Imaging in Astronomy

JOHAN BLEEKER
SRON Netherlands Institute for Space Research
Astronomical Institute Utrecht

August 12, 2022

Contents

1 Observing the spectral radiance of the sky: sensing impact

1.1 Characterisation of a celestial radiation field: radiometric definitions

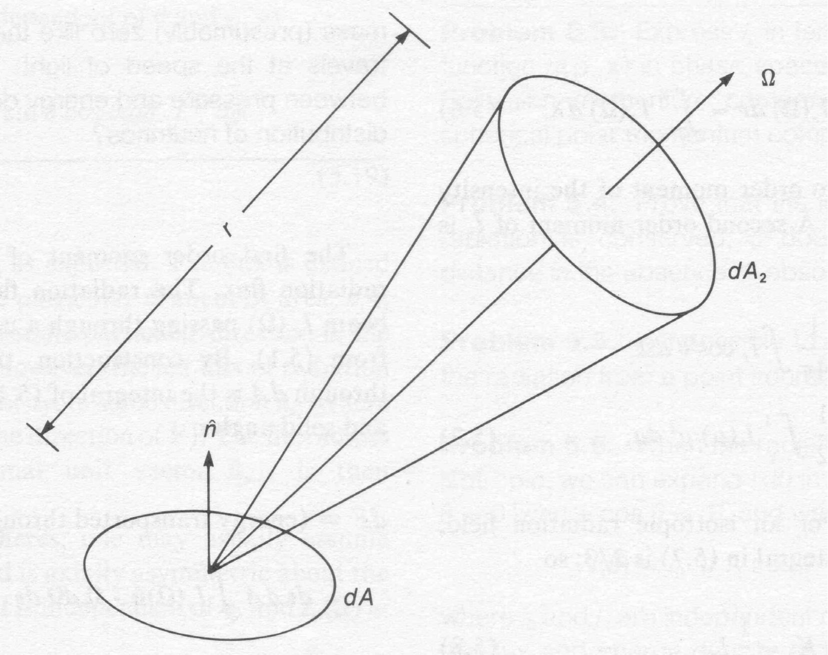

Figure 1: *geometry for defining intensity(radiance).*

The *spectral radiance* or *monochromatic intensity* is the basic quantity to describe a radiation field. It is defined as the amount of radiant energy per unit time, per unit area perpendicular to the beam, per unit solid angle, per unit wavelength (or frequency, energy). The relevant geometry is displayed in figure 1, the spectral radiance is expressed as:

$$I(\lambda, \vec{\Omega}, t) = \frac{dE}{\vec{n} \cdot \vec{\Omega}\ d\Omega\ dt\ d\lambda\ dA} \tag{1}$$

The *monochromatic radiation flux density* or *spectral irradiance* $F(\lambda, t)$ is derived by integrating $I(\lambda, \vec{\Omega}, t)$ over the solid angle:

$$\begin{aligned} F(\lambda, t) = \frac{dE}{dt\ dA\ d\lambda} &= \int\limits_{\Omega} I(\lambda, \vec{\Omega}, t)\ \vec{n} \cdot \vec{\Omega}\ d\Omega \\ &= \int\limits_0^{2\pi} \int\limits_0^{\pi} I(\lambda, \theta, \phi, t)\ \cos\theta\ \sin\theta\ d\theta\ d\phi \end{aligned} \tag{2}$$

The *total irradiance* $F(t)$ is subsequently obtained by integrating over all wavelengths. The *spectral(monochromatic) radiant flux* is the total amount of monochromatic radiant energy that is transported through a given area per unit time interval:

$$\Phi(\lambda, t) = \frac{dE}{dt\, d\lambda} = \int\limits_{\Omega} \int\limits_{A} I(\lambda, \vec{\Omega}, t)\ \vec{n} \cdot \vec{\Omega}\ d\Omega\, dA \tag{3}$$

Finally, the *radiant flux* represents the total amount of radiant energy that is transported through a given area, integrated over all wavelengths, per unit time:

$$\Phi(t) = \frac{dE}{dt} = \int\limits_{\Omega} \int\limits_{A} \int\limits_{\lambda} I_{\lambda}(\vec{\Omega}, t)\ \vec{n} \cdot \vec{\Omega}\ d\Omega\, dA\, d\lambda \tag{4}$$

For an isotropic radiation field from the upper hemisphere, e.g. an isotropically distributed sky background, the following relation holds:

$$F(\lambda, t) = \pi I(\lambda, \vec{\Omega}, t) \tag{5}$$

and for a point source at position $\vec{\Omega}_0$ with spectral radiance $S_p(\lambda, t)$ the spectral irradiance is:

$$\begin{aligned} F(\lambda, t) &= \int\limits_{\Omega} S_p(\lambda, t)\ \delta(\vec{\Omega} - \vec{\Omega}_0)\ \vec{n} \cdot \vec{\Omega}\ d\Omega \\ &= S_p(\lambda, t)\ \vec{n} \cdot \vec{\Omega}_0 = S_p(\lambda, t)\ \cos\theta_0 \qquad (6) \\ &= S_p(\lambda, t) \qquad \text{for normal incidence} \qquad (7) \end{aligned}$$

The spectral radiance I, which characterizes the radiation beam, has intrinsic statistical fluctuations. As a consequence, an astronomical measurement has to be treated as a stochastic process.

1.2 The sensing process: filtering of a stochastic signal

In observational astrophysics, the system response to an incoming radiation field can be characterised by filtering, arising from the individual elements making up the telescope configuration, of a stochastic process described by the monochromatic intensity $I(\nu, \vec{\Omega}, t)$. Generally, the time dependent output of the telescope is described by

$$X(t) = S(t) + N(t) \tag{8}$$

in which $S(t)$ represents the outcome of the filtering of the signal source and $N(t)$ represents the sum of all (filtered) noise components like background radiation, disturbances arising from the operational environment and intrinsic noise in the detection system, like dark current. The measuring process of the source signal $S(t)$ can be symbolically written as a series of consecutive convolutions, defined by the angular and the spectral response functions of the observational instrument:

$$S(t) = \int\limits_{\Delta\nu} \left[R(\nu) * \int\limits_{\Delta\vec{\Omega}_{FOV}} \left[I(\vec{\Omega}, \nu, t) * P(\vec{\Omega}, \nu) \right] d\vec{\Omega} \right] d\nu \tag{9}$$

$P(\vec{\Omega}, \nu)$ represents the collecting power of the telescope. which depends in general on the frequency (or wavelength, or energy). It is a function of the telescope off-axis angle in the field of view $\vec{\Omega}_{FOV}$ and contains the point spread function $H(\vec{\Omega}, \nu)$, which quantitatively describes the angular resolution (field position dependent). $\Delta\vec{\Omega}_{FOV}$ gives the solid angle over which the convolution $I(\vec{\Omega}, \nu, t) * P(\vec{\Omega}, \nu)$ is to be integrated. The choice of $\Delta\vec{\Omega}_{FOV}$ entirely depends on the number of image elements in the field of view, which may be as large as 10^8 in modern systems, and on the objective of the particular observation (e.g. ultimate angular resolution or high quality spectrum over a relatively large part of the field of view). Accordingly the integral can be done over the whole field of view $\vec{\Omega}_{FOV}$, which may cover a large part of the sky in the case of wide-field cameras, or it may be done over just one image pixel. The integration of the signal after the second convolution with $R(\nu)$ covers the spectral range of interest $\Delta\nu$, which is part of the total bandwidth ν. Again, this may vary from a very narrow range (e.g. measuring the line profile of a single spectral line) to a very broad range (in case of photometry). The number of frequency elements can therefore range from 1 (e.g. in the case of a bolometric detector) to approximately 10^6 in a high-resolution spectrograph. It is important to remember continually that the term *frequency* covers here three types of Fourier pairs.Every measurement or observation implies bandwidth limitations on each of these frequencies.

- The pair $I(t) \Leftrightarrow I(f)$ refers to time resolution, the frequency f relates to *temporal frequency.*

- The pair $I(\vec{\Omega}) \Leftrightarrow I(\vec{\zeta})$ refers to spatial resolution, the frequency $\vec{\zeta}$ in the Fourier domain has to be interpreted as *spatial frequency.* This refers to structures in the image contrast.

- The pair $I(\nu) \Leftrightarrow I(s)$ refers to spectral resolution, in this case s is a Fourier frequency which relates to a *spectral frequency.* A spectrum containing a large number of sharp features, like narrow emission and absorption lines, possesses much power in high spectral frequencies; a featureless continuum contains only low spectral frequencies.

The normalised value of the Fourier transform of a particular instrument response function, e.g. $R(s)$ or $H(\vec{\zeta})$, is called the *Modulation Transfer Function* (MTF) and describes the frequency dependent filtering of the source signal in the Fourier domain. The MTF refers either to the amplitude/phase transfer function of the signal or to the power transfer function, in practice this will be explicitly clear from the specific context in which the MTF is employed.

1.3 Observation filtering

The sensing (observation) process inevitably filters the sky image in at least three different ways:

- *Through* the finite exposure length, which leads to a finite number of photons incident during the measurement time. Consequently we always deal with a *sample* of the parent distribution representing the true image. This introduces measurement noise and leads to an imperfect restitution of $I(\vec{\Omega}, \lambda, t)$, relative to a measurement with infinite signal-to-noise ratio.

- *Through* the finite size of the telesope or antenna apertures, that imposes a fundamental restriction on the attainable image quality due to diffraction. In addition, the focussing properties and the imperfections in the realisation of the optical surfaces lead to geometrical aberrations that may become dominant over the fundamental restrictions imposed by diffraction.

- *Through* the radiation beam crossing the Earth's heterogeneous and turbulent atmosphere. This filtering applies of course solely to ground based observatories, in modern telescope systems this can partly be compensated for by employing adaptive optics.

In the following sections we shall treat in some detail the effects on image formation arising from diffraction, geometrical aberration and atmospheric "seeing".
At wavelengths shorter than ~ 0.1 nanometer, the application of focussing optics becomes untenable due to extremely low reflection efficiencies of the optical surfaces. Imaging can still be accomplished by employing beam modulation techniques (e.g. coded mask telescopes) and, at gamma-ray wavelengths, by exploiting the directionality properties of the photon interaction processes. In the latter case, the optical element and the detection system have become one and the same device. A discussion of these techniques is presented in section 9 of this booklet and is elaborated with reference to recently operated space observatories.

2 Time filtering

2.1 Finite exposure and time resolution

In practice, the measurement or registration of a stochastic process always takes place over a finite period T and with a certain resolution ΔT, i.e. the minimum time bin for a data point. The limitation in measuring time T corresponds to a *multiplication in the time domain* of a stochastic variable $X(t)$ with a window (block) function $\Pi(t/T)$. This function is described as follows,

$$\begin{aligned} \Pi\left(\frac{t}{T}\right) &\equiv 1 \qquad \mathrm{f}or \qquad |t| \leq \frac{1}{2}T && (10)\\ \Pi\left(\frac{t}{T}\right) &\equiv 0 \qquad \mathrm{f}or \qquad |t| > \frac{1}{2}T && (11)\\ & && (12) \end{aligned}$$

Consequently a new, time filtered, stochastic variable $Y(t)$ is introduced:

$$Y(t) = \Pi\left(\frac{t}{T}\right) X(t) \tag{13}$$

The limitation in time resolution always arises in practice due to the frequency-limited transmission characteristic of any physical measuring device.
Suppose, as an example, the measurement is taken at time t within the measuring period T with a temporal resolution ΔT. This corresponds to an integration of the stochastic variable $Y(t)$ between $t - \Delta T/2$ and $t + \Delta T/2$, divided by ΔT (a so called running average). It follows that

$$Z(t) \equiv Y_{\Delta T}(t) = \frac{1}{\Delta T}\int\limits_{t-\Delta T/2}^{t+\Delta T/2} Y(t')dt' = \frac{1}{\Delta T}\int\limits_{-\infty}^{+\infty} \Pi\left(\frac{t-t'}{\Delta T}\right) Y(t')dt' \tag{14}$$

This equation can also be expressed in terms of a convolution in the time domain:

$$Z(t) = \frac{1}{\Delta T}\Pi\left(\frac{t}{\Delta T}\right) * Y(t) = \frac{1}{\Delta T}\Pi\left(\frac{t}{\Delta T}\right) * \Pi\left(\frac{t}{T}\right) X(t) \tag{15}$$

which represents a low-frequency (or 'low-pass') filtering of the stochastic variable $Y(t)$. The values μ_T and $R_T(\tau)$ for an ergodic process obtained from a finite measuring period T will therefore slightly differ from the true values μ and $R(\tau)$. The error introduced by measuring the sample average μ_T rather than the true average μ is the subject of the next paragraph.

2.2 Error in a sample average

We wish to determine the accuracy with which the approximate value μ_T approaches the real value μ. An illustration of this is given in Figure 2. To do so we start by noting that determining the average corresponds to convolution in the time domain with a block function. We denote this

$$X(t) \rightarrow \boxed{\tfrac{1}{T}\Pi\left(\tfrac{t}{T}\right)} \rightarrow X_T \tag{16}$$

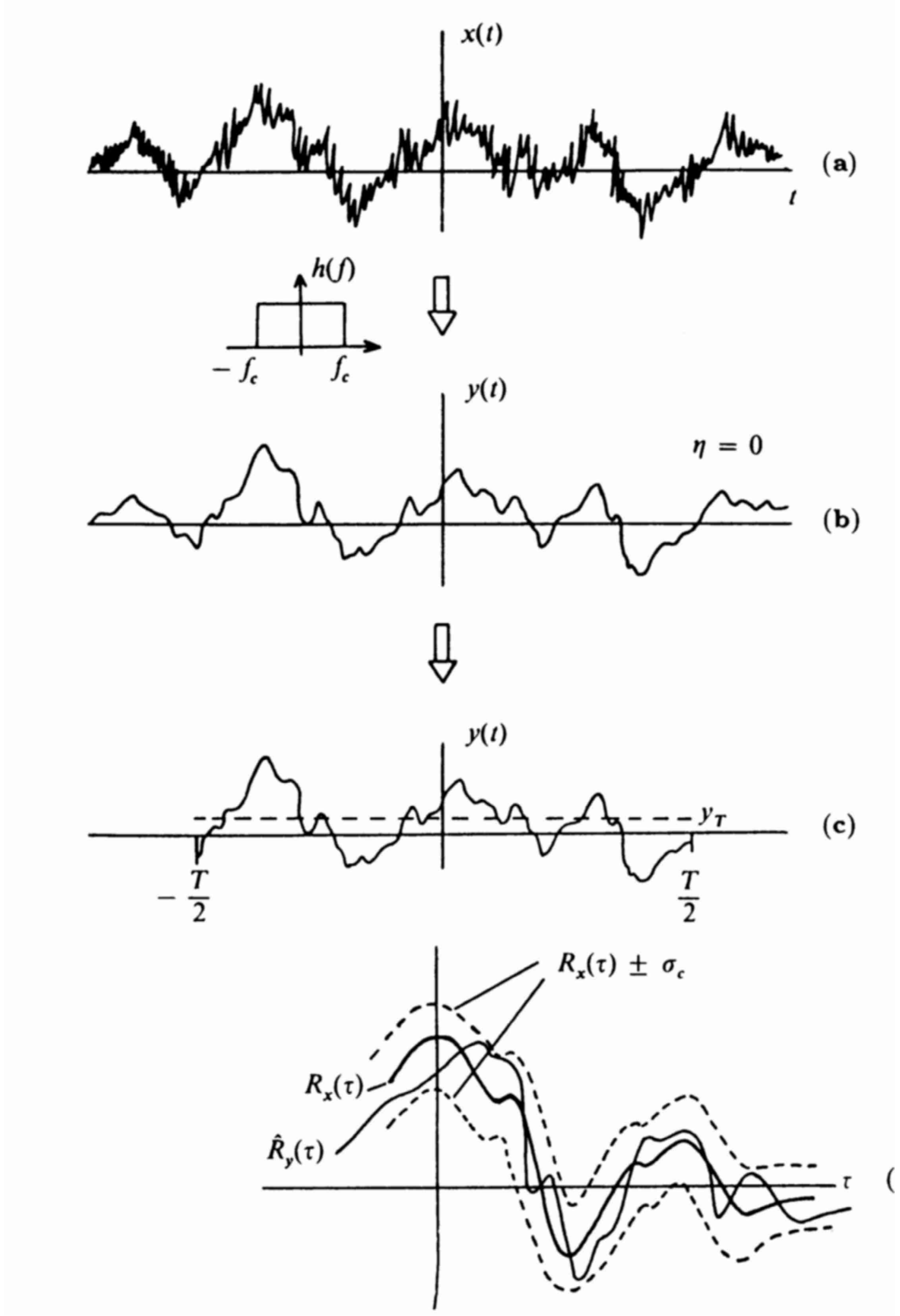

Figure 2: *Illustration of the errors in average and power spectrum of a stochastic process $x(t)$. a) a realization of the process b) filtered with a low-pass filter (high frequencies are removed): $y(t)$ c) measurement during finite time interval T, and the average during this interval (dashed line, value y_T) d) the autocorrelation functions of the process $R_x(\tau)$ and of the measurment $R_y(\tau)$. Figure taken from Lena et al 1998.*

