the detector must be cooled, to reduce the average bound-electron thermal energy to well below the bandgap energy. For useful astronomy, ultraviolet detectors generally operate near room temperature, for visible wavelengths thermo-electric, dry-ice (CO$_2$) or liquid-nitrogen cooling is sufficient, while thermal infrared ($\approx$10μm) detectors generally operate below 10K and are cooled by liquid helium or specialised closed-cycle coolers. These requirements strongly influence design of the detector section of an instrument (and sometimes that of the entire instrument).

To obtain dark-signal measurements for subtraction, separate observations are used, under conditions identical to those of 'real' observations. If the dark signal varies with time, one may in extreme cases alternate between signal and dark exposures: 'chopping' by opening and closing a shutter (in practice, one often chops between source and empty sky; in that case, the dark signal is present in both and is not determined explicitly). In array detectors, one may (for interpolation between full determinations) use data from dark parts of the image, but final reference should always be to a separate 'dark frame' taken just before and/or after the 'science' observation.

2.3.5. Speed of response

There are several reasons for sometimes requiring fast response of astronomical detector systems:

1) The (astronomical) object itself varies (e.g. the Sun, pulsars, scintillation)
2) One wishes to scan (rapidly) an image or a spectrum with a single-point detector, in order to get a quick impression of it
3) One uses modulation to eliminate zeropoints or gain variation (chopping in the infra-red, polarimetry)

Until recently, CCDs (array, but slow response) and PMT or APD (fast response, but single-point) were the effective choice. In a development of the last decade at most, much faster CCDs (tens of frames/second or more) at reasonable readout noise are becoming available. This is leading to array instruments with certain AC capabilities, which in turn enables sophisticated observational procedures at high instrument efficiency. The price one pays for this is in the computing power required to extract the final astronomical information (e.g. category 3 above).
In addition to fast-readout CCDs, detectors such as CMOS arrays and CIDs may be used for special applications. In these two types of detector, one may choose to read out only selected parts of the image, thus obtaining high effective speed of response of the system whenever only a small fraction of the image is of interest (other reasons for preferring such detectors may be their dynamic range or the non-destructive readout they permit). See Figure 2-3.

![Figure 2-3: Detector array developments (Photonics Spectra May 2000)](image_url)

### 2.3.6. Saturation signal

All real-life detectors have some sort of limit to their capacity to handle large signals. This is called saturation (as in a saturated solution or market, no further input can be handled). In photoelectric array detectors, the size of the memory element (pixel capacitor) determines the value of the integrated saturation signal (in electrons); to obtain the saturation signal (in electrons/second), divide by the integration time used (which may be milliseconds to many minutes for different detectors and/or applications). The size of the pixel capacitor is commonly specified by its 'full well' capacity to store electrons; in any one type of detector, it is proportional to the pixel area.

Saturation is mostly a gradual effect (the upper end of the response curve in Figure 2-2); the first onset of saturation may be treated as a non-linearity to be corrected (however, in the servo interpretation mentioned above, the loop gain decreases for increasing saturation and consequently the correction will be increasingly inaccurate).

When extreme saturation occurs, the change can leak away to neighbouring pixels ("blooming"). "Anti-blooming" measures generally make this charge leak away to earth. Such loss of signal is mostly undesirable and astronomy tends to steer clear of such devices, preferring to use neutral-density filters or short exposure times to reduce the signal.

Active-column sensors promise to compete head-to-head with CCDs, even for demanding scientific applications. In addition, smaller pixel sizes can yield dynamic range equivalent to an active-pixel CMOS imager at significantly lower cost.

Both types of CMOS imagers also can do things that the CCD cannot, such as nondestructive readout. Reading and averaging the signal greatly reduce the noise. CCDs also cannot address pixels randomly. This capability, which enables a single pixel or smaller subframes to be read at higher speeds, is important in some applications. Combine this with non-destructive readout and cooling, and a very good imaging system for astronomy is created.

References  [www.photon-vision.com](http://www.photon-vision.com)
2.3.7. Dynamic range

By this term one denotes the ratio of the largest to the smallest signal that the system can usefully handle. The concept needs to be defined in more detail for each application; for optical detector systems, the usual definition is the ratio of integrated saturation signal to readout noise (or some multiple of readout noise; 5 x noise floor is customary in radio astronomy, 3 x in some other professions; it all depends on the confidence level one requires in order to claim a discovery, invest in a start-up company, send someone to prison, launch a nuclear counter-strike rocket, etc).

Astronomy deals with a very large range in signals, generally far too large for a single detector. (Optical) attenuators are almost always pressed into service to extend the effective system dynamic range. Since such attenuators must be calibrated, fixed rather than variable ones are generally in use. The stepsize in a set of (optical) attenuators will depend on the ratio of the detector dynamic range to the precision desired for the final results (which may depend on noise reduction by repeated observations, so fixing the attenuation stepsize is an aspect of system definition).

2.3.8. Spectral discrimination

In present-day operational detectors, in general the output signal does not by itself carry information about the colour (wavelength, frequency, energy) of the incident photons. One obtains such information by spectral discrimination in the optical part of the system.

An exception is the CCD used as an X-ray detector: the highly energetic X-ray photons produce output pulses of many electrons, the size of each output pulse being proportional to the energy of the input photon. This enables one to construct X-ray imagers with some spectral capability inherent in their detectors. A similar development is taking place for the optical part of the spectrum; see "The STJ page". If this effort succeeds in emerging from the laboratory, the development will strongly influence the design of optical instrumentation, leading to more complex electronic backends and an even greater parallel-processing advantage than that of the present array-detector systems.
3. Optical systems: pupils and images

Telescopes and their instruments are particular cases of optical systems (or optical 'instruments' in the general sense). In all optical systems, the twin concepts of image (or 'field') and pupil (or 'aperture') play a fundamental role. We shall examine these concepts in detail, so that in later chapters we can use them without any further explanation.

The purpose of a modern astronomical "telescope" (i.e. without an eyepiece; just the objective lens or mirror system) is to form an image at its focus: a focal-plane representation, in terms of position \((x,y)\), of a celestial scene which we specify in terms of angular coordinates \((\xi,\eta)\): see Figure 3-1.

Such a telescope at some point uses an aperture to define the part of the wavefront actually involved in producing the focal image. For reasons that will become clear, a synonym for such an aperture is 'pupil' and the aperture that puts limits on, or 'defines', the wavefront is the 'defining pupil'. In most optical telescopes, the defining pupil is the primary mirror (exceptions: e.g. a Schmidt telescope, in which the defining pupil is located at the corrector plate and many IR telescopes, where the chopping secondary is the defining pupil; in such cases the primary is oversized to accommodate the finite angular field laid down in the telescope specifications).

If this were all, as indeed it is in a classical 'photographic refractor', the pupil concept would not be so important. However, a modern optical instrument may have several intermediate images before the final image is recorded (example: the VISIR spectrometer (Figure 3-3), which has a reimager before the entrance slit, followed by a collimator, grating and camera combination to form the final spectrally-dispersed image on the detector). Each of the lens or mirror systems which transform one image into the next also produces an 'image' of the previous pupil in the system, i.e. a scaled replica of the defining pupil.
Figure 3-2: A 2-component optical system

A two-element optical system, illustrating the concept of a re-imaged pupil, or 'pupil image'. The location of this pupil is where the beams from ALL image points coincide (exactly in the case of perfect optics). Note that optical elements and images are now represented by vertical black lines, as is customary in optical diagrams. The 'optical elements' may be lenses, mirrors or complex systems. Optical aberrations are no longer visible, since a 'paraxial' approximation was used in the ZEMAX ray-tracing program. The re-imaged pupil is the usual location for the defining pupil in infra-red systems (a 'cold-stop').

Figure 3-3: The high-resolution arm of ASTRON's VISIR, a spectrometer for the thermal IR; see also Figure 7-18

The cold stop is situated within the "re-imager", which comprises the second, third, fourth and fifth mirrors after the telescope focal plane; a filter and calibration component are shown, close to the cold stop.
A simple case is shown in Figure 3-2, where the second lens might be the eyepiece of a visual telescope used for star-gazing; in this case, one would strive to make the replica "image" of the pupil coincide with the pupil of the stargazer's eye, hence the term 'pupil' as a synonym for 'aperture'. Alternatively, the second lens in Figure 3-2 might represent the VISIR re-imager (which is actually an optical system made up of 4 mirrors; see Figure 3-3), in which case the second image would be formed on the slit of the spectrometer and the pupil replica is formed within the re-imager optics system. Often, as in the case of the eyepiece, the second image 'is located at infinity' (a virtual image; location at infinity means that the rays from any single 'point' in it are parallel, just as they were when they struck the telescope objective, i.e. the image is once again specified in terms of the angles ($\xi, \eta$) rather than in linear coordinates ($x,y$)). The lens of the human eye, finally, projects (re-images) the 'image at infinity' on to the retina. The entire arrangement of objective + eyepiece is just a beam reducer to make the wavefront fit the pupil of the eye; the normal human eye is comfortable (at rest, even) with objects at infinity, so the infinite object distance is not changed in the beam reducer. The beam-reducer concept has been discussed above. Since the rays exiting from an eyepiece appear to come from an object at infinity, it is not really a surprise that -- somewhere along that exit beam -- conditions exist that mimic the objective, which also is illuminated by an object at infinity. One may wonder where along the optical axis that replica image of the pupil exists; the answer is: at the distance given by the normal lens equation $1/f=1/u+1/v$, with $f$ being the effective focal length of the lens or other optics, $u$ the distance objective <-- eyepiece and $v$ the distance eyepiece <-- replica pupil image. In fact, the replica pupil is a perfectly normal (in this case real) image of the objective, seen in the light selected by an aperture ('field stop') at the telescope image plane (focal plane): one may, for instance, point the telescope at a bright star, the Moon or daylight sky and examine the dust on the primary mirror by focusing a microscope on a replica pupil further downstream. Virtual images of the pupil can also exist: the secondary mirror of a Cassegrain telescope forms a virtual pupil (virtual image of the primary mirror) at some distance behind itself (further out of the telescope, see Figure 3-4) this distance is a little more than the focal length of the secondary, since the primary is somewhat nearer than infinity. As seen from any point within the telescope focal plane image, the rays appear to come from this virtual pupil replica.

**Note on terminology:** In the optical literature, one often finds the (focal) image plane referred to as 'field' (hence 'field stop', 'field lens'; 'integral-field spectrograph'). In radio synthesis, 'image plane' is usual, with 'field' reserved for the field of view, i.e. the image-plane area covered by the 'primary beam'. Don't let yourself be confused by the term 'images of the entrance pupil'; think of them as 'entrance pupil replicas'.

Field lens = a lens in an image plane; such a lens is used to direct the rays passing through the image into some other optical element. One may also come across 'field mirror'. An infinitely thin field lens re-images the 'image' on to itself, therefore does nothing except bend the rays; the practical case approximates this. See e.g. Heavens & Ditchburn (1991, sections 18.7 'Field' and 18.8 'Relay systems') for more detail.
3.1. The general case

In general, in a complex system consisting of \( N \) optical elements, there are \( N \) focal images, but also \( N \) pupil replicas (a classical periscope [Heavens and Ditchburn 1991, section 18.8 'Relay systems'] is a good example). Both images and pupils can be real or virtual; at a real pupil replica, one may manipulate the beam as if one were doing it at any of the other pupil replicas. Examples of this strategy are:

- Adaptive optics uses a 'rubber' mirror (i.e. one whose shape can be modified) to undo the effects of the atmosphere. Ideally, this mirror is positioned at an image of the atmospheric seeing layer one is trying to neutralise, and the correction mirror may be chosen as the defining pupil for the system.
A 'cold(-pupil) stop' in an infra-red system (e.g. VISIR, Figure 3-3) is used to cut out all stray radiation that bypasses the telescope optics. Such stray radiation is mostly from sky background and telescope structure, therefore approximates 300K black-body radiation; the cold stop trick replaces this by blackbody radiation of a much lower temperature (that of the cryogenic environment of the cold stop, about 35K in VISIR). The cold stop is at a pupil position, but it may or may not be the defining pupil.

3.2. Entrance and exit pupils

The first pupil replica in an optical system is known as the entrance pupil, the last as the exit pupil. In most astronomical telescopes, the primary mirror or objective lens is both the entrance pupil and the defining pupil, but in the two examples above, this may not hold:

- in adaptive optics, the defining pupil may be the rubber mirror, but the (somewhat inaccessible!) entrance pupil location is several pupil replicas upstream, viz. in the atmospheric seeing layer.
- in the infra-red, if the chopping secondary mirror is the defining pupil, then the entrance pupil is virtual, a long way behind the primary mirror and the strict concept of entrance pupil is of little practical use (the "virtual entrance pupil" is where the rays forming the virtual image by the primary mirror appear to come from; this virtual image forms the virtual object for the secondary mirror; see what I mean?). The cold stop then has to be oversized w.r.t. the strict optical replica of the defining pupil and a certain amount of (bright) sky radiation can spill past the secondary mirror, enter the instrument and strike the detector. If, on the other hand, the cold stop is the defining pupil, the telescope will be oversized and one loses some of the sensitivity one has paid for. The compromise, as good as possible an image of the secondary on the cold stop, is extremely sensitive to alignment (subject to flexure and other instabilities) and therefore unacceptable (this is engineering basics, but might not occur to every astronomer).

3.3. Pupil matching

For a system consisting of a telescope and an instrument, the exit pupil of the telescope should coincide with the entrance pupil of the instrument. To calibrate such an instrument from a laboratory source, the exit pupil of the laboratory source must be identical (when the calibration mirror is flipped in) to that of the telescope. The extent to which this is true will determine how well the calibration-light input matches that of operational conditions and thus how well the calibration exposure can be assumed to be equivalent to a science exposure (the ultimate limit to such equivalence is that scintillation produces moving random bright and dark patches on the telescope pupil, which cannot be fully reproduced for the calibration pupil; the only way to circumvent this limit is to calibrate during the science exposure itself, e.g. wavelength calibration of a spectrometer can be done by using a Fabry-Perot etalon as a mirror in the optical train, thereby adding artificial absorption lines to the spectrum being recorded). Telescopes like the original Multi-Mirror-Telescope (MMT) and ESO's VLT (in incoherent-array mode) have multi-aperture pupils and the design of calibration sources must duplicate these. Even single Cassegrain telescopes have a central hole in the pupil (the shadow of the secondary) and a proper calibration source mimics that (strictly speaking: the hole is not located at the telescope exit pupil, but somewhat closer in -- between the exit pupil and the virtual image formed by the primary; the smaller the instrument
field of view, the less this distinction matters. This was a test of your mastery of the concept of pupils and images; come back here when you have read the next section).

The variable-delay-line system of the VLT interferometer includes a variable-curvature mirror, in order to image the system pupil in a fixed place in spite of the variations of delay-line length.

### 3.4. Pupil/image duality

The concept of a pupil becomes meaningful as soon as a telescope or other optical instrument has a finite field of view in its image plane(s); if the image-plane field is limited to a point, the pupil position is indefinite, as will be clear from what follows.

*Figure 2 adapted to show that rays passing through any one point in the pupil contribute to every point in the image, just as any one point in the image is illuminated by all points in the pupil.*

*Figure 3-5: Similar to Figure 3-2, except that 3 rays passing through 1 point in the pupil have been marked (thick lines)*

For an unobstructed optical train, each point in the image is illuminated from the entire pupil and conversely from each point in the pupil rays pass through all points in the image. By taking away (blocking) part of the pupil, one reduces the illumination at all points in the image; thus one may intensity-modulate the entire image in synchronism by mounting a modulator exactly at a pupil (this holds even when different parts of the pupil are not modulated in exactly the same way; at each moment, the modulation at all points in the image is the beam-area-weighted average of the modulation at the pupil). A point to remember, though, is that position \((x,y)\) in the image is equivalent to angular direction \((\xi,\eta)\) in the pupil; therefore, components for which the action depends on the angle of incidence (such as interference filters) will have position-dependent effects in the image. This question is discussed in a later section.

The images and the pupils of a system are the places where the light beams are well organised and placing a 'stop' (=aperture) has well-defined effects. At all other positions, blocking part of the beam will block part of the image, but in an 'unsharp' way (the nearer the blocking action is to the pupil, the hazier the result in the image will be). This unsharp projection of a blocking component into the image is called 'vignetting' (vignetting is basically undesirable, but may be accepted as the price to be paid for some other advantage, since it can be undone -- disregarding signal-to-noise considerations -- by flatfielding the image during subsequent processing).

Dust on the optics is, of course, a beam-blocking agent as discussed above. Dust on optics at a pupil affects all parts of the image equally, dust at an image blocks just that part of the image,
while anywhere else the effect of the dust is that of very local vignetting. This can be a consideration in deciding where to place optics such as filters.

One rather specialised use of the word 'pupil' should be mentioned. As explained in the section on Fourier Transform (FT) pairs, the spatial autocorrelation of the field distribution in a pupil plane and the intensity distribution in an image plane are related by a FT. This fact is used in all aperture-synthesis telescopes (mainly radio, but optical/IR has started); the image in this case is the sky, at infinity, so 'position' within the image is in angular measure. One measures samples of the correlation in the pupil plane (often called the 'uv-plane') and computes the sky intensity (or, for polarization, Stokes parameter) distribution (this is, of course, the 'object', or the 'zeroth' image 'formed' by the 'telescope', at infinity). The 'pupil plane' in this sense is closely related to the optical-system analogue, but the set of N pupils and images has shrunk to N=1 for pupils and N=0 for images.

Because of the FT relation, controlling the sensitivity distribution in a pupil plane influences the point-spread-function (PSF); this holds for all pupils and images in a complex optical system. In radio technology this pupil-plane manipulation is referred to as control of the pupil illumination (the term originated in transmitter-receiver duality), in the optical domain it is called apodization. More detail in the section on FT pairs.

### 3.5. Image-to-pupil conversion

Occasionally one uses an image, or part of it, as an entrance pupil for a succeeding part of the system. Two examples:

- When using an optical fiber to link the telescope to an instrument, it is desirable (in order to stabilise against tracking errors of the telescope, atmospherically-determined image motion etc.) that the configuration of input rays to the fiber is no longer recognisable at the output end: the rays should be scrambled geometrically to yield a more or less constant output configuration for variable input configuration. A long and rough-walled fiber does this automatically, but for good optical efficiency and blue/UV transmission the fibers are kept short and smooth; this contradiction is resolved by the optical scrambler shown in Figure 3-6 (in which the reader will by now recognise the geometry of image-to-pupil conversion); it can be a perfect scrambler if its optical design is ideal -- a physical impossibility, of course: basic engineering again. More sophisticated designs than that of Figure 3-6 exist, with a lens on each of the fiber faces, but the function is the same as for any pupil/image exchange: turn positions into angles and vice-versa.

![Figure 3-6: Pupil-to-image (or v.v.) converter used in a fiber scrambler](image)

- In an "integral-field spectrograph", an image in the telescope focal plane is formed first. Then parts of that image are selected by a lenslet array (or similar arrangement) and are focused to small patches of light which serve as input 'slits' to a spectrograph. So the
telescope image is dissected by the pupils of the lenslet array, which forms multiple (tiny) images (of the telescope pupil) near the array focal plane (why 'near' ??). The geometry of the rays in the multiple images by the lenslet array therefore resembles that at the pupil of the telescope and is independent of the relative intensity distribution in the image formed by the telescope: one obtains average (intensity-weighted) spectra of the localised portions of the celestial extended source. This arrangement is similar to that of the photo-electric stellar photometer discussed below. The difference is in the astronomical image, which goes all the way to the edge of the selecting aperture (the lenslet) and therefore is not fully stabilised against atmospheric image motion. One may configure the lenslet array for non-overlap of the resulting spectra (e.g. by rotating the array w.r.t. the dispersion direction) or use a (probably fiber) converter to go from a matrix to a long-slit arrangement; these are technical details that depend on (e.g.) the wavelength range of interest. For details of a modern instrument of this kind, look at the OASIS description (click to 'TIGER mode') and its optical diagram (reproduced in our Figure 7-19).

3.6. Image or pupil on the detector?

An image in the focal plane of a telescope moves; this is an inescapable fact of life below the atmosphere. Detectors have non-uniform response, that is another fact of life. Combined, these lead to photometric uncertainties if one places an image on the detector. Sometimes, e.g. in a spectrograph, that is more or less inevitable and one accepts either this (spectro-)photometric instability or the loss of light at a stabilising aperture, such as a slit. In a classical photo-electric single-source photometer, however, one is free to place a pupil on the photomultiplier detector, merely selecting a single star (plus some background sky) in the image plane. The larger this image-plane diaphragm is, the greater the stability against the effects of image motion, but the larger the background signal; thus the need for a selection of diaphragms in a photometer: large for high-accuracy work on bright stars, small for a faint limiting magnitude. If it is impossible to locate a pupil replica on the detector, then one should take other measures to optimise according to circumstances; for CCD spectro-photometry of the highest photometric accuracy, I discussed such questions in papers at a specialised conference some years ago. Optical fibers (either singly or in various kinds of arrays) are powerful components for producing the conditions required by (unavoidably imperfect) detectors; the image-to-pupil converter is a good example of this.

3.7. Narrow-band filters at pupil or image?

As mentioned above, a position (x,y) within an image is equivalent to an angle (ξ,η) at a pupil. Therefore, for an interference filter or other Fabry-Perot etalon, the finite-field-of-view performance as an optical component will depend primarily on its spatial homogeneity if it is placed at an image position, or on its angle-of-incidence characteristics if placed at a pupil position.

In the latter case, spatial inhomogeneity is of secondary importance. Inhomogeneity of the (angle-of-incidence characteristics of the) filter can only enter into the measurement if scintillation (or for instance obscuration of the telescope aperture by the dome) causes the effective pupil to be highly inhomogeneous. So, if dependence of filter passband on position within the image is inherently negligible (as for many absorption filters) or can be tolerated, a pupil is the preferred
position for filters. Fabry-Perot instruments such as TAURUS (Figure 7-8) are the most extreme examples of narrowband filter imagers with the filter at a pupil position. The gratings or prisms of spectrographs are always placed at pupils, for in their case the dependence of passband on position in the (final, dispersed) image is the object of the exercise rather than an unpleasant surprise.

In a telecentric beam, the ray bundles for each point in the field seem to come from an infinitely distant pupil (the virtual image of the pupil shown in the figure). Hence the name: they come from a 'remote centre'.

Figure 3-7: The telecentric beam

In the case of a filter at an image position, angular characteristics can in general still affect the performance, because the beam for each image point appears to originate at the exit pupil of the preceding optical system and this pupil in general is at a finite distance, therefore is viewed at different angles from different points in the image (Figure 3-2). In a telecentric beam (Figure 3-7), this pupil is at infinity and for all image points the beam fills exactly the same spread of angles of incidence. A filter position in a telecentric beam allows one to average out to some extent any spatial inhomogeneity of filter transmission and bandpass without introducing systematic effects due to angle-of-incidence characteristics of the filter. The further from the image the filter is located, the better the averaging will be, but the larger the filter will have to be constructed; that, however, is detailed engineering.

N.B. A telecentric beam is also used in other applications. In metrology, its virtue is that slight defocus (e.g. because of depth structure of the object being measured) does not introduce errors into the measurements of position which are the objects of the exercise. Figure 3-8, taken from a technical advertisement, refers to this application.

3.8. How small can a pupil be?

Whenever an object (which may be an image formed by a previous stage of the optics) is imaged by an optical system, such as a lens, the angles of incidence of individual rays are scaled inversely proportional to the magnification (or de-magnification: m<<1): the smaller the image, the larger the angles of incidence of the rays forming it. This is a fundamental difficulty in recording large angular fields of view, as imaged by large telescopes, with small detectors such
as CCDs: the final camera will have to have an incredibly high numerical aperture (low F-ratio) or, in the hypothetical limit, the rays will strike the CCD at grazing incidence.

![Telecentric Lens](image)

**Figure 3-8: You can now read this advert AND understand it!**

There are 3 pupil replicas in this system; where are they?

The same holds for the particular set of images we have called 'pupil replicas', there is nothing special about these images except their function. So the smaller the magnification between a given pupil and another one, the larger will be all the angles of incidence at the 'other' pupil. Since we put optical components at some pupil positions, we have to ask ourselves what angles of incidence these optical elements will tolerate without unacceptable optical aberrations of one sort or another (the exact meaning of 'unacceptable' depending on the application). An example mentioned above is formed by interference filters: their passband broadens and shifts with angle of incidence of the rays passing through them and the application for which the astronomer intends to use them determines the maximum wavelength shift or passband broadening that can be tolerated. For a specified telescope size and field of view, this determines the minimum size such a filter can have, or, conversely, for a specified filter size it determines the field of view that can be imaged with the passband and telescope size the astronomer wants. If the 'optical component' at the pupil position is just an aperture, the angle of incidence limit is rather slack at $90^\circ$; while this may be so, one is never content to just pass rays through an aperture, they will have to be accepted by some further optics, which can generally not accept more than a few tens of degrees -- if that -- and it is in fact the acceptance angles of these further optics which will determine the minimum pupil size at that position (again, the 'acceptance angle' is of course determined by the acceptable level of aberrations, as derived from analysis of the astronomer's system requirements).
4. The Ubiquity of the Fourier Transform

An important resource in understanding astronomical instrumentation is the theory of the Fourier Transform (FT; this abbreviation will be used both as a noun: "a FT" and as a verb: "to FT"). This mathematical construct relates a function of time to its spectrum, and also the EM field distribution within a pupil to that within an image or the antenna pattern of a telescope to the size and shape of its aperture (pupil). Through a theorem on the FT of the 'auto-correlation function', it also relates the 'power spectrum' of a signal to the auto-correlation of the time function and equally the power distribution within an image to the (spatial) auto-correlation of the EM field within the pupil. The latter case is the basis of aperture synthesis, used almost exclusively when very large 'telescopes' are required (the "zeroth image" of the sky by the telescope is the sky itself, at infinity; the entrance and defining pupils are both identified with the array of antenna elements on the ground). The VLT application to spatial interferometry (VLTI) is the optical equivalent of radio systems like VLBI and the VLBA. 'Fourier spectrometers' (optical) and 'auto-correlation receivers' (radio) are applications of the time-function variant of the auto-correlation theorem for the FT.

Because of its overwhelming importance for all (radio and) optical astronomical instrumentation, the FT will be treated in some detail as a discipline in its own right. The treatment will be minimally mathematical, but complex numbers and the integral as a kind of summation are necessary tools. Hamaker (sections 4 and 26) treats complex numbers; integrals are introduced in Appendix A.

The basic textbook for astronomical applications of the FT is that by Ron Bracewell. The following review I wrote for another purpose, but it will serve here, too:

Ron Bracewell's book (ISBN 0-07-007015-6) really is superb. If you can afford only one book on instrumentation, buy this one. It exists in several editions (1965,1978,1986). To start with, any of these editions will do, as the first 17 chapters are identical in all 3 and contain all of the basic material. So don't wait for the most recent edition to turn up at your library, but get cracking using any copy at hand.

In the second edition (1978), a 29-page chapter on the Discrete Fourier Transform was added, a basic manual for the F(ast) F(ourier) T(ransform) routines that are used in actual computation. Also added was a section of supplementary problems, among which are "several that go beyond being mathematical exercises by inclusion of technical background or by asking for opinions". In the revision of 1986, Bracewell added a 26-page section on the Hartley Transform, a double-sided real transform that can (and does, with a gain in efficiency) substitute for the FFT in machine calculations. A 3-page biography of Fourier is the final and very readable addition to this revised edition.

In his Preface to the first edition, Bracewell indicates that the basic sequence of chapters is 1, 2, 3, 4, 5, 6, 8, with subsequent branching into any of the other chapters (there is some slight detail in the branching, do look for yourself). He suggests that "it is desirable to illustrate the theorems and concepts by dealing simultaneously with a physical topic......for which the student already has some feeling". In our case, this means that you take a subtopic that interests you and of which you have some expert knowledge (hi-fi audio, synthesis mapping, FT-spectrometer = Michelson interferometer = autocorrelation-spectrometer backend, gratingspectograph, carrier-wave telephony, atmospheric seeing,digital sampling, or whatever takes your fancy) and think up your own examples and applications as you go through the mathematics.

At all stages of mastering the FT via Bracewell's book, do examine the graphical representations he offers and make his technique of splitting a problem into manageable chunks your own (a good example: a diffraction grating -- or the Westerbork radio array -- is represented by a groove profile -- Westerbork: primary beam -- , convolved with an infinite-extent comb (Sha) function, the result multiplied by a top-hat function. So the FT is ........ multiplied by ........, finally convolved with .......; well, I'll stop,
you're in the driving seat now). These habits are essential to all instrumental astronomy (and, of course, elsewhere, but that is beyond my responsibility).

If you need more mathematical background on the Fast Fourier Transform and/or practical details of computation, one of the texts you could try is:


4.1. Areas of application of the F.T.

In astronomy, the FT is used for the design of instrumentation in 2 main areas of application. In the following subsections, I discuss how the physical situation in each of these 2 cases is naturally described by a mathematical formula that can be recognised as a form of the FT.

4.1.1. A signal and its amplitude/phase spectrum

An audio signal (sound wave or electronic signal) consisting of a single tone, periodic (repeating) but non-sinusoidal, may be represented by the sum of a fundamental tone and certain amounts of its 'harmonics'; the relative proportions of such harmonics determine the characteristic sound (timbre) of a musical instrument.

$$ A(t) = \sum_{n=1}^{N} a_n \cdot \cos(n.2\pi \nu \cdot t + \phi_n) $$

- $\nu'$ = fundamental (Du: grondtoon)
- $n\nu'$ = $n^{th}$ harmonic (Du: boventoon)

The set of $a_n$ is usually called 'the spectrum' of the signal (more accurately, 'the amplitude spectrum'), while the set of $\phi_n$ are 'the phase spectrum'. These spectra consist entirely of values at the frequencies $n\nu'$.

We now stretch $A(t)$ to (almost) infinite length and finally consider just one single cycle, so that we obtain an (almost) non-periodic (i.e. non-repeating) signal. As we do so, the value of $\nu'$ decreases and to compensate we increase N (almost) without limit. The amplitude and phase spectra now begin to look like continuous functions, so we pretend they are and then we split them into an (almost) infinite number of very thin slices $\delta\nu'$, each centred on a value of $n\nu'$.

Next, we change our notation, so that:

$$ n\nu' \to \nu $$
$$ a_n \to a(\nu).\delta\nu $$
$$ \phi_n \to \phi(\nu) $$
Finally, as in Appendix 6 we go all the way to the limit of $\delta N \to 0$, $N \to \infty$ and we change our notation a bit more to indicate we have done so: $\sum_{i=1}^{N} \to \sum_{i=0}^{\infty}$ and $\delta N \to dN$

We thus obtain:

$$A(t) = \int_{0}^{\infty} a(\nu).\cos[2\pi\nu t + \phi(\nu)].d\nu$$

(b)

which is the equivalent of the summation (a), for a non-periodic signal.

Now see Hamaker sections 4 and 26 if you are not familiar with complex numbers.

Using

$$e^{ix} = \cos x + i \sin x,$$

and therefore

$$e^{-ix} = \cos x - i \sin x,$$

N.B.: Some engineering texts use $j$ rather than $i$ for $\sqrt{-1}$.

we substitute for $\cos x = \frac{1}{2}(e^{ix} + e^{-ix})$ and obtain (temporarily using $\nu''$ for $-\nu$ and $\phi''$ for $-\phi$):

$$A(t) = \frac{1}{2} \left[ \int_{0}^{+\infty} a(\nu).e^{i[2\pi\nu t + \phi(\nu)]}.d\nu + \int_{0}^{+\infty} a(\nu).e^{-i[2\pi\nu t + \phi(\nu)]}.d\nu \right]$$

(c)

$$= \frac{1}{2} \left[ \int_{0}^{+\infty} a(\nu).e^{i[2\pi\nu t + \phi(\nu)]}.d\nu + \int_{0}^{-\infty} a(\nu'').e^{i[2\pi\nu'' t + \phi''(\nu'')]} - d\nu'' \right]$$

$$= \frac{1}{2} \left[ \int_{0}^{+\infty} a(\nu).e^{i[2\pi\nu t + \phi(\nu)]}.d\nu + \int_{-\infty}^{0} a(\nu'').e^{i[2\pi\nu'' t + \phi''(\nu'')]} .d\nu'' \right]$$

These 2 integrals have the same form, except for the limits of integration and the variable of integration. One may conjecture that they can be combined into:
\[ A(t) = \frac{1}{2} \int_{-\infty}^{+\infty} a(v) e^{i[2\pi v t + \phi(v)]} dv \]

\[ = \int_{-\infty}^{+\infty} \frac{1}{2} a(v) e^{i\phi(v)} e^{i2\pi vt} dv \]

\[ = \int a(v) e^{i2\pi vt} dv \] 

(d)

where \( a(v) \) is now a complex quantity; \( a \, dv \) or \( |a| \, dv \) represents the amplitude of the phasor \( a \, dv \) and \( \phi \) is its phase, both of them being functions of the frequency \( v \). The product \( a(v) e^{i2\pi vt} \) is again a complex quantity. One may separately integrate its real and imaginary parts and combine them into yet another complex quantity \( A(t) \); this is what the shorthand form d) implies.

N.B. 1: In general, for arbitrary \( a \), \( A \) would be complex, but in the present example \( A \) is real because we defined it as such in (b); the fact that \( A(t) \) is real means that in this case \( a(v) \) has certain symmetry properties which I shall not derive. I could have started with a representation of the signal as a sum of complex quantities

\[ A(t) = \sum_{n} a_n e^{in2\pi vt} \], in which case \( a \) would in general have lacked symmetry. In electronic engineering, one may actually represent the signal itself by a complex quantity; the so-called quadrature signal (imaginary part) can usually be constructed physically (e.g. by using 2 LO signals in quadrature during conversion to IF). In optical, non-heterodyne, practice this is not so easy and the phasars that we are familiar with should be interpreted as an abstract representation of the real-life signal; one takes the real part of a phasor (or a sum or integral of phasors) to obtain the actual signal. In that case, multiplication of phasors is not permitted; one should take real parts first (summation, integration and multiplication by a complex constant are permitted, they are linear processes).