In terms of the Fourier transforms, the averaging corresponds to a multiplication with a sinc-function. In general, we write the effect of the measuring apparatus on the signal in the Fourier domain as $Y(f) = X(f)H(f)$ and hence $Y^*(f) = X^*(f)H^*(f)$, and with $H(f)$ the *transfer function*. We can thus also write $|Y(f)|^2 = |X(f)|^2|H(f)|^2$. **Note:** in the literature the term transfer function is used both for $H(f)$, i.e. the signal transfer function, but also for $|H(f)|^2$, i.e. the power transfer function. So be prepared for the correct interpretation whenever you encounter the term transfer function in the literature! With this we take the Fourier transform of the autocorrelation, to find:

$$S_{X_T}(f) = |H(f)|^2 S_{X(t)}(f) = \mathrm{sinc}^2(Tf) \cdot S_{X(t)}(f) \tag{17}$$

Transforming back to the time domain, we write this as

$$R_{X_T}(\tau) = h(\tau) * h(\tau) * R_{X(t)} \tag{18}$$

and note that $h \equiv (1/T)\Pi(t/T)$ is a real function. We note that the convolution of a block with itself is a triangle, and introduce the notation:

$$h(\tau) * h(\tau) \equiv \rho(\tau) \equiv \frac{1}{T}\Lambda\left(\frac{\tau}{T}\right)$$

to rewrite Eq. 18 as

$$R_{X_T}(\tau) = \frac{1}{T}\Lambda\left(\frac{\tau}{T}\right) * R_{X(t)} \equiv \frac{1}{T}\int\limits_{-\infty}^{+\infty} \Lambda\left(\frac{\tau'}{T}\right) R_{X(t)}(\tau - \tau')d\tau' \tag{19}$$

For convenience we consider the case where $\mu = 0$, i.e. $R = C$, and find

$$C_{X_T}(\tau) = \frac{1}{T}\int\limits_{-\infty}^{+\infty} \Lambda\left(\frac{\tau'}{T}\right) C_{X(t)}(\tau - \tau')d\tau' \tag{20}$$

To compute the variance we set $\tau = 0$ and find

$$C_{X_T}(0) \equiv [\sigma_{X_T}]^2 = \frac{1}{T}\int\limits_{-\infty}^{+\infty} \Lambda\left(\frac{\tau'}{T}\right) C_{X(t)}(-\tau')d\tau' = \frac{1}{T}\int\limits_{-\infty}^{+\infty} \Lambda\left(\frac{\tau'}{T}\right) C_{X(t)}(\tau')d\tau' \tag{21}$$

where we have used the fact that C is even. Explicitly writing Λ we finally obtain

$$[\sigma_{X_T}]^2 = \frac{1}{T}\int\limits_{-T}^{+T} \left(1 - \frac{|\tau'|}{T}\right) C_{X(t)}(\tau')d\tau' \tag{22}$$

Two things are to be noted in this equation. First: the integral ranges from $-T$ to $+T$, i.e. over a range with length $2T$, but nonetheless the normalization factor is $1/T$. Second, the autocovariance is *always* limited in the frequency domain.

As an example we consider a specific form for the transfer function, viz.:

$$H(f) = \frac{1}{1 + 2\pi i f \tau_o} \tag{23}$$

At small frequencies $f \ll 1/(2\pi\tau_o) \equiv f_o$ the transfer is complete, i.e. $|H(f)| = 1$, and at high frequencies $f \gg f_o$ the transfer is inversely proportional to the temporal frequency, i.e. $|H(f)| = f_o/f$. The frequency f_o is the 'cut-off' frequency of the transfer function $H(f)$.

We will forego the full mathematics here, and merely conclude that the autocovariance with such a system drops exponentially with (the absolute value of) the time difference τ:

$$C_{X(t)}(\tau) = C_{X(t)}(0)e^{-|\tau|/\tau_o} \qquad \text{where} \qquad \tau_o \equiv \frac{1}{2\pi f_o} \tag{24}$$

This means that at times $\tau \gg \tau_o$ the correlation is virtually zero. By entering Eq. 24 into Eq. 22 and performing the integration, we get

$$\left[\sigma_{X_T}\right]^2 = 2\left[\sigma_{X(t)}\right]^2 \frac{\tau_o}{T}\left[1 - \frac{\tau_o}{T}\left(1 - e^{-T/\tau_o}\right)\right] \tag{25}$$

To get a feeling for the meaning of this equation we consider two limiting cases. First the one in which the duration of the measurement largely exceeds the correlation time, $T \gg \tau_o$. Eq. 25 then becomes

$$\left[\sigma_{X_T}\right]^2 = 2\left[\sigma_{X(t)}\right]^2 \frac{\tau_o}{T} = \frac{\left[\sigma_{X(t)}\right]^2}{\pi f_o T} \tag{26}$$

and we see that the variance of the measured signal is proportional to the variance of the incoming signal, and approximates zero when the duration of the measurement goes to infinity, and also when the number of frequencies over which one measures goes to infinity. The measured signal is then said to be *ergodic in the mean*. The last limit can be understood by noting that f_oT is the number of cycles during T with a frequency f_o, i.e. it gives the number of measurements; Eq. 26 thus is analogous to the equation which gives the variance of the average $\sigma_\mu^2 = \sigma^2/N$.

The other limit we consider is the one for which the duration of the measurement equals the correlation time, $T = \tau_o$. Eq. 25 in this limit becomes

$${\sigma_{X_T}}^2 = 2{\sigma_{X(t)}}^2 e^{-1} \simeq {\sigma_{X(t)}}^2 \tag{27}$$

This is also understandable in terms of determining the average, in the case where just one measurement is taken: $N = 1$. The moral we can draw from this example is that one must take good care that the duration of the measurement is much longer than the correlation time, $T \gg \tau_o$, if one wishes to avoid *large* errors in the estimates of the average and of the variance. Another moral is that we must take into account the errors in the average and in the variance whenever we are looking for really *small* effects.

3 Spatial filtering

3.1 Diffraction by a single aperture

3.1.1 The Huygens-Fresnel principle

Consider a point source S at great distance compared to the size of the aperture of an observing telescope, a condition that is mostly satisfied in observational astrophysics. The EM-wave incident on the telescope aperture can hence be described by a plane wave. Propagation of this plane wave beyond the aperture opening is then governed by the Huygens-Fresnel principle. This principle states that *every unobstructed point of a wavefront, at a given instant in time, serves as a source of secondary wavelets with the same frequency as that of the primary wave. The amplitude of the radiation field at any point* ***beyond****, is the superposition of all these wavelets considering their amplitudes and relative phases.*

3.1.2 Fresnel and Fraunhofer diffraction

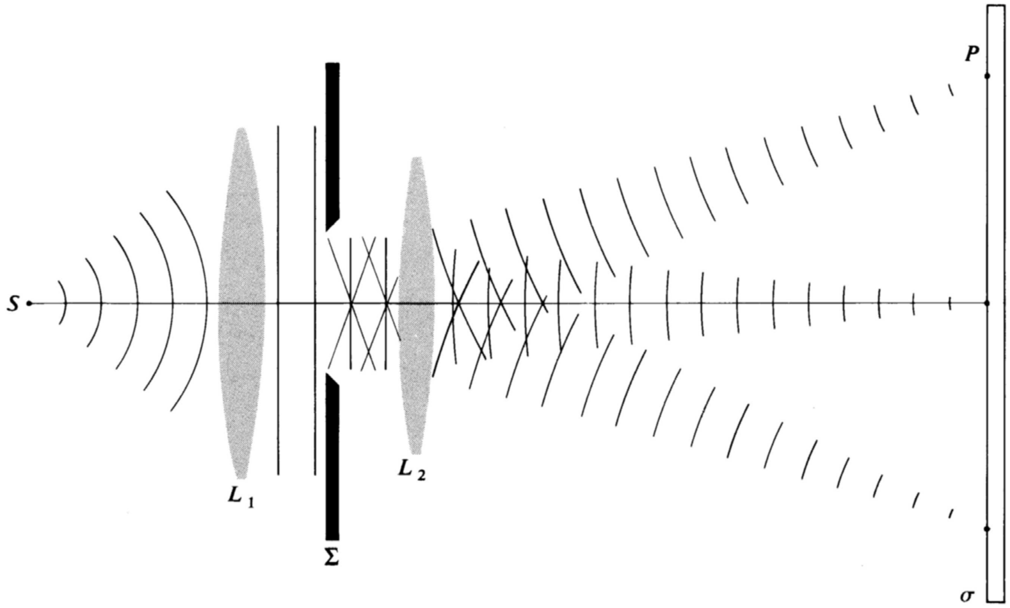

Figure 3: *Fraunhofer diffraction. Figure taken from Hecht 1987.*

Now imagine the aperture of the telescope as an opening in an opaque screen and consider a plane of observation very close behind this aperture. Under these conditions a clear and sharp image of the aperture is recognizable, despite some light fringing around its periphery (keep in mind that the idealized *geometric* aperture image corresponds to $\lambda \to 0$). If the plane of observation is moved further away from the aperture, the radiation starts to diverge and the image of the aperture becomes increasingly more structured as fringes become more prominent. This is known as **Fresnel** or **near field** diffraction. If the plane of observation is gradually moved out further, the fringes continually change, the projected pattern is now spread out considerably and bears almost no resemblance anymore to the actual aperture. Still further out, only the size of the pattern changes and not its shape. The radiation now propagates in spherical expansion. If R is the distance from the aperture, the wave amplitudes decrease proportional to R and the radiation power with R^2, in contrast to the plane wave incident on the aperture opening. In this region we have the **Fraunhofer** or **far field** limit for diffraction, which holds for the great majority of observations in astronomy. Putting it differently: taking a point source S and a point of observation P, both very far from the aperture opening, Fraunhofer diffraction applies as long as the incoming and outgoing (in a conical fashion) wavefronts approach being planar (i.e. differing therefrom by a small fraction of a wavelength over the extent of the diffracting aperture or obstacle).

Figure 3 shows the planar wave situation applicable to Fraunhofer diffraction, two lenses (L_1, L_2) have been inserted to reposition the radiation source S and the observation position P from infinity to a finite and physically feasible location.

A more quantitative way to appreciate this is that in the superposition of all wavelets at the observation point P (application of the Huygens-Fresnel principle defined above) the *phase* of each contributing wavelet at P, due to the differences in path traversed, is crucial to the determination of the resultant field amplitude in P. Now if the wavefronts impinging on and emerging from the aperture are planar, these path differences will be describable by a *linear function* of the two geometric aperture variables, say a 2-dimensional position vector $\vec{r}$. This *linearity in the aperture variables of the wavelet's phase differences* is the explicit mathematical criterion for the prevalence of Fraunhofer diffraction!

3.1.3 Point Spread Function (PSF) and Optical Tranfer Function (OTF) in the Fraunhofer limit

Consider the primary mirror, lens or radio dish of a telescope viewing the sky. The plane through the rim of the primary mirror (dish) is defined as the *aperture* or *pupil plane*. A position in this pupil plane can be specified by a 2-dimensional position vector $\vec{r}$. The position of a celestial source in the sky field being observed can be specified by a unit direction vector $\vec{\Omega}(\Omega_x, \Omega_y, \Omega_z)$, its components represent the direction cosines relative to a Cartesian coordinate system with its origin at the centre of the mirror. The x-axis is chosen perpendicular to the pupil plane and the y and z axes are located in the pupil plane, see figure 4. Obviously $\Omega_x^2 + \Omega_y^2 + \Omega_z^2 = 1$ should hold, so if Ω_y and Ω_z are given Ω_x is fixed apart from its sign. However the radiation is incident from the upper hemisphere, hence we have $\Omega_x \geq 0$. Consequently a source position in the sky can be described by a 2-dimensional rather than a 3-dimensional vector $\vec{\Omega}(\theta_y, \theta_z)$. The angular components θ_y and θ_z refer to two orthogonal angular coordinate axes across the sky field under observation, they are related to the direction cosines of $\vec{\Omega}$ through $\sin\theta_y = \Omega_y/(1-\Omega_z^2)^{\frac{1}{2}}$ and $\sin\theta_z = \Omega_z/(1-\Omega_y^2)^{\frac{1}{2}}$. In observational astronomy a particular sky field under observation is practically always strictly limited in angular size ($\leq$ square degree). Therefore, in good approximation, we can assume that $\theta_y \approx \sin\theta_y \approx \Omega_y$ and $\theta_z \approx \sin\theta_z \approx \Omega_z$.