N.B. 2: Between version (b) and (d) the range of integration has changed. However, (b) and (d) are equivalent; version (d) is generally preferred, because it is more compact and more symmetrical.

Apart from a few constants, the last form d) of our non-periodic signal is the so-called 'plus-i Fourier Transform'. \( a(v) \) represents amplitude and phase of the 'spectral component' at 'frequency' \( v \). The set of values \( a(v) \) is also called the 'spectrum', in the audio-engineering sense (this 'audio-engineering' concept of a spectrum that has both amplitude and phase applies to all electrical signals within electronic instruments; in astronomy, a 'spectrum' is something else, which we shall meet later, the 'power spectrum').

Fourier Transform (FT) theory states that we can invert this relation to obtain \( a(v) \) from \( A(t) \). This means that \( A(t) \) and \( a(v) \) are entirely equivalent and that we may manipulate the signal in either the time domain or the frequency domain (provided we can readily compute the transformation from one to the other; computing power is essential for practical applications). This is what makes the FT so useful in computer-aided instruments.

The generalised form of the FT we shall use (following Bracewell's notation; both \( f(x) \) and \( F(s) \) are in general complex quantities) is the following:
If

\[ F(s) = \int_{-\infty}^{+\infty} f(x)e^{-2\pi i sx} \, dx \]  

("minus-i transform")

then

\[ f(x) = \int_{-\infty}^{+\infty} F(s)e^{2\pi i sx} \, ds \]  

("plus-i transform")

We notice that, with

\[ A \leftrightarrow f \]  
\[ a \leftrightarrow F \]  
\[ t \leftrightarrow x \]  
\[ \nu \leftrightarrow s \]

our integral (d) indeed has the form of the plus-i FT.

x and s are called 'conjugate variables' and one speaks of representation in the x-domain when one uses f, conversely of the s-domain when F is used.

### 4.1.2. The EM field distribution within a pupil and an image

Before using the generalised notation to go into the useful properties of the FT, we examine one more area of application, the theory of antennas (single dishes and tied arrays). In single dishes, the EM field entering the pupil plane is redirected by the mirror on to the focal-plane antenna; the vector sum of all the pupil-plane phasors is automatically formed at this focal-plane antenna. In a tied array (see [Hamaker](#) section 7), the addition of the EM field components is done via electronic circuitry (usually after conversion to intermediate frequency). Here we examine the EM field in the pupil plane and do the addition algebraically; the components are phasors and complex-number algebra is used. The "total signal" we shall derive is entirely equivalent to that measured at the focus of a single-dish telescope or at the output of the electronic adder of a tied array.

**N.B.:** What happens in a single-dish radio antenna is really no different from what happens in an equivalent optical system (single lens or mirror). Parallel rays of 'light' are collected by the 'pupil' and brought to a 'focus', i.e. they are directed to one single point (along paths of equal length) and the EM field phasors are then summed at this focus. As we shall see below, the only distinction between radio and optical is in the ratio between pupil size and wavelength of the 'light'; this is the root cause of being 'diffraction-limited' and having appreciable 'sidelobes' in the radio case, while the optical regime generally lacks these properties (but beware active optics and/or the far infrared!) .

Consider the situation shown in Figure 4-1: a plane wave (from an object at infinity) of strength \( E.e^{i2\pi \nu} \) strikes the pupil plane of a telescope at an angle of incidence \( \theta \). We wish to determine the signal at the 'focal point' of the 'telescope' represented by this pupil and we want to know how this signal depends on the direction \( \theta \): we wish to determine the 'antenna pattern' of the
telescope. At any point \( x \), the phase of the wave (relative to \( x = 0 \)) is \( \frac{-2\pi}{\lambda} x \sin \theta \). So the contribution to the signal, from the direction \( \theta \), by an infinitesimal slice \( dx \), is:

\[
E. e^{i2\pi\nu t} \cdot e^{\frac{-i2\pi}{\lambda} x \sin \theta} \cdot dx
\]

If the pupil has non-uniform transmission \( F(x) \), then the total signal from the direction \( \theta \) accepted by the pupil is

\[
E_{\text{total}}(\theta) = E. e^{i2\pi \nu t} \int_{-\infty}^{+\infty} F(x) e^{\frac{-i2\pi}{\lambda} x \sin \theta} \cdot dx
\]

which we can rewrite as

\[
E_{\text{total}}(\theta) = \lambda. E. e^{i2\pi \nu t} \int_{-\infty}^{+\infty} F\left(\frac{x}{\lambda}\right) e^{\frac{-i2\pi}{\lambda} \left(\frac{x}{\lambda}\sin \theta\right)} \cdot d\left(\frac{x}{\lambda}\right)
\]

\[
= \lambda. E. e^{i2\pi \nu t} \cdot f(\sin \theta)
\]

Apart from a multiplying constant, this a minus-i FT between \( F(x/\lambda) \) and \( f(\sin \theta) \). In other words: a FT relates the "antenna pattern" \( f(\sin \theta) \) to the pupil transmission function \( F(x/\lambda) \). This is a very useful result, because it will allow us to use all the properties of the standard mathematical FT to predict the antenna pattern for any \( F(x/\lambda) \) we use in our system. The fact that \( x/\lambda \) is one of the conjugate variables of the FT means that all pupil-plane dimensions should be measured in units of \( \lambda \): an optical and radio pupil of the same size \( D/\lambda \) (i.e. pupil size expressed as a multiple of the respective wavelength) will have the same response \( f(\sin \theta) \) to radiation from directions \( \theta \).

Question: The Dwingeloo 25-m dish has (at 21 cm) a main beam of about 0°.5 and sidelobes at angles of 1 to a few degrees. What size would an optical telescope have to be \((\lambda = 0.5\mu m)\) for it to have the same sidelobe structure? What size is the theoretical main beam of the VLT 8-m dish at \( \lambda = 0.5\mu m \)?

The above presentation was for a 1-dimensional pupil. Most pupils have 2-dimensional extent. The FT then becomes a double integration, using 2 pairs of conjugate variables. This is more complicated, but there is absolutely no difference of principle.

Good examples of 2-D pupils and their FTs may be found in Reynolds et al: "The New Physical Optics Notebook: Tutorials in Fourier Optics", ISBN 0-8194-0130-7. Fig 3.3 on page 19 (reproduced here as Figure 4-4) is a demonstration that needs no explanation once you have mastered what is in the present document.
4.2. Basic properties of the Fourier Transform

As mentioned above, a convenient generalised form of the FT relations is that used by Bracewell throughout his book:

If

\[ F(s) = \int_{-\infty}^{+\infty} f(x).e^{-2\pi i sx}.dx \]  

("minus-i transform")

then

\[ f(x) = \int_{-\infty}^{+\infty} F(s).e^{2\pi i sx}.ds \]  

("plus-i transform")

Since Bracewell's book is widely available and can hardly be improved on, I shall quote from that book and refer you to it. The basic formulae shown here are found on page 7 of Bracewell, together with some alternatives that you may encounter in other books or papers.

4.2.1. Some often-used FT pairs

I shall not derive any FTs, but they can be found in Bracewell. In this section I shall mention a few pairs that are often found in technical literature. Refer to Bracewell p. 100, fig 6.1 (copied here as Figure 4-2) and to his 'Pictorial Dictionary of Fourier Transforms" (p 359 in the first edition, p 385 in the second, and p 411 in the "second edition revised). Bracewell shows imaginary functions as dashed curves in his figures; a complex function has real (solid curve) and imaginary (dashed curve) components.

The Gaussian \( e^{-\pi x^2} \) transforms into 'itself', i.e. \( e^{-\pi x^2} \). It is a function that has its maximum at 0, then decreases monotonically to 0 at \( \pm \infty \), without any 'sidelobe' structure. This function is often used in analysing complex spectral line profiles, such as that of the 21-cm hydrogen line.

The 'top-hat' or rectangle function \( \Pi(s) \) (=1 from –0.5 to +0.5, =0 elsewhere) transforms into the function called 'sinc(x)', i.e. \( \frac{\sin x}{x} \) (x measured in radians). The function sinc(x) also has a central maximum, after a while goes through 0 at both positive and negative x, then becomes negative and alternates between positive and negative at ever decreasing amplitude as \( x \to \pm \infty \).
Squaring $\text{sinc } x$ converts all the 'sidelobes' to positive values; its FT is the triangle function (drops linearly to 0 for both positive and negative $s$, from the central maximum). This is an illustration of the **auto-correlation theorem mentioned below**: if 2 functions are each other's FT, then the square of the modulus of the one FTs into the auto-correlation of the other.

If one takes the Gaussian and expands it without limit, in the $x$-direction, so that only the central value remains ($f(x) = \text{constant for all } x$), then its FT contracts to something that only has a non-zero value at $s=0$ and is called the 'impulse' function $\delta(s)$. This is an illustration of the **similarity theorem mentioned below**. By convention $\delta(s)$ is the 'unit impulse'; its integral is 1 (i.e. it is infinitely narrow, infinitely tall, the product of height and width being 1 and its 'shape' undefined). One uses the notation $a\delta(s)$ to denote an impulse ('point source') of strength $a$. 

---

**Figure 4-2: Buy Bracewell!**
The 'shah' function \( \text{III} \) (an infinite array of unit impulses; \( \text{III} \) is a Russian character) has itself as

the FT. Bracewell calls it the 'sampling function', because multiplying an arbitrary function \( f \) by

it produces an infinite number of equally-spaced samples of \( f \). He also calls it the 'replicating

function' because the mathematical operation of convolution with a function \( f \) produces an

infinite number of replicas of \( f \). See Figure 4-3 and Hamaker, sections 9 and 10; also Bracewell

fig 5.4 on p 77-79 and the Dictionary (p 359 in the first edition, p 385 in the second, and p 411 in

the "second edition revised").

Next is \( \cos \pi x \) (the pure sinusoid, the basic periodic function without any harmonics), which FTs

into 2 impulses at \( s = \pm 0.5 \). The fact that the 1 physical frequency is shown as 2 impulses in the

spectrum is a consequence of splitting the cosine into 2 complex exponentials and changing the

range of integration from \( 0 \rightarrow +\infty \) to \(-\infty \rightarrow +\infty \).

Finally, to emphasize that FTs of real functions are not all real themselves: the FT of \( \sin \pi x \)

consists of 2 imaginary impulses, a positive one at \(-0.5\) and a negative one at \(+0.5\). The FT of

\( \sin \pi x \) may be derived from that of \( \cos \pi x \) through the 'shift theorem' (Bracewell p 104; I shall not

describe it here) or the 'derivative theorem' (Bracewell p 117; similarly not mentioned here).

N.B. The 'sharp corners' on the outline of a function translate into relative preponderance of high

frequencies in its FT. For instance, the FT (sinc) of the top hat (\( \Pi \)) has many 'sidelobes', whereas

the utterly smooth Gaussian has no sidelobes at all. This is one reason for tailoring the

'illumination' of a radio dish by its feed antenna or 'apodisation' of a grating in a spectrograph.

One strives to minimise the effective extent of the FT and, in particular, to avoid sidelobes.

Two-dimensional equivalents are shown in Figure 4-4.
Note that the FTs have a 'spoke' perpendicular to every sharp edge in the aperture pattern and that large (or small) aperture dimensions lead to narrow (or wide) maxima in the spokes.

4.2.2. Addition

The FT of the sum of 2 functions is the sum of their FTs. This follows from the fact that the integral is itself a modified summation. The addition rule helps in visualising the FT of a function for which one knows the FTs of the constituents (Bracewell fig 6.5 on page 105).
4.2.3. Scaling (similarity; uncertainty relation)

If one contracts one of the functions of a FT pair by a certain factor, then the other expands by the same factor. Bracewell: "If \( f(x) \) has the FT \( F(s) \), then \( f(ax) \) has the FT \( |a|^{-1} F(s/a) \)" (Bracewell fig 6.2 on page 102). If one speeds up an audio tape (i.e. shortens the tape) by a factor 2, all the frequencies are a factor 2 higher, i.e. the spectrum has spread by a factor 2. The Bonn 100-meter dish sees 4 times sharper detail than the Dwingeloo 25 meter dish. The beam of the primary elements in SKA must cover most of the sky (≈ 1 radian). Hence they should be of order 1 wavelength in size, no larger. One recognises the 'uncertainty relation' discussed by Hamaker (section 10.2); such a relation is a consequence of the fact that the physical situation is properly described by a FT.

4.2.4. Convolution or smearing; the point spread function

The convolution of two functions \( f(x) \) and \( g(x) \) is another function \( h(x) \) defined by:

\[
h(x) = \int_{-\infty}^{+\infty} f(x') g(x-x') dx'
\]

It is written symbolically as \( h(x) = f(x) \ast g(x) \). \( \ast \) is the 'convolution operator'. To understand what this integral does, observe that it is very much like a (weighted) running average, which reads:

\[
h_{\text{w.a.}}(x) = \int_{-\infty}^{+\infty} f(x') g(x'-x) dx'
\]

![Figure 4-5: The convolution and the weighted average](image)

In fact, if \( g(x) \) is a symmetrical function of \( x \), there is no difference.

The convolution theorem (Bracewell p. 108) states that "If \( f(x) \) has the FT \( F(s) \) and \( g(x) \) has the FT \( G(s) \), then \( f(x) \ast g(x) \) has the FT \( F(s)G(s) \)". In other words: convolution of two functions is equivalent to multiplication of their transforms; this is why convolution is so important in astronomy (and signal processing in general). Acoustic or electronic signals can be filtered, i.e. their spectrum can be multiplied by a filtering function. This action is expressed as a convolution in the time-domain. Similarly for a source distribution in the sky, the so-called spatial frequencies from which it is constituted and the filtering by the antenna beam of the telescope or array.
The simplest to visualise is when $f(x)$ is a collection of point sources in the sky or a collection of infinitely sharp spectral lines and $g(x)$ is the antenna beam of the telescope or, in the case of a spectrum, the passband of a filter that one tunes through the spectrum. In these cases we can visualise what the convolution means, as long as the point sources or sharp lines are far apart: the convolution 'smears' each impulse into a replica of the convolving function $g(x)$. For this reason the convolving function is called the 'point spread function' (p.s.f. or even psf); it specifies how "the system" responds to a "point source input". To help in this, consider a single impulse at $x=x_0$, i.e. $f(x) = \delta(x-x_0)$. Then $f(x)g(x-x') \rightarrow \delta(x-x_0)g(x-x')$.

For any value of $x$, $h(x) = \int_{-\infty}^{\infty} f(x')g(x-x')dx'$ and the only contribution to this integral comes from $x'=x_0$; this contribution is therefore $g(x-x_0)dx'$: At $x=x_0$, $h(x) = g(0)$. At $x<x_0$, $h(x) = g(x-x_0)$, with $x-x_0<0$. $h(x)$ is therefore a replica of $g$, centred on $x=x_0$ and it is NOT a mirror image. Now see Bracewell chapter 3 p. 24 for much more detail.

In digital signal processing, convolution is a computation that is required often. One can compute the convolution directly. However, since it is often much easier to work in one domain than in the other, one may wish to go by way of FT, then multiplication, then FT back again (whether this is so depends on the number of data points needed to specify a particular function or its FT; it is for instance, much simpler and cheaper to multiply by a top-hat than to convolve with a sinc function).

When trying to obtain an insight into the shape of the synthesized beam of an interferometer array like Westerbork, or the psf of an optical grating spectrograph, the convolution is of great value. This is discussed in a separate section.

### 4.2.5. Auto-correlation

A central feature of FTs in astronomy concerns the 'power spectrum' of a signal, or the 'power image' of the sky. Astronomical signals (next section) are noise-like, i.e. absolute phase cannot be defined for them. For that reason, the amplitude/phase spectrum, or the amplitude/phase sky image, is no more than an abstract construct; it is irrelevant to the physical situation. Of much greater significance is the flux of energy in the astronomical signal, and in particular how that energy flux is distributed over the EM spectrum ('power spectrum') or distributed over directions in the sky ('power image'); in astronomical parlance, the word 'power' is omitted, but one should always realise that it is implicit. This question is discussed more fully in a separate section.

The autocorrelation theorem of FTs (Bracewell p 115) states that "if $f(x)$ has $F(s)$ as its FT, then the auto-correlation of $f(x)$ has the FT $|F(s)|^2$". $|F(s)|^2$ represents the power (amplitude squared) in the signal, at the frequency or from the direction represented by $s$. So the theorem states that, for the 2 applications above, spectra and images, the power FTs into auto-correlation. As astronomers, we are interested only in the power $|F(s)|^2$ (without any interest in absolute
phase), so we can choose to measure the auto-correlation instead and get all we want. The formula for the auto-correlation of \( f(x) \) (which itself is a function of \( x \)) is (Bracewell 40 et seq.):

\[
\int_{-\infty}^{+\infty} f^*(u) \cdot f(u + x) \, du
\]

**Note:** This is the general formula for the auto-correlation, when the signal \( f \) is complex. In electronic engineering, this is often the case; the so-called quadrature signal (imaginary part) can usually be constructed (e.g. by using 2 LO signals in quadrature during conversion to IF). In optical, non-heterodyne, practice this is not so easy and the phasars that we are familiar with should be interpreted as an abstract representation of the real-life signal: one takes the real part of a phasar (or a sum or integral of phasars) to obtain the actual signal. In that case, multiplication of phasars is not permitted (summation, integration and multiplication by a complex constant are permitted, they are linear processes), one should take real parts first; \( f \) in the formula above is then real and the \(*\) in the formula disappears. The important feature of the formula in all its re-incarnations is the multiplication: 'correlation' is more or less synonymous with 'multiplication'.

If from the signal \( f(x) \) we can somehow manufacture both \( f^*(u) \) and \( f(u + x) \), and integrate their product over the u-domain (or in practice over "the most important part" of it), we can by FT derive the \( |F(x)|^2 \) that we are interested in. This leads to the auto-correlation spectrometer as an alternative to the filter-bank (radio) or the FT-spectrometer as an alternative to the grating spectrometer (optical/infra-red), and to correlation interferometers (radio; Hamaker sections 12, 13) and their optical/infra-red equivalents (e.g. VLTI); these instruments will also be discussed more fully in a separate section. The important thing to remember from the present section is that, having established that FT relations exist between:

1. an electronic signal and its amplitude/phase spectrum
2. the EM field distribution in a (telescope) pupil and that in an amplitude/phase image (of the sky)

no further proof is necessary to deduce that FT relations exist between:

1. the (temporal) auto-correlation of an electronic signal and its power spectrum
2. the (spatial) auto-correlation of an EM field distribution in a (telescope) pupil and the power image (of the sky)

To implement an autocorrelation instrument, we have to carry out a multiplication and an integration. The multiplication in radio instruments is carried out digitally (1 or 2 bits, mostly) after conversion of the signal to intermediate frequency (IF). At optical wavelengths, conversion to IF is generally not available, so the operation must be carried out at signal frequency; this necessitates an analogue solution, which is found in some implementation of addition/subtraction followed by squaring, as explained below. The integration is with respect to time in the case of the spectrometer; in the case of the telescope pupil, the integration ought to be over the pupil's spatial coordinate \( u \) or \( u/\lambda \) (which would entail building many identical copies of the WSRT and instantaneously averaging their outputs!), but it is taken over the time instead (the assumption being that astronomical signals are stationary, i.e. that their statistical properties do not change over time, or at least not during the measurement).
4.3. Applications of the FT in astronomical instruments

4.3.1. Astronomical maps and spectra: power

Astronomical signals in general have no definable phase. If one were to try and define a phase, one would have to allow it to vary on a timescale which, though long with respect to the period of the signal, is short compared to any experiment one conducts or any measurement one makes (exceptions might be an astronomical maser signal or a signal which has been artificially made to vary slowly by passing it through an extremely narrow-band filter). In the chapter on polarization one may find a discussion on exactly what astronomical signals look like. Hamaker also discusses the matter (without polarization) in his section 9. The concepts that one distinguishes are:

1. true monochromatic radiation: sinusoid from \( t = -\infty \) to \( +\infty \). A single frequency with constant phase and amplitude. This is a theoretical concept only.
2. quasi-monochromatic radiation: sinusoid with slow variations of both amplitude and phase. Easy to visualise, but in nature only narrow spectral lines and maser signals qualify.
3. polychromatic radiation: sum of many different quasi-monochromatic components, each with its own frequency and independent variation of amplitude and phase.
   a) In its most extreme form, this constitutes a "white noise" signal: all frequencies from \( \nu = 0 \) to \( \infty \) are present in equal amounts. Truly white noise is just as theoretical as truly monochromatic.
   b) "band-limited noise" is the result of passing white noise through a filter (which may be a natural astronomical process). Inversely proportional to the width of the filter passband, there is a minimum time for amplitude and phase to change appreciably (or, put differently, a maximum time for which one may consider phase and amplitude to be defined: the "coherence time" or, multiplied by the speed of light, the "coherence length"); this is another example of the scaling law for FTs.

Since well-defined absolute phase is rare in astronomy, we use power \( a^2(\nu) \), or \( |a(\nu)|^2 \), as our measure of 'strength' of the radiation and never consider phase, except in a relative way (mostly within an instrument and in polarization concepts).

The 2 major kinds of dataset that observational astronomy obtains are:

1) spectra (distribution of power with frequency \( \nu \)) for a source at certain positions in the sky
2) images (distribution of power with angular position \( \sin \theta \) in the sky) at a certain frequency

As stated in another section, we may measure such power distributions either directly (by scanning a telescope or a filter as the case may be), or we may measure the auto-correlation function and FT it. This auto-correlation function will be a function of time \( t \) or of the spatial
coordinate $x/\lambda$ (the conjugate variables of $\nu$ and $\sin \theta$ respectively). The instrumental setups for direct and FT technique will be discussed in the next section.

4.3.2. Imagers and spectrometers in direct and FT technique

In the optical region, a real image of the sky is formed by a telescope, often followed by 're-imaging optics'. Such an optical image may be read by an array detector, it may be scanned by a single-point detector, or a fixed single-point detector may be used and the telescope or re-imaging optics may be scanned. One may even combine an array detector with telescope scanning, in which case multiple copies of the image can be built up simultaneously, to be used for internal relative calibration of the detector pixels, etc (it is also more efficient, of course). All such techniques are direct, in the sense that an image of the sky is detected directly (merely translated by the optics from an angular image at infinity to a real image at an accessible point in space).

Direct imaging in radio is limited mainly to single-point detectors and scanning the telescope. SCUBA and similar multiple front-ends are exceptional and are limited to small numbers of channels.

A direct optical spectrometer works by making a wavelength-dispersed image of a point object (or the spectrometer slit illuminated by an extended object) and then recording this image by any of the above methods for imagers. In radio, a narrow-band (IF) filter is the equivalent of the single-point detector. Filter banks are collections of such filters and 'filter bank spectrometers' are the usual direct-spectroscopy instruments in the radio region.

As stated above, a FT relation exists between a power spectrum or power image and the autocorrelation of the 'signal' or of the pupil EM field distribution, respectively. FT instruments use this relation for an alternative instrumental setup. Hamaker (sec 17.2, 17.3) discusses the RADIO AUTOCORRELATION SPECTROMETER; it is a multi-delay instrument, which obtains values of the autocorrelation at different delays simultaneously and therefore it is the FT equivalent of the filter-bank or of the array detector.

The OPTICAL AUTOCORRELATION SPECTROMETER is basically single-channel (although only half the light emerges at the output and 2 outputs are actually used): the Michelson or Mach-Zehnder interferometer of Figure 4-7 (or still other equivalent forms): The light to be auto-correlated is split into 2 equal components (A, B in Figure 4-7), which are then brought into coincidence again after one of them has suffered a variable delay. This is equivalent to forming $E(t)$ and $E(t + \tau)$ and ... (yes, what????) We added the EM fields by bringing them into coincidence, but we were supposed to multiply -- and then integrate – them. The answer is: we both add and subtract them and then, in the detector, we square the sum and difference:

$$A^2 + B^2 + 2AB \quad A^2 + B^2 - 2AB$$

The difference of these is $4AB$. Hey presto, our correlator!!!! One either alternates between the sum and difference conditions, so that the AC part of the detector signal represents the desired product, or one forms the
squares of the sum and difference simultaneously in a fringe system and the fringe amplitude represents the desired product).

So the technical details of the correlator are entirely different in the optical and radio regions (Figure 4-6).

Figure 4-6: Schematic representation of a correlator in the radio (at IF) and optical spectral regions

There are other differences. One could also split the optical signal and measure multiple samples of the autocorrelation, in parallel correlators, but signal-to-noise per channel would deteriorate and there would be no advantage; this is a consequence of optical noise being mainly shot noise within the signal itself, while radio noise is receiver noise (so that S/N per channel is not affected by splitting the signal after amplification). Therefore optical autocorrelation spectrometers ("FT spectrometers" in optical jargon) are single-channel and must use rapid scanning or difference/sum normalisation to obtain "interferograms" (auto-correlation as a function of delay) unaffected by scintillation and change of extinction. Optical FT spectrometers are used mainly in the IR, where they have S/N advantages over direct spectrometers. Occasionally, applications are in the visible (e.g. the Kitt Peak solar interferometer); the usual reason is that the sources of errors are different
from those in direct spectrometers (scattered light can be recognised by its fringe signature, since it is by definition of a different wavelength).

Figure 4-7: Optical auto-correlation spectrometers

To obtain the **radio spatial autocorrelation function**, one uses Westerbork-like systems, as discussed by [Hamaker](#) (sections 12, 13). Again, in the radio system one splits the signals (from telescope 1 in [Hamaker](#)’s figure 13.4) to obtain a number of samples of the complete autocorrelation function simultaneously.

The **optical spatial autocorrelator** uses optical paths rather than cables to carry the signals to the point of combination. Controlling phase behaviour is a major problem, which explains why some of the first 'optical' interferometers on large expensive telescopes are in fact for the infrared (Figure 7-7), where tolerances are easier to achieve. However, experimental visible-region interferometers exist, as dedicated instruments on arrays of small, inexpensive telescopes and in time the technique will be mastered. To keep phase variations under control, the optical paths must be in vacuum tunnels, and even then, servo systems are needed for detailed phase control; in addition, one must control polarization; for large telescopes, one also needs adaptive optics to undo the atmospheric effects. So the job is far from trivial and is best tackled in several development steps.

Apart from the admittedly huge differences in actual implementation, the radio and optical forms of aperture synthesis are almost identical. It is therefore no surprise that the Cambridge group, who in the 60s pioneered aperture synthesis in the radio spectral region, have now taken up the **optical equivalent** using VLBI techniques such as phase closure to do optical synthesis in spite of
atmospheric problems. Equally, it is no surprise to find ASTRON working on MIDI (even if by way of the "Roden group" and VISIR).

4.3.3. 'Spatial frequencies' in an image

Just as a function of time $A(t)$ can be represented by its FT $A(\nu_t)$ (the spectrum), we can represent an optical image $A(x)$ by its 'spatial spectrum' $a(\nu_x)$, where $\nu_x$ is a 'spatial frequency', expressed in something like cycles/mm (just as $\nu_t$ is in cycles/sec, for which we happen to have a name: Hz; if we need a name for cycles/mm, we can invent one). Again $a(\nu_x)$ is the FT of $A(x)$. Similarly, for a sky image $A(\theta)$ we can define spatial frequencies $\nu_\theta$ (in cycles/radian) and a spatial spectrum $a(\nu_\theta)$ which is the FT of $A(\theta)$. Previously, the analogue of $A(\theta)$ was called $f(\sin \theta)$, with its FT $F(\frac{\nu_\theta}{\lambda})$. Since the pupil transmission $F(\frac{\nu_\theta}{\lambda})$ and the spatial spectrum $a(\nu_\theta)$ are both FTs of $f(\sin \theta)$ or $A(\theta)$, $F(\frac{\nu_\theta}{\lambda})$ and $a(\nu_\theta)$ must be closely related.

![Figure 4-8: The concept of spatial frequency](image)

Let us take a single 2-element phased array as an example: the angular response $f(\sin \theta)$ is a sinusoidal pattern in the sky, with a maximum wherever the differential delay is a full period: $D \sin \theta = n\lambda$. This sinusoidal pattern is a single spatial frequency, so is represented in the spatial spectrum by an impulse. In other words, the array sees (selects) a single spatial frequency in the spatial spectrum, rejecting others. In other words again: the array is a narrow-band filter for the spatial frequencies that make up the sky distribution. This statement applies to the amplitude/phase distribution, which in astronomy is not very meaningful; however, one can make a similar statement about the sky power image, the spatial spectrum of this power image, and a 2-element correlation interferometer.

Questions:

1. Try to formulate the statement referred to in the last sentence above.
2. What is meant by the 'phase' of a spatial-frequency component?
This analogy can be extended further and made mathematically precise. For the present purpose, it is sufficient to have come across the concept (Figure 4-8) of spatial frequencies and arrays as filters for them. The concept explains why in a synthesis array 'all' spacings should be represented: one must determine the amplitude and phase of each component of the spatial spectrum; if not, the image is incomplete, lacks certain information. In particular, if the spacings go up to a certain maximum value, then spatial frequencies beyond a certain maximum are not present in the spatial spectrum recorded by the array, and the 'observed' image will be blurred to a certain extent (since the highest frequencies describe the finest detail, which the largest pupil apertures – here absent – are sensitive to).

Questions: A diffraction-limited IR telescope has an (amplitude/phase) psf which is the 2-dimensional FT of its pupil (including the 'hole' caused by the secondary mirror); in one dimension:

![Diagram](image)

It is a kind of "broadened narrowband filter" for spatial frequencies present in the IR sky.

1. What is controlled by the parameter D?
2. What happens to the pupil as d gets larger and larger at constant D?
3. What happens to the available pupil spacings as d gets larger and larger at constant D?
4. What happens to the image as d gets larger and larger at constant D?
5. What happens to the spatial spectrum of the image as d gets larger and larger at constant D?
6. What happens to the pupil spacings, spatial spectrum and image as d increases beyond about D/3?

A two-dimensional example of spatial frequencies is provided by Figure 4-9.
4.3.4. Improving the point spread function: filtering and CLEANing

The point spread function psf specifies the response of a system to a point source. The convolution of this function with the true source distribution (seen as a set of point sources) yields the observed source distribution or 'dirty map'. Can one reverse the convolution? The answer is "yes" (but, as always: "up to a point"). There are two basic methods, one of them operates in the FT domain, the other in the data domain. Both produce maps that are a trustworthy and reliable representation of the measured data, while removing the distractions of noise and spurious sources produced by sidelobes. The two methods may be combined, they eliminate different kinds of artefacts.

1. FT the observed distribution. The convolution operation turns into a product, so one may divide by the FT of the psf and obtain the FT of the true distribution. The only snag is that the FT of the psf has regions where it is zero or very small and dividing by it can exaggerate any noise that is present in the observed data at those 'frequencies' in the 'spectrum'. But one may construct an 'optimal filter' (which has a corresponding optimal psf as its FT) and restore (a useful part of) the true distribution without excessive extraneous noise. This is discussed very well in an old paper by Brault and White [Astronomy and Astrophysics, vol. 13, pp. 169-189 (1971!!)].

2. Alternatively, one may recognise in the dirty map the psf signature of a few dominant point sources. One may then assume the presence of point sources of the size of, and at
the positions indicated by, those dominant local maxima and subtract the corresponding scaled versions of the psf from the dirty map. The map is now a little less dirty, i.e. it has been cleaned a little. This procedure is repeated many times, each time at the highest points in the dirty map, until one finally ends up with a map consisting only of noise, without any obvious features with the shape of the psf. During this procedure one remembers where all these point sources were and finally one reconstitutes a map from this collection, using for the convolution a psf that has no secondary maxima (something like the Gaussian, but of finite extent) and adding back the noise that remained. This procedure is called CLEAN and, since its invention by Jan Högbom in the 60s, has become a standard component of all kinds of more complicated data treatment (of radio synthesis maps, CCD optical maps, spectra).

4.3.5. Westerbork and optical gratings

As a final example on how to argue about or play with FTs, let us consider the Westerbork tied array and an optical equivalent of it, the diffraction grating. In both cases a plane wave (from a point source that is effectively at infinity) strikes the "optical component"; in both cases the optical component is made up of equally-spaced elementary apertures and the array has a finite size (Hamaker figs 7.3 and 7.5). The difference is in how the signals are combined after striking the array: Westerbork uses cables and electronic adders, while in optics the spectrograph 'camera' focuses the reflected plane waves to add them at a point (ideally) in space. We may disregard this difference and mentally locate the addition at the array of apertures. We distinguish the following properties of the array:

1) A single element (telescope or a groove in the grating) has a certain amplitude/phase masking function $F_1(x/\lambda)$ and a corresponding angular response $f_1(\sin \theta)$ (where $F_1$ and $f_1$ are in general complex).

2) The elements are spaced a fixed distance apart but are identical. Therefore the array, if infinite in extent, could be represented by the convolution of a $\delta(x)$ with the groove profile $F_1(x/\lambda)$. $\delta(x)$ FTs to $\sin \theta$.

3) The array is not infinite, but has extent $F_1(x/\lambda)$, which FTs to $f_2(\sin \theta)$ (where $F_2$ and $f_2$ are also in general complex). Since $F_2(x/\lambda)$ is much wider than $F_1(x/\lambda)$, $f_2(\sin \theta)$ will be just as much narrower than $f_1(\sin \theta)$.

The total amplitude/phase profile or 'pupil transmission function' $F(x/\lambda)$ of the array is represented by:

$$F_2(x/\lambda) \times \left[ \delta(x/\lambda)^* \times F_1(x/\lambda) \right]$$
We have to FT this to obtain the amplitude/phase response as a function of $\sin \theta$ (see also Hamaker fig 7.4). The result of such a FT is:

$$f_2(\sin \theta) \ast \left[ \Pi (\sin \theta) \times f_1(\sin \theta) \right]$$

The product in square brackets represents an equally-spaced set of impulses (in $\sin \theta$), multiplied by $f_1(\sin \theta)$. The convolution by $f_2(\sin \theta)$ broadens each of these spikes into a replica of $f_2$.

![Figure 4-10: The point spread function of a grating such as Westerbork](image)

We identify $f_1$ with the 'primary beam' of 1 dish, or the 'blaze profile' of a groove in the grating. The spacing of $\Pi (\sin \theta)$ is inversely proportional to that of the dishes (grooves) $\Pi \left( \frac{\lambda}{x} \right)$. Finally, $f_2$ is the array beam, or grating 'instrumental profile'; its replicas are the 'grating responses' (Westerbork) or 'spectral orders' (spectrograph). Since $\frac{\lambda}{x}$ is conjugate to $\sin \theta$ and the $x$ pattern is built in, the $f_2$ replicas come at different $\sin \theta$ for different $\lambda$. In Westerbork, one is therefore reluctantly forced to work with moderately narrow bandwidth to avoid smearing the image, but in the optical spectrograph this $\lambda$-spread is actually the object of the exercise; life is never simple.

The scaling law for FTs implies that:

1) the larger the $F_i$ element is, the narrower will be the primary beam and the fewer grating responses or orders will fall within it.

2) the closer the elements are together, the further out the grating responses or orders will be.

3) the larger the total size of the array, the narrower will be the array beam or instrumental profile.

and vice versa.

Optimising these 3 properties of the array/grating is an important part of system design; such optimisation depends on the observing programmes contemplated and therefore cannot be done...
properly by engineers or astronomers by themselves (this should be self-evident, maar je zal ze de kost niet moeten geven....).
5. Polarization

This chapter deals with polarization, which is a property of EM radiation that is becoming more important in astronomy: it is used by astronomers to detect asymmetric or anisotropic situations in nature, and it is becoming increasingly irresponsible to neglect its effects in designing instruments (even if they are not intended for measuring polarization). The importance of polarization is not the same at all wavelengths, but at optical and radio wavelengths the above is certainly true. Polarization can be combined with all other ways of observing astronomical objects: imaging, spectroscopy of all kinds, aperture synthesis, VLBI, VLTI; as will be shown, it is 'merely' a matter of using 4 quantities instead of 'intensity' of the radiation, and for 'optics' a 4 x 4 matrix of conversion coefficients rather than a single transmission coefficient.

Note: The above statement ("it is merely...") holds for astronomers. For those who design instruments, there is more to it than that (see Jones matrices).

In the Introduction, the concept of the polarization of noise-like signals is discussed at some length. With some modifications, I have taken this from my book 'Astronomical Polarimetry' (Cambridge University Press, 1996; ISBN 0 521 47531 7), to which I refer the reader for a discussion of those details of astronomical polarimetry not covered in this course; all the references to the scientific literature can be found there, except for Johan Hamaker's 1999 and 2000 papers ("Self-calibration in the SKA: dealing with inherently strong instrumental polarization", in: Smolders and van Haarlem (eds): Perspectives in Radio Astronomy – Technologies for Large Antenna Arrays, NFRA, 1999 and "The full-coherency analogue of scalar self-calibration: Self-alignment, dynamic range and polarimetric fidelity": A&A Suppl. 143, 515-534, 2000).

5.1. Introduction: polarization of noise-like signals

In this section, the main concepts of polarized radiation will be introduced and discussed. These concepts apply at all wavelengths. Electromagnetic radiation will be treated as a continuous travelling-wave phenomenon. Quantum considerations can be postponed until the moment the radiation strikes a detector and is converted into an electrical signal. Ideal detectors are not sensitive to polarization, and, to the extent that a real-life detector can be seen as an ideal one preceded by polarization optics, quantum and polarization considerations can live side by side without the one influencing the arguments concerning the other. Of the electromagnetic wave, only the electric vector will be considered; the corresponding magnetic vector follows from Maxwell's equations.

Astronomical signals are noise-like. These noise-like variations of electric field strength (of the electromagnetic wave) may be passed through a narrow-band filter, so that a 'quasi-monochromatic' wave remains. Such a wave contains a very narrow band of frequencies and may be seen as a sinusoidal carrier wave at signal frequency, modulated both in amplitude and phase by noise-like variations. The highest frequencies (the fastest variations) in the modulating noise
determine the width of the sidebands around the carrier wave in the frequency spectrum. Any wide-band ('polychromatic') signal may be seen as the sum of many quasi-monochromatic signals, all with different carrier frequencies and generally each with its own amplitude and phase modulation. It might seem that the phase of such a composite noise-like signal is unimportant, certainly in astronomy where no calibration signal of absolute phase exists for reference. This simple point of view would hold for a scalar wave such as a longitudinal wave or a pressure wave. An electromagnetic wave, however, is transverse and has vector characteristics. The instantaneous electric field of the wave can be resolved into two components at right angles to each other (and to the direction of propagation). If the signal is noise-like in all respects, the two electric field components vary randomly, without any lasting correlation between them in phase or amplitude. However, if, for any frequency within the band, an amplitude and/or phase relation between the components persists for a time which is long compared to the vibration period of the wave, the resultant combined signal is less random than one might expect from pure noise, 'there is some method in the madness'. For at least part of the signal, it is then true to say that, as the two wave components pass through a fixed point in space, the tip of the vector that represents the instantaneous electric field of the combined wave traces out an ellipse, a circle or a straight line, rather than a completely random pattern. While tracing out this more organized pattern at the signal frequency, the electric field vector does vary slowly in amplitude and phase in a noise-like manner, i.e. the size (but not the shape, orientation or handedness) of the pattern varies slowly and so does the position of the tip of the vector within the pattern (at times it lags or leads a little with respect to the position it would have for strictly monochromatic radiation). The fact that such a long-term organized pattern is present within the short-term chaos is referred to as the polarization of a noise-like electro-magnetic wave. Corresponding to the extent to which the flow of radiant energy of the noise-like electromagnetic wave is represented by such a long-term organized pattern, the wave is said to be fully polarized, partially polarized or unpolarized. The shape of the pattern is specified by referring to linear, circular or (the general case) elliptical polarization.

These basic concepts will be refined and quantified in the sections that follow. The line of argument starts with an abstraction far removed from astronomy: a strictly monochromatic wave, 100% polarized. It then proceeds to quasi-monochromatic (which nature may provide in the form of line radiation from a cool low-pressure stationary source; alternatively a high-spectral-resolution instrument may select it, out of what nature offers) and finally to polychromatic, partially polarized, which is the usual type of signal met with in astronomy. Since these concepts are discussed here in considerable detail, I have referred to this section from elsewhere in the course.

5.1.1. Fully or 100% polarized radiation

5.1.1.1. Linear polarization

1. A monochromatic linearly polarized wave is the simplest concept. It has a transverse electric field with constant orientation, its strength at any one point in space varying strictly sinusoidally with time. The duration of this wave is infinite; it has constant amplitude and frequency for all time. Good approximations in real life are light from a well-stabilized
laser (very nearly monochromatic), filtered by a Polaroid, and the radiation from a dipole antenna driven by a sine-wave signal generator (stable single-frequency oscillator). The laser and the signal generator are assumed to be switched on for an infinite time.

2. We can conceptually convert this wave into one that is still 100% linearly polarized, but is quasi-monochromatic by allowing the amplitude and phase to vary slowly and often randomly. The faster these 'slow' variations are, the broader will be the range of frequencies contained within the wave (as described by Fourier transform theory). If we modify the wave in no other way, it is still 100% polarized: i.e. all of its energy is still transported by a transverse linear vibration with a well-defined orientation; all we have done is to distort, slowly and therefore only very slightly, the carrier wave sinusoid. Light from a filament lamp, filtered through a monochromator and a Polaroid, is a good approximation, as is the radiation from a dipole driven by a sine-wave generator with slowly varying phase and amplitude modulation, or from a radio-frequency noise source through a narrow-band electronic filter. If we rotate the Polaroid or the dipole 'very slowly' (i.e. slowly even compared to the 'slow' amplitude and phase variations), we still have 100% linear polarization of quasi-monochromatic radiation, but with variable orientation (we say that the position angle of the direction of vibration, the 'polarization angle', varies); this would not be allowed with strictly monochromatic radiation since rotation of the orientation of the polarization would modify the vibration, which therefore would not continue for infinite time and thus would no longer be monochromatic.

3. Fully or 100% linearly polarized polychromatic radiation is a superposition of quasi-monochromatic waves of many different frequencies; there is usually no stable phase relation between the electromagnetic field at different frequencies. Examples are light from a filament lamp filtered only through a Polaroid, or the radiation from a dipole antenna driven directly by a source of radio-frequency white noise. There is now no single dominant frequency, amplitude or phase, just a unique orientation of the otherwise often randomly vibrating electric field vector (‘random’ is in this context to be taken as: ‘random within the constraints of the mean flow of radiant energy and of the spectral bandwidth’). No clear dividing line exists between quasi-monochromatic and polychromatic radiation; it is a matter of convenience. One can regard polychromatic radiation as a sine wave modulated in phase and amplitude, but the wider the bandwidth (the faster the modulation), the less useful this concept becomes and the more attractive the polychromatic description is. In practice, the fractional bandwidth is the criterion; when it is small enough that one may neglect any frequency dependence of wave amplitude, phase, receiver gain, refractive index and such (e.g. many practical lasers and some radio applications), one uses a quasi-monochromatic description, but when functional dependence on frequency is important, a polychromatic description is more appropriate.

Two independent monochromatic linearly polarized waves, of the same frequency but with vibration directions at right angles to each other, can propagate through empty space and other homogeneous isotropic media, along the same path and at the same time. They are both solutions of Maxwell’s equations and only two independent solutions are possible: a linearly polarized wave of any other orientation can be seen as an in-phase combination of these two basic waves, the ratio between their amplitudes determining the position angle of the direction of vibration of the resultant. One may choose any two orientations at right angles as the base of the
representation; popular choices are horizontal/vertical, right-ascension/declination, latitude/longitude (ecliptic or galactic) or as dictated by the problem studied: parallel and perpendicular to the scattering plane, rotation axis of a magnetic star, symmetry axis of a double radio source, etc. In the strictly monochromatic case, the amplitude ratio of such basic polarized waves is necessarily constant for all time and the result is always 100% polarization. In the quasi-monochromatic case, the amplitude ratio may be constant or it may vary 'very slowly' without the (linear) polarization becoming less than 100%.

Linearly polarized radiation is sometimes said to be 'plane polarized'; the term is not as common as it used to be. The 'plane of polarization' is also an old term, which in fact used to refer to the direction of vibration of the magnetic field. Though often used nowadays to denote the direction of vibration of the electric field, the term is ambiguous in several ways and it is best avoided altogether; it is much better to refer to the 'direction of vibration (of the electric vector)'.

**Circular polarization**

1. A monochromatic circularly polarized wave can be seen as a combination of two monochromatic linearly polarized waves with vibration directions at right angles to each other, of equal amplitude and differing by ±90º in phase. The combined electric field vector has constant magnitude, but its orientation moves uniformly with time, making one revolution per wave period, rotating 'left' or 'right' according to the sign of the phase difference; all the radiant energy is associated with this circular pattern. Being monochromatic, the wave has infinite duration.

2. A quasi-monochromatic 100% circularly polarized wave differs from its monochromatic equivalent only by slow and often random variations of the amplitude and of the 'circular velocity'. The tip of the electric field vector still moves around the circle, on average at the wave frequency, but it now drifts around its mean position on the circle, while the size of the circle also changes slowly ('slowly' denoting a speed in keeping with the bandwidth of the signal). If one wishes, one can still regard the signal as the superposition of two quasi-monochromatic linearly polarized waves with 90º phase difference, now with mutually synchronized drifts in amplitude and phase; however, the concept of slow changes in the size of the circle and, generally independently of this, the drifts around its circumference is a much cleaner one. For the usual astronomical signals, the slow drifts in circle size and in position on the circle are both random.

3. Polychromatic 100% circularly polarized radiation is a superposition of quasi-monochromatic waves of many different frequencies, but all circularly polarized in the same way. There is in general no stable phase relation between waves at different frequencies. Alternatively, viewing the total signal as a phase- and amplitude-modulated carrier wave, changes in the size of the circle and of the position on its circumference are much faster than in the quasi-monochromatic case (but still 'slow', in keeping with the increased but always finite bandwidth of the polychromatic signal). All the radiant energy is still associated with the circular motion of the field vector tip, the circle having the same left- or right-handedness at all frequencies in the band. If one wishes to view this signal as the superposition of two linearly polarized polychromatic waves, the (faster, but still 'slow')
drifts in amplitude and phase of these two components must be synchronized, just as for the quasi-monochromatic case.

Two circularly polarized modes are possible, and they can propagate through empty space (and other homogeneous isotropic media) independently; their electric field vectors rotating in opposite directions; they are referred to as left-hand-circular (LHC) and right-hand-circular (RHC), although much confusion exists as to which of the two shows clockwise rotation of the electric field vector and about the point of view for examining the circle (facing the source or looking along the direction of propagation).

As noted, LHC and RHC may be seen as combinations of two linearly polarized waves of equal amplitude with +90º or -90º phase difference. However, LHC and RHC waves may themselves be used as the base for other polarization forms in the same way as two linear vibrations: two phase-correlated circularly polarized waves of equal amplitude add to give linear polarization, the orientation of which depends on the (constant) phase difference between the circular constituents (position angle of the direction of linear vibration = 1/2 phase difference).

Note: Linearly polarized radiation has a vector character: one must specify the orientation as well as the 'intensity' (flow of radiant energy) of the radiation. Circularly polarized radiation, however, requires only a single scalar, which can specify both the 'intensity' and the 'sense' (or 'handedness', LHC or RHC, specified by the sign of the scalar quantity).

5.1.1.3. Elliptical polarization

1. The most general form of polarization is elliptical, for which the tip of the electric field vector executes an ellipse at the signal frequency. The distinguishing parameters of the ellipse are orientation, axial ratio and handedness linear and circular polarization are special cases of this general form. A monochromatic elliptically polarized wave may be visualized as the sum of two unequal linearly polarized components with phase difference of ±90º, or as the sum of two linearly polarized components (which may or may not be equal) with a phase difference of something other than 0º or ±90º. This exercise in geometry is left to the reader. It is also possible to see elliptical polarizations as the sum of two unequal oppositely circularly polarized components (or even one circular and one linear component, though such a 'non-orthogonal' form is less useful) with a constant phase difference between them; these useful mental gymnastics are also left to the reader.

2. Quasi-monochromatic 100% elliptically polarized radiation is obtained by allowing the size of the ellipse to vary slowly, with similar slow variations of the position of the tip of the electric field vector within the ellipse (lag or lead with respect to the monochromatic equivalent). The elliptical pattern, however, is a constant of the wave. Alternatively, one may introduce (correlated) slow and often random variations of amplitudes and phases of the two linearly polarized waves which were used in the mental picture to construct the elliptically polarized wave.

3. Polychromatic 100% elliptically polarized radiation is a sum of quasi-monochromatic components, all with the same elliptical polarization. Its noise-like character is entirely
analogous to the case of circular polarization discussed above. In spite of all the amplitude and phase variations, the ellipse is again a constant of the wave.

Elliptical polarization is the most general form possible, linear and circular polarization being special cases; linear polarization has axis ratio 0, circular polarization has axis ratio 1. We have mentioned that two independent linear or circular polarization forms can be sustained in homogeneous isotropic media, these forms having 'opposite' characters to each other: two linear forms with directions of vibration at right angles, two circular forms of opposite handedness. By considering these limiting forms, one may suspect that 'opposite' elliptical polarization forms must have equal axial ratio, that the major axes of the two ellipses must be at right angles to each other and that the ellipses must be traced out in opposite directions (opposite handedness). This statement will be quantified below.

The position angle of the axes of the ellipse may be changing 'very slowly' with time, provided the radiation is quasi-monochromatic or polychromatic; similarly, the ellipticity of the ellipse may be changing 'very slowly'. This is entirely analogous to the case of a slowly rotating Polaroid discussed for quasi-monochromatic linear polarization.

5.1.2. The Stokes parameters

We now introduce the Stokes parameters, four quantities which all denote 'radiant energy per unit time, unit frequency interval, unit (detector or collector) area (and for extended sources: per unit solid angle)'; see Note on units below. This representation of polarized light was invented by Sir George Gabriel Stokes (1852); it was revived and introduced into astronomy by Chandrasekhar (1946). The absolute phase of the wave does not enter into the definition; addition of the Stokes parameters of beams of radiation represents incoherent superposition of these beams. The Stokes parameters are often gathered into a 4-vector $S$ with components labelled $S_0, S_1, S_2, S_3$, or $I, Q, U, V$ (which will be used in this course); or $I, M, C, S$. In terms of the properties of the polarization ellipse, they are defined by the equations shown in Figure 5-1. Other, equivalent, definitions exist. Some of these are more elegant, or are particularly useful for defining exactly what a polarimeter should measure. However, for handling noise-like signals in which phase is irrelevant but a 'polarization ellipse' is a suitable description for the method in the madness, the form of Figure 5-1 is appropriate: electric field amplitudes are squared, phase does not appear and, finally, the parameters of the polarization ellipse appear in the definition more or less in the functional form one might expect. Note that $I \geq 0$ but $Q, U$ and $V$ can be positive or negative. Note also that the double angles $2\beta$ and $2\chi$ enter into the definitions; the basic reason for this is the squaring involved in going from amplitude to radiant energy ('intensity').

Let us examine the form of the Stokes parameters for 100% linear, 100% circular and 100% elliptical polarization (Figure 5-1). For linear polarization, $\sin \beta = 0$ or $\cos \beta = 0$. In both cases $\sin 2\beta = 0$, so $V = 0$, while $Q = a^2 \cos 2\chi$, $U = a^2 \sin 2\chi$. 
The quantity \( a \) is related to the amplitude of the electric field vibration (so \( a^2 \) is concerned with 'intensity', or flow of radiant energy), the angle \( \chi \) is the orientation of the ellipse (in this case of the straight line) with respect to the chosen reference direction; \( \chi \) is called the 'azimuth' of the polarization, or the 'polarization angle'. The quantities \( Q \) and \( U \) are Cartesian components of the vector \( (a^2, 2\chi) \); note the doubling of the polarization angle in this true vector representation. A \( Q,U \)-diagram (or \( Q/I, U/I \)) is one of the common representations in astronomical polarimetry. Note that for the polarization ellipse the origin of \( \chi \) is chosen arbitrarily (e.g. towards zenith, equatorial North Pole or Galactic North Pole, along the symmetry axis of a double radio source), but in the \( Q,U \)-diagram the origin of \( 2\chi \) is by definition the \( Q \)-axis (which is conventionally drawn horizontally).

For circular polarization, \( \sin \beta = \pm \cos \beta \) or \( \sin 2\beta = \pm 1 \), \( Q = U = 0 \) and \( |V| = a^2 = I \).

General elliptical polarization is represented by non-zero values of \( Q, U \) and \( V \). Note that for 100% polarization (which is all we have mentioned so far) \( Q^2 + U^2 + V^2 = I^2 \). The axial ratio of the ellipse is \( \tan \beta \).

Note on units: I shall side-step the hornets' nest of radiometric quantities and units (magnitudes, candela per foot\(^2\)-angstrom, jansky, counts per second, etc), merely noting that in astronomy we deal with what in radiometry is called radiant intensity for point sources and radiance for extended sources (the difference being whether one integrates...
over the source solid angle or not). Questions of units are for photometrists; polarimetrists as such fortunately deal with relations between the four Stokes parameters, all of which (in any one application) represent similar physical quantities and are measured in the same units. Irrespective of the radiometric quantity being discussed or the radiometric units being used, the term (total) intensity is often used in polarimetry for Stokes I, and polarized intensity for Stokes Q, U and V or some combination of these. Readers are warned of this usage and are urged to be more specific in their own publications. Refer to Snell (1978) for an excellent overview of radiometric quantities and units; an overview of astronomical usage is given by Léna (1988, section 3.1). Berkhuijsen (1975) clarifies early confusion in radio-polarimetric terminology.

5.1.2.1. Alternative formulae for the Stokes parameters

One encounters other mathematical formulae for the Stokes parameters, which may be more useful in certain applications and which one should be ready to recognise. There is little point in trying to reproduce the conversion calculations, which are complicated and tedious. For references, see my book, section 4.3. The formulae are:

\[
\begin{align*}
I &= I_0 + I_{90} = \overline{E_x E_x^* + E_y E_y^*} = \overline{A_x^2 + A_y^2} = \bar{a}^2 \\
Q &= I_0 - I_{90} = \overline{E_x E_x^* - E_y E_y^*} = \overline{A_x^2 - A_y^2} = a^2 \cos 2\beta \cos 2\chi \\
U &= I_{45} - I_{-45} = \overline{E_x E_y^* + E_y E_x^*} = 2A_x A_y \cos \Delta = a^2 \cos 2\beta \sin 2\chi \\
V &= I_{\alpha} - I_{\beta} = i(\overline{E_x E_y^* - E_y E_x^*}) = 2A_x A_y \sin \Delta = \bar{a}^2 \sin 2\beta
\end{align*}
\]

where overlining denotes an average over time (sufficiently long to average over the 'noisy' variations of amplitudes and phases).

The last of these forms is the one we have used so far, specifying the Stokes parameters in terms of the properties of the polarization ellipse. As we shall see later, the first form is convenient for optical polarimetry (in which we measure differences of intensities by making our intensity-detector sensitive alternately to one or other orthogonal polarization form and recording the AC component of the detector signal). The second form, using products of amplitudes (which we construct in correlators), is more suitable for radio polarimetry. The third, hybrid, form is also occasionally encountered and is included for the sake of completeness.
5.1.3. Orthogonal modes and birefringence

For every polarization form (or 'mode'), we can define one with the same I but opposite Q,U,V; adding ±90° to β is the operation required (Figure 5-2). For such an opposite form, the axial ratio of the ellipse is the same as that of the original, but the axis is at right angles and the ellipse is traced out in the opposite direction. Such pairs of opposites are said to be 'orthogonal' to each other; they are independent solutions of Maxwell's equations, they can propagate independently through empty space and other homogeneous isotropic media.

In a homogeneous isotropic medium, all polarization modes have the same propagation velocity. In astrophysical magnetized plasmas, however (solar active regions, pulsars, the interstellar medium, the Earth's ionosphere), different polarization modes have different propagation velocities. For a given plasma and a given direction of propagation with respect to the magnetic field, there are always two orthogonal modes that can propagate through the medium without changing their polarization form (they are 'eigenmodes'; the terms 'normal modes' and 'characteristic waves' are also used). Although the polarization of these two modes remains unchanged, they do travel at different velocities, i.e. the medium has two refractive indices, one for each eigenmode; such a medium is said to be birefringent. Optical crystals are also birefringent media; the astronomical significance of this lies in their use for constructing accurate polarimeters.

If a medium is linearly birefringent, its eigenmodes have linear polarization. Radiation that has exactly the polarization of one of those two modes will not be changed, but for any other polarization angle, or for circular polarization, the polarization form will change as the radiation passes through the medium. One may think of the incident radiation as being resolved into the two eigenmodes, which propagate independently, each with its own velocity. On emerging from the medium, the components recombine with a phase difference induced by the medium; the exit polarization will thus be different from the polarization of the beam that went in. In optical
polarimetry, one makes use of slices of linearly birefringent crystal materials, so-called wave plates (e.g. 'quarterwave plate' for a phase difference of 90º, used for converting linear polarization to circular or vice versa).

Circular birefringence causes relative phase shifts between two circularly polarized eigenmodes. A linearly polarized signal impinging on such a medium should be thought of as being resolved into the two circular eigenmodes, each of which passes through the medium at its own velocity (i.e. with its own refractive index). On emerging from the medium, the two modes recombine to give linear polarization again, but the direction of vibration has been rotated by an angle of half the differential phase between the modes. In an ionized plasma with a magnetic field component along the line of sight, such rotation of the linear polarization is called 'Faraday rotation'. Circular birefringence also occurs structurally in certain optical crystals (e.g. quartz), due to a helical atomic arrangement. Such crystals can be used to construct circular retarders (or 'rotators', viz. of the direction of vibration of linear polarization). Many asymmetric organic molecules in solution cause the medium to be circularly birefringent or 'optically active'; this is used in laboratory techniques such as saccharimetry, but so far it has not been of professional significance to astronomers.

In general, birefringence will be elliptical, i.e. the eigenmodes are elliptical. This occurs in astrophysics (e.g. Jones and O'Dell 1977), but is not used much in technical work (at least not intentionally; oblique rays through crystal components designed for normal incidence inevitably have eigenmodes with some ellipticity).

**Note:** Rays with orthogonal states of polarization are sometimes denoted by 'o' for ordinary and 'e' (or 'x' in radio work) for extraordinary. This usage arose from polarization effects in crystals, such as calcite, in which the extraordinary ray is not refracted in the same way as in 'ordinary' isotropic materials, and it has been adopted for the description of radio wave propagation (IEEE 1969). Nowadays the terms o- and e-ray may loosely denote any pair of rays of orthogonal elliptical polarization, and may be encountered in descriptions of instruments and of magneto-ionic plasmas, such as the Earth's ionosphere or the Sun's corona; for linear polarization, 's' and 'p' (or \(⊥\) and \(∥\)) are also found.

## 5.1.4. Unpolarized radiation

What happens when we combine the electric fields of two equal (quasi-monochromatic or polychromatic) 100% polarized waves of orthogonal polarization forms, there being no long-term persistence in the phase relation between them ('incoherent sum', or 'intensity superposition')? For the sake of a clear and simple mental picture, we take as an example two linear polarizations at right angles. For a time short compared to the 'slow' variations of amplitude and phase discussed in previous sections, some definite polarization form will result (linear, circular or elliptical, depending on the momentary phase difference). However, a sufficiently long time average is part of the definition of the Stokes parameters (Stokes himself is quite clear about this, distinguishing between 'temporary intensities' and 'actual intensities'). After some time has elapsed, the 'slow' variations will have caused the polarization form of the incoherent sum to change to something else. During a 'sufficiently' long time interval, all possible forms of polarization will occur, all values of \(β\) and \(χ\) will occur, and the time-averaged Stokes parameters will be \(\left\langle a^2 \right\rangle, 0, 0, 0\), where \(\left\langle \cdots \right\rangle\) denotes a time average. We call such radiation unpolarized, since no single polarization form dominates or is conspicuously absent. We note
that the Stokes parameters of the two linear polarizations that went into the incoherent sum were \(I/2, Q, U, 0\) and \(I/2, -Q, -U, 0\) and that for such intensity superposition the Stokes parameters of the sum are equal to the sums of the Stokes parameters of the components (a mathematically inclined reader may prefer the proof of additivity of the Stokes parameters by, e.g., Collett (1993, section 4.6)).

The same argument may be used for incoherent addition of equal LHC and RHC polarized components. Again, the sum has \(Q = U = V = 0\). Unpolarized radiation can be seen as the incoherent sum of two beams, of any two 'orthogonal' polarizations \(I/2, \pm Q, \pm U, \pm V\). It does not matter what polarization forms one chooses, as long as the Stokes vector sum is \(I, 0, 0, 0\); in other words as long as the last three Stokes parameters are equal but opposite in the two components.

Paralleling the above definition of unpolarized radiation in terms of Stokes parameters, one may define it in operational terms, i.e. in terms of hypothetical measurements. One may set up 'polarimeters' (instruments that are sensitive to and measure some particular form of polarization) for any polarization form one likes to choose. If the measurement yields a null result no matter what polarization form one tries to detect, the radiation under scrutiny is said to be unpolarized.

**Note 1:** Unpolarized radiation is sometimes called 'natural radiation', but like Stokes' term 'common light' this term is ambiguous and is best avoided.

**Note 2:** Instead of averaging over time, one could use ensemble-averaging to define the polarization of noise-like signals. In fact, quasi-monochromatic and polychromatic radiation will usually be the result of ensemble-averaging of microscopic events, such as emission of quanta of line or continuum radiation, scattering of light by an ensemble of molecules, or pulses of synchrotron emission from an ensemble of relativistic electrons. In practice, therefore, averaging over time is more or less equivalent to ensemble-averaging, though in special cases (e.g. very fast time variations of polarization of light from a star) one may have to be more careful about the definitions one uses.

### 5.1.5. Partial polarization

Now that we have defined unpolarized radiation, the concept of partial polarization is simple: partially polarized radiation is the incoherent sum of an unpolarized and a fully polarized component. The Stokes parameters of partially polarized radiation are the sums of the Stokes parameters of the components; the \(I\) values (always positive) add, while the \(Q, U\) and \(V\) values are those of the fully polarized component. Therefore, for partially polarized radiation, \(Q^2 + U^2 + V^2 < I^2\). We call \(\sqrt{Q^2 + U^2 + V^2}/I\) the *degree of polarization*, generally denoted by \(p\).

One also encounters the degree of linear polarization \(p_{\text{lin}} = \sqrt{Q^2 + U^2}/I\), degree of circular polarization \(p_{\text{circ}} = V/I\) or other forms (e.g. \(p_X\) for \(Q/I\), \(p_Y\) for \(U/I\)).

Equally valid is a representation of partially polarized radiation as the incoherent sum of two generally unequal fully polarized components of orthogonal polarization (in this view, unpolarized radiation is just a special case of partial polarization).

In terms of the elliptical envelope of the electric field vector, one can visualize the field vector of partially polarized radiation as 'slowly' making the transition from one elliptical envelope to
another, with some ellipses occurring more frequently than others. More radiant energy is transported by one particular elliptical type of vibration than by any of the others; note that phase does not occur in this statement.

For obvious reasons, fully or 100% polarized radiation is sometimes said to be in a pure state of polarization; similarly, partially polarized and unpolarized radiation, as an incoherent sum of two pure states, may be referred to as being in a mixed state of polarization.

5.2. Designing instruments to include polarimetry

A widely-held misconception says that polarimetric instrumentation is specialist, complicated, difficult to construct and maintain. The present section should help to dispel that view.

5.2.1. Principles

There are two basic types of astronomical polarimeter: the optical and the radio. Optical polarimeters make use of modulation (and differencing), radio polarimeters of correlation. In principle, either method could be used at all wavelengths. However, an optical correlator is not a simple device and differencing does not work very well at radio wavelengths. The preference is thus entirely an engineering matter.

5.2.1.1. Modulation

In many astronomical applications, the polarized radiation flux is only a small part of the total. Under such circumstances, small errors in flux measurement could lead to large fractional errors in the degree of polarization. The problems include flexure of instruments or their supports, magnetic fields influencing detectors, dewing of optical surfaces and of feed antennas, etc.; a near-perfect cure exists in the form of the modulation technique.

The technique of modulation makes the measurement of degree of polarization or the normalized Stokes parameters $Q/I$, $U/I$, $V/I$ insensitive to many errors. Basically, modulation is a way of making a differential measurement very rapidly and is useful whenever the required information is represented by a small quantity superposed on a large, irrelevant background signal. For polarimetry, modulation is implemented as a rapid switching of the polarimeter sensitivity between two orthogonal states of polarization and measuring the ratio of the alternating (‘AC’) signal to the average (‘DC’) signal. This ratio is proportional to $Q/I$, $U/I$, $V/I$ or some combination of these, depending on the adjustment of the instrument. Clearly, such ratios are insensitive to external effects which multiply AC and DC by the same factor, such as time-varying gain of a radio receiver or scintillation in the atmosphere. Any (drift in the) zeropoint error will also be reduced; it will not affect the AC component, and the fractional error in degree of polarization will only be that of the I measurement.
Note: One often sees (linear degree of) polarization defined as \((I_0 - I_{90})/(I_0 + I_{90})\). This shows that a direct measurement of the numerator (differential or AC) and denominator \((2 \times \text{average or DC})\) yields the required result. For such alternative definitions of the Stokes parameters, see section 5.1.2.1.

A basic requirement for modulation techniques to work is that the modulation is faster than the time-constant of the errors which one tries to eliminate. For scintillation noise to be eliminated from optical polarimetry, modulation frequencies should be tens or hundreds of hertz; for elimination of slow gain changes (flexure of spectrographs, changing humidity in a radio feed, etc.), modulation may only be at millihertz frequencies (in this case, one could equally well call it differential measurement; the principle is the same).

5.2.1.2. Correlation

Radio techniques make use of polarized antennas in the focal plane of the telescope. Such a polarized feed antenna converts the free-space field of one particular polarization form into an electrical signal within a receiver system. It is possible to mount a pair of feed antennas of orthogonal polarizations within one focal-plane structure and to allow these to feed two entirely separate receivers (including the detectors, which square the amplitude signal to produce an output proportional to incident radiant energy). Depending on the polarization and/or orientation of the feed antennas, the difference of the 'energy' signals in these two channels represents Q, U or V, while the sum represents I. However, the two receivers, though as far as possible identical, in fact differ by small amounts which may vary with time, so that the looked-for small differences between large signals may be unstable. For that reason, the 'sum and difference' method is not the most widely used in the radio part of the spectrum. Instead, one makes use of the fact that a polarized wave generates signals with correlated (complex) amplitudes in a pair of orthogonal polarization channels and measures the correlation (basically a product of 2 amplitudes) in a special kind of multiplier. Such systems are not entirely immune from error, but the recorded polarized flux is proportional to receiver gain acting on the incident polarized flux rather than to small differences in gain acting on the total flux. The recorded signal is therefore more stable to fractional gain changes by a factor of order 1/p. However, stray signals that are correlated in the two channels will show up as spurious polarization; an example of this is mixer noise in one channel, introduced also into the other channel by non-orthogonality of the feed antennas or imperfections of the waveguide or horn in which they are mounted.