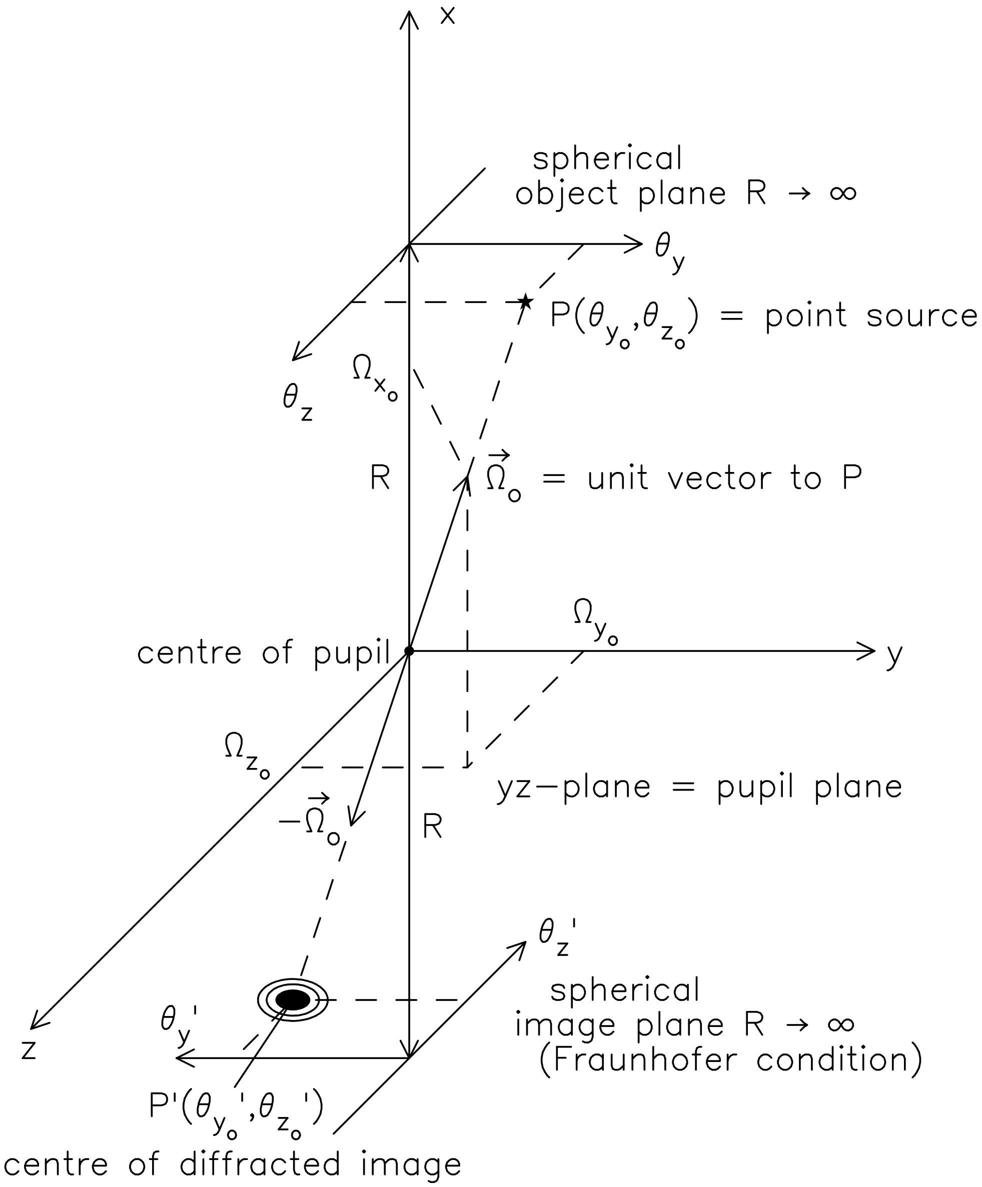

Figure 4: *Camera Obscura: sky and image coordinates in the Fraunhofer limit.*

To stipulate the essence of Fraunhofer diffraction, we reduce the telescope primary mirror to a single pupil in the yz-plane. If we consider the image plane to be spherical at a very large distance $R \to \infty$ from the pupil plane, a celestial point source at sky coordinates $\vec{\Omega}_0(\theta_{y0}, \theta_{z0})$ will produce a geometrical image located at $-\vec{\Omega}_0(\theta'_{y0}, \theta'_{z0}) = -\vec{\Omega}_0(-\theta_{y0}, -\theta_{z0})$ in the image plane. *In this approach the telescope has been reduced to a Camera Obscura.* This has fundamentally no impact on our analysis, since in reality the *actual telescope mirror only serves to retrieve the point source image* at $R = \infty$ to a practical distance: the focal length of the telescope. As a consequence of the diffraction by the pupil, the image of the celestial point source will not be limited to the direction $\vec{\Omega}_0$ of the geometrical image, but will also be blurred around this direction. (*Interalia:* we can omit the minus sign by simply reversing the pointing direction of

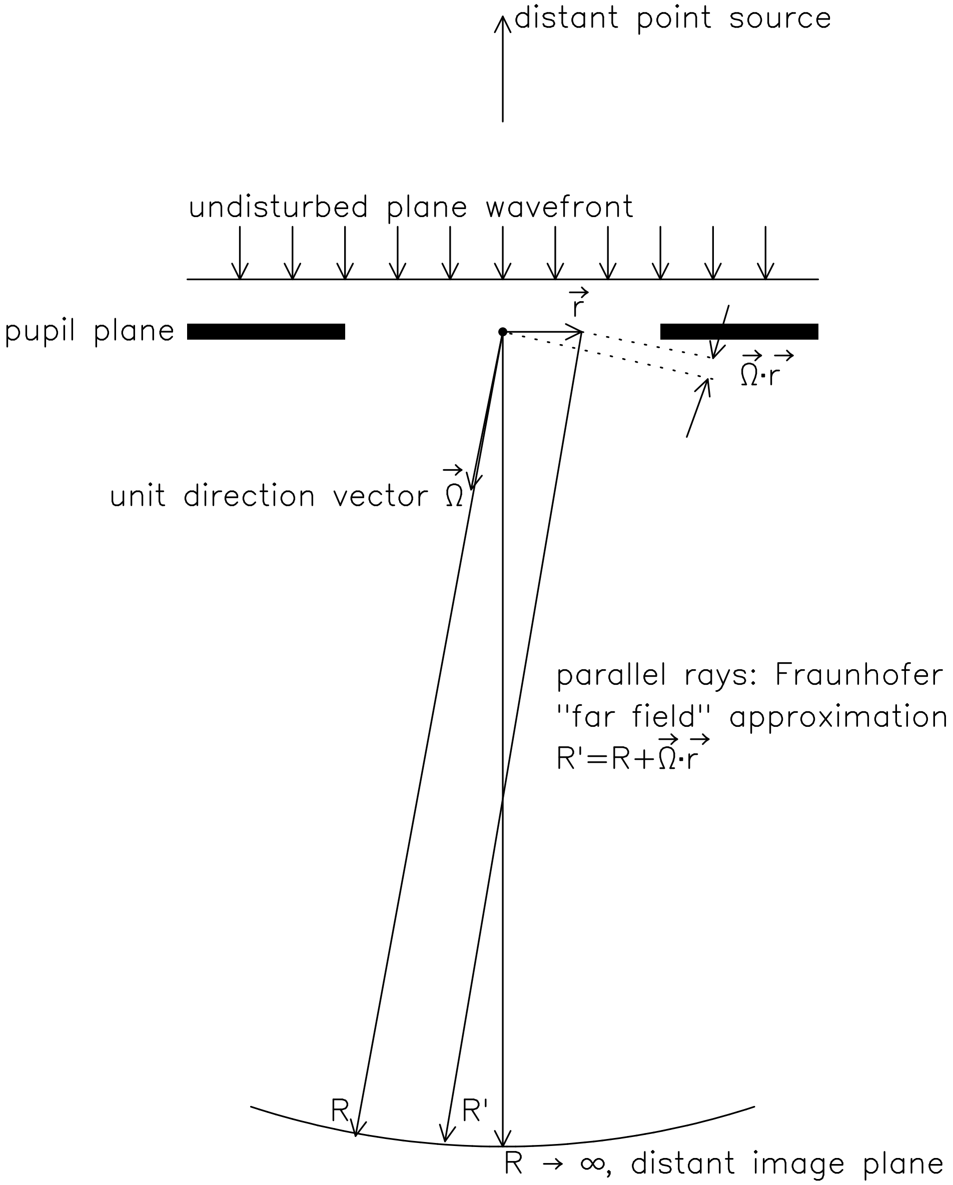
distant point source
undisturbed plane wavefront
pupil plane
$\vec{r}$
$\vec{\Omega}\cdot\vec{r}$
unit direction vector $\vec{\Omega}$
parallel rays: Fraunhofer
"far field" approximation
$R'=R+\vec{\Omega}\cdot\vec{r}$
R
R'
R → ∞, distant image plane

Figure 5: *Geometry for diffraction by the telescope pupil in the Fraunhofer limit.*

$\vec{\Omega}_0$, i.e. now pointing towards the image plane in the anti-source direction). Hence, if we have a celestial point source of *unit intensity* at a sky position $\vec{\Omega}_0$, the intensity of the diffraction image need to be described by a function $h(\vec{\Omega}-\vec{\Omega}_0)$: the Point Spread Function (PSF) of the diffraction-limited telescope. The Fourier transform of $h(\vec{\Omega}-\vec{\Omega}_0)$ is designated the Optical Tranfer Function (OTF), i.e. we have PSF $\Leftrightarrow$ OTF.
For the derivation, in the Fraunhofer limit, of the relation between the radiation field in the pupil plane and the emerging diffraction pattern described by the telescope's PSF consider the geometry in figure 5. A quasi-monochromatic celestial point source is located at the origin of the angular sky coordinates, i.e. along the direction of the x-axis ($\theta_y=0,\theta_z=0$). This source, at very large distance from the telescope aperture, produces a flat wavefront in the pupil plane. Taking the centre of the pupil as the origin, we have at every position $\vec{r}$ of the pupil the same electric field signal, i.e. a fully

coherent field distribution across the pupil, analytically expressed as:

$$\tilde{E}(t) = \tilde{E}_0(t) \cdot e^{i2\pi\bar{\nu}t} \quad \text{with the amplitude "phasor"} \quad \tilde{E}_0(t) = \mid \tilde{E}_0(t) \mid \cdot e^{i\phi(t)} \tag{28}$$

According to *Huygens-Fresnel*, the resulting field $\tilde{E}(\vec{\Omega}, t)$ in a direction $\vec{\Omega}$ on a distant image sphere with radius R is the superposition of the contributions from all positions within the pupil, each contribution with its own specific phase delay due to the difference in path length:

$$\tilde{E}(\vec{\Omega}, t) \;=\; \frac{C}{R} \int\!\!\int_{\text{pupil}} \tilde{E}_0 \left(t - \frac{R'(\vec{\Omega}, \vec{r})}{c} \right) e^{i2\pi\bar{\nu}\left(t - \frac{R'(\vec{\Omega},\vec{r})}{c}\right)} d\vec{r} \tag{29}$$

with $R'(\vec{\Omega}, \vec{r})$ the distance between the pupil position $\vec{r}$ and the image position in the direction $\vec{\Omega}$ (see figure 5), $d\vec{r} = dydz$, $1/R$ the amplitude damping factor due to the spherical expansion of the wavefield and C a proportionality constant. To evaluate the integral in equation (29) in the Fraunhofer limit, we make two important assumptions:

- The coherence length $l_c = c\tau_c$ is presumed large compared to the maximum path difference between the waves originating at different positions within the pupil (i.e. large compared to the pupil diameter). This imposes a requirement on the maximum allowable frequency bandwidth of the quasi-monochromatic source. With this condition fulfilled, the complex amplitude of the image in the direction $\vec{\Omega}$, $\tilde{E}_0\left(t - R'(\vec{\Omega}, \vec{r})/c\right)$, is independent of the pupil coordinate $\vec{r}$. Hence, the complex amplitude term in equation (29) can be placed in front of the integral.

- The far field approximation holds, i.e. the distance between position $\vec{r}$ in the pupil and the image position in the direction $\vec{\Omega}$ is linearily dependent on the pupil coordinate $\vec{r}$:
$$R'(\vec{\Omega}, \vec{r}) \;=\; R \;+\; \vec{\Omega} \cdot \vec{r} \tag{30}$$
with $\vec{\Omega} \cdot \vec{r} = \Omega_y y + \Omega_z z = \theta_y y + \theta_z z$ the scalar product between the unit image-direction vector $\vec{\Omega}$ and $\vec{r}$ (by definition 2-dimensional, since $\vec{r}$ has no x-coordinate).

Implementing these assumption, we can rewrite equation (29) as:

$$\begin{aligned} \tilde{E}(\vec{\Omega}, t) \;&=\; \left(\frac{C}{R} \int\!\!\int_{\text{pupil}} e^{\frac{-2\pi i \vec{\Omega}\cdot\vec{r}}{\lambda}} d\vec{r} \right) \left[\tilde{E}_0(t - \frac{R}{c})\; e^{2\pi i\bar{\nu}(t - \frac{R}{c})} \right] \\ &=\; \left(\frac{C}{R} \int\!\!\int_{\text{pupil plane}} P(\vec{r})\; e^{\frac{-2\pi i \vec{\Omega}\cdot\vec{r}}{\lambda}} d\vec{r} \right) \tilde{E}(t - \frac{R}{c}) \end{aligned} \tag{31}$$

where we have introduced the pupil function $P(\vec{r})$, with $P(\vec{r}) = 1$ inside the pupil and $P(\vec{r}) = 0$ everywhere else.

Note: The notion of the pupil function can actually be implemented in a more general fashion and makes it a versatile tool for describing the influence of an aperture on the incident radiation field, like transmission, reflection, absorption and/or phase shifts. For example, the pupil may act as a phase mask that introduces position dependent phase changes $\phi(\vec{r})$, these can be represented by putting $P(\vec{r}) \;=\; e^{i\phi(\vec{r})}$, rather than $P(\vec{r}) \;=\; 1$ inside the pupil boundary.

Equation (31) is now expressed in the form of a Fourier integral, however the integral implies a *scaled* Fourier transform of the pupil function $P(\vec{r})$ with conjugate variables $\vec{\Omega}$ and $\vec{r}/\lambda$.

Next, the proportionality constant C can be determined by using the fact that the total energy flux (= radiant flux) $\Phi(t)$ through the pupil needs to be conserved in the diffraction image, i.e.:

$$\int\int_{\text{image plane}} |\tilde{E}(\vec{\Omega},t)|^2 R^2 d\vec{\Omega} \;=\; \Phi(t) \;\;\Rightarrow\;\; \int\int_{\text{image plane}} |\tilde{E}(\vec{\Omega},t)|^2 d\vec{\Omega} \;=\; \frac{\Phi(t)}{R^2} \qquad (32)$$

Applying Parseval's theorem to the scaled Fourier transform given in equation (31) we also have:

$$\frac{1}{\lambda^2}\int\int_{\text{image plane}} |\tilde{E}(\vec{\Omega},t)|^2 d\vec{\Omega} \;=\; \int\int_{\text{pupil plane}} \frac{C^2}{R^2}|P(\vec{r})\;\tilde{E}(t-\frac{R}{c})|^2\frac{d\vec{r}}{\lambda^2} \;=\; C^2\frac{\Phi(t)}{R^2} \qquad (33)$$

Combining equations (32) and (33), we obtain $C = 1/\lambda$.