Synthesis telescopes like Westerbork employ many thousands of (digital) correlators as part of their essential hardware; these are exactly the kind of components one needs to do polarimetry. So, provided the feed antennas and RF and analogue parts of the IF receiver sections exist for 2 polarizations per element antenna, polarimetry in synthesis arrays is mostly a matter of being able to feed suitable inputs to the many correlators: for all interferometer spacings being used, one now needs to correlate each polarization of one dish with both polarizations of the other dish. Therefore 4 times as many correlators are needed per interferometer/delay-channel, so 4 times fewer are available for obtaining angular or spectral resolution. Needless to say, system design is not trivial when one has to foresee all the possible combinations of (yes-or-no)polarization/angular-resolution/spectral-resolution that astronomers may require from a future instrument (such as SKA).
5.2.2. Matrix methods

To express the vector character of the EM field, one needs more than a scalar. Various systems are in use, the two most usual being the 4 real numbers collectively known as the Stokes parameter vector and the 2 complex numbers that make up the so-called Jones vector. To relate the vector of the radiation that has passed through a medium or component to that of the radiation which went into it, one needs matrices. The matrices corresponding to Stokes parameters and Jones vectors will be introduced here. More detail in my book chapter 4) and its references. See also Hamaker 2000.

5.2.2.1. Mueller matrices

As discussed above, the four Stokes parameters denote the flow of radiant energy in specific vibrations of the electromagnetic field, and all four are expressed in the same units. We write them collectively as a four-element column matrix, which is generally referred to as the Stokes vector. When convenient, the vector will be written as a row matrix, but a column matrix is always intended.

When radiation propagates through a certain volume of space (which may be empty or contain some material medium), the polarization of the input and output radiation is represented by the input and output Stokes parameters. Within the volume, the state of polarization may be altered: in general, any elliptical polarization may be transformed into some other elliptical polarization form, or the radiation may be polarized by the medium. This can be represented by a transformation between the input and output Stokes parameters, and in general the transformation is linear. When the input and output Stokes parameters are arranged as 4-vectors, the transformation becomes a 4 × 4 matrix M:

\[
S_{\text{out}} = M \cdot S_{\text{in}}
\]

where

\[
M = \begin{pmatrix}
    m_{11} & m_{12} & m_{13} & m_{14} \\
    m_{21} & m_{22} & m_{23} & m_{24} \\
    m_{31} & m_{32} & m_{33} & m_{34} \\
    m_{41} & m_{42} & m_{43} & m_{44}
\end{pmatrix}
\]

and S stands for the Stokes vector (I,Q,U,V). Since Stokes parameters are real quantities, the elements of M are all real numbers; \( m_{11} \) must be positive (I is always positive), and the other elements can be positive or negative. The matrices are known as Mueller matrices, after H.Mueller who worked out their precise form for a number of optical components. The assumption of linear transformations amounts to assuming that there is no functional dependence of the elements of the matrix on the incident radiation (e.g. no processes such as squaring the amplitude in a mixer, frequency doubling, etc.; frequency conversion -- though involving a mixer stage -- can, as a single indivisible operation, be considered a linear transformation).
When the radiation travels through several media in succession, the output Stokes vector for one medium ('a') is the input Stokes vector for the next ('b'):

\[ S_{b,\text{out}} = M_b \cdot S_{b,\text{in}} \equiv M_b \cdot S_{a,\text{out}} = M_b \cdot M_a \cdot S_{a,\text{in}} \equiv M \cdot S_{a,\text{in}} \]

or

\[ M = M_b \cdot M_a \]

where \( M \) represents the combined action of the two media 'a' and 'b'; it is the matrix product of \( M_b \) and \( M_a \) (note the order: the first medium traversed comes last in the matrix equation).

For many purposes (such as reading this book), multiplication is the only matrix algebra one needs. It is specified by:

\[ m_{ij} = \sum_{k=1}^{4} (m_b)_{ik} \cdot (m_a)_{kj} \]

or, as a pictogram:

\[ m_{23} = \sum_{k=1}^{4} \begin{pmatrix} \rightarrow & \rightarrow & \rightarrow & \rightarrow \end{pmatrix} \times \begin{pmatrix} \downarrow & \downarrow & \downarrow \end{pmatrix} \]

Modern papers in both astrophysics and instrumentation increasingly use matrix methods of some sophistication (e.g. Hamaker 2000, in press); more insight into matrix algebra will certainly pay off in future work.

### 5.2.2.2. Jones matrices

When the phase of a polarized signal relative to some other polarized signal is important, the Stokes parameters are of no use; they deliberately ignore phase (except in a relative sense within each signal, as needed to specify the state of polarization). However, there are situations (such as when combining the beams of an interferometer) when phase does matter. Under such circumstances, one has to use Jones vectors and matrices, first introduced by R.C.Jones (1941); these represent the electric fields (and their transformations) of two orthogonal polarization forms (usually linear), including absolute phase if desired. Jones calculus is used in astronomy only for the design of instruments and only the basic ideas will be presented here. The important thing is to know when it is necessary to use Jones rather than Mueller calculus (i.e. when phase is important).

Jones calculus cannot handle partial polarization (mixed states of polarization). There are situations in which phase is important but polarization is partial. In such a case, one must formally separate the radiation into two (generally unequal) mutually incoherent fully polarized
components (pure states) of orthogonal polarizations (i.e. that of the partial polarization in the input signal and its opposite), treat each separately by Jones calculus and obtain the final result by incoherent formal recombination of the outputs.

The notation used for Jones calculus is the complex notation for sinusoidally varying quantities that we have met before. The notation is explained clearly by Hamaker (sections 4 and 26).

For astronomical purposes one usually converts to Stokes parameters and Mueller matrices at the end of the calculation (with the added complication in correlation-type interferometers that the observables are spatial cross-correlation products; these are complex quantities, viz. the spatial complex Fourier transforms of the real Stokes parameter sky distributions).

Jones represented a fully polarized signal as the vector sum of two electric fields at right angles, as we did in words above. In terms of complex amplitudes,

\[ \mathbf{E} = \mathbf{E}_x \mathbf{l} + \mathbf{E}_y \mathbf{m} \]

where \( \mathbf{l} \) and \( \mathbf{m} \) are unit vectors in the x and y directions. In the Jones calculus, \( \mathbf{E} \) is written as a column matrix, with complex components:

\[ \mathbf{E} = \begin{pmatrix} \mathbf{E}_x \\ \mathbf{E}_y \end{pmatrix} \]

With every state of 100% polarization one can associate such a column matrix, or 'Jones vector'. Distinction is made between 'full Jones vectors' (which include amplitude and phase of both components) and 'standard normalized Jones vectors' (for which the modulus is equal to unity). The standard normalized Jones vectors for horizontally (along x) and vertically polarized radiation are, respectively,

\[ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } \begin{pmatrix} 0 \\ 1 \end{pmatrix} \]

For linear polarization at position angle \( \chi \) the standard normalized Jones vector is

\[ \begin{pmatrix} \cos \chi \\ \sin \chi \end{pmatrix} \]

For general elliptical polarization, the standard normalized Jones vector is

\[ \begin{pmatrix} \cos \chi \\ \sin \chi \cdot e^{i\Delta} \end{pmatrix} \]

where \( \Delta \) is the phase difference between the x and y components (in the sense \( \phi_y - \phi_x \)). More symmetrically, whenever absolute phase is not important, this is rewritten:
Circular polarization is represented by \( (\chi = 45^\circ, \Delta = 90^\circ): \)
\[
\begin{pmatrix}
\cos \chi \cdot e^{-i\Delta/2} \\
\sin \chi \cdot e^{+i\Delta/2}
\end{pmatrix}
\]

The Jones vectors
\[
\begin{pmatrix}
m \\
n
\end{pmatrix}
\]
and
\[
\begin{pmatrix}
-m^* \\
m^*
\end{pmatrix}
\]
represent mutually orthogonal polarization forms. Any Jones vector may be expressed as a linear combination of any such pair of orthogonal Jones vectors; the standard normalized Jones vectors form a complete orthonormal set. In practice, pairs of orthogonal linear polarization forms ('horizontal' and 'vertical') are almost always chosen as the base of the complex vector space.

This completes the thumbnail sketch of the Jones vectors. To represent the action of a medium on polarized radiation, matrices are again employed. These matrices now have \( 2 \times 2 \) elements, which, however, are complex. A non-polarizing and non-retarding absorbing medium absorbs both components equally and causes no relative phase shifts; it has the matrix
\[
\begin{pmatrix}
t & 0 \\
0 & t
\end{pmatrix}
\]
Here \( t \) is the \textit{amplitude} transmittance \((0 < t \leq 1)\), while what we usually measure in astronomy is the \textit{intensity} transmittance \( t^2 \). For a linear polarizer with maximum transmittance for electric field vibrating along the \( x \)-axis, the matrix is:
\[
\begin{pmatrix}
t_x & 0 \\
0 & t_y
\end{pmatrix}
\]
where \( t_x^2 \) and \( t_y^2 \) are the intensity transmittances for an electric field vibrating along the \( x \)- and \( y \)-axes respectively \((t_x > t_y)\). The term \textit{diattenuation} is sometimes used to describe the phenomenon of \( t_x \neq t_y \).

Just as with the Mueller matrices, the order of the matrices is important (the matrices do not in general commute). Thus, one cannot easily establish the matrices for media which both retard and absorb within the same volume of space. For these applications, Jones developed an extension of his matrix calculus, essentially splitting the medium into an infinite number of infinitely thin layers; for an infinitely thin layer, one may think of the two actions taking place one after the
other, and the order is immaterial (the infinitesimal matrices do commute). The infinitesimal result is then integrated to obtain the result for the entire medium. The differential matrices employed are known in the optical literature as Jones N-matrices.

Another specialized development of the Jones calculus, suitable for polarization ray-tracing of complete optical systems, uses three-dimensional complex electric field vectors, to allow for rays inclined to the system optical axis. Optical components are represented by $3 \times 3$ matrices with complex elements; the advantage for computer ray-tracing is that one system of coordinates can be used for the entire optical system, rather than many local systems and the transformations between them. For analysis and optimization of systems including polarization optics such polarization ray-tracing is indispensable, and it is being incorporated into optical design software.

### 5.3. Polarimetric instrumentation in practice

In this section, instrumental principles will be discussed, with emphasis on system behaviour.

#### 5.3.1. Telescopes

The first optical element of an astronomical observing system is always a telescope (disregarding the atmosphere for the present discussion). It is important to realize that, in general, a telescope will modify the polarization of the radiation before the polarimeter measures it. It is equally important to have some general feeling for the conditions under which such modification is likely to be appreciable and how it can be minimized.

The guiding principle is symmetry; any departures from full symmetry will modify the polarization. The considerations below illustrate this, but full understanding will require mathematical treatment by Mueller or Jones calculus, with optical constants applicable to the wavelength of interest.

Oblique incidence on a mirror produces both diattenuation (polarizing action) and retardation (wave plate action). These effects are minimal at near-normal and, somewhat surprisingly, at grazing incidence; the largest effects occur at intermediate angles of incidence, the details depending on the values of the real and imaginary parts of the refractive index (which in their turn depend on the wavelength). In general, Coudé and Nasmyth telescopes will exhibit strong polarization modification, while prime-focus, Cassegrain and Gregorian systems will be relatively free from it.

Rotationally symmetric telescopes of large focal ratio (slow optical systems) show very little linear polarization of unpolarized incident radiation for an image on the optical axis, since the
polarizing action of different parts of the mirror(s) is basically radially oriented and the resultant averages out.

Figure 5-3: The LEST solar polarization-optimized telescope (alas, never funded, so never built)

Note that even rotationally symmetric telescopes are not ideal, they do not transmit the state of polarization of incident radiation to the on-axis focal image without any change whatsoever: a certain amount of net (pseudo-)depolarization of the incident radiation will remain after the averaging; the converted linear polarization averages out to zero, with the result that the diagonal elements are less than unity. This is not all, however: contrary to the lowest level of intuition, the off-diagonal elements do not all average to zero, either; what remains is (rotationally invariant) coupling between Q and U and separately between I and V (but not between these pairs of Stokes parameters; see McGuire and Chipman (1988), particularly J.O.Stenflo's foreword to that report, which makes clear that with a more sensitive intuition we should really have expected a result of this nature).

Note: When the telescope aperture is sufficiently small in terms of the wavelength of the radiation, the diffraction pattern will extend beyond the image determined by geometrical optics, and a single point in the focal plane (e.g. the on-axis image) will receive radiation from more than one direction in the sky: the spatial point spread function (antenna pattern) will have sidelobes, which have polarization properties of their own. This is a problem mainly in the radio region of the spectrum.
For off-axis images the rotational symmetry is broken; the angles of incidence on different parts of the mirror(s) are no longer symmetrically distributed, so that residual polarizing action will exist, larger for faster (lower focal ratio) optical systems. Schmidt et al. (1993) discuss a 'software beam-switching' system for a wavelength of 2.8 cm, and in their very instructive figure 4 show (linear) polarization antenna patterns for each of the four offset focus antennas. The focal-plane antennas of the VLA and VLBA radio telescopes are offset, in order to allow front-ends for several frequencies to be mounted permanently. This introduces asymmetry, instrumental polarization and a different primary antenna pattern for the two orthogonal polarization channels of the receivers (the latter feature causes problems in obtaining the highest polarization accuracy from these synthesis arrays; the Westerbork telescope does not suffer from this, but requires more complicated operations to change frequency); the matter is discussed and references are given in Spoelstra (1992). An optical case is discussed by Sánchez Almeida and Martínez Pillet (1992), who conclude that only fast, wide-field systems are likely to show measurable effects in the optical region (cf. Schmidt et al. (1992), section 2.1, on the original Multi-Mirror-Telescope).

The design of 'polarization-free' telescopes is becoming an important part of astronomical engineering. Any oblique reflection that is not symmetrical about the optical axis should be avoided or made innocuous by preceding it with a polarization modulator; the LEST design is an example of the latter approach. Alternatively, but less fundamentally, some kind of compensation may be used. The ideal solution is a full-aperture modulator as the first optical element of the telescope; the only such system that I know of is WISP, a far-ultraviolet Schmidt telescope of 20-cm aperture with a rotatable wave plate as its entrance window (Nordsieck et al. 1994b).

Even when the telescope is rotationally symmetric and we are observing the complete seeing disc as a single entity, on-axis, we may expect some polarizing action by the mirror surfaces, owing to technical imperfections constituting a lack of symmetry of the internal structure of the surface. Early aluminium coatings for optical mirrors sometimes showed strong polarization, which could also depend strongly on wavelength. This was traced to asymmetric and too rigorous cleaning of the surface before coating and/or to oblique incidence of the aluminium atoms during the coating process; it is generally not a significant problem in modern telescopes. Similar effects are to be expected from internal details of the mechanical support structure or the reflecting surface of radio mirrors; at short X-ray wavelengths, structure in the 'mirror' at the atomic-lattice level may be expected to cause trouble, while at longer wavelengths harmful residual structure could remain after diamond-machining to produce the complicated shapes required for X-ray mirrors. The perennial questions will be to what extent a mirror is ideal and what else it does.

The analysis of polarimetric errors usually deals with the entire optical system, i.e. including the telescope. Indeed, in synthesis telescopes it would be difficult to say where the 'telescope' ends and the 'instrument' begins.

5.3.2. Optical polarimeters

Optical polarimetric instrumentation has a long history of development. Early polarimeters had errors at the level of a few tenths of a percent at best, and polarization signals were small, so
that polarimetry was very much a specialist craft. B. Lyot was the first to obtain very high accuracy by devising a modulator and using it on the Sun. For stars, the signals were generally so small that photon shot noise was appreciable, and there was little incentive to design sophisticated systems of unavoidably smaller throughput.

The situation has changed drastically within the last decade or two. Larger telescopes are available, CCD detectors now offer thousands of parallel channels of potentially very good accuracy, and improved modulators of high transmission have been devised. The higher signal levels have meant that greater resolution (spectral, temporal, spatial) can be used, and this has had the effect of increasing the degree of polarization provided by nature (less smearing of polarizations from neighbouring resolution elements); the end result is that (i) many more situations within astronomy can usefully be tackled by polarimetry without exceptional cost in telescope time and (ii) 'common-user' polarimetry is becoming available in the optical/near-infrared wavelength region (the 'CCD domain'). The latest development is that such polarimetry is becoming available in the 1-5 μ region, where the detector arrays are improving fast and modulators similar to those at optical wavelengths can be constructed; as in all infrared systems, much of the detailed design work is concerned with cooling as much of the system as possible, and the resultant equipment may look very different from its visible-wavelength counterpart.

All optical and near-infrared polarimeters can be analysed by Mueller matrix calculus (or Jones calculus when an interferometer is part of the system). It is sufficient to split the instrument into its simplest components, multiply all the matrices together and inspect the top row of the resultant matrix. If one expresses the response of a real-life detector (which depends on polarization of the light striking it) as that of an ideal detector (which responds only to I), preceded by a fictitious optical element which one includes in the matrix train, the top row of the matrix for the total optical train specifies the output I signal (and hence the detector output) for any input Stokes vector, and describes both the intended mode of the polarimeter and its errors. For an example, see Tinbergen (1973). If (as is usual in a modulation polarimeter) the polarization state of the light striking the detector does not vary with the state of the modulator, it is not necessary to know the details of the fictitious optical element within the detector. Those details enter only as a fixed gain factor, whereas we are interested in the quotient AC/DC, which represents the normalized Stokes parameters Q/I etc.; the fictitious optical element is therefore in general omitted, or included within an arbitrary gain constant.

5.3.2.1. Modulators

Polarization modulation is essential to accurate polarimetry in the optical spectral region. The technique used most often is to vary a retarder within the instrument. Radiation of both orthogonal polarizations passes through the same components for most of the instrument, and one strives to modulate only the polarization preference, leaving the Stokes I sensitivity constant; this generally means either switching a retarder from one state to another by some means, or rotating a constant retarder. Modulation may be included in the Mueller and Jones matrices of the system.

Two modulators are shown in Figure 5-4. They are examples only, but serve to illustrate the basic principles. The general form of their Mueller matrices, including the analysing polarizer but excluding component 1, is:
For the circular modulator
\[
\frac{D}{A} = f(t), \quad \left| \text{Alt}(B \text{ and/or } C) \right| = \varepsilon, \quad \left| \text{Alt}(A) \right| \approx \varepsilon^2
\]
while for the linear modulator
\[
\frac{(B \text{ and/or } C)}{A} = f(t), \quad \left| \text{Alt}(D) \right| = \varepsilon, \quad \left| \text{Alt}(A) \right| \approx \varepsilon^2
\]
where \( f(t) \) represents an alternating function of time with amplitude close to unity, \( 0.5A \) is the transmission for unpolarized light, "Alt(x)" is short for "the part of x that alternates with the same frequency as \( f(t) \)" and \( \varepsilon \) is a small quantity (preferably of order 1%). The modulating function \( f(t) \) is a square wave for the circular modulator and a sine wave for the linear one; including component 1 of Figure 5-4 in the instrument would reverse this and also exchange the values of \( B \) (and/or \( C \) and \( D \)).

Figure 5-4: Two polarization modulators

In the optical region, the only good polarizers available are linear polarizers (commercial 'circular polarizers' are always a combination of a linear polarizer with a quarter wave plate). Until the beam passes through the analyser, its Stokes \( I \) is constant and all that happens is that the
polarization form of the polarized part of the beam is switched between the two eigenmodes of the analyser (which in this case transmits one, absorbing the other, like a Polaroid). The function of the analyser is to convert the modulation of the polarization into modulation of Stokes $I$, which can be detected reliably by standard electronics.

Component 2 is the actual modulator. As the circular modulator rotates, the halfwave sections reverse the sense of circular polarization; this is a $+/-$ type of modulation, leading to a square wave of known phase in the output of the analyser. In the linear modulator, the direction of vibration beyond the rotating halfwave plate rotates as well (twice as fast as the halfwave plate), leading to a sine wave in the analyser output; the phase of this sine wave corresponds to the polarization angle at the input. In both cases, the size of the modulated Stokes I component in the output beam, as a fraction of the total Stokes I signal, corresponds to the degree of polarization of the input light.

![Diagram of photo-elastic modulator](image)

**Figure 5-5: The photo-elastic modulator**

Rotating components generally limit modulation frequencies to about 100 Hz, which is insufficient to suppress all scintillation noise. 'Electro-optic crystals' (or 'Pockels cells') can modulate faster, as can 'photo-elastic' (or 'stress-birefringence') modulators. The former type consists of a crystal that changes its birefringence when an electric voltage is applied to it and is generally operated as a square-wave modulator. The latter type (Kemp 1969) is a piece of glass (or fused silica for ultraviolet transmission) in mechanically resonant oscillation and thus with time-varying stress-birefringence (Figure 5-5).
Since the last component is a linear polarizer, the Mueller matrix of the photo-elastic modulator has the same general form as before; in this case

\[
\begin{align*}
A &= 1 \\
B &= 0 \\
C &= C_0 + \text{even harmonics of the (mechanical) resonance frequency} \\
D &= \text{fundamental and odd harmonics of the resonance frequency}
\end{align*}
\]

Tuned amplification or synchronous demodulation in the electronics is then used to select the desired periodic term (generally, but not always, the fundamental). The photo-elastic modulator is the most nearly perfect polarization modulator known; its maximum birefringence is only about one part in 40,000, so the oscillations do not influence the optical path of the light beam in any significant way (and therefore there is no significant spurious periodic signal which could be interpreted as being polarized); it can also be perfectly transparent over a wide range of wavelengths and will tolerate large angular width of the beam. The only flaw of the photo-elastic modulator is that it cannot easily be constructed in achromatic form (however, it can readily be tuned in wavelength).

More complex modulators, for 'full Stokes polarimetry' (i.e. simultaneous determination of all four Stokes parameters), are required for solar and specialized stellar work.

### 5.3.3. Radio polarimeters

We have seen that, in optical polarimeters, orthogonal polarization signals follow the same path through the optical system. Birefringent components are used to generate a distinction between these two signals, but the same system gain applies to both of them separately, so that, apart from the system gain as a multiplying factor, the difference between them is reproduced faithfully. The gain can be calibrated reliably by observing a polarized source. Radio receiver channels are normally not birefringent and they lack the capability to process a pair of polarization signals within one physical channel; two channels are required, and these channels have different gains, the ratio of which inevitably varies somewhat with time (and in astronomy we are particularly sensitive to this, since we usually have a small polarization difference between two large signals, each representing about half the total radiation). One could calibrate the time-varying ratio, but this would only transfer the problem to the stability of the calibration arrangements (both the source and the coupling circuitry). Rather than determining a difference or ratio of two large and spuriously varying intensity signals, in radio-polarimetry one determines the polarized component directly, by measuring the correlation between the electric fields of orthogonal polarization forms. This is possible because in radio receivers one has available an electrical signal which represents wave amplitude and phase rather than 'intensity'. The correlation technique has been introduced above. Polarization processing is carried out mainly at intermediate frequency.

**Note:** Another peculiarity of radio techniques is that the photons are of such low energy that even the smallest signals contain very many of them and photon shot noise is vastly exceeded by the noise of the first stage of electronic
processing. This means that after amplification one may divide a radio signal into a number of identical copies and process these independently without worrying about introducing extra photon noise by having split the energy of the signal over several channels. One has indeed split the signal, and therefore fewer photons are available in each channel, but the effect of this on total noise is negligible. In synthesis instruments, one correlates simultaneously the output of each (telescope/polarization) channel with all other channels, and achieves a gain in observing speed by this massive parallelism. In contrast, if one were to divide an optical signal into many parts, one would increase the fractional noise in each channel and there would be no net gain in total speed; this is because the best optical detectors have very low noise compared to that of the energetic optical photons, and dominant noise in most optical systems is therefore 'noise-within-signal' (photon shot noise) rather than amplifier or detector noise (however, CCDs have only recently attained 'one-electron' readout noise, and infrared arrays do have significant readout noise).

Figure 5-6: A polarization receiver of the 1960's, to clarify basic principles

A 'correlator' is a standard component in radio astronomy; aperture synthesis depends on correlation-type interferometry. Basically, a correlator multiplies the instantaneous voltages in two receivers. Figure 5-6 shows how, for linearly polarized feed antennas, this can be used to measure linear polarization, with rotation of the feed antenna to obtain both Q and U. Unpolarized, and in this case circularly polarized, signals do not lead to an output in the correlator channel.
Figure 5-7: Arbitrary-polarization feed antenna

The feed antenna system can be made sensitive to other polarization forms, in particular to circular polarization. Figure 5-7 shows how elliptically-polarized feed antennas can be conceived and actually constructed as a network linking two linearly-polarized antennas. Seen as a transmission system fed from terminal A, this is a hardware implementation of an earlier section where all polarization forms were conceptually generated from two correlated linear polarization signals in certain proportions and with a 90° phase difference between them.

As a receiving system, it produces an output at A whenever the dipoles receive correlated signals which correspond to the polarization form selected. When \( \beta = \pm \pi/4 \), the system is adjusted for transmission (and reception) of circularly polarized radiation; \( \beta = 0 \) or \( \beta = \pi/2 \) denotes adjustment for linear polarization.

5.3.3.1. Polarimetry by synthesis arrays and VLBI

Aperture synthesis and VLBI systems use correlators in one form or another to determine the correlated part in the signals received by different telescopes. Such correlations provide information about the angular distribution of radiation over the sky, since they are values of the spatial auto-correlation function and the relation of that to the angular power distribution in the sky is a Fourier transformation.

It is a relatively simple extension, in principle, to cross-correlate (in four combinations) two orthogonal polarizations of the field received by each of two different telescopes and thus to
determine values of the spatial auto-correlation function of Stokes $Q$, $U$, or $V$ as well as $I$. Aperture synthesis and VLBI techniques as such are not the subject of this book. The polarimetry extension may be summarized as follows:

In correlating the outputs from two elementary telescopes of a synthesis or VLBI installation, there is a choice of two orthogonal polarizations in each; therefore four instantaneous combinations can be made, which is just sufficient to determine all four of the Stokes parameter visibilities $I$, $Q$, $U$, $V$. The time series of $I$, $Q$, $U$, $V$ visibilities can be Fourier transformed into maps of their corresponding sky distributions $I$, $Q$, $U$, $V$, after which sky distributions of degree and angle of polarization can be derived.

For a more detailed description, the reader is referred to Christiansen and Högbom (1985, sections 7.13, 7.14) and to Thompson et al. (1986, p. 150). The mathematics is all in terms of complex numbers to represent the electric fields and the currents or voltages into which they are translated. This leads to complex correlator outputs, representing 'complex visibilities' containing amplitude and phase of the sinusoid resulting from a source moving through the interferometer antenna pattern. These complex visibilities $I$, $Q$, $U$, $V$ are Fourier transformed into real sky distributions.

The hardware necessary to measure a complex visibility includes 'fringe-stopping' (which reduces the frequency of the sinusoid to exactly zero for the 'fringe-stopping-centre' of the field and to near zero for other points) and 'complex correlators' consisting of two simple correlators in quadrature, i.e. with 90° phase delay in one of the inputs of one of them.

A single-dish correlation polarimeter may be regarded as a special case of a two-element polarization interferometer, viz. one with orthogonally polarized channels and zero antenna spacing. This concept is useful when tracing much of the relevant mathematics; note that $I$, $Q$, $U$, $V$ are now synonymous with $I$, $Q$, $U$, $V$. Single-dish polarimeters are more sensitive to local interference than are widely spaced correlation interferometers, since the two receiving dipoles are in the same location and spurious local signals are likely to be correlated.

VLBI polarimetry is basically the same as polarimetry with any synthesis instrument, the only difference is in the way the link between the telescopes is implemented.

**5.3.4. Infra-red and (sub)mm polarimeters**

Hough et al. (1994) report on a successful infrared (1-2.4 μ) polarimeter using a warm rotatable halfwave plate outside the cryostat (Figure 5-8). A retarder is not a lossy component, hence its temperature should be irrelevant; to the extent that this is true for a practical optical component (i.e. imaginary part of refractive index negligible for the wavelength range admitted to the detector), this arrangement should work, considerably simplifying construction, hence allowing more sophisticated polarimetric arrangements. It remains to be seen how far into the infrared this technique can be pushed (interestingly, it has been used successfully at 1.3 mm, in a hybrid instrument using an 'optical' waveplate modulator and a 'radio' twin-feed focal-plane receptor).
Fig. 6.16 An infrared imaging polarimeter with the retarders outside the cryostat, from Hough et al. (1994); IRIS is the Anglo-Australian Telescope’s multi-purpose infrared imager/spectrometer. The fact that the (often compound) retarders do not have to be subjected to temperature extremes allows a simpler and more reliable design with less stress-birefringence; having the retarders accessible is also more convenient operationally.

Figure 5-8: An infrared imaging polarimeter
6. The Earth's atmosphere

Both outer space and instruments are clean, but the atmosphere is filthy. This is an exaggeration, but it is not all that far from the truth. While we can be fairly sure of conditions over light-years out in space and we can test our instruments until we know almost exactly what they do, the Earth's atmosphere smears our images, attenuates the light, rotates the plane of polarization, depolarises our signal and makes point sources twinkle. Astronomers have found some remedies against these effects, but real progress is being made only by instrumental developments of the recent hi-tech era. The central theme of hi-tech approaches is to both map the 'instantaneous' atmosphere using a calibration source and correct for the atmospheric effects on short time scales; one then integrates the corrected data. Computers are essential, as are adaptive optics, controlled-phase arrays, array detectors, relatively large telescopes (because of the short time scales), etc. Did I say: hi-tech? Do I hear anyone say: large projects, lots of engineers? Well, such is the nature of modern astronomy; cultivate your garden if, like me, you don't particularly like it........

6.1. Atmospheric dispersion

Everywhere except in the zenith, the atmosphere bends the rays of light as they travel through it. The effect is similar (in reverse) to what happens at a water/air interface (the kink in the stick one pushes into the water), even though the atmosphere constitutes a continuous change of refractive index rather than a step change and is not a plane-parallel layer. This bending is worse for blue than for red, since the refractive index of air is larger for blue than for red. The end result is that in red light stars appear to be lower in the sky than in blue light. For narrow-band observations, one could compute the effect and correct the telescope pointing for any difference in median wavelength between telescope autoguider and science instrument. However, for really wideband work, one ends up with a coloured smudge instead of the crisp star one hoped to get from the active and adaptive optics; in particular, it will be impossible to force all the light into the entrance slit or fiber-pickup for the wide wavelength range that some modern spectrometers can handle.

6.2. Seeing, scintillation and extinction

These 3 terms cover the observable effects of the atmosphere on optical signals. The term "extinction" denotes the absorption and scattering of light by dust particles and molecules, "scintillation" is the twinkling of the light from point sources caused by layers high in the atmosphere (the light being deflected so that it misses the telescope some of the time), while by "seeing" we mean the smearing of our images to much worse resolution than a diffraction-limited
telescope would produce (seeing is mostly caused close to the Earth's surface, until recently mostly within the telescope dome!).

1) Extinction is caused by several atmospheric components (such as Sahara dust on La Palma, volcanic dust worldwide after a big eruption, the molecules of dry air, water vapour in the infra-red, ozone in the near ultra-violet etc). Some of these components are constant for many hours at a time and over the whole sky, others may be variable on quite short timescales or over short distances in the sky. Again, some components of extinction are almost neutral, others may be variable on quite short timescales or over short distances in the sky. Lastly, to the extent that the Earth is flat, extinction varies with \( \sec \zeta \), \( \zeta \) being the 'zenith angle' \( (90^\circ - \text{elevation}) \).

Extinction includes both scattering and true absorption. In true absorption, the light energy disappears as such; it is converted into heat. In scattering, the light energy is redirected and contributes to the (sky) background, but it does not disappear.

Extinction is independent of telescope size.

![Typical atmospheric transmission in the mid-IR](image)

**Figure 6-1:** The almost-catastrophic effects of extinction in the thermal IR (optimal Chilean mountain site, overhead!)

2) Scintillation is caused by inhomogeneities in the atmosphere at a height of some kilometres. Phase gradients due to the pressure gradients of air turbulence cause mild deflections of the direction of travel of the wavefront, so that what the telescope intercepts is, at any one time, brighter or fainter than the average intensity. The pressure disturbances are moved around by upper air winds and one can deduce upper-air wind speed and direction by comparing scintillation in 2 apertures; this is potentially a way of correcting for scintillation.

The patches of brighter or fainter light are fairly small, so that small telescopes are more troubled than large ones (which average the effects of a number of elementary patches). For 5-metre telescopes and larger, scintillation is generally not an issue (but this does depend on the precision required and on other kinds of noise, therefore on bandwidth).
3) 'Seeing' is the historical term for the fact that the image produced by a telescope larger than 10 (or a few tens of) centimetres is not as sharp as one would expect from its size and optical quality, but is fuzzy, the fuzziness varying with time and weather conditions. A short exposure (milliseconds) of such an image reveals a large number of 'speckles' which are of the size expected for the telescope. These speckles move around, disappear and reappear on short timescales. The speckle image is in fact a result of interference between wavefronts of randomly moving and (dis)appearing subapertures defined by the atmospheric turbulence 'bubbles' in the first few tens to hundreds of metres above the telescope. The speckle image of a point source is in fact the instantaneous p.s.f. of the telescope+atmosphere; its smallest details (in angular measure) are inversely proportional to telescope diameter (in units of $\lambda$) and its total size is inversely proportional to the size of the 'bubbles' in the turbulent atmosphere (cf. Figure 4-10). The optical telescope is like a SKA tied-array version with randomly varying positions of the 'stations' or 'tiles' and with random phase changes in all of the connections; the saving feature is that some of the sources in the sky are bright enough to yield an image in a time short with respect to all the random changes, so that an 'instantaneous psf' can be defined and sometimes derived from the data.