Substituting C in equation 31 and writing the Fourier integral as a "true" (i.e. *non-scaled*) Fourier transform in the conjugate variables $\vec{\Omega}$ and $\vec{\zeta} \equiv \vec{r}/\lambda$ (beware: $d\vec{\zeta} \equiv d\vec{r}/\lambda^2$), results in:

$$\tilde{E}(\vec{\Omega},t) \;=\; \left[\left(\frac{\lambda}{R}\right)\int\int_{\text{pupil plane}} P(\vec{\zeta}\,)\; e^{-2\pi i\vec{\Omega}\cdot\vec{\zeta}}\; d\vec{\zeta}\,\right]\tilde{E}(t-\frac{R}{c}) \;=\; \tilde{a}(\vec{\Omega})\tilde{E}(t-\frac{R}{c}) \qquad (34)$$

The 2-dimensional conjugate vector variable $\vec{\zeta}$ has the dimension $\textit{radian}^{-1}$ and represents a specific ***spatial frequency on the sky*** expressed in periods/radian. The **dimensionless function** $\tilde{a}(\vec{\Omega})$ is called the amplitude diffraction pattern:

$$\tilde{a}(\vec{\Omega}) \;=\; \left(\frac{\lambda}{R}\right)\int\int_{\text{pupil plane}} P(\vec{\zeta}\,)\; e^{-2\pi i\vec{\Omega}\cdot\vec{\zeta}}\; d\vec{\zeta} \quad \text{with} \;\; a(\vec{\Omega}\,[0,0]\,) = \left(\frac{\lambda}{R}\right)\left(\frac{\textit{pupil area}}{\lambda^2}\right) \qquad (35)$$

The **dimensionless quantity** $a(\vec{\Omega}\,[0,0]$ entails the product of the amplitude damping term $1/R$, *scaled to* λ, and the geometrical pupil area, *scaled to* λ^2.

Equation (35) shows a most important result regarding the diffraction phenomenon: in the Fraunhofer limit the amplitude diffraction pattern, with the proper normalisation factor at $\vec{\Omega}\,[0,0]$, can be obtained by taking the Fourier transform of the pupil function, i.e.:

$$\tilde{a}(\vec{\Omega}) \;\;\Leftrightarrow\;\; \left(\frac{\lambda}{R}\right)P(\vec{r}/\lambda) \qquad (36)$$

This Fourier pair shows that the diffraction image depends on the size of the pupil, expressed in the number of wavelengths λ.

The derivation of the amplitude diffraction image $\tilde{a}(\vec{\Omega})$ was done for a quasi-monochromatic point source at the origin of the angular sky coordinates along the direction of the x-axis. A similar point source at an arbitrary position $\vec{\Omega}_0$ on the sky yields a diffraction image $\tilde{a}(\vec{\Omega}-\vec{\Omega}_0)$, i.e. the same diffraction image but displaced to the geometrical image position $\vec{\Omega}_0$. This is an obvious result, and straightforward to derive.

Equation (35) expresses the distribution of the amplitude diffraction pattern. Taking the square of the absolute value of this expression yields the intensity diffraction pattern, defined as the telescope's Point Spread Function (PSF).

Intermezzo: Power flux density transported by an EM-wave

The energy streaming through space in the form of an electromagnetic wave is shared between the constituent electric and magnetic fields.
The energy density of an electrostatic field (e.g. between plates of a capacitor) $\rho_{\vec{E}} = \epsilon_r \epsilon_0 |\vec{E}|^2/2$ (dimension Joule/m^3), with $|\vec{E}|$ the magnitude of the electric vector (dimension V/m) and ϵ_0 the vacuum permittivity ($8.8543 \cdot 10^{-12}$ Asec/Vm). Similarly, the energy density of a magnetic field (e.g. within a toroid) equals $\rho_{\vec{B}} = |\vec{B}|^2/(2\mu_r\mu_0)$ (dimension Joule/m^3), with $|\vec{B}|$ the magnitude of the magnetic vector (dimension Tesla$\,=\,$Vsec/m^2) and μ_0 the vacuum permeability ($4\pi \cdot 10^{-7}$ Vsec/Am).
The wave equation for a **plane electromagnetic wave** traveling along the x-direction in vacuum is given by:

$$\frac{\partial^2 E(x,t)}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 E(x,t)}{\partial t^2} \quad \text{and} \quad \frac{\partial^2 B(x,t)}{\partial x^2} = \frac{1}{c^2}\frac{\partial^2 B(x,t)}{\partial t^2} \tag{37}$$

for the electric field wave and the magnetic field wave respectively. The magnetic field wave travels in a plane perpendicular to the electric field, both the electric field and the magnetic field directions are perpendicular to the direction of propagation (x). The plane wave solution can be expressed by a harmonic function, using a complex scalar representation:

$$\tilde{E}(x,t) = E_0 e^{i\cdot 2\pi(\nu t - x/\lambda)} \quad \text{and} \quad \tilde{B}(x,t) = B_0 e^{i\cdot 2\pi(\nu t - x/\lambda)} \tag{38}$$

Consistency with Maxwell's equations requires that for the EM-wave holds $\rho_{\vec{E}} = \rho_{\vec{B}}$. Hence, from the above, we have $B_0 = E_0/c$.
The flow of electromagnetic energy through space associated with the traveling EM-wave is represented by the Poynting vector $\vec{S} = (1/\mu_0)\vec{E}\,\mathbf{x}\,\vec{B}$, a vector product that symbolizes the direction and magnitude of the energy transport per unit time across a unit area (e.g. in units Watt m^{-2}). The vector magnitude $|\vec{S}| = |\tilde{E}||\tilde{B}|(\sin\phi)/\mu_0$ equals $|\tilde{E}||\tilde{B}|/\mu_0$, since the magnetic field is perpendicular to the electric field ($\phi = \pi/2$). Representing the *actual* wave signal by taking the real part of expressions (38) we get:

$$|\vec{S}| = E_0 B_0 \cos^2 2\pi(\nu t - x/\lambda) = \epsilon_0 c E_0^2 \cos^2 2\pi(\nu t - x/\lambda) = (\epsilon_0/\mu_0)^{\frac{1}{2}} E_0^2 \cos^2 2\pi(\nu t - x/\lambda) \tag{39}$$

The *average* power flux density for an *ideal monochromatic* plane wave, $\overline{I(t)}$ equals $\overline{|\vec{S}(t)|}$:

$$\overline{I(t)} = (\epsilon_0/\mu_0)^{\frac{1}{2}} E_0^2 \overline{\cos^2 2\pi(\nu t - x/\lambda)} = (\epsilon_0/\mu_0)^{\frac{1}{2}} \frac{E_0^2}{2} = 2.6544 \cdot 10^{-3} \frac{E_0^2}{2} \tag{40}$$

expressed in Watt/m^2 for E_0 in Volts/meter.
An idealised monochromatic plane wave is represented in the time domain by an infinitely long wave train and is by definition *fully polarised.* As has already been discussed, an *unpolarised, quasi-monochromatic*, radiation field from a thermal source can

be described by a complex expression for the electric field $\tilde{E}(t)$, comprising a harmonic oscillation at an average frequency $\bar{\nu}$ modulated by a slowly varying envelope, accomodated by the phasor $\tilde{E}_0(t)$, i.e. $\tilde{E}(t) = \tilde{E}_0(t) \cdot e^{i(2\pi\bar{\nu}t)}$. The average power flux density for this wave then follows from the expectation value of the product $\tilde{E}(t)\tilde{E}^*(t)$:

$$\overline{I(t)} = (\epsilon_0/\mu_0)^{\frac{1}{2}} \mathbf{E}\left\{\tilde{E}(t)\tilde{E}^*(t)\right\} = 2.6544 \cdot 10^{-3}\, \mathbf{E}\left\{|\tilde{E}_0(t)|^2\right\} \tag{41}$$

Since we are primarily concerned with *relative power flux densities generated by these traveling waves within the same medium*, we can disregard in what follows multiplication with the numerical constant in expression (41), since this (deterministic) quantity is only of relevance for assessing the *absolute numerical value* of the power flux density and bears no influence on the description of the stochastic nature of the signals. In practical computations, this constant should of course be applied!

End intermezzo: Power flux density transported by an EM-wave

The brightness distribution, $s_\lambda(\vec{\Omega}, t)$, of a cosmic source under observation is convolved with the PSF to obtain the diffraction-limited source image $d_\lambda(\vec{\Omega}, t)$. Similarly, the Fourier transform of the source brightness distribution, $S_\lambda(\vec{\zeta}, t)$, is multiplied with the Optical Transfer Function (OTF) of the telescope to obtain the spatial frequency spectrum of the diffraction-limited source image $D_\lambda(\vec{\zeta}, t)$. Hence we have:

$$\begin{aligned} d_\lambda(\vec{\Omega}, t) &= \int\!\!\int_{\text{source}} h_\lambda(\vec{\Omega} - \vec{\Omega}')\, s_\lambda(\vec{\Omega}', t)\, d\vec{\Omega}' \quad \text{and} \\ D_\lambda(\vec{\zeta}, t) &= H_\lambda(\vec{\zeta}) \cdot S_\lambda(\vec{\zeta}, t) \end{aligned} \tag{42}$$

In the above relations we have added an index λ to stipulate the wavelength dependence of the PSF and the OTF.

Multiplying (35) with its complex conjugate yields the mathematical expression for the PSF $h_\lambda(\vec{\Omega})$:

$$\text{PSF} = h_\lambda(\vec{\Omega}) = \tilde{a}(\vec{\Omega}) \cdot \tilde{a}^*(\vec{\Omega}) = \left|\, \tilde{a}(\vec{\Omega})\, \right|^2 = \left| \left(\frac{\lambda}{R}\right) \int\!\!\int_{\text{pupil plane}} P(\vec{\zeta})\, e^{-2\pi i \vec{\Omega}\cdot\vec{\zeta}}\, d\vec{\zeta} \right|^2 \tag{43}$$

To obtain an expression for the OTF $H_\lambda(\vec{\zeta})$, taking the general case of a complex pupil function, we can use the following relations (verify this yourself!):

$$\tilde{a}(\vec{\Omega}) \Leftrightarrow \left(\frac{\lambda}{R}\right) P(\vec{\zeta}) \quad \text{and} \quad \tilde{a}^*(\vec{\Omega}) \Leftrightarrow \left(\frac{\lambda}{R}\right) P^*(-\vec{\zeta}) \tag{44}$$

Hence, by applying the convolution theorem, we get:

$$\tilde{a}(\vec{\Omega}) \cdot \tilde{a}^*(\vec{\Omega}) \Leftrightarrow \left(\frac{\lambda}{R}\right)^2 \left[P(\vec{\zeta}) \ast P^*(-\vec{\zeta})\right] \tag{45}$$

In the common case of a *centrally symmetric pupil*, the autoconvolution of equation (45) is just the autocorrelation, and so we can write:

$$\text{OTF} = H_\lambda(\vec{\zeta}) = \frac{1}{R^2} \int\!\!\int_{\text{pupil plane}} P(\vec{r})^*\, P(\vec{r} - \lambda\vec{\zeta}) d\vec{r} \tag{46}$$

with the vector variable of the integral rescaled to the pupil coordinate $\vec{r}$. (The complex conjugate of the pupil function in equation (46) has of course no meaning in case the pupil function is real)

Equation (46) characterizes the response function of a low pass filter for spatial frequencies: the transmission continually decreases with increasing spatial frequency, at a certain value $\lambda\vec{\zeta}_{max}$, the mutually displaced pupil functions in (46) will no longer overlap and higher spatial frequencies will not be transmitted.

Important notion: *a finite pupil acts as a low pass filter for the spatial frequencies in the brightness distribution of the celestial object observed. Given a fixed size of the pupil: the longer the wavelength the lower the cut-off frequency ($\lambda|\vec{\zeta}\,| = constant$).*

Moreover it is worth noting that, although non-circular pupils are rarely encountered in astronomy, expressions (43) and (46) are applicable for *arbitrary* pupil geometries.

3.1.4 Circular pupils

Circular pupils play a central role in astronomy, so this warrants a special description. Given the circular symmetry, the pupil function can be expressed as a function of a scalar variable ρ with the aid of the 2-dimensional *circular* box function $\Pi(\rho)$, where $\Pi(\rho) \equiv 1$ for $|\rho| < 1/2$ and $\Pi(\rho) \equiv 0$ for $|\rho| \geq 1/2$. The Fourier transform of this function involves the first order Bessel function $J_1(x)$:

$$\Pi(\rho) \;\;\Leftrightarrow\;\; \frac{1}{2}\left[\frac{J_1(\pi\theta)}{\theta}\right] \tag{47}$$

in which the scalar variable θ, replacing the angular direction vector $\vec{\Omega}$, represents the circular symmetric diffraction angle. Taking a telescope diameter D and applying the scaling law for Fourier transforms, we have:

$$\Pi\left(\frac{\rho}{D}\right) \;\;\Leftrightarrow\;\; \frac{1}{2}\left[\frac{DJ_1(\pi\theta D)}{\theta}\right] \tag{48}$$

Substituting for ρ the spatial frequency variable (now also a scalar) $p = \rho/\lambda$, the amplitude of the diffracted field follows from equation (36):

$$a(\theta) \;\;\Leftrightarrow\;\; \left(\frac{\lambda}{R}\right)\Pi\left(\frac{\lambda p}{D}\right) \;\;\Rightarrow\;\; \left(\frac{\lambda}{R}\right)\left[\frac{1}{4}\pi(D/\lambda)^2\right]\left[\frac{2J_1(\pi\theta D/\lambda)}{\pi\theta D/\lambda}\right] \tag{49}$$

Hence, introducing the reduced variable $u = \pi\theta D/\lambda$, we arrive at the expression for the *diffraction-limited* PSF for a circular telescope:

$$PSF \;=\; |a(\theta)|^2 \;=\; \left(\frac{\lambda}{R}\right)^2\left(\frac{1}{4}\pi(D/\lambda)^2\right)^2\left[\frac{2J_1(u)}{u}\right]^2 \tag{50}$$

The first term in the expression at the right-hand side is the normalisation of the PSF for undiffracted light, i.e. for $\theta = 0$. It involves the **dimensionless** quantity that accomodates the attenuation factor (*scaled to* λ) of the radiation energy due to the spherical expansion of the wavefield in the Fraunhofer limit and the geometrical area of the circular aperture, *scaled to* λ^2.