The most efficient way to produce bad seeing is to build a dome round one's telescope and then to put electronics and people in it; the worst seeing I ever experienced was on a bitterly cold and windy night when I was doing polarimetry on very bright stars, so that I could afford to use a very large aperture round the star image; I had the electric floor heater switched on (which actually was only there for the comfort of technical staff working on the equipment during the day!). Scientific investigations carried out by the Royal Greenwich Observatory (†) as part of the preparations for the La Palma WHT included similar experiments with electric heaters.

If one compares speckle images of different stars taken at the same moment, then one finds they are only identical if the 2 stars are close to each other in the sky (this area is called the 'isoplanatic patch'; it is only a few arcseconds in the visual, increasing to arcminutes in the infra-red; like speckle restoration, adaptive optics basically only works over the isoplanatic patch area).

The Dutch Open Telescope (DOT) on La Palma (Figure 6-2) has been designed to reduce seeing in solar images to a minimum. The dome has been done away with, the telescope has been raised some 15 metres above the surface and allows the wind to blow away all convection bubbles that arise in spite of precautions (such as painting all structures white). At the focus, a (water-)cooled diaphragm isolates the portion of the solar image to be investigated and the heat of the rest of the solar image is removed. Somewhere in the optics, an interference filter transmits just the portion of the spectrum that is of interest and, again, the rest of the energy is removed (reflected back to the Sun in this case). The electronic camera system records and stores short exposures, so that later selection of the best momentary images is an option. Finally, post-processing speckle restoration techniques produce images of the solar photosphere that have unique spatial resolution.
6.3. Faraday rotation

Faraday rotation is rotation of the plane of linear polarization when an EM wave propagates along a magnetic field and free electrons are present in the medium as well. Faraday rotation is proportional to the square of the wavelength, so that (long-wave) radio astronomy is most vulnerable, the Earth's magnetic field being strong enough to produce measurable rotation within the ionosphere, at a height of about 100 km and up.
When Faraday rotation is small, the polarization map one measures is not quite what is actually there in the sky. When the rotation is large, differential rotation within the passband or within the antenna beam produces depolarization. In synthesis arrays, differential rotation can produce decorrelation; for LOFAR this problem has to be taken into account from the start in system design.

At optical wavelengths, extremely strong magnetic fields are required and Faraday rotation is not an atmospheric problem.

6.4. Undoing the effects of the atmosphere

6.4.1. Atmospheric dispersion correction

Using prism(s) in the general neighbourhood of the focal plane (Figure 6-3), one may compensate for the bending of the light rays by the atmosphere. Such compensation is approximate, since the wavelength-dependence of the refractive index of air is not identical to that of fused silica, but for the very small angles involved the arrangement is satisfactory. At the ESO FORS website one may find a link to "our paper presented at the 1996 Landskrona conference", which discusses 3 different designs for ADC's; I lifted Figure 6-3 from that paper.

![Figure 6-3: Atmospheric dispersion correction in the FORS imager/spectrometer](image-url)
6.4.2. Eliminating extinction effects

Astronomers have 5 main methods in their arsenal to combat extinction; they may be combined in some cases:

1) Select a 'photometric' site, such as the Chilean mountain locations.
2) Assume that at least some stars are constant in their light output. At levels of 1% this is true of many stars, at 0.1% or less there are very few stars that can be relied upon, but one may use averages of, say, 'all the stars on a CCD frame' as the virtual star one assumes to be constant. Observations of such stars at the same time as the programme stars then allow one to calibrate the atmosphere and correct for it. 'Differential photometry' using a nearby comparison star for reference is a standard procedure.
3) Make a ratio measurement, if the ratio is what is astrophysically significant; polarization and spectral absorption-line strength are the classic examples, but the same principle applies in simultaneous multi-passband photometry: atmospheric extinction variations at different wavelengths are correlated, so that "instantaneous photometric colours" (=ratios of simultaneously-measured intensities at different wavelengths) are more stable than magnitudes. Even in the presence of 50% variations in signal strength, observations with accuracy 0.01% are feasible in the most favourable cases.
4) Use the $\sec \varepsilon$ variation. A standard procedure is to observe 'extinction standards' at large zenith distances in between observations of one's programme stars. Specialised methods allow this to be exploited to better than 0.1%. Painstaking work over years with this method allows one to identify the stars which are constant enough to be used in method 2.
5) Go to space (the ultimate photometric site) for primary calibration of the network of standards, repeating the previous steps.

6.4.3. Avoiding scintillation effects

Scintillation manifests itself as an extremely time-variable form of extinction of a broad-band kind (varies only slowly with wavelength), so that a very rapid (preferably instantaneous) ratio measurement is the correct solution here. In practice one splits the light with a beamsplitter, or spreads it into a spectrum, or alternates rapidly between the quantities one wishes to compare. The first 2 alternatives use different detector pixels for the quantities to be compared, so that the ratio of pixel sensitivities comes into the measurement and must be eliminated by calibration or by experimental procedures such as moving the image around on the detector. If one alternates rapidly between the quantities to be compared, the same detector pixel is used for both and the sensitivity drops out of the ratio; however, one is using the system at only 50% efficiency, since half the time one is observing the comparison quantity which is of no inherent interest (moving the image around on an array detector eliminates this disadvantage).

Since scintillation is due to traveling disturbances which change only relatively slowly, one could in principle surround a large telescope by a ring of smaller ones on the same mounting and detect the disturbances as they arrive. Having measured wind speed and direction from the small-telescope results, one would correct the results of the central telescope for the atmospheric effects deduced from the interpolated results in the smaller telescopes. Up till now, this has not been
worth the trouble, simple averaging over the pupil of the large telescope has been sufficient (large telescopes are used in particular for long exposures on faint sources, so photon noise and extinction noise tend to dominate over scintillation noise).

6.4.4. Undoing the effects of seeing

There are two approaches to correcting for the effects of seeing. Each has its virtues and it is conceivable that future systems will use adaptive optics to correct the errors for a 'central' area within the image, after which optimized speckle algorithms may extend the area of good correction to outside the isoplanatic patch. Adaptive optics requires considerable online hardware and computing, while speckle methods use large amounts of computing after data-taking and require high signal/noise in short-exposure images. Control of the final p.s.f. is better with speckle methods, although CLEANing of the image can always be used if the recorded image is oversampled. Adaptive optics is the essential method whenever the input aperture of an instrument has to be small to limit the amount of background sky radiation.

6.4.4.1. Adaptive optics

The principle of this approach is to assume that a particular source within the field of view is actually a point source and to manipulate the instrument optics in such a way that one reproduces a point source in the image (compatible with the diffraction-limited psf of the telescope, as given by FT theory); this requirement calls for the instrument optics to be manipulated to counteract the manipulations by the atmosphere, in a servo arrangement. If there is no suitable point source in the field, it is sometimes created by firing a powerful laser beam at the layer of sodium atoms high up in the atmosphere.

As in all servos, one needs a sensor, a reference and an actuator. The sensor is the same as for the active primary mirror of the telescope: the low frequencies of the error signal are used to drive the primary mirror, the high frequencies are used for the adaptive system. The sensor records the wavefront from the assumed point source by an auxiliary instrument; it generally uses a rapid-readout CCD as its detector. In a computer algorithm, the wavefront is compared with that expected in the absence of an atmosphere, and the various aberrations (of the atmosphere considered as an optical component) are derived several hundred times per second. Of the rapid errors, those of position are corrected by a tip/tilt mirror and those of wavefront shape are corrected by a deformable mirror of some kind (such as piezo-driven thin silica disc or a membrane) placed roughly at a pupil-replica position (actually at a position optically conjugate to the atmospheric layer that produces most of the seeing phase changes; a so-called multi-conjugate adaptive-optics system can correct for the phase changes produced in several discrete atmospheric layers; multi-conjugate systems are still mostly pie in the sky – excuse the pun – but the Gemini adaptive-optics page has a link to it [which failed when I tried it in August 2000]; the page also has links to various other AO projects and is a good starting point).
6.4.4.2. **Speckle methods**

A short-exposure speckle image shows high-spatial-frequency components which constitute a random array of diffraction-limited images (the p.s.f. of the telescope+atmosphere is a randomly-placed assembly of replicas of the p.s.f. of the undisturbed telescope). These spatial-frequency components have recognisable and fairly stable amplitude, but their phase varies randomly. A meaningful measurement is therefore the *power* spectrum of these spatial frequencies. This power spectrum may be obtained as the (2-dimensional) FT of the 2-D spatial auto-correlation of each individual speckle image, and time-integration of the power spectrum is then the way to reduce noise. This does not yield a true image of the object, but for double stars it is sufficient to be able to deduce separation and magnitude difference (and, with 180° ambiguity, the orientation).

Other FT-based manipulations of the speckle images can yield most of the *relative* phase information ("where are the objects, relative to the brightest source in the image?" as in VLBI) and can include full-polarization treatment. In solar speckle, one finally has to stitch together a large number of fractional images, each only the size of the iso-planatic patch for LOFAR, something similar will be necessary, perhaps by techniques which are not too different from those in the optical domain.

Figure 6-4 shows how good earthbound imaging can be, if done properly by cascading different methods. Bear in mind that at most sites the 'very best' *nighttime* seeing 1 or 2 decades ago used to be of order 1 arcsec (the smallest scale divisions in the inset picture); "sub-arcsec seeing" was then the holy grail. The DOT achieves good primary *daytime* seeing by suitable placement and construction, then image selection is practised to obtain the top picture and finally data massaging by 'piecewise speckle restoration and rubbersheeting' produce an image of very superior resolution (better than 0.3 arcsec, by the looks of it) and a very high degree of credibility. And all this with a telescope actually following the Sun, whereas the nighttime astronomers looking for sub-arcsec seeing would not even permit the dome to be opened during the day, for fear of deforming the telescope mirror or charging up the heat reservoir formed by the telescope structure or the dome interior; this shows the value of proper engineering (it is fair to comment that the enormous signal provided by the Sun has helped, in that high spatial frequencies were amplified by factors up to about 1000 during speckle restoration; but it is good engineering to design the system so as to take advantage of this happy circumstance: high-speed and high-S/N data recording is an essential link in the chain).

More pictures of sunspots at best-ever resolution can be downloaded from the [DOT database](#).
Figure 6-4: DOT imaging and data-processing of a sunspot
6.4.5. Derotation of Faraday-degraded data

There is only one way of removing the Faraday rotation produced by the ionosphere; that is to measure it and then to use the data for correction of all results taken under equivalent circumstances. One needs a source of known linear polarization, or the assumption that the intrinsic polarization (the polarization as it was before the wave struck the ionosphere) of a celestial source does not vary with wavelength (not always a good assumption, since there may be interstellar or internal source Faraday rotation). In the latter case, one uses the $\lambda^2$ variation of the rotation to find the polarization at $\lambda = 0$.

Since the ionosphere changes only relatively slowly and has a fairly large spatial scale, models of the ionosphere with a small number of adjustable parameters play a useful role. Measurements of signals from calibrated celestial sources or from GPS satellites can then be used to determine the variation of those parameters with time and position, after which the model is suitable as a basis for correction of measurements.

In practice, removal of Faraday rotation is as extensive and specialist a subject as correcting optical photometry for extinction. It is beyond the limits of this course.
7. Optical-astronomical Instruments

7.1. Telescopes revisited

Telescopes have evolved considerably within the last decade or so. Let me paraphrase the conclusions of Jan Willem Pel's Zenit paper (in Dutch), on the ESO VLT now being completed (comments by JT):

The VLT is special, because it has been designed within one coherent strategy (system engineering as it should be, of course):

- A perfect site; dry, very little cloud, extremely good seeing.
- Optimally-designed domes, which do not spoil the naturally good seeing.
- Single mirrors of the largest size and best precision compatible with state-of-the-art techniques.
- Active-optics systems in the mirror support cells, to retain the excellent mirror characteristics under varying telescope attitudes and temperature regimes.
- An extensive suite of cameras and spectrometers with large and sensitive array detectors for the wavelength range 0.3 to 28 μm (Figure 7-1); this in itself is revolutionary: far too often in the past, all the money was spent on the telescope and little was left to exploit its characteristics properly until the telescope itself was obsolescent.
- Adaptive-optics systems to achieve diffraction-limited performance.
- Beam transport, in underground tunnels, for all telescopes (prime 8-metre and auxiliary 1.8-metre) for aperture synthesis with very clean synthesized beams.
- Survey telescope for all kinds of preliminary work with the same excellent site characteristics.

![Figure 7-1: The initial instrumentation suite for ESO's VLT](image-url)
Modern telescopes are all reflectors, with just possibly a refractive field-flattener lens or corrector plate. However, this does not mean that all telescopes are equal or similar; far from it. Depending on the scientific project, one may require a particular telescope design (optical or mechanical). Also, as telescopes grow larger in size, one has to relax some tolerances on mechanical stiffness and then to compensate for this somehow (optical feedback on mirror figure, telescope flexure model, etc). This initial subsection is intended as a superficial guide through the engineering discipline of "large telescope design".

The primary aim of a telescope is to collect as much light as possible into as sharp an image as possible, with an image scale that matches available detectors. This historical aim has produced the Cassegrain design (paraboloid/hyperboloid), or alternatively the Gregorian (paraboloid/ellipsoid). Imaging is 'perfect' on-axis, but the (angular) field of good definition is small. By modification to the mirror surfaces, it is possible to increase the field of good definition considerably, but the tolerances on mirror spacing and alignment are then much stricter (so that a 'Ritchey-Chrétien system' has a much more expensive mechanical environment than a mere 'Light Collector'). The increased field of a R-C system, however, can be put to good scientific use in wide-field imaging and multi-object spectroscopy (if, that's to say, instruments are designed for such a large field of view and the scientific aim calls for many closely-spaced celestial objects). Even larger fields of good definition are obtained by Schmidt telescopes. In this 50-year old design, the primary mirror is spherical and oversized, while the defining and entrance pupil is at the corrector plate upstream. This plate modifies the wavefront before it strikes the primary mirror, in such a way as to counteract the spherical aberration by the primary mirror, over a large field of view. The focal plane is inside the telescope (which is acceptable for a small filter-plus-detector unit). A 'Schmidt-Cassegrain' has both a corrector plate and a secondary mirror, to make the final focus more accessible (ASTRON: see Ralf Ottow's private telescope). The corrector plate of Schmidt systems is slightly chromatic (unless executed as an achromatic doublet) and in sizes larger than about 1 metre tends to sag too much under its own weight.

A Cassegrain-type telescope for infra-red use generally has a 'chopping secondary' mirror, to alternate between a source and empty sky; this is a necessity because of the strong (thermal) background generated by the 250-300 K atmosphere. The chopping secondary is usually the defining pupil. The JCMT (sub)mm telescope on Hawaii also has such a chopping secondary.

The mechanical design of precision optical telescopes is usually of the 'Serrurier-truss' type pioneered for the 5-metre Palomar telescope (the '200-inch'). This design does allow the mechanical tube to sag under gravity, but by proper proportioning one may ensure that primary and secondary mirror suffer only translation (and by equal amounts!), so that the optical system remains aligned.

For the largest telescopes, the alt-azimuth mount (axes horizontal and vertical) is preferred over the equatorial (axes towards the celestial Pole and at right angles to that), in order that the gravity vector acting on the telescope tube is confined to a single plane and mechanical design of instruments can be more straightforward. The telescope tube then rotates with respect to the sky and one requires an image rotator or an instrument rotator to compensate (such devices, and the Alt-Az drive itself, are computer-driven systems and therefore are a development of the last 2 decades or so, as far as optical-precision drives are concerned); in some polarimetric applications, one actually puts the telescope rotation to good scientific use in distinguishing between
instrumental polarization and true source polarization (as we did in Dwingeloo for 75-cm polarized Galactic emission in the 1960s).

Modern large telescopes use thin deformable primary mirrors, the actual operational shape of which is determined by an auxiliary optical system (wavefront sensor) and computer analysis, hundreds of computer-driven actuators forcing the mirror to conform to what is required (Figure 7-2). Such a system is known as 'active optics'. In addition to such active optics (which act on the primary mirror itself to correct slow errors and also to adjust the mirror to the optimal shapes for either Cassegrain/Nasmyth or Coudé), modern telescopes can have adaptive optics to correct for atmospheric effects. Such systems have a tip/tilt mirror to correct rapid image motion (too fast for the telescope drive system) and one or more deformable mirrors to correct for rapid wavefront deformations (each deformable mirror at a position conjugate to the atmospheric layer that causes the wavefront deformation). Active/adaptive optics constitutes another specialism of astronomical (and military, of course: spy satellites) engineering and are extensively reported in proceedings of conferences on telescope design for the 21st century.

**ACTIVE OPTICS AT THE VLT**

<table>
<thead>
<tr>
<th>No. of active, axial supports:</th>
<th>No. of lenslets in WFS:</th>
<th>Update interval:</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>30x30</td>
<td>30 sec (typ.)</td>
</tr>
</tbody>
</table>

**Wavefront Sensor (WFS):** Shack-Hartmann  
**Correction scheme:** Modal  
**Coma and focus correction:** on M2

Figure 7-2: The active support system for the primary mirrors of the VLT Unit Telescopes
An alternative to the single primary mirror is the segmented design used in the 10-metre Keck Telescope and needed for future 25-metre and larger telescopes (which are being studied now, both Earthbound and for space, just as seriously as SKA in the radio domain). In such segmented systems, capacitive sensors (one plate on each of a pair of neighbouring segments) and servo-driven active support of the individual hexagon segments are used to achieve effectively a single large primary mirror. Manufacture of the individual hexagons is a complex matter: they are deformed by predetermined amounts during polishing as an on-axis mirror, so that they will take on the desired off-axis shape when released afterwards (numerically controlled diamond-machining is an alternative, of course, but is also complicated). Figure 7-4 illustrates engineers' tentative projections towards an optical equivalent<sup>1</sup> for SKA: a ≈ 100-metre diffraction-limited telescope for the visible called OWL for OverWhelmingly Large telescope.

**Press comment**: At least the OWL acronym has the benefit of bringing to mind an appropriate mental image. Still, we wonder how much more extravagant astronomers can become in their nomenclature. What's next – the Boy It's Gargantuan (BIG) telescope? Or the Extremely Gigantic Observatory (EGO)?

---

<sup>1</sup> equivalent in the sense of "as yet equally speculative and preliminary"; no scientific or technological equivalence is implied.
After a telescope has been built and good optics have been achieved by active/adaptive methods, there is always some residual flexure in the supporting structure, leading to pointing errors. Using the known positions (Hipparcos) of a large collection of stars well-spread over the sky, the variation of pointing error with altitude and azimuth is determined and analysed into a 'pointing model'. That model is then integrated into the computerized telescope drive, thus leading to an almost 'perfect telescope with perfect pointing' (all to certain final, non-zero, tolerances, of course; to the engineer, tolerances are as important as, even part of, the intended functions of the equipment; the astronomer tends to ignore tolerances, preferring a holistic attitude of: "things are never as good as they are made out to be"; in a well-organized project, of course, tolerances are agreed beforehand and are based on certain scientific aims, they are subsequently used as design aims and are finally verified during commissioning and monitored during routine operation).

The progress from simple 'as-good-as-you-can-make-it' engineering feats of the early 20th century to the sophisticated giant structures of today and tomorrow has been achieved by deliberate relaxation of construction principles and tolerances of parts of the design, compensating for this by auxiliary instrumentation (largely by active feedback using passive defining components, but also by concentrating on essentials, such as recognising that the mechanics may be allowed to sag, as long as the optics stay aligned). The end result is that the cost of telescopes has risen considerably less steeply than the cost \( \propto \text{diameter}^{2.7} \) that had been predicted from previous experience; for telescopes such as OWL to become feasible, this trend must continue (to put it mildly).
7.2. Correlator interferometry and aperture synthesis

There is one more telescope property one may relax in aiming for high spatial resolution: the filled aperture. In a development similar to that of radio techniques some 30 years ago, future optical very-high-resolution telescopes will often be of the aperture-synthesis type, i.e. the spatial autocorrelation of the EM field will be sampled by combining, in correlation interferometers, the beams from pairs of individual telescopes. While radio waves are converted to (IF) electric currents or voltages and then transported by cable, optical signals must be transported at signal frequency, i.e. as light; conversion to electricity is done in the detector, which can be configured to function as a correlator. Transport of optical signals must be done by optics (such as mirrors or optical fibers); generally this means that 2 orthogonal polarization forms are transported through the same optical 'channel' (in contrast to the radio domain, where 2 orthogonal polarizations are picked up by orthogonal antennas and 2 corresponding electronic signals are transported through separate cables). This situation means that in optical interferometry care must be taken to combine within the correlator identical polarization forms from each telescope of the pair; failure to do so will in general lead to loss of correlator output. The technical difficulties posed by the Earth's atmosphere, the short wavelength of optical signals (which sets extremely high stability tolerances for mechanical structures and also means that instrumental noise has a different character), the absence of suitable amplifiers, and this polarization problem have been the cause of the 30-year delay between radio and optical aperture synthesis.

A very useful description of an early optical correlation interferometer instrument (excluding the collector telescopes, the adaptive optics and the delay lines) for wavelengths close to the visible is given in the AMBER report and for the thermal IR (Figure 7-7) in MIDI documents. The instrumental complexity for even a basic 3-baseline interferometer system is staggering. Nevertheless, the actual beam combiner instrument system is relatively small (1 \times 1 \times 0.5 metre, roughly, in AMBER).

As amply explained in Hamaker sections 12-14, the basic element needed for aperture synthesis is a two-element correlator interferometer consisting of 2 collecting apertures, a correlator (=multiplier) and transport of the signals (with relative phases left intact) from the apertures to the correlator. If 3 rather than 2 elements are available simultaneously, phase-closure techniques are possible. If some of the collecting apertures can be moved, then samples of the spatial autocorrelation function may be obtained one after the other; if the apertures are fixed, one needs a number of them at a sufficient number of different spacings and orientations to obtain an acceptable psf of the synthesized telescope after 12 hours of observations. The design of ESO's VLTI (which is the first large optical/IR telescope facility expressly designed with aperture synthesis in mind) is a hybrid: there are 4 large (8-metre) fixed telescopes and will be up to 8 auxiliary telescopes of smaller (1.8-metre) aperture, which can be set up at a number of locations (like the WSRT's movable telescopes); see Figure 7-5. The VLTI is a pioneering project which will materialize over decades; a good description of the present (early 2000) plans is given in the

\[\text{As in radio-astronomy, there is a need for filled-aperture instruments as well as the diluted-aperture systems that aperture synthesis provides (roughly: diluted apertures for high resolution of almost-point-sources, filled apertures for faint extended sources). Hence the development of OWL as well as VLTI.}
\[\text{Perhaps this is not surprising, considering the complexity of systems like the WSRT, but we tend to think of optical systems as basically simple and monolithic, even if they are expensive through size or choice of materials.}\]
first paper (invited paper by Glindemann et al) at the Munich SPIE Conference of March 2000 on Astronomical Telescopes and Instrumentation and the project's development and successes will be recorded at its website (Recent Publications and Links on that site is a very useful resource, through which the other important optical interferometer projects may be tracked down).

Figure 7-5: The layout of the ESO VLT Interferometer; up to 8 auxiliary telescopes can be accommodated

In spite of the basic similarity of radio and optical interferometers, the technical details differ vastly. Some of the more important differences:

1. Because of the atmospheric effects, the collecting telescopes, even if 'perfectly' fabricated to a 'perfect' design, do not operate in diffraction-limited mode. To achieve this, even to an incomplete extent, an adaptive-optics system is needed at each telescope.
2. The tolerances on the surface of the telescopes are related to optical wavelengths, which are of order 1 million times smaller than radio wavelengths. An active-optics system is needed to support the largest mirrors against deformations under their own weight.
3. Heterodyning with any useful bandwidth is not available at optical wavelengths, so signal transport has to be at signal frequency (which is of order 1 million times higher than radio signal frequency; which, in its turn, is considerably higher than the local-oscillator signal base frequency which determines the tolerances on the cables or other links in a radio installation).
4. Signal transport is by 'optical' techniques, which does allow one to push 2 orthogonal-polarization signals through 1 'channel'; the price one pays is that one has to keep track of polarization changes within the channel and make sure that such changes are the same for both telescopes.
5. It is quite impossible by passive means to stabilize the signal-transport optics against unwanted random differential phase changes. The frequency range and amplitude of incident-phase changes is very large, necessitating complex arrangements to effect the

---

A telescope is said to be diffraction-limited if its psf is limited purely by its physical size and FT theory. Generally unavoidable fabrication errors and the requirement of finite field of view cause departures from this happy condition at visible wavelengths, but in the IR it is a practical aim to build a diffraction-limited large telescope.
compensating phase changes in the 'delay lines'. A carriage is used for the largest-amplitude (and, fortunately slowest) phase changes, with some sort of motor drive; for faster changes, a piezo-driven or loudspeaker-coil system is used, with smaller, lighter mirrors where necessary.

6. It is quite impossible to determine the absolute interferometer baseline to the required accuracy (a fraction of a wavelength) by geodetic means, so that the absolute position of objects in the sky is undetermined (as in VLBI). It is possible to do differential astrometry (position measurements) from 1 isoplanatic patch to another, but only by arranging for parallel channels for the 2 patches and actually measuring the instrumental differential optical-path-difference between the 2 patches; laser metrology over the actual signal path from telescope to correlator is necessary and complicates the system considerably (see 2000 paper no 12 by Lévêque for a VLTI study of the subject).

7. To detect the random phase-difference changes that need to be corrected, various auxiliary sensing instruments (called 'fringe trackers' etc) are needed, according to the circumstances of the observation (e.g. bright point source available within the isoplanatic patch?) and the required science (image shape or image position?).

8. In spite of the delay-line optics moving over large distances as one points the telescopes to different parts of the sky (65 metres in total, in the VLTI case!), the entrance pupil for the final instrument should be in a fixed place. This requires a variable-power (variable-curvature) lens or mirror at an image position within the optical system of each telescope.

9. The correlator is based on a completely different principle from its radio equivalent. Radio correlation is accomplished by 1- or 2-bit digital multiplication at IF. Optical correlation is at signal frequency, is analogue and is accomplished by forming sum and difference of the amplitude signals, squaring these in the detector and then taking the difference.5

10. Gaining system efficiency by multiplexing (i.e. duplicating processing channels, so that parallel processing of related signals may be carried out) is based on entirely different principles in radio and optical. Radio photons are of such low frequency and hence energy that the signal may be considered to be continuous (with the noise determined entirely by the shot noise or thermal noise of the first electronic component). This means that one retains phase information during electronic processing, may subdivide the signal after some amplification without increasing system noise and process each of these derived signals independently; thus, one may often increase efficiency of a system by moving towards simultaneous use of all possible non-redundant collector-antenna spacings, delays and polarization combinations, without adding extra front-end components. In the optical case, the noise is mainly due to the signal itself, so that in subdividing the signal (at signal frequency, without amplification) one loses at least as much by increasing the fractional noise as one gains by parallel processing. The exception to this is if one uses spectral demultiplexing to process the different frequencies in the broadband signal in parallel: generally the best way to do this is to use a 'spectrograph' (disperser of the 'white' input image) as the last element in the optical system, just before the detector (see Figure 7-6 and Figure 7-7).

5 The actual way one forms the sum and difference in practice may differ, depending on wavelength, auxiliary instrumentation and instrument optics. Often a fringe system is imaged in some way on to the detector, where the amplitude of the fringes is proportional to the 'difference of squares' referred to above. In such a case, the position ('spatial phase') of the fringe system is related to the phase difference and the spatial frequency of the fringes may be used to distinguish one superimposed correlator signal from another in multi-baseline systems (Figure 7-6).
Figure 1: Preliminary optical layout of AMBER. An afocal system (1) is used to compress the incoming beams. Babinet-like prisms (2) intend to correct the difference in polarization which might be introduced in the different beams by non-homogeneities of the coatings. A second set of prisms (3) may correct chromatic effects such as the atmospheric dispersion and/or differential refraction. The key feature (4) of the AMBER instrument is an off-axis parabola that feeds a short optical fiber acting as a spatial filter and isolating a single coherent mode. Second off-axis paraboloids produce parallel beams which are reflected on beamsplitters (5), the first part of the beam combiner. The two closest beams are almost tangent and the third one is two beam diameters away (center to center) from the second one. These beams produce superimposed Airy disks and each pair of pupils will produce a set of fringes. Due to the non-redundant spacing of the pupils, this set of fringes can be discriminated by their spatial frequency. An anamorphic system (7) made of a pair of cylindrical mirrors in an afocal combination compresses the beam orthogonally to the fringes. The flat mirrors (6,8) convey the photometric beams used for calibration of atmospheric fluctuations. Cold stops on a wheel (9) and a filter wheel (10) are located at a pupil position. The light is then dispersed by a grating (11). A compact spectrograph design requires two chamber mirrors (12). The detector (13) will probably be 1024 x 1024 HAWAII Rockwell array. With three telescopes, the fringes are analyzed within a strip of 12 x n pixels, where n is the number of spectral channels (n < 1024). The photometric beams are slightly dispersed by a fixed prism (14) to take into account the chromatic variations of the Strehl ratio and the spatial filter efficiency. For some objects, we plan to use the 2 non-vignetted field available in the VLTI laboratory. This is achieved by replacing the spatial filter unit by an afocal system without spatial filter (4a, 4c) which maintains the direction of the output beam but divides its diameter by two. The figure roughly respects the proportion between the elements. The size of the spectrograph is 45 cm x 30 cm.

Figure 7-6: The AMBER near-IR beam combiner for the VLTI

For 'chamber mirrors', read: 'camera mirrors'
'Imagers', as the term is normally used in astronomy, are instruments that produce a recorded image of the sky. Optical imagers may be very simple, just a telescope and detector, or they may include focal-plane filters, pupil-plane filters, telecentric-beam filters, a choice of scale-changing optics, a polarization modulator, etc. The filter positioning options are mainly to ensure that the imperfections of available filters have minimal influence on instrument performance, while the choice of image scales is to allow the detector array to oversample the instrument psf in some cases (e.g. for image restoration algorithms or for good photometry or polarimetry) and in others to have as large a field of view as possible.

Imagers (cameras) are the oldest astronomical instruments, in that the eye is an imager. However, the eye can distinguish colours, if the source is bright enough; so is it a spectrometer? Not really, in the sense that astronomers use the term 'spectrometer'; rather, colour perception by the eye is a function of the detector assembly, as in X-ray CCDs and optical STJs (the fact that the retina may contain structures tuned to certain wavelength ranges – i.e. optical filters – does not invalidate this way of looking at the eye as an optical instrument; it just means there may be more ways than one for dividing it into conceptual modules).

Imagers in other wavelength regions are very different. Hard X-ray and γ-ray imagers use masks to cut off radiation from directions other than the preferred. Soft X-ray imagers can use a (grazing-incidence) telescope and a CCD detector in a quasi-optical arrangement. SCUBA, with its multiple feed antennas in the telescope focal plane, is a radio imager in the 'optical' sense. A synthesis array like the WSRT, at any one wavelength, may be thought of as a
'radio imager', but equally it could be classified as a spectrometer, or a hybrid instrument; perhaps 'system' would have been the best description, but that was not evident at the time it was built and baptized, so 'telescope' it is!

Examples of modern imagers are shown in Figure 7-8, Figure 7-9 and Figure 7-10. Note the very different optical elements employed, the optical layout and mechanical construction. Note that certain optical elements are in a pupil plane, others in an image plane. Try to understand why each component is placed as it is.

Figure 1. TAURUS: optical configuration (schematic).

Figure 7-8: The TAURUS imager (named for its shape!) for images through a F-P filter of very narrow passband
As may be seen from the optical ray diagram, most of the 'manipulation' optics are located near the pupil. The prism-like component is a grism, for straight-through spectroscopy, the double prism is a polarization beamsplitter. Note that these components are on 'filter-wheels', so that they can be combined in different ways for different purposes. There is also a choice of image scale through exchange of collimators. The mechanical drawing, derived from an AUTOCAD design drawing, gives a good impression of the complexity of a modern instrument, even if it is "only" an imager (in fact, it is much more, of course; that is why it is so complex).

Files kindly made available for use in these lecture notes by Messrs Fürting, Schöffner and Seifert, of the FORS team, Heidelberg.
Figure 7-10: The VISIR imager for the thermal IR, for 2 different image scales on the detector

Note that this instrument, also for ESO's VLT, is similar in principle to FORS (Figure 7-9); it has a telescope focal plane as input, there is a pupil inside the instrument for filters etc. and there is a final image on the detector array. The image scale is here determined by exchange of camera optics; the same constraints hold as with the collimator for FORS; the optics are exchanged, but pupil and focal images stay put. The complications for this instrument are in its cryogenic operation: it has to be adjusted at room temperature, then be able to work at about 30K, and it must warm up safely if the mains power were to fail.

7.4. Photometry

Photometers are optical instruments in which high accuracy of signal strength measurement is the prime design aim. Until fairly recently, they were mostly stellar photometers, designed to observe point sources and were distinct from imagers (photographic cameras). Nowadays, the distinction between imagers and photometers is one of emphasis only. The extent to which the atmospheric extinction effects can be compensated, calibrated, corrected, etc, define the accuracy one aims for in photometer design. Whatever the atmosphere allows, astronomers can use for some purpose or other; in fact, for the most accurate photometry, astronomers have elected to go into space and for the accuracy of those instruments, "the sky is the limit" (interpret as you wish). So, once again, we are back to the old adage: "Just make as good an instrument as you can" (I'm not really serious; it is of course up to the astronomer to at least attempt to estimate what accuracy she will need to solve a certain problem; and to indicate what other properties she is willing to forego in the quest for extreme photometric accuracy).