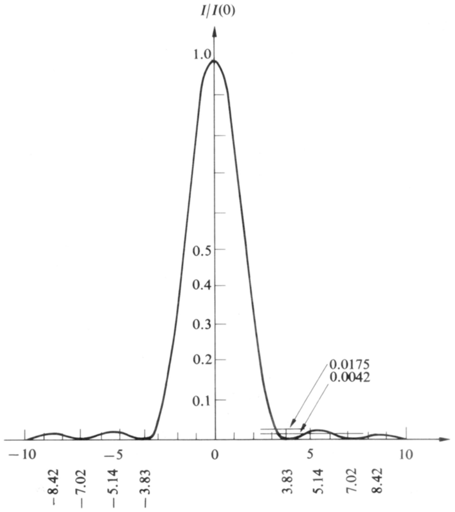

Figure 6: *The Airy brightness function normalised to unit intensity at* $\theta = 0$. *Figure taken from Hecht 1987.*

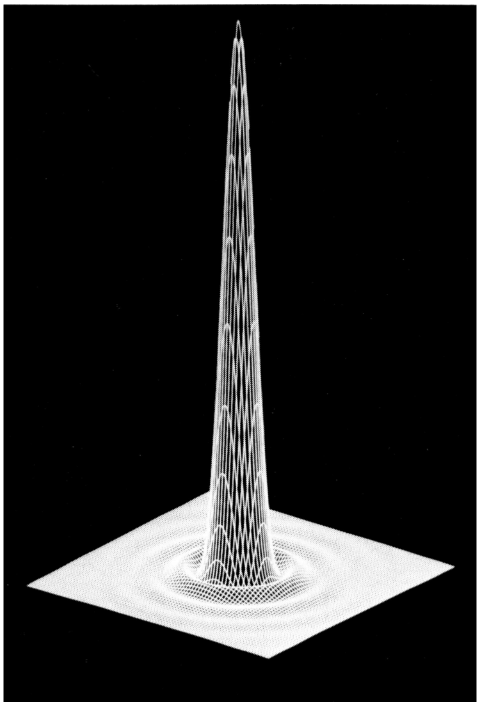

Figure 7: *The 2-dimensional Airy brightness function for the diffracted intensity. Figure taken from Hecht 1987.*

The second term is often called the *Airy brightness function* and has a ring-like structure. We can designate this term as $h_\lambda(\theta)$, the point source response function normalised to unit intensity at $\theta = 0$, i.e.:

$$h_\lambda(\theta) \; = \; \left[\frac{2J_1(u)}{u}\right]^2 \; \equiv \; \left[\frac{2J_1(\pi\theta D/\lambda)}{\pi\theta D/\lambda}\right]^2 \tag{51}$$

Figure 6 and figure 7 show the Airy diffraction patterns in 1- and 2-dimensional display. The FWHM of the central peak is roughly equal to λ/D *radians*, around this main peak ring shaped secondary maxima are present that decrease monotonously in strength. The OTF can be derived from the autocorrelation of $\Pi(\lambda p/D)$:

$$\begin{aligned} OTF &= \left(\frac{\lambda}{R}\right)^2 \left[\Pi\left(\frac{\lambda p}{D}\right) * \Pi\left(\frac{\lambda p}{D}\right)\right] \\ &= \frac{1}{2}\left(\frac{D}{R}\right)^2 \left[\arccos\left(\frac{\lambda p}{D}\right) - \left(\frac{\lambda p}{D}\right)\left(1 - \left(\frac{\lambda p}{D}\right)^2\right)^{\frac{1}{2}}\right] \end{aligned} \tag{52}$$

for $0 \leq p \leq D/\lambda$.

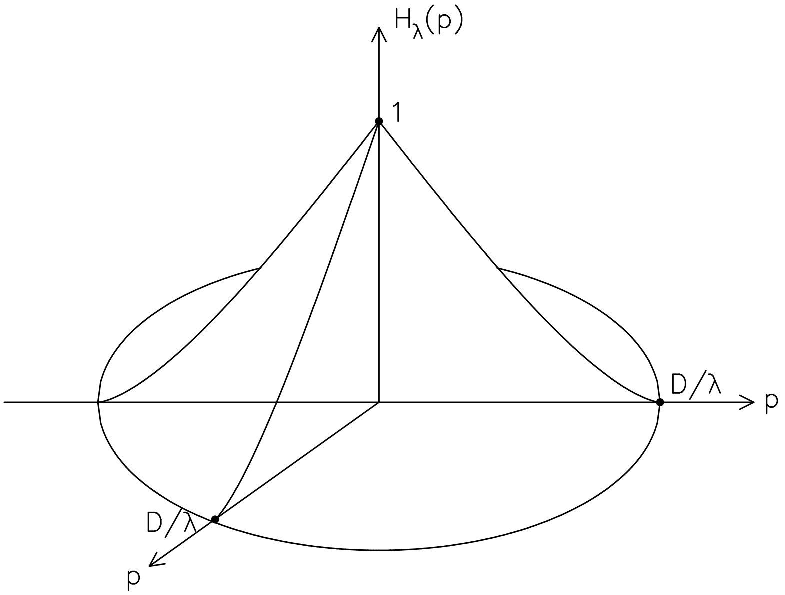

Figure 8: *The Optical Tranfer Function (OTF) for a circular aperture with diameter D. The 3d-shape is sometimes referred to as the Chinese Hat function.*

Normalization to unity response for zero frequency ($p = 0$), by *dividing* with the telescope area and *disregarding* the $1/R^2$ attenuation factor resulting from spherical expansion of the diffracted field in the Fraunhofer limit, yields the spatial frequency response function $H_\lambda(p)$ of the telescope:

$$H_\lambda(p) = \frac{2}{\pi}\left[\arccos\left(\frac{\lambda p}{D}\right) - \left(\frac{\lambda p}{D}\right)\left(1 - \left(\frac{\lambda p}{D}\right)^2\right)^{\frac{1}{2}}\right] \tag{53}$$

Figure 8 shows the OTF resulting from self-convolution of the circular pupil function: the shape is referred to as the "Chinese hat", the transmission of spatial frequencies decreases almost linearly up to the cut-off frequency D/λ, beyond this frequency the transmission of the telescope equals zero.

3.1.5 Complex pupils

In expression (31) the pupil function was introduced in its simplest form, i.e. $P(\vec{r}) = 1$ inside the pupil and $P(\vec{r}) = 0$ everywhere else. We already noted at that point that the concept of the pupil function is much more general, we shall elaborate shortly on this here. $P(\vec{r}) = 1$ implies that all points within the pupil emit the same field amplitude $\tilde{E}(t) = \tilde{E}_0(t) \cdot e^{i2\pi\bar{\nu}t}$. If we now more generally introduce within the pupil a complex pupil function $\tilde{P}(\vec{r})$, the same treatment of the Fraumhofer diffraction holds as before. However, each pupil point now emits its own specific field amplitude and phase determined by the complex pupil function:

$$\tilde{E}(\vec{r}, t) \; = \; \tilde{P}(\vec{r})\tilde{E}_0(t) \cdot e^{i2\pi\bar{\nu}t} \tag{54}$$

The derivation of the amplitude diffraction pattern remains unchanged (verify this youself) and we have (see also equation(35)):

$$\tilde{a}(\vec{\Omega}) \; = \; \left(\frac{\lambda}{R}\right) \int\int_{\text{pupil plane}} \tilde{P}(\vec{\zeta}\,)\; e^{-2\pi i \vec{\Omega}\cdot\vec{\zeta}}\; d\vec{\zeta} \tag{55}$$

Complex pupil functions can be generated in several different ways, we give here two examples.

- *An optical mask inserted in the aperture.*
  Consider a plane wave incident on the pupil plane. Cover the aperture with a foil with a position dependent transparency (determines the amplitude of $\tilde{P}(\vec{r})$) and a position dependent thickness or refractive index, which determines the phase of $\tilde{P}(\vec{r})$.

- *Microwave phased-array antenna's.*
  Antenna's for radar waves and microwave communication often comprise a flat plane, filled with a large number of radiators, like waveguide exits, dipoles etcetera. They radiate according to expression (54). The radiation field emitted from the pupil $\tilde{P}(\vec{r})$ can be regulated electronically, in particular its phase. The radiation beam generated by the collection of individual radiators is defined by the distribution of radiation power over the sky direction vector $\vec{\Omega}$. If we have $\tilde{P}(\vec{r}) = 1$ this distribution is given by the Airy function centered around $\vec{\Omega} = 0$, i.e. perpendicular to the radiator plane. If we select a pupil function with a linear phase dependence, $\vec{\Omega} = e^{-2\pi i\vec{\Omega}_0 \cdot \vec{r}/\lambda}$, the radiation beam rotates towards sky direction $\vec{\Omega}_0$, commensurate with the Fourier *shift theorem.* In this way the direction of the radiation beam can be very rapidly controlled electronically in two angular dimensions without any mechanical rotation devices!

3.1.6 Rayleigh Resolution Criterion

The image of two, equally bright, point sources with an angular separation θ is the incoherent superposition of two identical Airy functions. The limiting angle at which the two sources can be separated has been fixed at the value where the maximum of the one Airy function coincides with the first zero of the other Airy function. The first zero of $J_1(u)$ occurs at a value of the reduced variable $u = u_0 = 3.832$, this relates to an angle $\theta = \theta_0 = 1.22\lambda/D$. This angular value is often used to specify the *angular resolving power* or discriminating power of the telescope pupil and is commonly referred to as the *Rayleigh criterion*. This criterion is only approximate and much less quantitative than the PSF $h_\lambda(\theta)$ for a circular pupil. Moreover, in certain cases it is possible to resolve two point sources closer than θ_0, for example in the case of the two components of a double star, if the measurement of the image is made with excellent signal to noise ratio (larger than a few hundred). In such a case the diffracted image profile $I(\theta)$ can significantly differ from the profile of a single point source $h_\lambda(\theta)$ even if the angular separation of the two components is less than θ_0. On the other hand, if the signal to noise ratio is poor, or if the two sources have greatly different brightness, the value of θ_0 might be largely insufficient to reliably extract separate source contributions. The latter case is certainly applicable to crowded source fields with high brightness contrasts between the observed field sources. In that case the value of θ_0 can be severely compromised and the actual resolving power will be significantly reduced in a case specific fashion!

3.2 Other limits to image quality

3.2.1 Optical aberrations

A common and severe aberration of both lenses and mirrors is *spherical aberration*. This aberration arises from the fact that lens or mirror annuli with different radii have different focal lengths. In the case of a spherical mirror this aberration can be completely eliminated for rays parallel to the optical axis by deepening the spherical mirror surface to a paraboloidal surface.

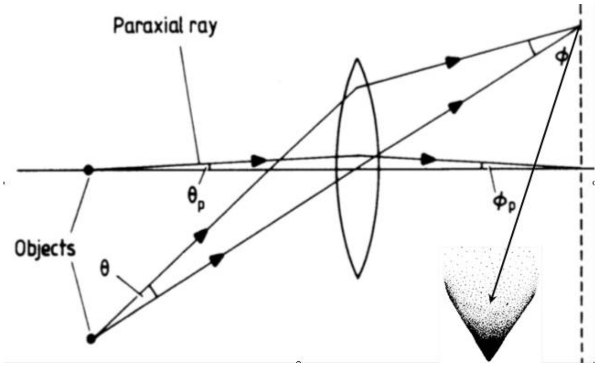

Figure 9: *Coma error for an off-axis object. The characteristic "pear-shaped" ray distribution near the off-axis geometrical image point is indicated in the lower right-hand corner. Figure taken from Kitchin 1998.*

However a paraboloidal mirror suffers from another aberration called *coma*. Coma causes the images for objects that are not located on the optical axis to consist of a series of circles which correspond to the various annular zones of the mirror surface, these circles are progressivily shifted towards or away from the optical axis giving rise to a characteristic pear-shaped image blur (see figure 9).

The figure also shows the condition that needs to be obeyed to reduce coma to zero in an optical system, the *Abbe sine condition*:

$$\frac{\sin\theta}{\sin\phi} = \frac{\theta_{parax}}{\phi_{parax}} = \text{constant} \tag{56}$$

where the angles are defined in figure 9.
The severity of the coma aberration at a given angular distance from the optical axis is inversely proportional to the square of the focal ratio of the telescope. Therefore, the effect of comatic aberration can be significantly reduced by employing as large a focal ratio as possible. Examples of optical designs with large focal ratios comprise the Cassegrain and Ritchey-Chretien systems, we shall dwell on these shortly in the following paragraph since they constitute the most common format for large telescopes in astronomy.
The configuration of the Cassegrain system is sketched in figure 10. It is based on a paraboloidal primary mirror and a convex hyperboloidal secondary mirror, the near focus of the secondary hyperboloid coincides with the the focus of the primary paraboloid. The Cassegrain focus is the distant focus of the secondary mirror. The major advantage of the Cassegrain system is its *telelens* characteristic: the secondary hyperboloid

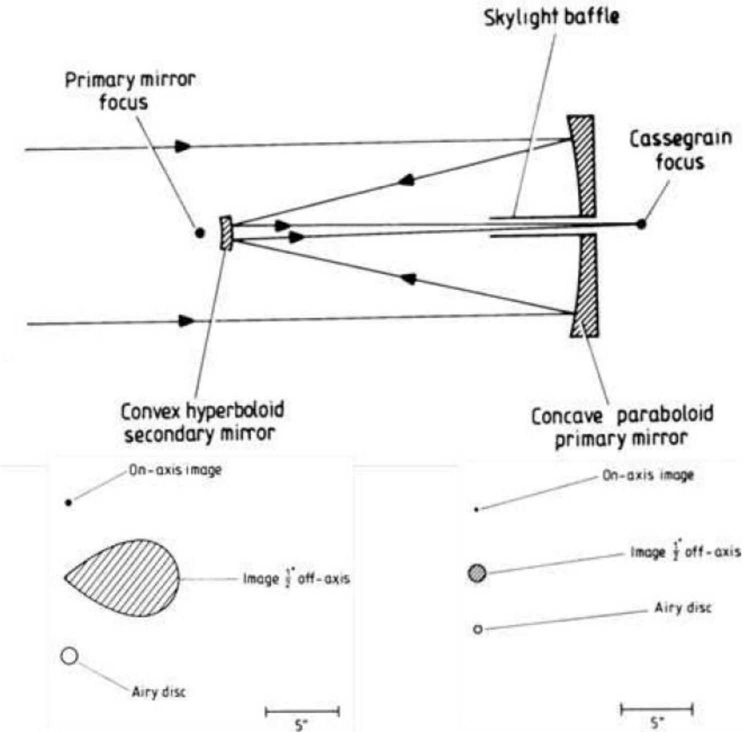

Figure 10: *Upper panel: Cassegrain configuration with parabolic primary and hyperbolic secondary mirror. Lower left: comparison between the on-axis Airy-disk and the coma-dominated image at an off-axis angle of 0.5 degrees. Lower right: improved off-axis performance for a two-hyperboloid Ritchey-Chretien configuration. Figure taken from Kitchin 1998.*

expands the beam from the primary mirror so that the focal length of the Cassegrain system becomes several times that of the primary mirror. The coma aberration is consequently substantially reduced to that of a single parabolic mirror with a focal length equal to the effective focal length of the Cassegrain. The beam expanding effect of the secondary mirror makes that Cassegrain telescopes normally work with focal ratios between $f12$ to $f30$, although the primary mirror is only $f3$ or $f4$! Figure 10 shows the images for a 25 cm primary in a $f4/f16$ configuration. Displayed are the theoretical on-axis geometrical image point, the on-axis Airy disk arising from diffraction by the 25 cm primary mirror, and the coma-dominated pear-shaped image blur at an off-axis angle of 0.5 degrees (the angular scale of 5 arcseconds is given for reference to the actual angular sizes of these images).

A large improvement in image quality can be obtained if the Cassegrain is modified to a Ritchey-Chretien design. The optical design is the same as for the Cassegrain configuration with the exception that the primary mirror is hyperboloidal rather than paraboloidal and a stronger hyperboloid is used for the secondary. With this design both spherical aberration and coma can be corrected, resulting in an aplanatic system. The improvement in the image quality is also displayed in figure 10 for a 50 cm Ritchey-Chretien telescope with the same effective focal length as the 25 cm Cassegrain. In fact, the improvement relative to the Cassegrain system is larger than the comparison of the images displayed in figure 10 suggests, since a *50 cm Cassegrain* would have its off-axis image twice the size (i.e. four times the area) and the on-axis Airy disk half the size (i.e. a quarter times the area) shown in the figure. The optics employed in the Hubble Space Telescope (2.4 m primary mirror) is a Ritchey-Chretien design.