Historically, (stellar) photometers were specialized instruments, in which a diaphragm passed only a single star image in the telescope focal plane. Using only this selected light, a lens ('Fabry lens') then imaged the telescope pupil (section 3.6) on to the detector (Figure 7-11), in order to reduce atmospheric noise in the photometric measurement (a randomly moving image on a non-uniform detector would yield unnecessary random noise in the photometry).
As detectors have become more uniform and computer methods are allowing better data-taking and more extensive calibration, the trend has been to record the image itself, allowing simultaneous raw photometry of many sources. Such quantitative imaging allows one to use 'ensemble photometry', in which almost the entire collection of stars recorded in a frame is used as a virtual reference star for extinction correction of the object(s) one is interested in (this method is useful for variable stars; in that case one is interested in the variations of the light output of a star rather than in its exact average value). Nevertheless, one could still use the older principle of imaging the pupil on to the detector in order to achieve even higher accuracy of what astronomers regard as 'absolute' photometry (it is always relative to some standard star, in fact). I have explored questions related to this in two papers ('New Techniques' and 'Transformations and Modern Technology') in IAU Colloquium 136 (Dublin, 1992) "Stellar Photometry — Current Techniques and Future Developments" (CUP 1993, C.J.Butler and I.Elliott eds); Figure 7-12 and Figure 7-20 are taken from the oral presentation of those papers, which are as relevant now as they were in 1992.

There are a number of other points one could (and should) consider with respect to the optimum design of photometers:

1. Photometers generally include optical filters to limit the bandwidth of the recorded light. Where in the optical path to place such filters is a non-trivial question.

2. In combating the variations of the atmosphere (in this case mainly extinction, but also the artefacts caused by image motion on a non-uniform detector), the accuracy requirements will refer to different lengths of time and to variations of other external circumstances (such as telescope pointing direction and ambient temperature). Therefore, design aims for a photometer should always contain stability specifications, as well as accuracy specifications. For increased efficiency, beamsplitter assemblies of various designs and multiple detectors may be used in a photometer. In such a case, the measurements at several wavelengths are taken through the same atmosphere and one may make use of the correlation between extinction variations at several wavelengths to increase accuracy of the final result. Such an instrument will therefore have more stringent accuracy specifications, as well as being more complicated.
3. If one does want stellar images on the detector, then there is a trade-off between, on the one hand, image definition (sharpness) and faintness limit, and on the other hand stability of the photometric measurement (Figure 7-12).

![Figure 7-12: A hypothetical CCD multi-channel photometer](image)

Right: the correct way to image a point source on to the pixels of a CCD, if one wishes to obtain accurate photometry in spite of image wander. Left: the result of combining a low-dispersion spectrometer with a Lyot-type band-definition filter to obtain from a single CCD multiple simultaneous photometric measurements in Nyquist-sampling passbands.

### 7.5. Polarimetry

Optical polarimetry consists of the following (more detail here):

- Preserving the polarization characteristics of the radiation in the telescope (generally, by using a rotationally symmetric optical system upstream of the polarization modulator).
• Using a polarization modulator to convert the fractional polarization into fractional intensity modulation.
• Choosing any photometer, spectrometer or other 'photometric' instrument one cares to use. Its polarization characteristics are irrelevant, except that they may influence its efficiency (optical transmission).
• Determining the fractional intensity modulation of the detector signal (e.g. by tuned amplifier and synchronous demodulator, or by a CCD with framerate synchronised to the modulator frequency).

Depending on detector capabilities and sources of noise (shot noise for faint sources, scintillation or image motion for bright ones) the modulation frequency can be from milliherz to tens of kiloherz. If the modulation is slow, a 2-beam polarizer and ratio measurement are imperative to eliminate the higher frequencies of the noise (since a 2-beam polarizer is always desirable to collect all the light, this is mainly a matter of how one uses the information present in the 2 beams).

Designing the modulator -- for wavelength range, detector properties and optical beam characteristics -- is a job for a polarization specialist. Compromises are generally made w.r.t. spectral range, modulation efficiency, stray modulation, transmission, accuracy, achievable modulation frequency, etc.

### 7.6. Spectro(polari)metry

As mentioned above, at optical wavelengths converting a photometric instrument into a polarimetric one is a matter of adding a polarization modulator. This also holds for spectrometers, so that, everything said below concerning spectrometry will also apply to spectro-polarimetry.

Note: astronomers generally denote spectrometers by the historically-determined term "spectrograph" (στραφέιν: photographic detectors just recorded or 'wrote', they did not really 'meter' much!)

Spectrometers are instruments for determining the power spectrum of an astronomical source and exist in 2 basic kinds:

**FT spectrometers** for the visible do exist, but they are special-purpose devices for fundamental work on a few bright sources (the Sun, notably; the Kitt Peak Spectral Atlas of the Sun was obtained with an FT spectrometer, as a check on and a complement to the earlier Utrecht Atlas made with a classical spectrograph). Some way into the IR, where some detectors are very noisy, FT spectrometers have an efficiency advantage ("multiplex advantage") over a direct spectrometer. The basic FT spectrometer is a Michelson or Mach-Zehnder interferometer with a variable delay in one arm (Figure 4-7). One records a number of samples of the temporal autocorrelation function of the signal and FTs this function to obtain the power spectrum. Except for design and mode of operation of the correlator, such a spectrometer is analogous to the autocorrelation spectrometers used at radio frequencies (Hamaker chapter 17).
Figure 7-13: Various dispersors used in optical direct spectrometers (schematic representation only).
Almost all optical spectrometers are of the direct kind: they form the (power) spectrum as an image on an array detector (or the image may be scanned by a slit followed by a single-point detector). The image on the detector is a wavelength-dispersed (power) image of the light coming through the input aperture (usually a slit, so that some sky is exposed next to the object; astronomical spectroscopists talk of "the slit", never of "the aperture"). The dispersing element is a prism, a grating, or a back-to-back combination of these called grism (grating+prism). The common feature of these devices is that they deflect light by different amounts for different wavelengths, so that a parallel beam of white light is split (dispersed) into a fan of parallel beams, each of a certain wavelength.

A prism merely shifts the direction of the beam (a prism produces a phase shift proportional to aperture coordinate x, the FT of which situation is a displaced version of the FT of the aperture without the prism; the directional shift is largest for the blue end of the spectrum, since the wavelength is smallest for blue light and the refractive index is larger for blue than for red light).

Gratings spread the light into a number of orders, which are numbered according to the number of wavelengths of delay between one groove and the next. Figure 7-13 only indicates where the various orders will show up; their intensity will depend on the groove shape, which determines the blaze profile (antenna primary beam in radio synthesis terms), as discussed in the section on Fourier Transforms.

To make deflection of parallel beams (point source at infinity) work for a point source in the focal plane of the telescope, we must produce the white parallel beam by using a lens or mirror (system) whose focus is at the source in the telescope focal plane. To convert the dispersed parallel beams into a spectrum, we need another lens or mirror. This completes the basic design of the classical spectrometer: entrance aperture, collimator, dispersor, camera, detector. If the dispersor is placed in front of the telescope objective (so-called objective prism or grating), the collimator can be dispensed with (the white beam from any given object is parallel already). Spectrometers may be of any size from the large astronomical Coudé or Nasmyth instruments to sub-miniature ones constructed as 'integrated optics'; Figure 7-14 shows an intermediate-size design for mass-production methods.

Figure 7-14: Miniature spectrometer for mass production

Question: Can you spot the error in the bottom-left diagram of Figure 7-14? The top diagram is (more or less) correct.
An echelle (=stepladder) grating is a special kind of diffraction grating in which the groove is very asymmetric. The delay $b$ between one groove and the next is many wavelengths and the light is reflected by the short side of the groove, the elementary mirrors of width $a$ \( \text{Figure 7-15} \). Thus many different wavelengths \( (b/n \text{ with a number of different values of } n) \) undergo constructive interference in the same direction and have to be disentangled. One way to do this is to pass one 'order' (small range of wavelengths with one value of $n$) and reject others, by a filter. A more efficient way is to employ "cross-dispersion" (by prism or another grating): dispersion at right angles to the echelle dispersion. One then obtains the total spectrum as a series of bands (\text{Figure 7-16}) on the (2-D) detector, each band corresponding to 1 order, covering a certain range of wavelengths (with some wavelength-overlap at the edges of the orders).

As explained in the FT section, the 'blaze profile' of a grating is analogous to the 'primary beam' in a WSRT-like tied array and is the FT (or its $\mathcal{F}$) of the groove's 'pupil transmission function' (in the reflective-grating case: 'pupil reflectivity distribution'). In the case of an echelle, this is a top-hat function of width $a/\lambda$, which FTs into the blaze profile $\text{sinc}^2(\theta a/\lambda) = \text{"blaze profile"}$. This 'blaze profile' is the range of directions within which near-maximum light is returned by the elementary mirrors (their 'primary beam' in antenna terms); constructive interference also occurs in some directions outside this range, but there is (virtually) no intensity outside the blaze profile (hence the name).
We have tacitly assumed that the input aperture is infinitesimal. In practice, the entrance slit has a certain size (roughly to match the size of the seeing disc produced by the telescope; one may choose narrower slits to obtain a certain resolution irrespective of image motion, but a smaller slit than the seeing disc reduces efficiency drastically). To accommodate this small field and treat all points within it in an identical way, the dispersive element is placed at a pupil position (=telescope entrance pupil, as re-imaged by the collimator).

In so-called 'long-slit' mode, simultaneous spectra are recorded of a number of points in the sky (the 'points' along the slit). The resultant 2-D image has a spectral dimension and a spatial one. Long-slit mode is an alternative to cross-dispersed mode in filling the 2-D detector with astronomically useful information. If long-slit mode is made available, there must be provision to orient the slit on the sky arbitrarily, either by rotating the image or by rotating the instrument (since modern large telescopes are generally Alt-Az mounted such a facility is also necessary for imaging exposures longer than a few seconds).


A special way to use a long slit is to feed the individual points along the slit by a flexible optical-fiber light-transfer system. This allows the observer to record a number of point source spectra simultaneously and is yet another way to use the expensive large telescope and available detector area efficiently.

The above should be sufficient background to understand a Users' Manual for a modern spectropolarimeter such as UVES (Figure 7-17) or VISIR (Figure 7-18). Engineering details of the components used in a practical instrument are discussed in another section.

Exercise for the reader: In Figure 7-17, go through the list of functions and see if you can identify the purpose of each module.

---

(almost identical, anyway: all beams do strike the same area of the dispersor, but at slightly different angles; for 'identical' to apply, the blaze profile should therefore be considerably wider than this range of input angles)
### ROTATOR FUNCTIONS

100 Rotator structure  
105 Telescope shutter (TESH)  
110 Flat field unit (FFC)  
120 Thorium-Argon unit 1 (TAC1)  
130 Thorium-Argon unit 2 (TAC2)  
140 Calibration mirror slide (CALM)  
150 Iodine cell unit (IODC)  
160 Image slicer unit (IS)  
170 Image derotator unit (DROT)

### BLUE SPECTROGRAPH FUNCTIONS

300 Blue filter unit (BFIL)  
310 Blue folding mirror (BFMI)  
320 Blue intermediate spectrum mirror (BISM)  
330 Blue main collimator (BMC)  
340 Blue transfer collimator (BTC)  
350 Blue echelle unit (BEG)  
360 Blue cross disperser (BCD)  
370 Blue F/1.8 camera unit (BCA)  
375 Blue camera tilting unit (BCT)  
380 Blue exposure meter (BEXP)  
390 Blue flat field lamp (BFFL)

### RED SPECTROGRAPH FUNCTIONS

400 Red filter unit (RFIL)  
410 Red folding mirror (RFMI)  
420 Red intermediate spectrum mirror (RISM)  
430 Red main collimator (RMC)  
440 Red transfer collimator (RTC)  
450 Red echelle unit (REG)  
460 Red cross disperser (RCD)  
470 Red F/2.5 camera unit (RCA)  
475 Red camera tilting unit (RCT)  
480 Red exposure meter (REXP)  
490 Red flat field lamp (RFFL)

### PRE-SLIT FUNCTIONS

200 Pre-slit filter unit (PFIL)  
210 Pre-slit plate  
220 ADC unit (ADC)  
230 Depolarizer unit (DPOL)  
240 Pupil stop unit (PS)  
250 Mode selector unit (MODE)  
260 Doublets (DOUB)  
270 Slit viewer units (SV)  
280 Blue slit unit (BSL)  
285 Red slit unit (RSL)

### TABLE AND FUNCTION TOOLS

600 Instrument table  
610 Table support  
622 Red baffle  
623 Blue baffle  
630 Cable supports

---

**Figure 7-17:** UVES, ESO’s VLT-Nasmyth-focus echelle spectrometer for ‘visible’ light
The spectrometer has 3-mirror off-axis collimators, which in a second pass serve as cameras; the highly-folded design is possible in the IR only because gold mirror coatings are very nearly 100% efficient; the low- and high-resolution arms are of similar design, except that the high-resolution grating is an echelle, for which cross-dispersion is provided by grisms in the filter-wheel; the re-imager provides a cold stop (near the filters); the bleed-off mirror in the VISIR imager optical train is not shown, but its effect on the rays is (top left, just after the telescope focal plane, which is where the present raytrace starts). An exploded optical diagram of this instrument is shown in Figure 3-3.
7.7. Astrometry

Astrometry concerns itself with determining positions in the sky, i.e. measuring angles. The scientific need for such data can be for pointing a telescope, but, as precision and accuracy have increased, it has increasingly become possible to detect transverse motion of objects by measuring the change in their positions over time. The accuracy achieved has increased in less than 50 years from arcseconds (in all-sky astrometry; of order 0.1 arcsec in small-angle astrometry) to 'milli-arcseconds and micro-arcseconds' in present and planned projects; since even at tenths-of-arcseconds' accuracy some objects were already found to move when observed over sufficiently long time intervals, the increased accuracy is really causing a revolution in astrometry-based astronomical science.

Traditional 'absolute' astrometry was by setting the crosswires of a telescope eyepiece on a star, then reading the telescope circles and an accurate clock. In so-called 'meridian circles', the telescope swung in a N-S vertical plane, so that only the declination circle needed to be read; reading the declination circle and timing the transit of a star were later automated by PMT and computer, but the principle remained the same. Photographic plates at the focus of a (more or less normal) telescope were pressed into use for relative position measurements, which were then used in conjunction with the, less accurate but absolute, meridian-circle astrometry. With the progress in CCD manufacture and assembly of large arrays of CCDs, astrometry is now entirely electronic, mostly together with specialised telescopes, interferometers, instruments and software. Nowadays, systems like the WSRT and VLBI arrays are used for radio astrometry and the combination of optical and radio measurements on the best point sources (quasars) yields the most reliable connection of (optical) astrometry to geodesy; inverting the procedure, i.e. assuming constancy of position in the sky for quasars, VLBI yields data on continental drift.

The approaches used in modern astrometry are to engineer a stable angle standard (for large angles and all-sky 'absolute' astrometry) or a stable baseline (for small-angle relative astrometry at very high precision). Both approaches may ultimately find their way to space, but for some time yet there will be room for Earth-based instruments such as the VLTI and WSRT (large-angle optical astrometry is already space-based, with the success of the HIPPARCOS mission). The reason for preferring space as the locale is the absence of gravity: engineering a stable angle (such as the top angle of a prism or the angle between 2 halves of a split mirror) is then mainly a matter of avoiding or releasing stresses during component manufacture and of sufficiently good thermal shielding in orbit, so that the temperature of the angle standard is constant and homogeneous. For a stable baseline, space may be less essential, since baseline length is to be determined by an auxiliary laser interferometer, anyway; it may be just as 'easy' to do this on Earth as in space. However, the orientation of a long baseline on Earth can in practice be changed only by using the Earth's rotation, which means that errors of period 1 day (e.g. due to temperature effects [or ionospheric effects for radio astrometry]) might cause problems, so that space may be preferable after all. Experience with the VLTI's PRIMA instrument (which in its turn is based on experience with smaller instruments) will probably indicate whether Earth or space is the way to go for even longer baselines in small-angle optical astrometry.

Since the design of astrometric systems is largely a mechanical and thermal problem and the optics used depend very much on which properties of the installation need to be
7.8. Hybrid instruments

Astronomical requirements are not always satisfied by the simple instruments discussed above. One may, for instance, require spectra of a large number of galaxies within a cluster, or one may wish to make a spectral survey of an extended object rather than a point source. If neither a standard imager-and-filter combination, nor a standard long-slit or fiber-fed spectrometer can perform efficiently, a hybrid instrument may be constructed, optimized for the specific scientific purpose.

The classical example of such an instrument is the use of an objective prism (see alternative concept). This turns each star in the field into a wavelength-dispersed star (='spectrum') on the detector. If stars are too close together, their spectra will overlap (the cure is to rotate the prism and take another exposure). This simple arrangement is not suitable for faint sources, since the stellar spectrum is spread out, whereas the sky background remains at the same level (redistributed, true; but just as strong). In a later development, a mask at the telescope focus has pinholes where individual stars or galaxies are, thus eliminating most of the skylight for the spectrometer that follows; preparation of a mask for each field to be investigated is a necessary part of the observing routine. A multi-fiber feed is the most recent development in multi-object spectroscopy; it can be 100% efficient in detector use and does not require mask preparation machines: the fiber positioner has taken over that function and performs it online; see the 2DF project for a good example.

In an 'integral-field spectrograph', the telescope focal plane image is dissected by an array of microlenses. Each microlens is the entrance pupil of an elementary mini-‘telescope’ for a single channel through the spectrometer. The bright spots at the foci of the microlenses are the input 'sources' for the spectrometer, which is optimised for the array of microlenses in use (and v.v.). The spectra produced are each an average over the part of the primary image covered by a particular microlens. Detector use can be 100% efficient, but sampling of the image is in a regular array rather than exactly spot-on features of interest. This is ideal for determining 2-D velocity fields in images of galaxies, a sufficiently large scientific field to make a specialised instrument worthwhile. See Figure 7-19 for a very instructive practical example.
If spectrally-selective detectors such as STJs can be produced in large arrays, then in future low-resolution spectroscopy will largely be done by 'imagers with STJs', and the cross-dispersed mode of echelle spectrometers may become obsolete. We shall then be able to construct multiple-fiber-fed echelle spectrometers with STJs. Such instruments will have even larger electronic racks than at present, but detector and telescope use will be even more efficient than at present, so for future very large telescopes this will be a likely development. The corollaries will be detector operation well below 1K, even more stringent requirements for instrument design and operation, and an even more specialized workforce at the observatories. The link with military and space technology will be even stronger than at present. Whatever your wavelength range, astronomy of the 21st century will be big science, big projects, loyal teamwork, management issues; like it or lump it ('cultivate your garden' if lump it is your choice; there are plenty of interesting gardens to cultivate in other sciences).
7.9. Calibration

In a simple-minded view of instrumentation, one designs and builds an instrument to measure something, then goes ahead and observes, finally converting the raw observations into astronomical results. In this simple-minded view, it is a linear, sequential, process: having designed the instrument, one knows its properties and can compute results from observational data.

In a more sophisticated view, one admits that the instrumental parameters are not fully known, but one assumes (hopes, prays) that they will be stable, that they will not vary with time or telescope pointing direction. So, every once in a while, one foregoes an opportunity to make a 'science' observation, substituting a 'calibration' observation to establish zeropoints, gains, image scales, polarization balance, non-linearities, etc. The frequency of such calibrations is determined by one's expectation of the stability time constant (and other stability properties) of the instrument; a certain amount of trade-off between this and other instrument properties is a part of system design. Calibration in this sense is usually part of the astronomers' task, depending strongly on the observational programme and requiring expert knowledge of the instrument (eh, yes!).

In a final view, one admits that instrument properties will unavoidably vary, but determines their time constants and other variabilities. One calibrates frequently, possibly 'continuously', perhaps readjusting the instrument in the light of the calibrations in a servo-like arrangement. In such a servo arrangement, calibration becomes part of the instrumental system and often becomes the responsibility of the operational staff ('night assistant' and daytime technical staff), or it may be fully automatic (and can go wrong, but assessing that risk is part of system design).

The guiding principle of all calibration efforts should be that the instrument must be used in absolutely equivalent manner for the science and the calibration observations. Since this is not always convenient, actual practice often deviates from this, to a degree that one considers "safe". In particular, one often uses brighter sources and shorter exposures for calibration than for science; this means that system (particularly detector) linearity must be assumed or non-linearity allowed for.

In practice, 3 broad categories of calibration can be distinguished:

1. Use of an auxiliary light source. The advantage is that source properties can be controlled (e.g. intensity, polarization, wavelengths of spectral lines). However, it can be a problem to ensure that the source uses the instrument in exactly the same way as an astronomical object does. The calibration source optics must, for instance, have a (usually virtual) exit pupil which is an 'identical' (to what tolerance??) copy of that of the telescope, and signal intensity should not be too different from that for an astronomical object, while for broadband systems the spectrum of the source must resemble that of the science observations to be calibrated. See Figure 7-20 for one solution (schematic only).
Figure 7-20: The calibration beam must approximate the telescope beam.

One solution to this problem. The fiber light guide was included so as to be able to use bulky laboratory light sources without having to encumber the telescope with them. Note the central obstruction in the calibration beam pupil, simulating the effect of the secondary mirror of the telescope. As far as I know, no telescope actually has a calibration system even remotely resembling this sketch.

2. The way to reduce these problems is to use a celestial source. Many properties will be the same automatically in this case, but in particular spectra of celestial sources differ considerably from one another. This can for instance result in "colour terms" in the calibration of photometers or in the determination of atmospheric extinction, whenever observing bandwidths are inappropriately large.

3. The best way to calibrate, if it is feasible, is to use the science exposure itself. An example is wavelength calibration of a spectrometer by using a reflecting Fabry-Perot etalon (i.e. use it as a mirror; it acts as a very-narrow-band band-blocking interference filter) during the science exposure. Such an etalon inserts extra 'absorption lines' into the spectrum; if one can control its properties by suitable fabrication (e.g. known small temperature coefficient and effective temperature control during operation), such
monitoring of wavelength stability is near-perfect. Of course, such a procedure assumes that calibration and science features can be distinguished in the composite record. Another example is differential (i.e. intensity ratio) photometry, in particular ensemble photometry. One uses most of the stars in a frame to deduce 'instrumental' properties such as atmospheric extinction, field-dependent gain, colour-dependent gain, etc. and then uses these data for the 'science' objects in the field. Often, such "self-calibration" depends on a (self-)consistency assumption; WSRT selfcal, for instance, depends on the assumption that, apart from a number of point sources, the sky is empty and therefore imperfectly known instrumental parameters may be manipulated ('massaged') to indeed yield a mostly empty sky (such an assumption would not hold under all circumstances; and in particular it would never hold in the thermal IR, except for the differential image obtained by chopping). The challenge in using self-calibration is usually to develop the actual 'reverse' procedures/algorithms that establish the instrumental properties from the observation; a minimum requirement is that one should not try to calibrate more properties than there are 'variability factors' evident in the observations (the inversion problem should be overdetermined).

On final remark: as soon as any calibration procedure has become routine, it may be possible to feed the result back into the instrument and to correct its performance online. Adaptive optics vs speckle is a case in point. The advantage is that the instantaneous composite instrumental system is of higher quality, the disadvantage is that later the correction algorithm may turn out not to have been optimal and one may not be able to improve it after the fact (servo use may – need not, but may – have inherently invalidated later improved self-calibration, e.g. by not recording the applied corrections with the data). The choice between (self-)calibration and feedback is thus a system-level choice of some sophistication: one must automate some corrections, in order not to overburden the (self)calibration task, but in doing so one probably throws away some (hopefully marginal) final quality; where does one draw the line?

(Question: Is a committee the best place to a) discuss and b) to decide on such a dividing line? Give a reasoned answer in half a page A4, including detailed examples based on your own experience. File your report).

7 The 'most perfect' absolute calibration of this kind I ever heard of (except that the light source was a laboratory one, hence pupil illumination may have been inappropriate) was for a Coudé spectrometer at the McDonald Observatory. The S/N ratio of calibration observations was such that the extremely shallow lines of the atmospheric absorption bands due to the air within the spectrometer itself could be detected: guaranteed to have zero Doppler effect. One could include the reflective F-P during such observations, thus transferring the calibration to its lines.
8. Optical Components in Engineering

This chapter lists some of the basic properties of the engineering components one uses to build optical instruments. Detailed design of such components is beyond the bounds of this course; I have listed suitable reference books at [http://www.strw.leidenuniv.nl/dai/texts/txt.html](http://www.strw.leidenuniv.nl/dai/texts/txt.html).

A very useful general-purpose service is the Photonics Dictionary.

Some commercial firms provide useful material in their catalogues; one of the best in this respect is Melles-Griot; select one of the 5 chapters of Optics Guide for explanatory chapters; they take some time to load, if you have the paper or CD-rom version, use that), the catalogue (including the Optics Guide) is also on paper and CD-rom. Another useful catalogue is that of the [Linos Photonics combination](http://www.steegreuter.de/). Then click 'English', 'Catalog' and 'Information to Technical Optics', finally click the Adobe Acrobat logo to download the 0.5MB pdf file.

8.1. Basic imaging components

8.1.1. Transmission optics

Lenses and prisms are the optical components originally used in astronomy. They were (and are, mostly) made by grinding and polishing. Alternative methods of manufacture are molding or other replication techniques, chemical leaching (or, historically, for very small sizes for high-power eyepieces, by melting the end of a glass thread into a bead).

Single lenses suffer from chromatic errors (even on-axis), since lens power depends directly on refractive index and this is a function of wavelength. "Achromats" are made from 2 or more components of different materials and can have approximately wavelength-independent behaviour (the wavelength range depending on the tolerances on imaging quality: wavelength range for astronomy say 2:1, for electronic-chip manufacture by laser-illuminated mask printing: < 0.1%). A Schmidt telescope corrector plate may be achromatized for the best optical results over a large spectral range; whether this is worthwhile will depend on the increase in telescope flexure or in sagging of the plate due to the added weight of the achromatic combination.

Prisms are chromatic on purpose (unless they are used as substitutes for mirrors: right-angle deviation by internal reflection). Nevertheless, 2 materials may be used, to make them transmit the 'central' wavelength without change of direction (a grism is an alternative to such a direct-vision prism; see Figure 7-13).

A lens of a given power, made of a glass of given refractive index, may still be constructed in an infinite number of ways: the sum of reciprocals of the radii of curvature of the faces is defined, but one of the individual radii may still be chosen freely. The best value to choose depends on the
use one is going to make of the lens: what is the ratio of object distance to image distance ('conjugate ratio')? The aim is to minimize the average angle of incidence on the surfaces, because the smaller this angle is, the smaller the 'aberrations' (imaging imperfections) will be. The rule therefore is that the more highly-curved surface of an asymmetric lens should be on the side of the more remote of the conjugates (object and image).

Off-the-shelf catalogue achromats are usually constructed for infinite conjugate ratio (object or image at infinity) and their imaging is not perfect at a conjugate ratio of 1 (object and image at equal distances). The best imaging with catalogue achromats (at the expense of some extra light loss) in the case of a finite conjugate ratio is obtained with a back-to-back pair, with a collimated...
(parallel) beam in the space between the two achromats; this is a much cheaper solution than designing and manufacturing a lens specially for the finite conjugate ratio and will therefore be good engineering if it satisfies the tolerances.

Question: for such a back-to-back pair with a finite conjugate ratio, are the more highly-curved faces of the constituent achromats turned inward or outward?

<table>
<thead>
<tr>
<th>Material</th>
<th>Usable Transmission Range</th>
<th>Index of Refraction</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>BK7</td>
<td>BK7</td>
<td>1.52 @ 0.55 μm</td>
<td>Excellent all-around lens material provides broad transmission with excellent mechanical characteristics</td>
</tr>
<tr>
<td>LA5N9</td>
<td>LA5N9</td>
<td>1.86 @ 0.55 μm</td>
<td>High refractive-index flint glass provides more power with less curvature needed</td>
</tr>
<tr>
<td>SF11</td>
<td>SF11</td>
<td>1.79 @ 0.55 μm</td>
<td>High refractive-index flint glass provides more power with less curvature needed</td>
</tr>
<tr>
<td>F2</td>
<td>F2</td>
<td>1.62 @ 0.55 μm</td>
<td>Material represents a good compromise between higher index and acceptable mechanical characteristics</td>
</tr>
<tr>
<td>BK71</td>
<td>BK71</td>
<td>1.57 @ 0.55 μm</td>
<td>Excellent all-around lens material, but has weaker chemical characteristics than BK7</td>
</tr>
<tr>
<td>Optical Quality Synthetic Fluoride Silica (OQFS)</td>
<td>OQFS</td>
<td>1.46 @ 0.55 μm</td>
<td>Material provides good UV transmission and superior mechanical characteristics</td>
</tr>
<tr>
<td>UV-Grade Synthetic Fluoride Silica (UVGFS)</td>
<td>UVGFS</td>
<td>1.46 @ 0.55 μm</td>
<td>Material provides excellent UV transmission and superior mechanical characteristics</td>
</tr>
<tr>
<td>Optical Crown Glass</td>
<td>OPTICAL CROWN</td>
<td>1.52 @ 0.55 μm</td>
<td>This lower tolerance glass can be used as a mirror substrate or in non-critical applications</td>
</tr>
<tr>
<td>Low-expansion Borate glass (LEBG)</td>
<td>LEBG</td>
<td>1.48 @ 0.55 μm</td>
<td>Excellent thermal stability, low cost, and homogeneity makes LEBG useful for high-temperature windows, mirror substrates, and condenser lenses</td>
</tr>
<tr>
<td>Sapphire</td>
<td>SAPPHIRE</td>
<td>1.77 @ 0.55 μm</td>
<td>Excellent mechanical and thermal characteristics make it a superior window material</td>
</tr>
<tr>
<td>Zinc Selenide</td>
<td>ZINC SELENIDE</td>
<td>2.40 @ 10.8 μm</td>
<td>Zinc selenide is most popular for transmissive IR optics, transmits visible and IR, and has low absorption in the red end of the spectrum</td>
</tr>
<tr>
<td>Calcium Fluoride</td>
<td>CALCIUM FLUORIDE</td>
<td>1.339 @ 5 μm</td>
<td>This popular UV excimer laser material is used for windows, lenses, and mirror substrates</td>
</tr>
</tbody>
</table>

Figure 8-2: Properties of glasses and crystals for transmission optics (Melles-Griot)

Characteristics of common materials used for construction of transmission optics are displayed in Figure 8-2, Figure 8-3 and Figure 8-4. A website listing more detailed properties of a number of materials is that of Crystran. Further data in the Handbook of Optics.
8.1.2. Reflective optics

In a general way, mirrors are equivalent to lenses, except that their behaviour is the same for all wavelengths within the range of reflectivity of their coating. For the IR, mirrors also have the
advantage of higher efficiency than lenses (very important as soon as one needs more than 2 or 3 components in succession). The disadvantage of mirrors is that they block part of the light when used on-axis (e.g. Cassegrain telescope); this can often be solved by using off-axis configurations.

Such off-axis solutions introduce other aberrations (astigmatism; different focus positions for fans of rays within the plane containing the optical axis and fans at right angles to that plane); the design of systems with several off-axis components must be optimized through modern ray-tracing programs, intuition is insufficient.

In the soft X-ray region, mirrors are used at grazing incidence (Figure 8-6, Figure 8-7 and Figure 8-8), since it is only under those conditions that any reflectivity is obtained. Curved mirrors, such as paraboloids, take on ring shape under those circumstances, the space around the optical axis is empty; for reasonable efficiency, such rings are nested.

Optical mirrors for astronomical use are always coated on to some substrate such as glass, fused silica, Zerodur, aluminium alloys or beryllium (the choice depending on mechanical, thermal or other properties of the substrate material – such as ease of manufacture). The coating is chosen for its optical properties, such as wavelength range, polarization properties at oblique incidence, etc. If the coating is incompatible with the substrate (e.g. peels off after cooling to liquid nitrogen temperatures), an intermediate coating of some other material may solve the problem. The optical properties of coatings are the subject of another section.

### 8.1.3. Spherical/aspheric/off-axis optics

The surfaces of most optical components are parts of spheres (a flat surface being regarded as a sphere of infinite radius of curvature). Spherical surfaces are simple to produce by classical grinding and polishing. However, different annular zones of a 'spherical' optical component focus the rays in slightly different places (Figure 8-5) and the on-axis image quality deteriorates as more and more off-axis rays are included (i.e. as the defining pupil of the system is opened up); older photographic camera lenses yield the sharpest images when stopped down to very small apertures (this is no longer true for modern photographic objectives: their design has been optimized for intermediate apertures and performance worsens for both large and small apertures).

---

8 Almost all materials become reflective at grazing incidence, as may be verified by glancing along a piece of finely-ground glass.
The following drawing shows the computer plot of an achromat \( f = 160 \text{ mm}; \theta = 31.5 \text{ mm} \) Part No. 32 2270, the aperture of which has been stopped down to its maximum aperture to 20 mm diameter. The difference in the spherical aberration is apparent in the detail which, in this case, has been plotted with a magnification of 7500 instead of 750.

Coma can also be influenced. The following illustration is a computer plot for two beams at an oblique incidence of \( \pm 5^\circ \). By stopping the aperture down to 20 mm \( \theta \) (f-stop 6), coma for these beams is nearly eliminated, as the (250 x) detail clearly shows.

The same distance as above with a magnification factor of 7500

One should note, however, that this plot is only theoretical since an achromat is diffraction-limited over the entire aperture, i.e., the resolution in the focus of a completely opened achromat is limited by the Airy diffraction disc and not by the diameter of the minimum spot size of the spherical aberration.

Stopping down the aperture results in an increase in the depth of field, as the light bundle becomes smaller on the image side. This is also apparent from the computer plot.

Figure 8-5: The effect of spherical aberration (LINOS Spindler and Hoyer catalogue)
Whenever spherical aberration is dominant (extreme examples: mirror telescopes with small angular field of view and single fast-focal-ratio lenses used for focusing semiconductor laser beams; work out for yourself why chromatic aberration – and, if you know what it is, astigmatism – do not play a role here), on-axis imaging can be improved dramatically by slight modifications to the spherical surface. Paraboloids, hyperboloids and ellipsoids of revolution (rotationally symmetric 3D versions of the quadratic conic-section curves other than the circle) are used. Whenever larger angular field of view is required, further modifications to the surfaces are made and all such optics are designated "aspheric". The improved imaging is always obtained at a certain cost: manufacturing is more complicated, and tolerances on both manufacture and assembly (for multi-component systems such as telescopes or photographic objectives) are tighter, so that mechanical and thermal design of the mounting and operational environment become more critical.
In contrast to spherical optics, aspheric optics inherently have an axis of symmetry. Classical aspheric mirrors are used on-axis, i.e. both the active part of the mirror and the angular field of view are close to that symmetry axis. Whenever the circumstances do not permit this (such as: light loss due to central obscuration is not acceptable; or: reflection is appreciable only at grazing incidence), the on-axis condition for the active part of the mirror is relaxed (Figure 8-7). In general the image quality is only retained in the central part of the image and complete design to a specified image quality over a specified field of view is an optimization exercise of some complexity, using modern design software. Again in general, manufacturing and operational
tolerances for off-axis systems are tighter than for on-axis, so one should not adopt such solutions without thorough systems analysis. The most extreme example of using an off-axis part of the mirror (the complete mirrors are rotationally symmetrical and strictly speaking they are on-axis systems) is probably formed by the multiple nested rings of Wolter systems (Figure 8-8) used in soft X-ray satellites; manufacture of these systems stretches the state of the art to its utmost, is largely proprietary and extremely expensive.

8.1.4. Cylindrical optics

Sometimes we need to distort an image, e.g. when we want to squeeze a round patch of light on to an elongated detector, without losing light or effective detector area. In such a case we use optics that are curved in one direction and flat in the other; such optics are called 'cylindrical', the cylinder perhaps distorted into a conic section. An imaging system using cylindrical optics is shown in Figure 8-9.

Figure 8-9: Cylindrical optics for applying controlled distortion (Kingslake)

8.1.5. Non-classical imaging optics

Classical lenses, mirrors and prisms make up most of the optical component production volume and most of what is useful to astronomy. However, other types of component do exist and have niche applications. Some of these are:

1. Fresnel components (mainly lenses). These may be regarded as derived from classical components as follows: the component is split up into zones from which constant-thickness slabs of material are removed (conceptually), while retaining the original surface shape within each zone (Figure 8-10). A prism thus becomes a series of parallel prismatic strips,
each strip with the same inclination of the surface as the original classical prism; a lens becomes a series of concentric zones, with the surface parallel to, and of the same shape as, the classical lens from which it was derived. Thus, for every zone, the rays are given the same tilt as they would have been in the classical equivalent; the only difference is that the distance to the image has been given a different offset for each zone. This makes some difference in the image quality, but for a given image distance the effect may be corrected by a slight change in the tilt of the surface within each zone (Figure 8-10). As long as the zones are large compared to the wavelength, this geometric arrangement works like a normal lens for all wavelengths and therefore for white light. Fresnel components, being of much reduced thickness, are light-weight and can have very high transmission. They are mostly produced in optical plastic by replication, but the very large ones used in lighthouses are made of glass to withstand the heat of the lamps. Fresnel mirrors can be made by coating the plastic component. For more detail, visit http://www.fresneltech.com/html/products.html and download their pdf brochure.

Figure 8-10: Fresnel lenses (Fresnel Technologies)

2. Fresnel zone plates. Though they also carry Fresnel's name, these are a different kettle of fish entirely. Conceptually, they arise as follows: from the desired image point, one looks back at the optical-component plane and divides up the incident wavefront at that point into zones of roughly equal phase (i.e. within a range of $\pm \lambda/4$). One then masks off every alternate zone, so that only zones of roughly equal phase remain and multi-wave constructive interference will occur at the intended image point. Since neighbouring transparent zones differ in path length to the image by one wavelength, focusing at the intended image position only works for one particular wavelength; for longer wavelengths this condition occurs closer to the zoneplate and v.v. Zone plates are therefore narrow-
band components and their scope within optical astronomy is not enormous. They have been used in astronomy to focus X-rays of too short a wavelength to handle by grazing-incidence mirrors.

3. If one does not mask off the alternate zones, but adds half a wavelength's thickness of transparent material to them (on a thin-disk substrate), then the entire wavefront contributes to the intensity at the image point. This kind of component can be fabricated by chipmaking techniques (mask projection, photosensitive layer and etching; Figure 8-11). The result can go by various names such as binary optics or diffractive optics. One can go a step further and subdivide the zones into narrower ones, going from one-bit to multi-bit phase masks deposited on the substrate. Like the original zone plates, however, these are narrowband components and as such not all that useful to astronomy (except in the form of diffraction gratings). Using advanced techniques based on chipmaking technology, there are developments towards arrays of programmable micro-mirrors (Figure 8-25), to make things like variable-groove-density gratings possible; these may find their way into astronomy in future.

4. Holography has been pressed into use to fabricate optical components. The optical component consists of a hologram, which produces phase changes in the illuminating wavefront depending on position and direction, leading to a reconstructed virtual image in 3 dimensions. Such components are being used in 'head-up displays' for aircraft pilots who must see relevant computer data superimposed on the landscape scenery. Mostly these holographic components are narrowband, but this need not be; holographically produced diffusers exist, which are claimed to be quite wideband: the hologram can be manipulated to produce diffusers with any specified psf, so that the collecting optics may be illuminated optimally and yet the source has become diffuse. The only important function of such diffusers within astronomy would appear to be photometric calibration.
8.1.6. Fibers

A ray of light into one of the right-angle faces of a 45-45-90° prism will be reflected at the 45° face; this is called total internal reflection. The reflection is total in the sense that no light is lost and internal in the sense that the light is 'trying' to leave the inside of the prism. Such total internal reflection takes place at all interfaces where the ray would pass from the medium of higher refractive index to that of lower index; it only happens at large angles of incidence (far from normal incidence) and the transition angle depends on the refractive index difference.

One may exploit this to transport light around corners and to moving assemblies without losses (except for absorption in the glass). If one surrounds a glass rod (core) by a cladding of lower refractive index, then light entering one of the end faces will bounce off the walls of the rod at grazing incidence and will emerge at the other end of the rod almost unaffected. If one constructs the rod as a thin fiber, then this fiber may be bent round relatively sharp corners without affecting the light transmission: in particular, one end of the fiber may be fixed, while the other can be moved around. This latter feature is extremely useful at the input to a spectrometer (which is a -- fixed -- slit); one may use fibers to pick up light from stars (or other almost-point-sources such as remote galaxies) spread at random over the focal plane and record the spectra of all of them simultaneously.

The absorption losses in optical fibers are not negligible; after all, optically a fiber consists of a very thick slab of glass (tens of metres in the case of fibers from a telescope prime focus to a fixed Coudé spectrometer). The material used most in astronomy is therefore fused silica, for its low losses in the blue; for near-IR fibers, fluoride glass is used for the core. The cladding is often a doped form of fused silica (the light wave does penetrate some way into the cladding and its losses should therefore not be too large either); the change of refractive index can be a step ('step-index') or a more continuous change ('gradient-index'). As long as the core diameter is very large compared to the wavelength, the fiber acts like the macroscopic rod we used to introduce the concept and the actual value of the diameter is unimportant (varies in practice between say 10 μm and 1 mm, with correspondingly large range in 'minimum bend radius'). When the core diameter is of order the wavelength, the waveguide properties become important and the diameter is important ('monomode fibers', as opposed to 'multimode'). In such monomode fibers, asymmetries may be built in (elliptical or stressed core), so that they become birefringent (2 independent polarization channels are feasible); such fibers are important in sensor construction (and, incidentally, in AMBER, see Figure 7-6).

The smoothness of the core-cladding interface is very important when long lengths of fiber are used. If the interface is a perfect cylinder, then the rays of light emerging at the output will have the same inclination to the fiber axis as they had at the input; the azimuth will in general be different, due to the internal reflections. However, if the interface has random bumps and dents in it, then the ray pencil at the output will be more spread out (in angle) than at the input. This is known as 'focal ratio degradation' and is an important design parameter, since the size and shape of collecting optics will depend on it. A similar effect will occur whenever the fiber is bent with too small a radius of curvature.

Optical fibers are increasingly important in signal transmission and information processing; that is of course where the commercial interest lies and the interest for projects like SKA. For optical
astronomy, however, much of the technical/commercial literature is just noise and in particular most commercial test equipment is useless for astronomy.

Fibers are also assembled into coherent bundles (or image guides) and incoherent bundles (or light guides). An alternative to a fiber bundle light guide is a liquid-filled flexible tube. Fiber bundles are occasionally used in astronomy to transfer images or larger patches of light from a moving to a stationary (and v.v.) part of the equipment. Light guides may be specified to be 'randomized', so that the intensity distribution of an image may be rearranged into a pattern guaranteed to be 'random' (photometric scrambler). Alternatively, a light guide may be configured to rearrange a circular image (such as a star) into a narrow strip (such as the entrance slit of a spectrometer); in other words, it may function as an image slicer (which classically would be constructed from tiny mirrors or sophisticated arrangements of slots in, and cover glass on, a prism face). Finally, fibers may be tapered by special treatment. Tapered coherent bundles are used as compact and rugged optical systems for (de)magnifying images (e.g. as input to a CCD in military and medical equipment); a classical optical system is more light-efficient, so they are not normally used in astronomy, except as a compact interface between an image intensifier and a CCD in 2-D photon-counting detector systems (where the system noise is a function only of the image-intensifier performance and moderate light losses are therefore of no consequence).

8.2. Thin films and other coatings

Optical coatings are used to improve the performance of components (Anti-Reflection coatings) or to separate the 'substrate' and 'optical' functions of a component (mirrors and filters). Design and manufacture of coatings is a specialist branch of optical engineering, with its own design software and manufacturing facilities such as vacuum-deposition chambers and thickness-monitoring equipment.

8.2.1. Anti-Reflection coatings

Anti-Reflection (AR) coatings are used on transmission optics to improve performance. All refractive-index discontinuities cause (partial) reflections (similar to impedance mismatch in radio-frequency work); for transmission optics such reflections represent undesirable losses and in addition the reflected light itself can, after a second reflection or scattering at the instrument structure become a nuisance as stray light on the detector: either in the form of stray images ("ghosts") or as an increased, probably non-uniform, background. The IR spectral region in particular is plagued by reflection losses (tens of % per reflection for air-to-component surfaces), since common IR optical materials such as germanium and silicon have very high refractive index values. AR coatings are therefore mandatory on IR transmission optics.

Since the cause of the partial reflection is the discontinuity in refractive index, the cure is to try to turn the discontinuity into a 'slow' change, i.e. one that proceeds over an interval of several wavelengths rather than a small fraction of a wavelength.
The most radical solution is to modify the surface of the component in such a way that the refractive index does indeed vary slowly with depth, from the 'outside' index to that of the component's bulk material. An example of this is when a glass component has "microscopic bubbles" (or perhaps "nanoscopic" would be a better term) within it, relatively many at the surface and progressively fewer as one goes further into the material. Such a structure may, for instance, be prepared by producing a mixed-glass (glass alloy) surface and subsequent leaching or chemical etching, or by laser etching of the surface; the techniques are still under development. In general, the surface that results has very low reflection losses over a wide spectral band (more than an octave), but is inherently extremely vulnerable mechanically\(^9\); this is due to the microscopic cavities which, at the surface, take up tens of % of the volume. Therefore, within astronomy, such AR coatings can only be considered for components inside completely sealed instruments or for easily-replaced components within, say, a laboratory setup.

More robust components are obtained by vacuum deposition of thin films of other materials on to the optical component. Such materials are chosen for their refractive index, spectral transmission range, mechanical hardness and melting point. Some coating designs (for example simple 1-layer coatings) are tuned to a particular wavelength, others work over up to about 1 octave with somewhat relaxed tolerances on residual reflection. A much-used material for AR coatings is magnesium fluoride. For the visible (400-700 nm) there is a range of additional materials to use in the more complex coating designs; AR coatings with residual reflection of less than 1% over somewhat more than an octave in wavelength can be achieved. In the IR, the choice of coating material is more restricted and the reflectivity of the bare substrate is very large, so that a solution to the problem of reflected light is more urgent but also more difficult and a reflective optical system may be more effective. Coating materials for use in the UV are scarce, many optical materials do not transmit below about 350 nm; again, this often leads to a reflective optical system. The performance of coatings varies with angle of incidence; this must be allowed for in the design and extremely-curved component surfaces or very-wide-angle beams or fields of view are circumstances that inevitably degrade the performance of AR coatings. At large angles of incidence, the coating behaviour will also depend on the polarization of the light. Figure 8-12 contains examples of readily available AR coatings, Figure 8-13 concerns an application in CCD technology.

\(^9\) Such components cannot, in general, be cleaned after manufacture; like diffraction gratings, only more so.
Figure 8-12: Classical broadband AR coatings (Melles-Griot); note dependence on angle of incidence

**Figure 5.20** HEBBAR™ coating for visible /078
- Industry-standard multilayer AR coating for 415 to 700 nm
- Excellent performance with HeNe and visible diode lasers
- Optimized for normal incidence
  - $R_{avg} < 0.4\%$, $R_{dm}<1.0\%$
  - Damage threshold: 3.8 J/cm² ±10\%,
    10-nsec pulse (230 MW/cm²) at 532 nm

**Figure 5.22** HEBBAR™ coating for near-infrared /077
- Covers popular TIRapphire and diode laser wavelengths: 750 to 1100 nm
  - $R_{avg} < 0.4\%$, $R_{abs}<0.6\%$
  - Damage threshold: 6.5 J/cm² ±10\%,
    20-nsec pulse (260 MW/cm²) at 1064 nm

**Figure 5.21** HEBBAR™ coating for visible /079
- Optimized for 425-670 nm at 45-degree incidence
- Perfect for plate beamsplitting applications
  - $R_{avg} < 0.6\%$, $R_{dm}<1.0\%$
  - Damage threshold: see /078 (similar specifications)

**Figure 5.23** HEBBAR™ coating for near-infrared and diode wavelengths /075
- Optimized for performance from 660 to 835 nm
- Versatile for use with most diode lasers from visible to near-infrared wavelengths
  - $R_{avg} < 0.5\%$, $R_{abs}<1.0\%$
  - Damage threshold: see /078 (similar specifications)

**Figure 5.24** HEBBAR™ coating for diode lasers /076
- Optimized for diode laser wavelengths, from 780 to 850 nm
  - $R_{avg} < 0.25\%$, $R_{dm}<0.4\%$
  - Damage threshold: see /078 (similar specifications)

**Figure 5.25** HEBBAR™ coating for ultraviolet /074
- Excellent broadband coverage for 300 to 500 nm
- Covers HeCd and argon laser lines
  - $R_{dm}<1.0\%$
  - Damage threshold: 3.2 J/cm² ±10\%
8.2.2. Mirror coatings

The most common mirror coatings are silver (classical and good, but vulnerable unless overcoated), aluminium (less good, but also less vulnerable) and gold (for the IR). These coatings (unless they are overcoated=protected) have the virtue that they are easily removed and replaced (as practiced routinely for telescope mirrors).

All-dielectric multilayer coatings can be designed to be highly reflective over a limited wavelength range. Such coatings are used for limited-wavelength-range components such as lasers and (F-P type, interference) filters.

The two processes may be combined, usually in proprietary designs, to yield coatings matched to particular applications; it always pays to search the coating-manufacturers' catalogues to find out the state of the art, whether you plan to use an off-the-shelf coating or aim for more optimization.

Figure 8-15 and Figure 8-16 indicate the range of choice of reflective coatings; in general, each type has its advantages and disadvantages for a particular application and often none of them is quite satisfactory to the astronomer (who always wants/needs more). Note the dependence on angle of incidence.

Note also the dependence on polarization for oblique reflections. The behaviour with polarized light is even more complicated than suggested in the figures: in addition to polarizing action, oblique reflections cause retardation, resulting in polarization conversion. In correlation...
Interferometry in particular, this can be a source of errors and of loss of correlation. On the other hand, the polarization behaviour can be exploited to construct cube polarizers (Figure 8-14). Interference filters are basically F-P etalons of a special kind. The coatings used in them are mirror coatings; since the filter is only required to work well over a small bandwidth, the same holds for the 'mirror' qualities of the coatings. In eliminating the unwanted F-P responses, other spectrally-selective coatings (low-pass dichroic) mirror coatings are used, in addition to high-pass glass filters. An interference filter is thus a fairly complex assembly.

8.2.3. Dichroic coatings

These coatings ideally are mirrors over a particular spectral range and fully transmissive at all other wavelengths (hence the name, 'two-coloured', nothing to do with 'dichroism' in polarization); this explains names such as 'cold (or hot) mirror', 'UV cold mirror', 'heat filter'. Technically, dichroic mirrors are used in projection systems (to flush out the heat, while using the light), tri-channel colour systems, etc. In astronomy, their main use is to divide the spectrum into red and blue parts (UVES, Figure 7-17), or to separate visible light for active/adaptive-optics metrology from the IR used for scientific purposes. If dichroic mirrors are used obliquely, the transition wavelength is a function of linear polarization (Figure 8-14); this is useful in polarization beamsplitters, but mostly it is a nuisance.

![Figure 8-14: Polarization-dependence of dichroic-mirror transition wavelength (Melles-Griot)](image-url)
Metallic High-Reflection Coatings

We offer eight forms of standard metallic high-reflection coatings formed by vacuum deposition. These coatings, which can be used at any angle of incidence, can be applied to most optical components. Simply append the coating suffix number to the component product number (see figures 5.32 through 5.39).

Metallic reflective coatings are delicate and require care during cleaning. Dielectric overcoats substantially improve abrasion resistance, but they are not impervious to abrasive cleaning techniques. Clean, dry pressurized air can be used to blow off loose particles, then clean, denatured water, a mild detergent, and alcohol can be used. Gently cleaning with a swab is recommended.

ALUMINUM (016)

- The most widely used metallic mirror coating
- Provides consistently high reflectance throughout the near-ultraviolet, visible, and near-infrared regions
- $R_{ave} > 90\%$ from 400 to 1200 nm
- Damage threshold: 0.2 J/cm² ± 10%, 10-nsec pulse (12 MW/cm²) at 532 nm; 0.3 J/cm² ± 10%, 20-nsec pulse (14 MW/cm²) at 1064 nm

Aluminum, the most widely used metal for reflecting films, offers consistently high reflectance throughout the visible, near-infrared, and near-ultraviolet regions of the spectrum. While silver exhibits slightly higher reflectance than aluminum through most of the visible spectrum, the advantage is temporary because of oxidation tarnishing. Aluminum also oxidizes, though more slowly, and its oxide is tough and corrosion resistant. Oxidation significantly reduces aluminum reflectance in the ultraviolet and causes slight scattering throughout the spectrum.

![Figure 5.32 Aluminum coating /016](image)

![Figure 5.33 Protected aluminum coating /011](image)

PROTECTED ALUMINUM (011)

- The best general-purpose metallic reflector for visible to near-infrared
- Protective overcoat extends life of mirror and protects surface
- $R_{ave} > 87\%$ from 400 to 800 nm
- Damage threshold: 0.3 J/cm² ± 10%, 10-nsec pulse (21 MW/cm²) at 532 nm; 0.5 J/cm² ± 10%, 20-nsec pulse (22 MW/cm²) at 1064 nm

Protected aluminum is the very best general-purpose, metallic coating for use as an external reflector in the visible and near-infrared spectra. Unless we specify otherwise or you specifically request a different coating, our mirrors are coated with protected aluminum. Protected aluminum is coated with a dielectric film of silicon dioxide ($SiO_2$) of half-wavelength optical thickness at 550 nm. The protective film arrests oxidation and helps maintain high reflectance. It is durable enough to protect the aluminum coating from minor abrasions.

![Figure 5.37 Protected silver coating /038](image)

Figure 8-15: Mirror coatings commonly used in astronomy (Melles-Griot)
Figure 8-16: More mirror coatings (Melles-Griot); gold can be used far into the IR
8.3. Systems

Traditionally, optical instruments were built up from *components* (lenses, mirrors, prisms, etc). In the last decade or so, mass-fabrication methods have reached out to optical manufacture and one may buy complete systems as single components. The dividing line is not very clear: is a coated achromatic doublet lens a component or a system? What about a photographic telephoto lens of 5 or more component lenses? And a 'zoom telephoto' lens, with a motorized zoom control, automatic diaphragm and electronic shutter? One can buy such a 'system' in the camera shop as a single 'component' for the 'camera system' that one bought previously. Since they are produced for laboratories of all kinds, off-the-shelf systems may be more useful in the laboratory, when setting up test or alignment systems for the astronomical instrument one is building, than for incorporation into the instruments themselves. However, the principles on which their design is based may be very useful for astronomical design work, *mutatis mutandis*.

Systems are generally optimized for a certain *conjugate ratio* and one should not use them at widely different conjugate ratios. Of course, it is always possible to reverse the system and use it at the inverse of the design conjugate ratio (rays of light may be reversed in classical optics); the classical advice to mechanically-inclined amateur photographers, who want to photograph very small objects but cannot afford a macro lens, used to be: "reverse your standard 50 mm lens and attach it to the camera by, say, a 50 cm tube; then move this assembly backwards and forwards to focus on your object".

8.3.1. Zoom/telephoto/macro

These systems are mainly for 'photographic' applications, but they may come in useful in astronomical projects. They are complex assemblies of some 10 (or even more) lenses and as systems they are very highly corrected for particular applications; they perform extremely well for those applications, but for very different applications they may not be at all suitable. In general, unless explicitly stated, their transmission range is only about 400-700 nm, which is rather limited for astronomy.

A macro lens is intended for *conjugate ratio* reasonably close to unity, i.e. for relatively large images of nearby objects. Telephoto systems produce an image larger than one would expect from the distance at which it is formed (the 'back-focal distance'). Basically, they consist of a convex lens with, some distance behind it, a concave lens of lesser power; however, each of these are compound lenses in order to achromatise them, allow a large field of view and a certain focusing range. A zoom system contains 3 or more modules, each consisting of a number of components. One or more of the central modules can be moved and this alters the scale of the image, more or less without changing the position of the image.
8.3.2. Telecentric

A telecentric beam is one for which the pupil is at infinity; this means that, over the entire image field, the ray pencil converging on any image point is parallel to the optic axis ([Figure 3-7](#)). A commercial system is shown in [Figure 3-8](#).

8.3.3. Microscope objectives

Normally, an optical microscope consists of an objective, which forms an enlarged real image of the object at a height of about 25 cm, and an eyepiece, which puts the (virtual) image at infinity and matches the output pupil to the (entry) pupil of the human eye. The design of such an objective is optimized for the image distance of 25 cm and a conjugate ratio depending on its nominal magnification (for which the image distance of 25 cm is assumed).

Most microscope objectives are for looking through and therefore the spectral range is limited. All-reflecting microscope objectives do exist and can be used well into the UV or the IR (the coatings are usually optimized for one or the other application).

8.3.4. Eyepieces

Eyepieces are optical systems for allowing the human eye to look at a (normally real) image in reasonable comfort (eye at rest, focused at infinity) and with reasonable efficiency (exit pupil matched to the pupil of the human eye: same size and place). Telescopes, microscopes, binoculars and other instruments all have eyepieces of one sort or another; 2 examples are shown in [Figure 8-17](#). Eyepieces differ in their power, field of view and facilities such as reticles.

![Figure 8-17: Two examples of eyepieces from the technical literature](#)

Traditional eyepieces from the LINOS (Spindler & Hoyer) catalogue and recent publicity aimed at stargazers.
8.3.5. Condensors and diffusers, non-imaging optics

These items, which may be systems or just single components, all have to do with illumination. Proper imaging is not required, in fact, since most light sources are inhomogeneous, a certain amount of diffuse imaging is often an advantage. The relevance to astronomy is in producing suitable beams of calibration light, mimicking the 'natural' beam from a telescope.

Diffusers are the components of choice for softening the image of the light source, since they exist with different widths of their psf (the diffuse – in angular measure – patch of light into which they spread an incident pencil of parallel rays). The most recent addition to the available range of diffusers is the holographic diffuser of proprietary design, which (if we can believe the manufacturer) can be tailored to almost any psf for almost any given wavelength range.

Condensors are often single (possibly aspheric) lenses; their function is to collect as much light from the source as possible and to concentrate ('condense') it in the imaging optics. Some condensors, to be used with effectively point source lamps, have fairly good optical quality, but mostly they act to some extent as diffuser as well as fulfilling their primary function. Fused-silica condensors are intended for applications into the UV.

The term 'non-imaging optics' is used for flux collectors, such as are required to concentrate as much sunlight as possible on to solar cells. Mostly such systems consist of mirrors, but transmitting elements may be incorporated if the gain in efficiency is worth the cost. They are optimized for both the (angular) size of the source and for the size of the detector they serve. The astronomical relevance would seem to be about nil.

8.3.6. Etc

As optical technology expands further, driven by the needs of information processing, of instrumentation for non-invasive testing and monitoring, and of intelligent-robot design, more and more 'systems' will become available for astronomy, either as they are or with slight customization, or with in-house modifications or additions. It pays, therefore, to scan the commercial/technical journals; they regularly report on new technology that may be of use to astronomy.

8.4. Manipulation

This category contains a hotchpotch of components that 'do things' to a light beam. They may pass only certain wavelengths, or modulate the properties of the beam (make them vary with time). The only thing I have excluded is purely spatial manipulation ('optics' in the restricted sense of the word: lenses, mirrors etc).
Figure 8-18: Spectral transmission curves of certain absorption-type glass filters (Schott)

Figure 8-19: Spectral transmission curves of certain scattering-type glass filters (Schott)
8.4.1. Filters

Optical filters, like their electronic counterparts, pass certain wavelengths and block others. Being optical components, their behaviour and manufacturing tolerances do depend on where they are located: at a pupil or in an image, or elsewhere still. Optical filters are of 2 basic kinds:

- 'Glass' filters (including plastics). The filtering action depends on the material; the filter is normally just a slab. Mostly the filtering action depends on true absorption in the material (Figure 8-18), but some 'edge' filters depend on multiple scattering (Figure 8-19): these latter need a certain minimum thickness of material to function well; their different 'edge' wavelengths are produced by different heat treatment of the same or similar optical material: glass edge filters are all 'high-pass', i.e. they transmit the longer wavelengths\(^{10}\); low-pass edge filters have to be constructed as thin-film components. Glass filters are relatively cheap and most kinds live for ever, almost.

- Thin-film 'interference filters'. These are simple solid forms of the Fabry-Perot interferometer, similar to a resonant cavity in acoustics or a coil/condensor combination in electronics (the light bounces back and forth a number of times within the filter layer, on the average, before it is allowed out; the passband is inversely proportional to the time delay, which in its turn is proportional to the cavity depth and the number of times the light bounces back and forth, i.e. depends on the reflectivity of the partial mirrors that define the cavity). As in electronics, one may couple 2 or more such cavities resonant at slightly different frequencies, thus broadening the passband into something approaching 'square' (Figure 8-20); if this is not done exactly right, if the materials age after manufacture or moisture penetrates the filter, the passband will drift and become skew over time. The central wavelength of an interference filter passband is a relatively strong function of angle of incidence (Figure 8-21), so that passband shape depends on the angular spread of the beam at the filter location. Also, and not surprisingly given their construction, the passband is sensitive to temperature (Figure 8-21).

\[10\] Note that the usage is opposite to electronics, where 'high-pass' filters pass the higher frequencies.

Figure 8-20: Passband shape and the number of coupled cavities (Omega Optical)

'o optical density' is the negative log of transmission
8.4.2. Polarization modifiers

Almost all optical polarizers are for linear polarization (unlike the radio situation, where a helical antenna can receive or transmit a circularly polarized signal directly). This means that only linear polarization can be produced or analysed directly; all other forms have to be modified from or into linear polarization before they can be handled. So in addition to polarizers/analysers we need polarization converters in order to construct polarimetric instruments; we also need them to construct polarization-insensitive photometric instruments.
8.4.2.1. Polarizers

The most familiar type of polarizer is the Polaroid sheet. It is available in several qualities and for several wavelength ranges, and it is generally convenient to use, being just a thin slab of material. Its great disadvantage is the low transmission: <50% by virtue of transmitting only the wanted component, but usually much less than that if one needs high rejection of the unwanted component. For astronomy, whenever possible, one uses calcite (or quartz, magnesium fluoride or other materials still) to construct -- preferably 2-beam -- polarizers (Figure 8-22) with total transmission (in the 2 beams) close to 100%.

Figure 8-22: Crystal polarizers, for visible, UV and near-IR (Halle)

8.4.2.2. Polarization converters

All polarization converters are in one form or another waveplates (also known as 'retarders'); they introduce a phase difference between 2 orthogonal (usually linear) polarizations. The most common types of waveplate are the quarter-wave plate, which can transform linear polarization into circular (and v.v.) and the half-wave plate, which can be used to reverse the sense of circular polarization, or to change the polarization angle of linear polarization.
Simple waveplates are just thin slices of birefringent crystal (mica, quartz, magnesium fluoride, cadmium selenide, etc). Such waveplates work correctly only for one wavelength. Achromatic waveplates can be made by combining 2 materials, in the same way as one achromatizes a lens. Another possibility is to combine 3 suitably chosen and oriented slices of the same material; this is something one can only do with polarization components. It is possible to combine the 2 prescriptions; the resultant components are known as "superachromatic", work well for at least an octave (Figure 8-23) and are very expensive.

**Figure 8-23:** Performance of a super-achromatic halfwave plate, as a waveplate and as a linear depolarizer

![Image of waveplate performance](image)

**Figure 8-24:** Fresnel (a,b) and other rhomb 'waveplates'; (b) has optimal coatings

![Image of Fresnel and rhomb waveplates](image)
A different kind of wave-plate altogether is the "Fresnel (him again!) rhomb" and similar devices (Figure 8-24), which rely on the difference in phase change at total internal reflection. This phase difference is a function of angle of incidence and refractive index. By choosing the angle of incidence correctly, one may arrange for a quarterwave phase difference. Since the refractive index is a rather slow function of wavelength, such 'retarders' are inherently quite achromatic.

8.4.3. Time-dependent devices/modulators

In some cases, one wishes to determine details of an image or a spectrum that are so faint that instrumental or atmospheric instabilities make a good measurement quite impossible. Sometimes it is possible to arrange the measurement in such a way that the detail can be determined differentially, i.e. with respect to some reference quantity (spectral continuum, sky background). Such a differential measurement is best made by alternating between the unknown and the reference quantity, at a rate faster than any of the instabilities that plague the direct measurement. This is generally done by altering one of the components of the instrument periodically and measuring the same-period AC component of the detector signal (sometimes as a fraction of the DC signal). The periodically-altered component is referred to as a 'modulator' and the process of isolating the AC detector signal is called 'demodulation'; these go hand in hand, of course.

The 'chopping secondary' mirror of an IR telescope is such a modulator, as are the polarization modulators described earlier. Without modulation, it would be impossible to detect faint (or, in the thermal IR, any) point sources in the IR sky, nor would it have been possible to detect degrees of polarization as weak as $10^{-5}$.

When optical precision detectors were single-channel PMTs, operated in analogue mode, it was a simple matter to separate the AC component from the DC: the AC was passed through a condensor. Now that CCDs and other array detectors have taken over (with enormous gain in efficiency), demodulation can be somewhat more involved. One way is to employ high-readout-rate CCDs, if their higher readout noise can be tolerated. Another way is to shift the 2 images to and fro on the CCD in the form of 2 charge images that represent unknown and reference signal, and finally, when enough total signal has accumulated, to read out both images as one. As before, the aim is to modulate faster than any drifts or instabilities in the system (including the detector); to quantify this, one has to specify a maximum instability error during one modulation cycle, or alternatively, to measure the instability spectrum and deduce what modulation frequency one should employ.

Apart from the chopping secondary, tilting mirrors may be used to scan a spectrum across a detector or, in the supermarket, to scan a laser beam across the barcode. It is because of this latter application that very superior 'scanners' (i.e. mirror drives) are on the market as fully developed systems. Such scanners may use a rotating or an oscillating mirror and in the latter case, they may be mechanically resonant; all these various devices have their virtues in particular cases. The latest in tilting-mirror technology is the development of arrays of micro-mirrors (the largest reported array being $1024 \times 1024$; Figure 8-25).
Such arrays are being developed for optical switching in communication networks, so they may find a use in SKA before they are ever employed in optical astronomy; but who knows (with an array of micro-lenses, their low duty-fraction may perhaps be circumvented)? Sophisticated modulation methods will become feasible (different periods for different parts of the image, for instance).

A so-called 'liquid crystal' is another time-dependent optical component that may be pressed into the service of astronomy. The term describes organic liquids that, on the application of an electric field, become birefringent. They can therefore be employed in time-dependent waveplates, which are often employed as polarization modulators and may be turned into electronic shutters. Most of the liquid crystals are rather slow and therefore less useful than they might be, but more recent developments include faster-acting varieties. Liquid crystals are already being produced in addressable-array form; they can therefore be used in online image-processing.
8.5. Design

8.5.1. Optical systems and ray-tracing

Optical systems are generally designed in several steps. Once a system concept has been established, the first step is an analytic design. After that, numerical verification and optimization are used to produce the 'instrument design', in the optical sense. The final step is to design a mounting that will preserve the optical properties (i.e. not spoil them by permitting or inducing deformations or birefringence). Increasingly, the component mountings are designed as part of the optical component; examples of this approach are found in plastic optics (usually made by replication) and diamond-machined metal mirrors (Figure 8-27); in such cases, the final, computer-aided, design may be transferred directly to numerically-controlled machines for manufacture.

8.5.1.1. Analytic

Analytic design at the lowest level uses thin-lens theory, i.e. is based on the formula one remembers from high school:

\[
\frac{1}{u} + \frac{1}{v} = \frac{1}{f}
\]

where \(u\) is the distance of the object from the centre of the lens, \(v\) is the distance of the image from the centre of the lens and \(f\) is the 'focal length' of the lens; \(f\) is determined by the curvature of its surfaces and the refractive indices of the lens material and of the material surrounding it. The same thin-lens theory, with minor modifications, holds for mirrors.