A Cassegrain system can however be improved considerably by the addition of *corrective optics* just before the focus. This corrective optics comprises lens assemblies whose aberrations oppose those of the main Cassegrain system, they involve aspheric surfaces and/or the use of exotic materials like fused quartz. Hence, images can be reduced to less than the size of the *seeing disk* (see next paragraph) up to fields of view of the order of one degree.

3.2.2 Atmospheric degradation: speckle images

The intrinsic quality of modern optical telescopes in terms of their imaging performance can be made very close to the diffraction limit, certainly when taking into account recently developed sophisticated control systems for optimizing the *figure* of the mirror surface by employing active corrections for bending under gravity of the mirror mass and temperature gradients over the area, the so called *active optics*. However, in ground-based telescope systems the image quality is limited by turbulent motions in the atmosphere, giving rise to a randomly varying value of the refractive index $n(\vec{r}, t)$ in the air columns over the pupil area. Horizontal scales for these fluctuations range from several *centimeters* to several *meters*. If an undistorted flat wave front enters the atmosphere, three main effects caused by atmospheric turbulence can be identified in the pupil plane of the telescope:

- Variations in amplitude of the wavefront (i.e. lighter and darker brightness patches to the "eye") corresponding to concentration or spreading of the wavefront energy (*scintillation*).

- Variation in the angle of the mean tangent plane to the wavefront causing *angular motion* of the image. Characteristic wavefront slopes are of the order of a few μm's/meter.
- Reduction of spatial coherence of the wavefront across the pupil plane due to random fluctuation of the phase, that leads to *smearing* of the image, resulting in image sizes much larger than from diffraction alone.

These effects can be described by introducing a new, *instantaneous*, pupil function which is random and complex:

$$\tilde{P}(\vec{r},t) \;=\; P(\vec{r})\psi(\vec{r},t) \tag{57}$$

in which $P(\vec{r})$ represents the simple *geometrical* pupil function introduced in equation (31) and $\psi(\vec{r},t)$ represents a wavefront that randomly varies in amplitude and phase. During a very short exposure, of the order of a few milliseconds, one may consider the wavefront frozen in its instantaneous shape. If we neglect for the moment the amplitude fluctuations (i.e. scintillation), we can write:

$$\psi(\vec{r},t) \;=\; e^{-i\phi(\vec{r},t)} \tag{58}$$

in which the phase of the wave, $\phi(\vec{r},t)$, is a random variable whose spatial statistical distribution is determined by the properties of the randomly varying value of the refractive index in the turbulent cells of the overlaying atmosphere. If we consider an atmospheric layer of thickness Δh that is large compared to the scale size of the turbulent cells, so that Gaussian statistics apply (i.e. the central limit theorem), the phase shift produced by the refractive index fluctuations is:

$$\phi(\vec{r},t) \;=\; k \int\limits_{h}^{h+\Delta h} n(\vec{r},t,z)dz \tag{59}$$

with $n(\vec{r},t,z)$ the refractive index random variable, z the vertical coordinate and $k = 2\pi/\lambda$ the wave number. This randomly varying phase shift describes the effects of angular motion and smearing of the image, the amplitude variations may often be neglected if the turbulence is not very severe.
Hence we have now an instantaneous, random, pupil function:

$$\tilde{P}(\vec{r},t) \;=\; P(\vec{r})e^{-i\phi(\vec{r},t)} \tag{60}$$

which can be used to derive the point source response of the telescope due to atmospheric turbulence. The *instantaneous image* of a point source is obtained from the Fourier transform of the *autocorrelation of the instantaneous pupil function*. Consequently, the image will be an intensity distribution in the focal plane of the telescope, that randomly changes shape with a frequency of 200 to 300 times per second, determined by the coherence time τ_c of the atmosphere. For exposures shorter than this

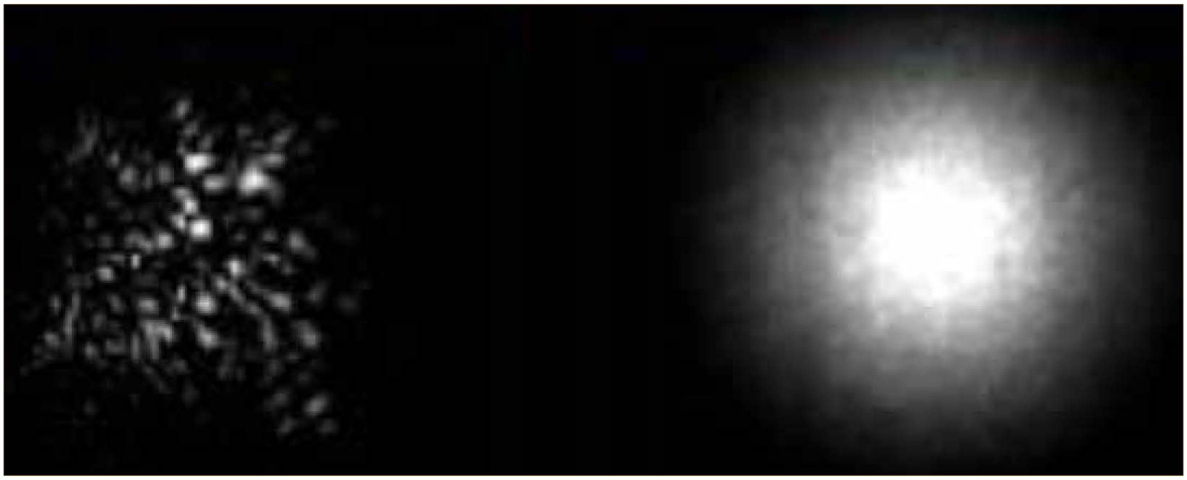

Figure 11: *Left panel: short exposure speckle image. Right panel: long exposure showing the smoothed "seeing" disk of the observed point source.*

coherence time, the wavefront is considered frozen in its momentary shape, the associated momentary intensity distribution comprises an image with a "speckle" structure as shown in figure 11. Every speckle has a diffraction limited PSF of λ/D, e.g. 0.1 arcseconds for a telescope diameter of 2 meters at $\lambda = 1\mu$m. In good approximation one may consider each speckle to be the image resulting from rays leaving perpendicular "equal-slope" areas on the wavefront in the direction of that specific speckle (see sketch of ray paths in figure 12). The speckle pattern randomly fluctuates on the same time scale as the wavefront due to the random phase fluctuation across the pupil plane, moreover the speckle pattern as a whole jitters as well following the variations of the

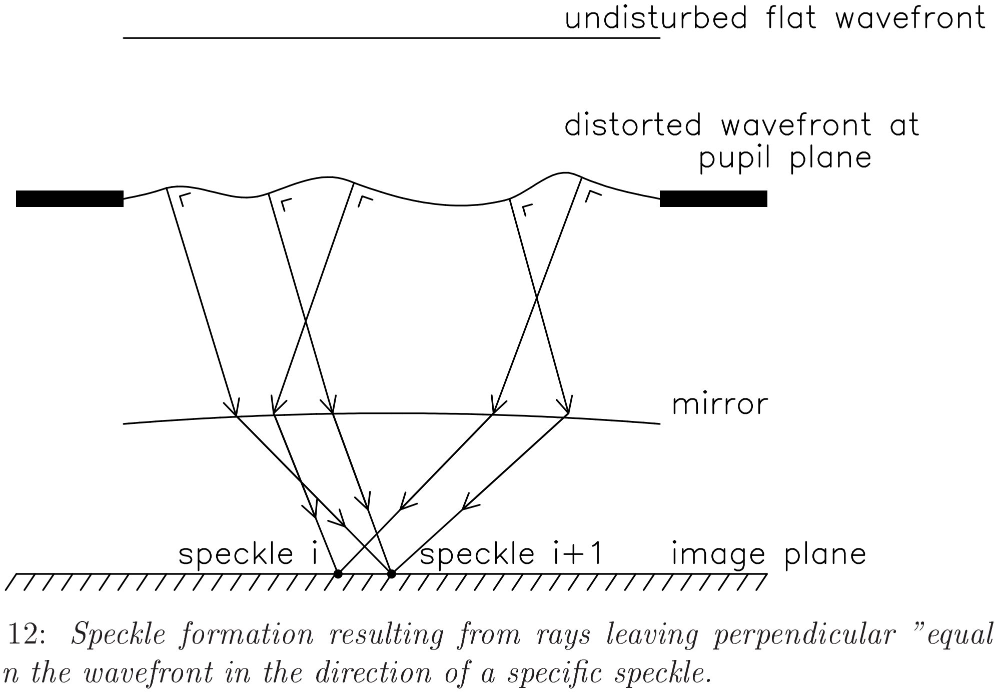
undisturbed flat wavefront
distorted wavefront at pupil plane
mirror
speckle i
speckle i+1
image plane

Figure 12: *Speckle formation resulting from rays leaving perpendicular "equal-slope" areas on the wavefront in the direction of a specific speckle.*

mean tangent plane to the wavefront (the angular motion).
The *technique of speckle imaging* requires taking many short exposures, which "freeze-out" the effects of the atmosphere. All short exposures are Fourier transformed and averaged in Fourier space to preserve the high-resolution (i.e. diffraction limited) information present in the individual speckle images. The averaged spatial frequency spectrum is then subjected to an inverse Fourier transform, resulting in the final image. In the picture shown in figure 13, the left frame shows a 100 millisecond exposure on the bright T-Tauri star V807 Tau, the speckle structure caused by the atmosphere is clearly evident. The central frame displays a 40 seconds exposure on the same source, the speckle structure has averaged out to a smooth fuzzy source image. The right panel shows the result of Fourier-processing of the series of speckle images, demonstrating that this star is actually a binary!

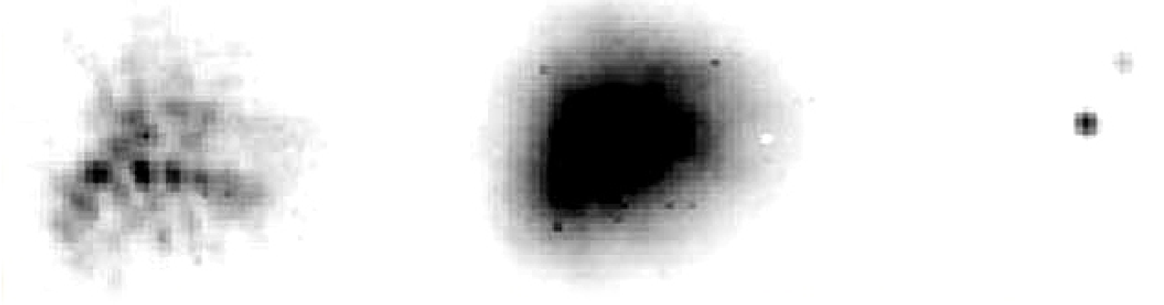

Figure 13: *Left frame: a 100 millisecond exposure on the bright T Tauri star V807 Tau, the speckle structure is clearly seen. Middle frame: a 40 second exposure on the same source, the speckles have averaged to a smooth fuzzy source. Right panel: Speckle-processed image revealing the binary nature of the source.*

3.2.3 Seeing: the Fried parameter

Figure 14 shows a comparison of a sky image taken by a ground based telescope, suffering from atmospheric seeing, with the image of the same sky field obtained from outer space with the Hubble Space Telescope, in which the atmospheric influence is eliminated and the quality of the image is solely determined by the diffraction limit of the HST primary mirror. As is evident from figures 11, 13 and 14, long exposures combine speckle images into a severely broadened image, this broadening is called *seeing*. The resulting smoothed PSF, the so-called *seeing disk*, is slightly broader than a single speckle pattern and amounts to $\approx$ one arcsecond. This seeing-value of one arcsecond corresponds to the diffraction limit of a 10 cm diameter telescope at $\lambda = 0.5\,\mu$m. The merit of using larger optical telescopes is, therefore, not primarily sharper images but sensitivity owing to their much larger light collecting power!
The observed intensity at each point of a long-exposure image is simply the *time-averaged* instantaneous intensity $\bar{I}(\vec{\Omega}) = \mathbf{E}\left\{I(\vec{\Omega}\,,\,t)\right\}$:

$$\bar{I}(\vec{\Omega}) \;=\; \mathbf{E}\left\{s(\vec{\Omega}) \,*\, h(\vec{\Omega},t)\right\} \;=\; s(\vec{\Omega}) \,*\, \mathbf{E}\left\{h(\vec{\Omega},t)\right\} \tag{61}$$

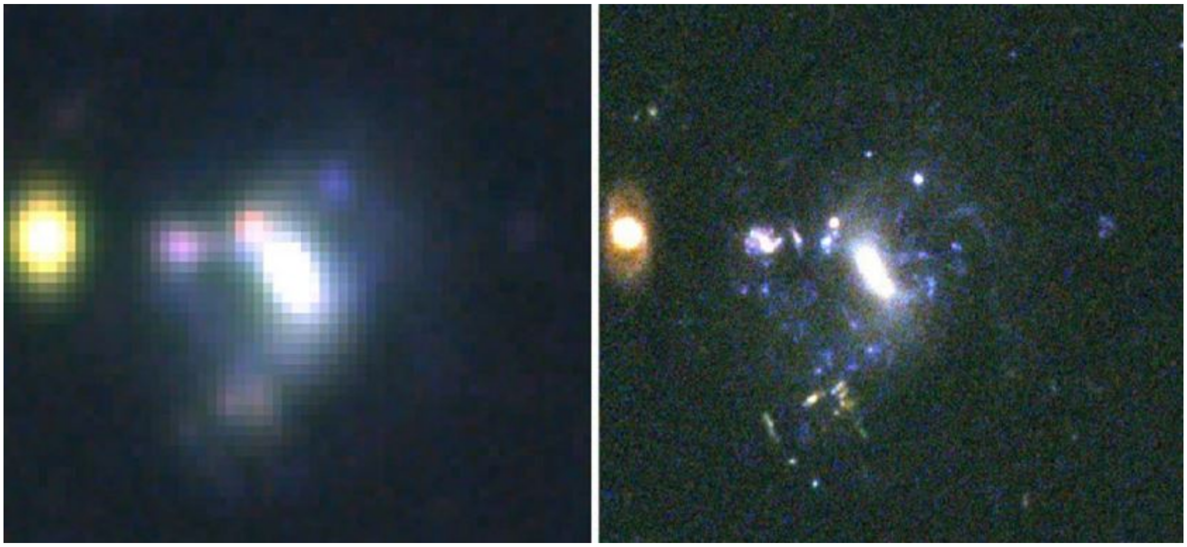

Figure 14: *Influence of atmospheric seeing on a sky image. Left: image obtained by a ground-based telescope, resolution 1.1 arcseconds dominated by atmospheric turbulence. Right: the same sky field imaged by the 2.4 meter Hubble Space Telescope, resolution diffraction limited to 0.05 arcseconds. Credit Space Telescope Science Institute.*

with $s(\vec{\Omega})$ the intensity of a constant quasi-monochromatic source and $h(\vec{\Omega}, t)$ the randomly variable PSF.
The spatial frequency spectrum of the seeing disk in long exposure images follows from the computation of the normalised mean OTF $\bar{H}(\vec{\zeta}\,) = \mathbf{E}\left\{H(\vec{\zeta}, t)\right\}$ which, according to equation (46), equals the normalised mean autocorrelation of the pupil function in the case of centrally symmetric pupils:

$$\begin{aligned}\bar{H}(\vec{\zeta}\,) &= \mathbf{E}\left\{\frac{1}{\int\int_{\text{pupil plane}} P^2(\vec{\zeta}\,')d\vec{\zeta}\,'}\int\int_{\text{pupil plane}} P(\vec{\zeta}\,')e^{-i\phi(\vec{\zeta}\,',t)}P(\vec{\zeta}\,'-\vec{\zeta}\,)e^{i\phi(\vec{\zeta}\,'-\vec{\zeta},t)}d\vec{\zeta}\,'\right\} \\ &= \frac{1}{\int\int_{\text{pupil plane}} P^2(\vec{\zeta}\,')d\vec{\zeta}\,'}\int\int_{\text{pupil plane}} P(\vec{\zeta}\,')P(\vec{\zeta}\,'-\vec{\zeta}\,)\mathbf{E}\left\{e^{-i[\phi(\vec{\zeta}\,',t)\,-\,\phi(\vec{\zeta}\,'-\vec{\zeta},t)]}\right\}d\vec{\zeta}\,'\end{aligned} \tag{62}$$

with $\vec{\zeta}\,' \equiv \vec{r}\,'/\lambda$, $\vec{\zeta} \equiv \vec{r}/\lambda$ and $d\vec{\zeta}\,' \equiv d\vec{r}\,'/\lambda^2$ the *scaled* pupil position vectors and differential area, normalised to the wavelength λ. The phase fluctuations $\phi(\vec{\zeta}, t)$ over the wavefront in the pupil plane can be regarded as resulting from a large number of perturbations in the overlaying atmosphere, which are mutually phase independent. According to the central limit theorem the distribution of the random variable $\phi(\vec{\zeta}, t)$ will be *Gaussian with zero mean* over both time and space. If x is a real Gaussian random variable, it is straightforward to show (verify this yourself!) that:

$$\mathbf{E}\left\{\exp(ix)\right\} = \exp\left[-\frac{1}{2}\,\mathbf{E}\left\{x^2\right\}\right] \tag{63}$$

and consequently we have for the expectation value in the integrand of equation (62):

$$\mathbf{E}\left\{\exp\left[-i[\phi(\vec{\zeta}\,',t)\,-\,\phi(\vec{\zeta}\,'-\vec{\zeta},t)]\right]\right\} = \exp\left[-\frac{1}{2}\,\mathbf{E}\left\{\left[\phi(\vec{\zeta}\,',t)\,-\,\phi(\vec{\zeta}\,'\,-\,\vec{\zeta},t)\right]^2\right\}\right]$$

$$= \exp\left[-\frac{1}{2}\,Q_\phi(\vec{\zeta}\,)\right] \tag{64}$$

where we have introduced a *structure function for the phase*:

$$\begin{aligned} Q_\phi(\vec{\zeta}\,) &= \mathbf{E}\left\{\left[\phi(\vec{\zeta}\,',t) - \phi(\vec{\zeta}\,' - \vec{\zeta}\,,t)\right]^2\right\} \\ &= 2\left[\mathbf{E}\left\{\phi^2(\vec{\zeta}\,',t)\right\} - \mathbf{E}\left\{\phi(\vec{\zeta}\,',t)\phi(\vec{\zeta}\,' - \vec{\zeta}\,,t)\right\}\right] \end{aligned} \tag{65}$$

This structure function for the phase distribution over the wavefront can be derived from the structure function of the refractive index for homogeneous and isotropic turbulence. We shall forego this calculation here and refer to the treatment by F. Roddier: The Effects of Atmospheric Turbulence in Optical Astronomy (Progress in Optics XIX, 281, 1981). The phase correlation on the perturbed wavefront expressed by equation (65) can be characterised by a *wavelength dependent correlation length* $r_c = \lambda\zeta_c$. A detailed computation involving the afore mentioned structure function of the refractive index yields:

$$Q_\phi(\vec{r}) = 2\left(\frac{|\vec{r}|}{r_c}\right)^{5/3} \quad\Longrightarrow\quad Q_\phi(\vec{\zeta}\,) = 2\left(\frac{\lambda|\vec{\zeta}|}{r_c}\right)^{5/3} \tag{66}$$

with:

$$r_c = \left[1.45\,k^2\int\limits_0^\infty C_n^2(z)\,dz\right]^{-3/5} \tag{67}$$

in which the wave number $k = 2\pi/\lambda$ and where $C_n^2(z)$ represents the *refractive index structure constant*, that strongly depends on the altitude z, typical values range from $10^{-14}\,\mathrm{m}^{-2/3}$ near the ground decreasing to $10^{-17}\,\mathrm{m}^{-2/3}$ at an altitude of 10 km.
The common case of a *circular pupil* with a diameter $D \gg r_c$ yields for the mean value of the normalised OTF:

$$\bar{H}(p) = \exp - \left(\frac{\lambda p}{r_c}\right)^{5/3} \tag{68}$$

where we have again used the *scalar* spatial frequency variable p following the limitation to circular symmetry. The point source response is the Fourier Transform of $\bar{H}(p)$.
It is customary to express the degraded resolution in terms of the diameter D_F of a diffraction limited circular pupil which would give an image of the same angular extent as the seeing disk. D_F is the so-called *Fried parameter*, which can be computed from:

$$\int\limits_0^\infty \bar{H}(p)p\,dp = \int\limits_0^{D_F/\lambda} \frac{2}{\pi}\left[\arccos\left(\frac{\lambda p}{D_F}\right) - \left(\frac{\lambda p}{D_F}\right)\left(1 - \left(\frac{\lambda p}{D_F}\right)^2\right)^{\frac{1}{2}}\right] p\,dp \tag{69}$$

For the lefthand integral in equation (69):

$$\int\limits_0^\infty p\left[\exp - \left(\frac{\lambda p}{r_c}\right)^{5/3}\right] dp \tag{70}$$

we can introduce the variable $x = [(\lambda p)/r_c]^{5/3}$ and rewrite this as:

$$\frac{3}{5}\,\frac{r_c^2}{\lambda^2}\int_0^\infty x^{1/5}\,e^{-x}\,dx \;=\; \frac{3}{5}\,\frac{r_c^2}{\lambda^2}\,\Gamma\left(\frac{6}{5}\right) \;=\; 0.55\,\frac{r_c^2}{\lambda^2} \tag{71}$$

Introducing a change of variable $x = (\lambda p)/D_F$, we can rewrite the righthand integral of equation (69) as:

$$\left(\frac{2}{\pi}\,\frac{D_F^2}{\lambda^2}\right)\left[\int_0^1 \left(x \arccos x \;-\; x^2\sqrt{(1\;-\;x^2)}\right)dx\right] \;=\; \left(\frac{2}{\pi}\,\frac{D_F^2}{\lambda^2}\right)\left(\frac{\pi}{16}\right) \;=\; \frac{D_F^2}{8\,\lambda^2} \tag{72}$$

Hence, the Fried parameter relates to the phase correlation length r_c on the perturbed wavefront as:

$$D_F^2 \;=\; 4.4\,r_c^2 \qquad \Longrightarrow \qquad D_F \;=\; 2.1\,r_c \tag{73}$$

and, implementing expression (67), we find:

$$D_F \;=\; \left[0.423\,k^2\int_0^\infty C_n^2(z)\,dz\right]^{-3/5} \;=\; 0.185\,\lambda^{6/5}\left[\int_0^\infty C_n^2(z)\,dz\right]^{-3/5} \tag{74}$$

Subsequently, expressing the phase structure function $Q_\phi(\vec{r})$ and the mean OTF $\bar{H}(p)$ for a *circular aperture* in terms of the Fried parameter D_F, we can rewrite equations (66) and (68) as:

$$Q_\phi(\vec{r}) \;=\; 6.88\left(\frac{|\vec{r}|}{D_F}\right)^{5/3} \qquad\qquad \bar{H}(p) \;=\; \exp\left[-\,3.44\left(\frac{\lambda p}{D_F}\right)^{5/3}\right] \tag{75}$$

From this we see that the spatial frequency cut-off $p \approx D_F/\lambda$ ($\bar{H}(p) \approx 0.03$), consequently the angular resolution $\Delta\theta \approx \lambda/D_F$. This value is often called the *seeing angle* or simply the *seeing*. A typical value for D_F in the visible range of the spectrum ranges between 10 and 20 cm. So if for instance the seeing at the Keck 10-m telescope is 10 cm, the image quality is no better than that provided by a 10-cm amateur telescope. From expression (74) it is clear that the *Fried parameter is highly chromatic*: $\sim \lambda^{6/5}$. This means that the coherence area rapidly increases towards the infrared: a D_F of 20 cm at 0.5 μm increases to almost 1.2 meters at 2.2 μm.

3.2.4 Real time correction: principle of adaptive optics

Under certain conditions, it is possible to restore spatial frequencies, filtered by the atmosphere, in real time. This is the aim of *adaptive optics*, which has developed very rapidly since 1985.

An atmospheric compensation system employing active optics contains three main components: a sampling system, a wavefront sensor and a correction system. The sampling system provides the sensor with the distorted wavefront, in astronomy this normally entails a partly reflecting mirror, which typically diverts $\approx$ 10 percent of the radiation to the sensor, allowing the bulk of the radiation to proceed to form the main image. Many

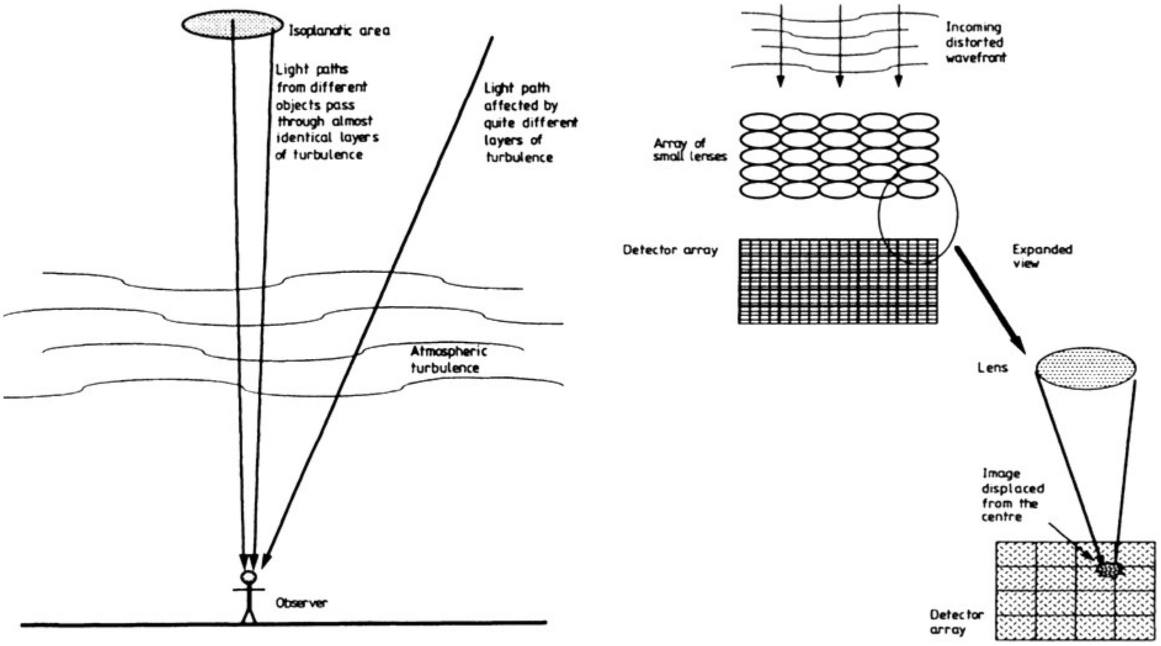
Isoplanatic area
Light paths from different objects pass through almost identical layers of turbulence
Light path affected by quite different layers of turbulence
Atmospheric turbulence
Observer
Incoming distorted wavefront
Array of small lenses
Detector array
Expanded view
Lens
Image displaced from the centre
Detector array

Figure 15: *Left: the isoplanatic area. Right: The Hartman array sensor for generation of the error signals caused by turbulence induced "folds" in the flat wavefronts arriving from a distant point source. Figure taken from Kitchin 1998.*

adaptive optics systems use a *guide star* rather than the object of interest to determine the distorsions of the wavefront. Inevitably, the guide star must be very close in the sky to the object of interest, or its wavefront will have undergone a different atmospheric distorsion. The region of the sky over which the images have been similarly affected by the turbulent atmosphere is called the *isoplanatic area or patch*. This can in practise be as small as a few arcseconds. The notion of the isoplanatic area is displayed in the left panel of figure 15. This small angular size of the isoplanatic area means that only very few objects have suitable guide stars, therefore artificial guide stars have been produced. This is being accomplished by the production of a so-called *optical echo*: a laser is tuned to one of the sodium D-line frequencies and excites the free sodium atoms in the atmosphere at a typical height of 80 - 90 km. The glowing atoms appear like a star-like point close to the object of interest. Laser-produced guide stars possess two inherent problems. Due to the relatively low height of the laser-produced guide star, the light path is slightly conical and may still differ substantially from the light path traversed by the radiation from the object of interest. Secondly, the outgoing laser beam is affected by atmospheric turbulence as well, i.e. the guide star moves with respect to the object of interest, again resulting in a blurred image during long exposures.
A wavefront sensor detects the distorsions in the incoming wavefront provided by the beam splitter. Figure 15 shows the so-called Hartmann sensor, frequently used in astronomical adaptive optics systems. It employs a two-dimensional array of small focussing lenses, each of which providing an image onto an array-detector. In the absence of wavefront distorsions, each image is *centered* on an array-element. Distorsions in the wavefront will displace the images from the detector centres, the degree of displacement

and its direction is used to generate error signals which are fed to a correction mirror. Since the atmosphere changes on a timescale of the order of ten milliseconds, the sampling, sensing and correction has to occur in a millisecond or less. The simplest systems only correct for the overall tilt of the wavefront, i.e. the mean tangent plane change of the wavefront that causes angular motion of the image. This is accomplished by suitably tilting a plane or segmented mirror placed in the light beam of the telescope in the direction opposite to the angular motion. A similar technique is the so-called *shift and add* methodology, in that case multiple short exposure images are shifted until their brightest points are aligned and then added together. More sophisticated techniques also involve fine scale displacement corrections by employing a thin mirror capable of being distorted by piezo- electric or other actuators placed undernearth the deformable mirror surface. The error signals of the sensor elements are then used to distort the mirror in the opposite manner relative to the distorsions of the incoming wavefront. Currently operating systems using this approach can achieve diffraction-limited performance in the near-infrared for telescopes of 3 to 4 meter diameter, i.e. ≈ 0.2 arcseconds at 2 μm.

The efficiency of an adaptive optics system is measured by the so-called *Strehl ratio*, which is the ratio of the intensity at the centre of the corrected image to that at the centre of a perfect diffraction-limited image of the same source. Strehl ratios of 0.6 to 0.8 are currently being achieved.