This simple theory can be used to estimate the layout of the system, locate all the intermediate images and pupils, the approximate size of the components and the fields of view required from each of them. Fundamental problems with the design will probably show up at this stage. A useful primer is the chapter 'Fundamental Optics' of the 'Optics Guide' in the Melles-Griot catalogue.

The next step is to replace the components with realistic versions. Lenses necessarily have a certain thickness, single lenses are 'chromatic' (their focal length is a function of wavelength), spherical surfaces do not give perfect images, etc. Classical aberration theory takes all of these effects into account; many books (some more than 50 years old and still valid) exist on this subject, which is as complex and specialist as any mechanical or electronic design specialism.
The analysis is usually split into a 'paraxial' or 'Gaussian' stage (in which lenses may be thick and single lenses are replaced by achromats, but all rays considered lie very close to the optical axis of the system) and the stage of minimization of aberrations by slight modifications to the paraxial design. This latter stage nowadays is mostly taken over by optimization software.

Although the basics of this craft have not changed appreciably for many decades, new materials, new manufacturing techniques (such as diamond machining and computer-controlled polishing) and new mathematical techniques (such as computer-aided mathematical analysis) mean that the subject keeps developing and that new types of optical system are invented regularly.

8.5.1.2. Numerical

Once an optical system has crystallised into a paraxial design and one has some idea of its performance in real life, one may set up a numerical simulation by a ray-tracing program. Within such programs, one may specify a large number of individual rays (from different points in the image and at various angles). The software traces the path (and in the most modern versions computes the polarization) of each ray. The software can compute the position of best definition and generate a 'spot diagram' showing where all the rays pass through a plane at that position. One may then specify an optimization criterion and the set of properties the program may vary to reach an optimum, and examine the final result that one may expect to obtain after 'error-free' manufacture. One may also investigate the effect of various manufacturing errors on the final image, with or without re-optimization after the most critical components have been manufactured and their performance has been measured. ASTRON has extensive expertise, mostly gained during the VISIR and MIDI projects, in using such software (ZEMAX) and testing its predictions on both prototype and production models (e.g. using a wavefront interferometer through a cryostat window to test components during cooling and in stable 30 K conditions).

8.5.2. Other design activities

8.5.2.1. Specialised optical design

The design sketched above is a description of how a system behaves when the ray approximation is valid. However, some properties of certain optical components depend on the wave aspects of EM radiation, on mechanical properties, or on properties in yet other areas. Examples of such 'optical components' are:

1. The performance, as a narrow-band filter, of a resonant cavity such as the air space within a classical Faby-Perot interferometer, or the strength of its 'sidelobes' in the spectrum. Since such interferometers are the basis of interference filters (which contain up to about 5 such cavities, optically coupled for manipulating the passband shape), design software also exists in this area.
2. The design of thin films for all kinds of purposes, such as anti-reflection, dielectric-mirror, beamsplitter or polarizer coatings is another specialism with its own analytical methods and software.

3. Yet another area is the design of gratings, which is an EM wave problem of great complexity dealing with diffraction and interference (analogous to the design of feed antennas for radiotelescopes); here also, software can help the designer to verify his inspired guesses, optimize or even suggest new types of design.

4. Optical scanners are generally electro-mechanical assemblies involving a mirror, a galvanometer-type angular drive and special servo electronics to optimize the system. Special optical components may also be part of such a system, e.g. to ensure that the scanned (laser) beam is always focused on a straight line (the bar code on a flat surface, for instance) rather than on a circular arc. Design of such systems is only partly an 'optical' activity.

5. Polarization design used to be another such specialized area, but the trend is now to include polarization in raytracing and in filter and grating design. This generally involves replacing scalar treatment by (Jones) vector/matrix treatment; in this it is similar to the development of radio synthesis methods from scalar into full-polarization treatment. In thin-film design, up till now polarization has been included by carrying out the calculation for 2 refractive indices, 2 reflection coefficients, etc. Just recently, it was reported that birefringent materials have been drafted into thin-film manufacture, for 'broadbanding' certain types of coating even more than is possible by 'classical' means; this means that full-polarization treatment must of necessity have been carried out as part of the thin-film design:

High-Reflection Mirror Is All-Polymer Film (News item in Photonics Spectra July 2000)

ST. PAUL, Minn. -- A new generation of multilayer polymer film developed by 3M Corp. shows advanced control of optical reflection at interfaces. In its commercial release last year, the film attracted attention for its high reflectance of both incident polarizations at all incident angles over a wide wavelength band in the visible range. In the most recent iteration, 3M's 448-layer (N.B.) film demonstrates reflectivity greater than 99 percent for both polarizations and extends its reflectivity into the infrared between 800 and 1200 nm.

"We were surprised by the incredible reflectivity of these films," said Andrew Ouderkirk, a corporate scientist at the company's Film/Light Management Technology Center. "They provide lower cost and weight along with the best reflective characteristics of interference mirrors, then further add the ability to independently treat the polarization components of light."

Rather than depending upon large index differences between layers of isotropic media, 3M tailors the refractive indices along all three axes of the film's layered media. By adjoining birefringent polymer layers to isotropic polymer layers with uniform indices of refraction, the company constructed a material with an imaginary Brewster's angle; i.e., the interfacial reflection for p-polarized light is constant or increasing with incident angle, as opposed to the usual situation where the reflection decreases to zero at Brewster's angle.
The X- and Y-indices in the birefringent material differ from the index in the isotropic layer, while the Z-index is the same as in the isotropic layer. Light with incidence normal to the interface will see a change of index in both polarizations. As the angle of incidence increases, the s-polarized light continues to see the same index difference, as does the in-plane component of the p-polarized light. However, the projection of the p-polarized light along the Z-axis sees no index difference between the two layers, so it is unaffected by the transition between layers. Therefore, there is no angle at which the reflectance for p-polarized light goes to zero.

The result is a material that can exhibit the high reflectivity of dielectric mirrors without the high variation of reflectance with incident angle. Manipulating the material indices allows the p-polarized reflectance to be tailored to either match the reflectance for s-polarized light or to differ from it in well-controlled ways. The company's research team further describes the propagation effects of the 448-layer extruded polymer films in the Nov. 27, 1999, issue of Science. Ouderkirk calls this class of interference mirrors "giant birefringent optics." Invented in the early 1990s, the films found use in LCDs and handheld computers. In the future, the technology could be applied in components for long-distance light transmission, inexpensive laser resonators and optoelectronic filters or as thermoformed dichroic films (and astronomy?? JT).

In such cases, the 'optical designer' generally takes the result of other design work for granted, as a set of parameters describing the behaviour of the component or subsystem. In industrial or professionally-engineered scientific equipment (such as ASTRON builds for ESO etc), the trend is towards hierarchical and modular design, with only very specific and special properties of the required system being designed in-house in any detail. This is no different from the situation in radio-astronomical technology. However, it is relatively new for optical astronomy and adjustments in attitude and project management are still taking place, both at astronomical and at technological institutes.

8.5.2.2. The 'environment' of optical components

Optical components are vulnerable, yet they are required to perform their extremely delicate tasks in environments that can be very hostile, a far cry indeed from the usual temperature-controlled laboratory with its stable optical bench. Fortunately, astronomical optics do not have to resist salt water spray or oil-well temperatures, but other environmental factors pose quite enough problems to keep a small army of engineers on their toes; think of detector cooling (to 100 K for CCDs, <10 K for thermal-IR detectors and <0.5 K for STJs), instrument cooling (≈ 30 K for thermal IR), earthquake-tolerant design for Chile and Hawaii, and operation on an instrument rotator at a Cassegrain focus of a moving telescope. These factors pose non-trivial problems of mechanical and thermal stability during operation, but also of transitional conditions (e.g. initial cooldown, which is largely controlled; emergency warm-up, caused by electrical mains failure or by operation of alarm switches designed to guard against potential instrument damage or personnel injury). Add to these aspects the requirements of serviceability, modular design to allow design-team members to work in parallel on different subsystems, the desired 10-year (or more) lifespan of instruments (say with >95% system reliability) and the availability of spares. A project time of order 5 years from first serious design to commissioning then seems short\(^\text{11}\) to anyone but the

\(^{11}\) I have not even mentioned such trivialities as User and Servicing Manuals, or the bureaucratic chores inevitably generated by any large organisation (which, unfortunately, we can't do without, given the scale of our observational ambitions).
most engineering-blind astronomer eager to get his hands on the new facilities (or rather on the remote-control keyboard).

To take account of these factors, auxiliary design activities are needed, mostly mechanical, thermal, electronic and iterative combinations of these. To take just one telling example: an IR detector is the coldest part of its instrument during operations. Important questions for the design of its environment are:

1. How does one obtain such a low equilibrium temperature in view of the fact that there are conductive heat leaks from other parts of the instrument, radiation from sky, telescope and instrument, and that the detector itself dissipates appreciable power during readout (which must often – but not for all applications – be done at a high rate of some tens of times per second to avoid saturation)?
2. How does one balance the contradictory requirements of high cooling capacity and large temperature differences?
3. What compromise does one strike between wanting temperature sensors everywhere to monitor the detector and the rest of the instrument, and the fact that sensors have connecting leads which are efficient heat leaks?
4. In an emergency (probably with a power failure on one's hands), how does one prevent frozen-out residual gases (or air entering the cryostat from outside) from freezing on to the detector (the coldest surface in sight) and damaging this extremely expensive unit irreparably? (protective window? 10-20% light loss in the IR, even with the best AR coatings; ghosts and scattered light; imagine the uproar!)
5. How does one prevent the residual vibrations of the cooler from generating microphonic signals in the detector and yet preserve good enough thermal contact between them to reach the required low temperatures?

I am not suggesting that these problems are insoluble, but they must be formulated, faced and finally solved by a formidable array of technical expertise and specialised software (none of the software items apparently designed to cooperate with any of the others; fortunately the experts are better at it). The VISIR spectrometer is a good and fairly well-documented illustration of the size and scope of a modern instrument project for a modern telescope.

8.6. Manufacturing techniques

Like everything else in optics, manufacturing techniques are changing fast, both responding to and stimulating the transformation from a rather old-fashioned and time-honoured branch of engineering into the flashy commercial activity it is in danger of becoming. The following list of topics is not exhaustive, is just meant to whet your appetite and to help you realise the possibilities.
1. Traditionally, glass and crystal optics are ground and polished\textsuperscript{12}, using carborundum and similar abrasives, polishing powders (e.g. "jewelers' rouge"), water and other fluids. Traditionally pedal power and elbow grease are required and these are still used for fine finishing of small one-off optics, but at the other end of the scale all sorts of automatic machines now exist for mass production (the ASTRON workshop has a small cross-section of such traditional machines, in modern reincarnations). \textit{Grinding and polishing} are primarily suitable for spherical surfaces and flats\textsuperscript{13}. By suitably \textit{prestressing} the fixed workpiece during processing, one may after release of the stress end up with (mildly) \textit{aspheric and off-axis} components; iterative computer analysis and intensive testing are required, but this technique has been used successfully for the primary mirror segments of the Keck Telescopes. Iterative optical testing and \textit{local polishing} have been raised to a fine (computer-assisted) art in producing the highly special telescope optics such as RC \textit{primaries and large Schmidt corrector plates}. Traditional grind-and-polish production requires skilled craftsmen for success (Figure 8-26).

![The Gemini primary mirror is cleaned after edge grinding at Corning Inc. Courtesy of Gemini Observatory/AURA/NOAO/NSF.](image)

\textbf{Figure 8-26: Traditional glassworking !}

2. Diamond-tool lathes and milling machines are increasingly used in the manufacture of mirrors and molds for \textit{replicated optics}. These machines are generally for softer materials such as aluminium, that require only little polishing (in contrast, milled or ground glass has micro-cracks; the cracked layer has to be polished away, which tends to spoil the figure obtained in the previous step). The advantage of diamond-tool machines is that they can generate a much more general class of surfaces than the classical methods. In particular, extreme \textit{off-axis components} are fabricated relatively simply (by specifying the mathematical equation describing the surface). Another advantage is that the optical component and its mounting can be combined into a single monolithic unit (Figure 8-27).

\textsuperscript{12} \textit{Grinding} breaks small lumps out of the surface of the component and leaves a micro-cracked surface; \textit{polishing} causes plastic flow and seals the surface (restores the surface tension).

\textsuperscript{13} Special polishing 'stroke' regimes allow one to obtain mildly aspheric components such as the paraboloids and hyperboloids of revolution used in optical telescopes.
The optical surface after production is rather rough by optical standards, so that these components are best for the IR; for the visible, they require "post-polishing" (by classical methods, or by a modified diamond-tip regime that produces plastic flow rather than cutting). Diamond-tool machines are expensive; they incorporate air bearings and interferometric metrology. Operation of such a machine can be more automatic than the classical grind-and-polish routine; the craftsmanship is now needed at the programming stage.

3. Replication is a collection of techniques used in high-volume production. There are several optical plastics and 2-component epoxies that are suitable, but higher-temperature molding of glassy materials is also possible. The process is generally less accurate than the previous ones and the components often have internal stress birefringence leading to polarization problems. The molds (e.g. of stainless steel) used in replication are themselves made by classical or diamond-tool methods. So-called Fresnel lenses for illumination optics are usually made by replication, using one of the optical plastics. Mass-produced gratings are often the result of a replication process (Figure 8-28). Even high-quality aspherics such as are needed for CD laser optics, when mass-produced, are made by modified replication methods.

4. Cementing is often used to join components together in a multi-component lens ('cemented doublet' is accepted jargon). The optical cement has refractive index close to that of the glass components, so that reflection losses become negligible. Some cements are 2-component epoxies, others are cured by exposure to UV light, while some (oils and greases) remain fluid or are thermoplastics. The transmission range of the cement must satisfy the requirements of the system. In some cases the cement must remain slightly flexible, to avoid stress birefringence in the components; in others it must be temporary only, to be removed by heating the assembly. The ASTRON workshop has expertise of visible and UV-transmitting cements for both permanent and temporary joints.

5. Optical coatings are generally applied in a vacuum tank, although some are deposited from solution or produced in the top layer of the component's surface by leaching or etching. Thickness monitoring is required (e.g. by electronic monitoring of the resonance frequency of a quartz crystal, within the vacuum tank, exposed in the same way as the workpiece). In some cases spectral performance is measured after each layer of a multilayer coating, in order to correct manufacturing errors by adjusting the design slightly online, before depositing the next layer. ASTRON can deposit simple mirror coatings, but anything more complicated is specialist work and has to be contracted out.

6. Diffraction gratings are produced by either of 2 methods:
   - direct ruling by a diamond-tooled ruling engine (followed by replication in some cases: Figure 8-28).
   - holographic production of the line pattern, exposure of photosensitive material and subsequent etching (Figure 8-29).

The groove profiles obtained are different, and the holographic method can not only produce straight lines but also patterns that allow correction of some of the aberrations (in that way very compact spectrometer systems become possible). Grating production is very

14 At one point, for instance, Philips manufactured aspheric lenses for CD-players by hot-pressing low-precision glass blanks and then casting a high-precision epoxy coating on to these, reusing the precise and expensive casting molds many times (whether this is still their method of manufacture, I don't know).

15 The two component surfaces normally have equal curvature, so some design freedom is lost compared to the air-spaced component.
much a specialist craft, certainly for the outrageously large gratings required by
astronomy.
7. Optical fibers are made by first producing an ingot of the correct composition (core
material and cladding), then heating it to softening point and drawing it out into a thin
fiber while retaining its internal structure (don't ask me how, this is very much a specialist
business). Flexible coherent fiber bundles (image guides) are mostly made by winding
many turns of a single fiber on to a large wheel, cementing the assembly in one small
section on the circumference and then cutting through this section after the cement has
cured. Sometimes small square multifibers (e.g. $5 \times 5$) are assembled first, in order to
shorten and ease the handling problems in the final operation; the choice depends on the
required minimum bend radius.
8. Recently, semiconductor chipmaking technology has been pressed into service, with mask
projection and leaching/etching steps as part of the process. Such techniques are used to
produce, for instance, arrays of tiny addressable angle-scanning mirrors (Figure 8-25) and
also holographic optical elements (HOEs). As familiarity with the techniques grows and
optical methods spread into information processing, this area of production technology is
bound to expand; while it started from 'traditional' semiconductor production techniques,
it is branching out into new methods (such as the undercutting-by-etching, necessary to
produce programmable-tilt mirrors), in order to deal with 'miniaturized classical optics'
(very small components, but large compared to the wavelength) as well as 'optical
waveguide components' (size of order of the wavelength) and HOEs ('large' components,
with details whose size is of order one wavelength).

![Figure 8-27: An ASTRON diamond-machined single-point-mounted ('wineglass') mirror for IR use](image)
The material is a special aluminium alloy, thoroughly annealed after manufacture; the mirror surface is gold-coated; note the solid hinges in the base.
REPLICATION

Precision replication is a process that results in the transfer of the three-dimensional topography of a master grating to another substrate, allowing reproduction of a master in full relief to extremely close tolerances.

Replication involves the following steps:

1. Vacuum deposition of an extremely thin separation layer onto the master grating.
2. Vacuum deposition of aluminum on top of the separation layer.
3. Adding a thin coat of formulated epoxy to the aluminum layer.
4. Carefully positioning a second substrate, with no coating, on the epoxy-covered master.
5. After curing, the replica is separated from the master by a mechanical or thermal process.
6. The replica is then cut, cleaned, tested and packaged.

Figure 8-28: Grating manufacture by replication (Optometrics)

Figure 8-29: Schematic of holographic-grating production (Optometrics)
8.7. Test and measurement

Complementary to being able to design and manufacture (or contract out) optical components, it is absolutely necessary to be able to verify that one has actually obtained what one required, to the tolerance called for in the system design. This holds for components one produces in-house, but it applies even more to components or features obtained from external contractors*. Optical test and measurement has a traditional base, but it is expanding fast into specialized instruments for the hi-tech components one requires in modern astronomy. The ASTRON optics lab must (and does) participate in this development, but a large laboratory like ESO's is more like the true Mecca. Examples of indispensable test equipment include (very far from an exhaustive list, but a basic set, comparable with 'oscilloscope, multimeter, laboratory oscillator and Wheatstone bridge' for an electronics lab):

1. Fine-mechanical tools used during grinding such as thickness gauges, angle gauges, spherometers; these must be CLEAN, so can't be shared with a metalworking shop.
2. Simple optical tools such as magnifying glasses, microscopes, sodium lamp (to see interference fringes and thus estimate optical surface shape).
3. Alignment laser: used for quick alignment of a multi-component setup; can be used on flat and on curved surfaces; return beam is bright enough to observe, used for angular alignment; transmitted beam can be seen on the component, so serves as a rough guide on lateral alignment.
4. Auto-collimator: projects a parallel beam into a test setup. Return beams from (partial) reflections by flat surfaces in the test setup are compared to the input beam for parallellism (by observing crosswires and their reflections). Used for angular alignment to arcsecond accuracy. Auto-collimators exist in computer-assisted form, for monitoring purposes.
5. Precision rotator mount: used for verifying angles of prisms and cubes.
6. Wavefront interferometer. Expensive instrument for characterizing the optical surface to an accuracy of a small fraction of a wavelength. Produces a coherent (laser) beam (basically a flat wavefront, but this may be modified by auxiliary – very expensive – optics); forms and analyses an interference pattern between the input and return beams of the component under test.
7. Transmission/reflection spectrometer. For verifying filter passbands, mirror characteristics, etc. Generally this will require additions to a commercial system, in order to be able to scan across the component, tilt it, and vary the polarization of the beam. The wavelength range of astronomical interest starts at about 300 nm and continues to about 25 μm; this range of more than 6 octaves cannot be covered by one instrumental setup, but must contain exchangeable modules for parts of the wavelength range.

Desirable additions for modern instrumentation would be:

a. (Spectro)-polarimeter. For investigating the general polarization properties of a component. Again, commercial systems are usually too specialized. Probably the best course is to integrate polarimetry with the spectrometer.

* Harvey Butcher, in his previous incarnation as head of Roden's Kapteyn Sterrenwacht Werkgroep (†), fully realised this and laid the basis for what are now ASTRON's optical test and measurement capabilities.
b. Fiber test equipment. Commercial equipment is not suitable; it concentrates on fibers as substitutes for copper cables. For optical applications one wants to know things like spectral transmission (including the near UV), acceptance angle, focal-ratio degradation, polarization conversion. Almost certainly, the test setup will have to be an in-house development, perhaps integrated with the spectrometer/spectropolarimeter above.
9. One for the road.

Obviously, as optical technology expands further into computer-aided modifications of classical optics, into mass-produced molded optics, into true mini- and micro-optics and into optical-waveguide structures, "design" will split up into a large number of specialist areas, from basic system design (performed in astronomy mostly by technical astronomers) down to very specialist application design of things like optical waveguides, specialist coatings, or subsystems with extreme reliability, yet with tolerance of cryogenic, vacuum or Antarctic conditions, etc. Clearly, there will be room for generalists like the technical astronomer, for academic optical engineers and for specialized optical designers of one kind or another. If you are an extreme individualist, optics (like astronomy) is no longer the science/technology for you; however, if you are a good teamworker (or even prefer teamwork to a solitary existence), then the technology for optical astronomy offers as stimulating an environment as any, with strong links to mechanics, cryogenics, vacuum technology, electronics, computing, and of course academic optics.

For astronomers who may be reading this, I offer the following message:

In driving your car, you do not really need any technical knowledge. However, a certain amount of overall understanding helps you to get the best out of it. The present lecture notes, though not written expressly for you, may help you to drive your favourite telescope and instrument more intelligently than a mere keyboard operator would. And, of course, you just may come up with some brilliant improvement.

Let's hope for the best!

* Time, gentlemen, please. No comment intended on drunken driving; Dutchmen have bicycles.
A. The integral as a summation

For the discussion of Fourier Transforms, we need the concept of an integral. This appendix is a basic introduction for those who have not met the concept before.

If we represent a function $f(x)$ graphically:

then we call the area between the curve for $f(x)$ and the $x$-axis "the integral of $f(x)$", which is written as $\int f(x) \, dx$. This notation arises as follows:

We approximate $f(x)$ by a stepped function in $x$–intervals of width $\delta x$. The area under the curve is then approximately:

$$\sum_{1}^{N} f(x) \delta x$$

To improve the approximation, we allow $\delta x$ to decrease and $N$ to increase, keeping the product $N \delta x$ constant (the interval $x_a$ to $x_B$). In the limit, as $\delta x \to 0$ and $N \to \infty$, the stepped function becomes equal to $f(x)$ itself, and the sum is exactly equal to the area under the curve between
To indicate to the initiated that we have passed from the discrete steps $\delta x$ to the limit, we replace the discrete $\delta x$ by its symbolic limit form $dx$ and the sum sign $\sum$ by an integral sign $\int$

$$\int_{x_A}^{x_B} f(x) dx$$

Areas of positive $f(x)$ count positively, areas of negative $f(x)$ negatively. For many simple functions the integrals can be worked out from first principles and books exist listing integrals of all sorts of special functions, derived by clever mathematicians.

**Example:** Let us construct $f(X) = \int_0^X \cos x \, dx$, in other words the area underneath the curve of $\cos x$ vs $x$ between $x = 0$ and some value $x = X$ (i.e. $x_A = 0, x_B = X$). Draw for yourself the curve of $\cos x$ and split the "area under the curve" into narrow strips of width $\delta x$ as in the figure above. Now work out the value of the integral for various values of $X$. From $x = 0^\circ$ to $90^\circ$ this area counts positive, then from $x = 90^\circ$ to $270^\circ$ it counts negative, from $x = 270^\circ$ to $450^\circ$ it counts positive again, etc. So the value of the integral is 0 for $X = 0^\circ$, increases (at a decreasing rate) from $X = 0^\circ$ to $90^\circ$, then decreases again until at $X = 180^\circ$ it has equal positive and negative contributions (so equals 0) and from $X = 180^\circ$ to $360^\circ$ it is negative, again becoming 0 at $X = 360^\circ$. At $X = 90^\circ$ and $270^\circ$, $\cos x$ passes through 0, in other words the rate of increase of the integral is 0 and the value of the integral is at a maximum or a minimum.

Putting these facts together into a curve representing $f(X)$ vs $X$ (best plotted underneath $\cos x$ vs $x$), you will see that it looks remarkably like $\sin x$. And that is in fact the answer, as the "clever mathematicians" just referred to can prove.

Since an integral is a special kind of sum and a complex number (Hamaker sections 4 and 26) is a special kind of sum of 2 real numbers, the integral of a complex function is another special kind of sum and it can be split into the sum of a real and an imaginary part:

$$\int f_c(x) dx = \int \left[ f_r(x) + i f_i(x) \right] dx$$

$$= \int f_r(x) dx + i \int f_i(x) dx$$

Hence we know what we mean by the integral of a complex function and we can compute it (or at least construct it).
Graphically:

So far, we have integrated \( f(x) \) between \( x_A \) and \( x_B \). If \( f(x) \) drops to zero fast enough as \( x \to \pm\infty \), we may move \( x_A \) to \(-\infty\) and \( x_B \) to \(+\infty\) without the value of the integral running away to \( \infty \). This leads to the notation for the total area under the curve, from \( x=-\infty \) to \( x=+\infty \):

\[
\int_{-\infty}^{+\infty} f_c(x) \, dx
\]

Since we know what we mean (do we? see Hamaker sections 4 and 26) by the product of 2 complex numbers, we understand "the integral of a product of 2 complex functions of \( x \)":

\[
\int_{-\infty}^{+\infty} f_{1c}(x) \cdot f_{2c}(x) \, dx
\]

If we now take \( f_{2c}(x) \) to be \( e^{-2\pi i x} \) and write \( f_{1c}(x) \) simply as \( f(x) \), we have:

\[
F(s) = \int_{-\infty}^{+\infty} f(x) \cdot e^{-2\pi i x} \, dx
\]

which is the formula for the Fourier Transform in complex form.
This is all we really need to know, in order to understand what the Fourier Transform is about. There is of course a lot more to "integration" and its converse "differentiation". The mathematical subject is called "differential and integral calculus" and itself requires a book or a course of lectures for even a basic exposition. However, for a background course in optical instrumentation, the above is all we need: we know what the magic $\int$ sign denotes and that, if we wished, we could work out all the details. So we are now happy to accept what the experts tell us about Fourier Transforms, their properties and the consequences for instrumentation ("we don't understand the detailed design of an electric motor, but we do use a vacuum cleaner rather than a broom and dustpan"; come to think of it, do we ever wonder why dust grains do not slip through between the tufts of the broom?). So "Happy Transforms to You"!
# B. Glossary

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
<th>Figure</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADC</td>
<td>Atmospheric Dispersion Compensator (or: Compensation)</td>
<td>6-3</td>
</tr>
<tr>
<td>AIPS++</td>
<td>Advanced Image Processing System</td>
<td></td>
</tr>
<tr>
<td>AMBER</td>
<td>Astronomical Multiple-BEam Recombiner (VLTI instrument)</td>
<td>7-6</td>
</tr>
<tr>
<td>ANS</td>
<td>Astronomische Nederlandse Satelliet</td>
<td></td>
</tr>
<tr>
<td>APD</td>
<td>Avalanche Photo Diode</td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>Anti-reflection (coating)</td>
<td>8-12, 8-13</td>
</tr>
<tr>
<td>ASTRON</td>
<td><a href="https://www.astron.nl">Netherlands Foundation for Research in Astronomy</a> (‘NFRA’)</td>
<td></td>
</tr>
<tr>
<td>CCD</td>
<td>Charge-Coupled Device (in astronomy: the most common array detector)</td>
<td></td>
</tr>
<tr>
<td>CFHT</td>
<td>Canada-France-Hawaii Telescope</td>
<td></td>
</tr>
<tr>
<td>CHANDRA</td>
<td>X-ray satellite, previously known as AXAF</td>
<td></td>
</tr>
<tr>
<td>CID</td>
<td>Charge Injection Device (alternative for CCD in some applications)</td>
<td></td>
</tr>
<tr>
<td>CMOS</td>
<td>Complementary Metal-Oxide-Semiconductor (an electronic-chip technology)</td>
<td></td>
</tr>
<tr>
<td>COMPTEL</td>
<td>COMPton TELescope (γ-ray telescope)</td>
<td></td>
</tr>
<tr>
<td>DAI</td>
<td>Dictionary of Astronomical Instrumentation</td>
<td></td>
</tr>
<tr>
<td>DIY</td>
<td>Do It Yourself (Dutch: DHZ)</td>
<td></td>
</tr>
<tr>
<td>EM</td>
<td>Electro-Magnetic (radiation)</td>
<td></td>
</tr>
<tr>
<td>ESA</td>
<td>European Space Agency</td>
<td></td>
</tr>
<tr>
<td>ESO</td>
<td>European Southern Observatory</td>
<td></td>
</tr>
<tr>
<td>EUV</td>
<td>Extreme Ultra-Violet</td>
<td></td>
</tr>
<tr>
<td>FORS</td>
<td>FOcal Reducer/low-dispersion Spectrograph (VLT imager: 2 of them)</td>
<td>7-9</td>
</tr>
<tr>
<td>F-P</td>
<td>Fabry-Perot (etalon), a multi-path interferometer, basis of <em>interference filters</em></td>
<td></td>
</tr>
<tr>
<td>FT</td>
<td>Fourier Transform</td>
<td></td>
</tr>
<tr>
<td>HOE</td>
<td>Holographic Optical Element</td>
<td></td>
</tr>
<tr>
<td>ICCD</td>
<td>Intensified CCD</td>
<td></td>
</tr>
<tr>
<td>IF</td>
<td>Intermediate frequency</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>Infra-Red</td>
<td></td>
</tr>
<tr>
<td>IRAS</td>
<td>Infra-Red Astronomy Satellite</td>
<td></td>
</tr>
<tr>
<td>ISIS(POL)</td>
<td>The Cassegrain spectrometer at the WHT (and its polarimetry option)</td>
<td></td>
</tr>
<tr>
<td>IUE</td>
<td>International Ultraviolet Explorer (long-lived satellite for UV research)</td>
<td></td>
</tr>
<tr>
<td>JCMT</td>
<td>James Clerk Maxwell Telescope (mm/submm) on Hawaii</td>
<td></td>
</tr>
<tr>
<td>LETG</td>
<td>Low Energy Transmission Grating X-ray spectrometer aboard CHANDRA</td>
<td></td>
</tr>
<tr>
<td>LOFAR</td>
<td>Low Frequency Array (radio telescope in feasibility-study phase)</td>
<td></td>
</tr>
<tr>
<td>MFFE</td>
<td>Multi-Frequency Front End (at WSRT)</td>
<td></td>
</tr>
<tr>
<td>MIDI</td>
<td>Mid-infrared Interferometric Instrument (VLTI instrument)</td>
<td>7-7</td>
</tr>
<tr>
<td>MOSFET</td>
<td>Metal-Oxide-Semiconductor Field-Effect Transistor</td>
<td></td>
</tr>
<tr>
<td>NFRA</td>
<td>See ASTRON</td>
<td></td>
</tr>
<tr>
<td>OASIS</td>
<td>Optically Adaptive System for Imaging Spectroscopy</td>
<td>7-19</td>
</tr>
<tr>
<td>PFA</td>
<td>Police Research</td>
<td></td>
</tr>
<tr>
<td>PUMA</td>
<td>PULsar MACHine (at WSRT)</td>
<td></td>
</tr>
<tr>
<td>QED</td>
<td>Quod Erat Demonstrandum ('which we set out to prove')</td>
<td></td>
</tr>
<tr>
<td>SAX</td>
<td>Italian-Dutch satellite for X-ray astronomy</td>
<td></td>
</tr>
<tr>
<td>SCUBA</td>
<td>Submillimetre Common-User Bolometer Array (at the JCMT)</td>
<td></td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>SKA</td>
<td>Square Kilometre Array (radio telescope in feasibility and development phase)</td>
<td></td>
</tr>
<tr>
<td>SRON</td>
<td>Stichting Ruimte Onderzoek Nederland</td>
<td></td>
</tr>
<tr>
<td>STJ</td>
<td>Superconducting Tunnel Junction (a developmental optical detector)</td>
<td></td>
</tr>
<tr>
<td>TAURUS</td>
<td>Fabry-Perot imager at the WHT. Prototype looked like a bull: Figure 7-8</td>
<td></td>
</tr>
<tr>
<td>UES</td>
<td>Utrecht Echelle Spectrometer (at the WHT)</td>
<td></td>
</tr>
<tr>
<td>UV</td>
<td>Ultra-Violet</td>
<td></td>
</tr>
<tr>
<td>UVES</td>
<td>Ultra-violet and Visible Echelle Spectrograph for the VLT Figure 7-17</td>
<td></td>
</tr>
<tr>
<td>VISIR</td>
<td>VLT Imager-Spectrometer for the IR Figure 3-3, Figure 7-10 and Figure 7-18</td>
<td></td>
</tr>
<tr>
<td>VLA</td>
<td>Very Large Array (US aperture synthesis telescope)</td>
<td></td>
</tr>
<tr>
<td>VLBA</td>
<td>Very-Long-Baseline Array (US VLBI array)</td>
<td></td>
</tr>
<tr>
<td>VLBI</td>
<td>Very-Long-Baseline Interferometry</td>
<td></td>
</tr>
<tr>
<td>VLT</td>
<td>Very Large Telescope; ESO, Paranal (Chile)</td>
<td></td>
</tr>
<tr>
<td>VLTI</td>
<td>VLT Interferometer; the spatial-interferometry mode of the VLT Figure 7-5</td>
<td></td>
</tr>
<tr>
<td>WHT</td>
<td>William Herschel Telescope (La Palma 4.2-metre)</td>
<td></td>
</tr>
<tr>
<td>WSRT</td>
<td>Westerbork Synthesis Radio Telescope</td>
<td></td>
</tr>
</tbody>
</table>