4 Band-limited sensing systems: Nyquist frequency

Let us consider the general case of a signal $S(x)$ which has been subject to the instrument response $R(x)$, so that the resulting measurement $M(x)$ follows from

$$M(x) = S(x) * R(x) \tag{76}$$

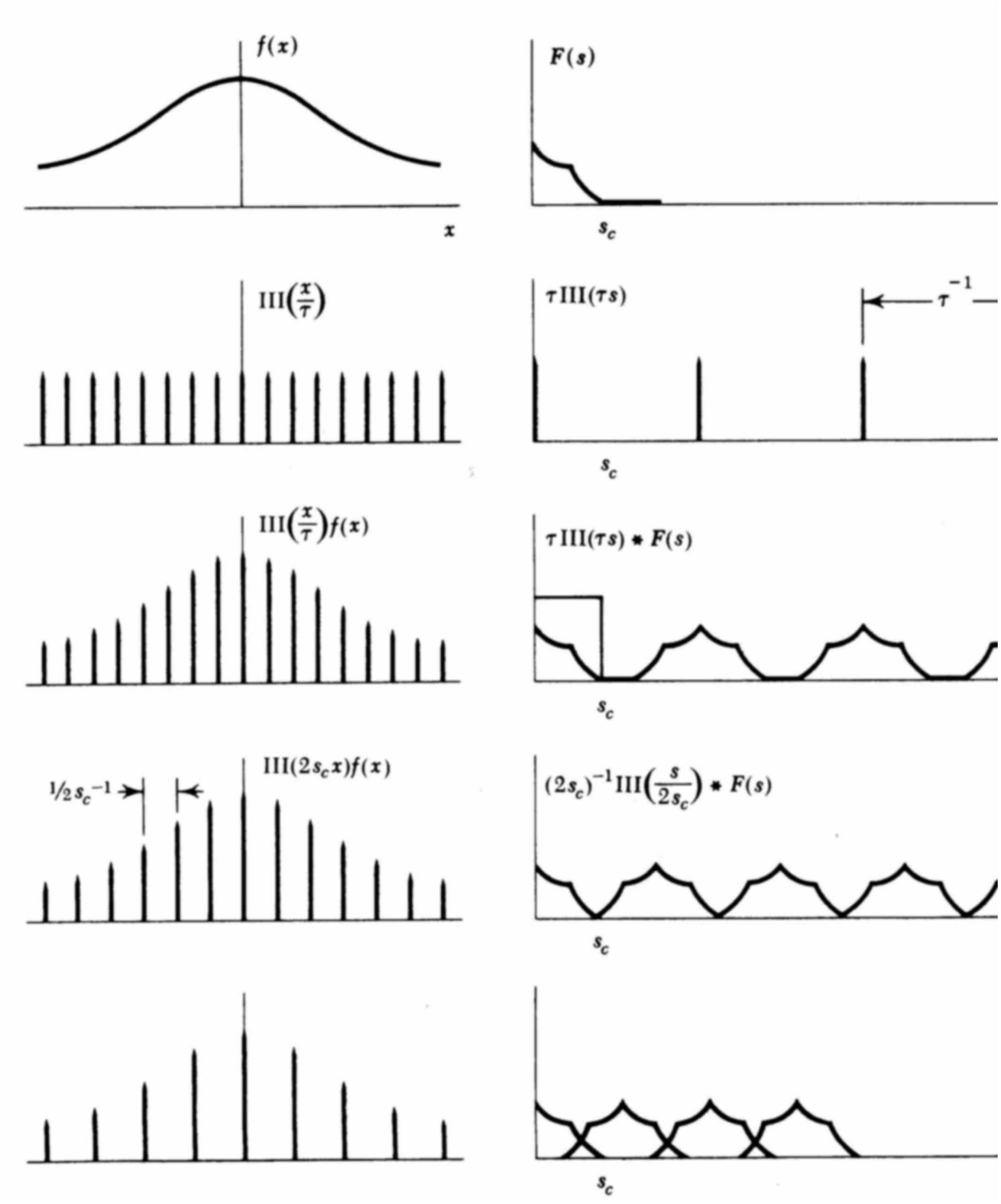
$f(x)$
x
$F(s)$
s_c
$\mathrm{III}(\frac{x}{\tau})$
$\tau\mathrm{III}(\tau s)$
τ^{-1}
s_c
$\mathrm{III}(\frac{x}{\tau})f(x)$
$\tau\mathrm{III}(\tau s) * F(s)$
s_c
$\frac{1}{2}s_c^{-1}$
$\mathrm{III}(2s_c x)f(x)$
$(2s_c)^{-1}\mathrm{III}(\frac{s}{2s_c}) * F(s)$
s_c
s_c

Figure 16: *Illustration of sampling optimally, compared with undersampling and oversampling. Left in the time domain, right in the Fourier domain. Figure taken from Bracewell 1986.*

Because of the finite frequency response of the instrument, $M(x)$ is always limited in bandwidth, i.e. the Fourier transform $M(s) \Leftrightarrow M(x)$ is a bandwidth-limited function. This function is characterised by a maximum cut-off frequency s_{max}, also called the critical or Nyquist frequency (s_c). In the case of a gaussian response the frequencies will never be distributed purely gaussian, since no physical system can transmit the tail frequencies up to ∞. Nyquist (and Shannon) established a theorem for optimum sampling of band limited observations. This theorem states that no information is lost if sampling occurs at intervals $\tau = 1/(2s_c)$. Let $M(x)$ subsequently be sampled at regular intervals, $M(x) \rightarrow M(n\tau)$ with n an integer and τ the sampling interval. To describe the sampling process quantitatively we introduce the *shah* function, also called the comb of Dirac, which constitutes a series of δ functions at regular distances equal to 1:

$$\perp\!\!\perp\!\!\perp(x) \equiv \sum_{n=-\infty}^{\infty} \delta(x-n) \tag{77}$$

This shah function can be extended to arbitrary distances by noting $a\perp\!\!\perp\!\!\perp(ax) = \sum_n \delta(x - n/a)$.
The sampled signal $M_s(x)$ can now be expressed as

$$M_s(x) = \sum_n M(n\tau)\delta(x-n\tau) = \frac{1}{\tau}\perp\!\!\perp\!\!\perp\left(\frac{x}{\tau}\right)M(x) \tag{78}$$

The Fourier transform $M_s(s) \Leftrightarrow M_s(x)$ equals

$$M_s(s) = \perp\!\!\perp\!\!\perp(\tau s) * M(s) = \frac{1}{\tau}\sum_n M\left(s - \frac{n}{\tau}\right) \tag{79}$$

This expression shows that, except for a proportionality factor $1/\tau$, $M_s(s)$ represents a series of replications of $M(s)$ at intervals $1/\tau$. Because $M(s)$ is a bandwidth-limited function with a cut-off frequency of say $s = s_c$, we can recover fully the single (i.e. not repeated) function $M(s)$ from this series by multiplication with τ and by filtering with the gate function $\Pi(s/2s_c)$:

$$\Pi\left(\frac{s}{2s_c}\right)\tau\perp\!\!\perp\!\!\perp(\tau s) * M(s) \Leftrightarrow 2s_c \mathrm{sinc} 2s_c x * \perp\!\!\perp\!\!\perp\left(\frac{x}{\tau}\right)M(x) \tag{80}$$

$M(x)$ can be reconstructed exactly if the series of $M(s)$ functions in the frequency domain touch without overlap. This is the case if we sample at $\tau = 1/(2s_c)$, which therefore is the optimum sample interval. Performing the convolution we fully reconstruct $M(x)$, i.e.:

$$M(x) = \int_{-\infty}^{+\infty} \mathrm{sinc}\left(\frac{x-x'}{\tau}\right)\sum_n M(n\tau)\delta(x'-n\tau)dx' = \sum_n \mathrm{sinc}\left(\frac{x-n\tau}{\tau}\right)M(n\tau) \tag{81}$$

We can check this result easily at a sampling point $x = j\tau$, with $\mathrm{sinc}(j-n) = 1$ for $j = n$ and $= 0$ for $j \neq n$:

$$M(x) = M(j\tau) \tag{82}$$

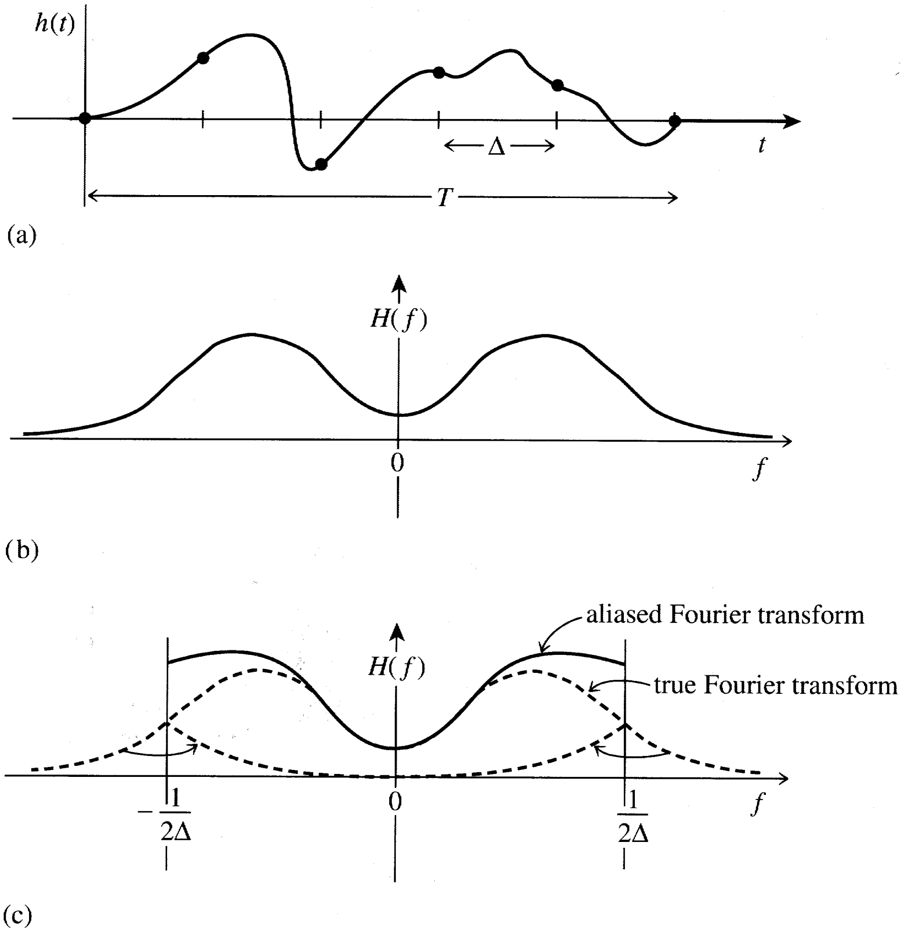

Figure 17: *The function $h(t)$ shown in the top panel is undersampled. This means that the sampling interval Δ is larger than $\frac{1}{2f_{max}}$. The lower panel shows that in this case the power in the frequencies above $\frac{1}{2\Delta}$ is 'mirrored' with respect to this frequency and produces an aliased transform which deviates from the true Fourier transform. Figure taken from Press et al. 1992.*

Note: Eq. 80 shows that the calculation of intermediate points from samples does not of course depend on calculating Fourier transforms. The equivalent operation in the x-domain entails the convolution of $2s_c\mathrm{sinc}2s_cx$ directly with $\perp\!\!\!\perp\!\!\!\perp(x/\tau)M(x)$. Notice that the omission of the $1/\tau$ factor in Eq. 78 ensures the proper normalization in the s-domain! Expression 81 shows that a superposition of a series of *sinc*-functions with weight factors $M(n\tau)$, i.e. the sample values, at intervals τ exactly reconstruct the continuous function $M(x)$. In fact the *sinc*-functions provide the proper interpolation between the consecutive sample points, this is the reason why the *sinc*-function is sometimes referred to as the *interpolation function.*

Thus, the use of a discrete Fourier transform causes no loss of information, provided that the sampling frequency $\frac{1}{\tau}$ is twice the highest frequency in the continuous input function (i.e. the source function convolved with the response function). The maximum frequency s_{max} that can be determined for a given sampling interval equals therefore

$\frac{1}{2\tau}$. If the input signal is sampled too slowly, i.e. if the signal contains frequencies higher than $\frac{1}{2\tau}$, then these cannot be determined after the sampling process and the finer details will be lost (see Fig. 16). More seriously however, the higher frequencies which are not resolved will beat with the measured frequencies and produce spurious components in the frequency domain below the Nyquist frequency. This effect is known as *aliasing* and may give rise to major problems and uncertainties in the determination of the source function, see figure 17 in the case of a time function.

5 Coherence and Interference

5.1 The Visibility function

The coherence phenomenon is directly coupled to correlation, and the degree of coherence of an EM-wave field $\tilde{E}(\vec{r},t)$ can be quantitativily described by employing the auto- and cross-correlation technique for the analysis of a stochastic process.
The electric vector of the wave field at a position $\vec{r}$ at time t, $\tilde{E}(\vec{r},t)$, is a complex quantity, denoting the amplitude and phase of the field. To assess the coherence phenomenon, the question to be answered is: how do the nature of the source and the geometrical configuration of the situation relate to the resulting phase correlation between two laterally spaced points in the wave field?
This brings to mind Young's interference experiment in which a primary monochromatic source S illuminates two pinholes in an opaque screen, see figure 18. The pinholes S_1 and S_2 act as secondary sources, generating a fringe pattern on a distant observation plane Σ_O. If S is an idealized monochromatic point source, the wavelets issuing from any set of apertures S_1 and S_2 will maintain a constant relative phase; they are precisely correlated and therefore mutually fully coherent. On the observation plane Σ_O a well-defined array of stable fringes will result and the radiation field is spatially coherent. At the other extreme, if the pinholes S_1 and S_2 are illuminated by separate thermal sources (even with narrow frequency bandwidths), no correlation exists; no fringes will be observable in the observation plane Σ_O and the fields at S_1 and S_2 are said to be incoherent. The generation of interference fringes is seemingly a convenient measure of the degree of coherence of a radiation field. The quality of the fringes produced by an interferometric system can be described quantitativily using the *Visibility function* V:

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{83}$$

here I_{max} and I_{min} are the irradiances corresponding to the maximum and adjacent minimum in the fringe system.

5.2 Young's dual beam interference experiment

To assess the mutual coherence between two positions in a radiation field in a quantitative fashion, consider the situation displayed in figure 18, with an extended narrow bandwidth radiation source S, which generates fields $\tilde{E}(\vec{r}_1,t) \equiv \tilde{E}_1(t)$ at S_1 and $\tilde{E}(\vec{r}_2,t) \equiv \tilde{E}_2(t)$ at S_2, respectively. *Interalia: no polarization effects are considered, and therefore a scalar treatment using* $\tilde{E}(\vec{r},t)$ *will suffice.*

If these two positions in the radiation field are isolated using an opaque screen with two small apertures, we are back to Young's experimental set-up. The two apertures serve as sources of secondary wavelets, which propagate out to some point P on the observation plane Σ_O. The resultant field at P is:

$$\tilde{E}_P(t) \;=\; \tilde{C}_1\tilde{E}_1(t-t_1) + \tilde{C}_2\tilde{E}_2(t-t_2) \tag{84}$$

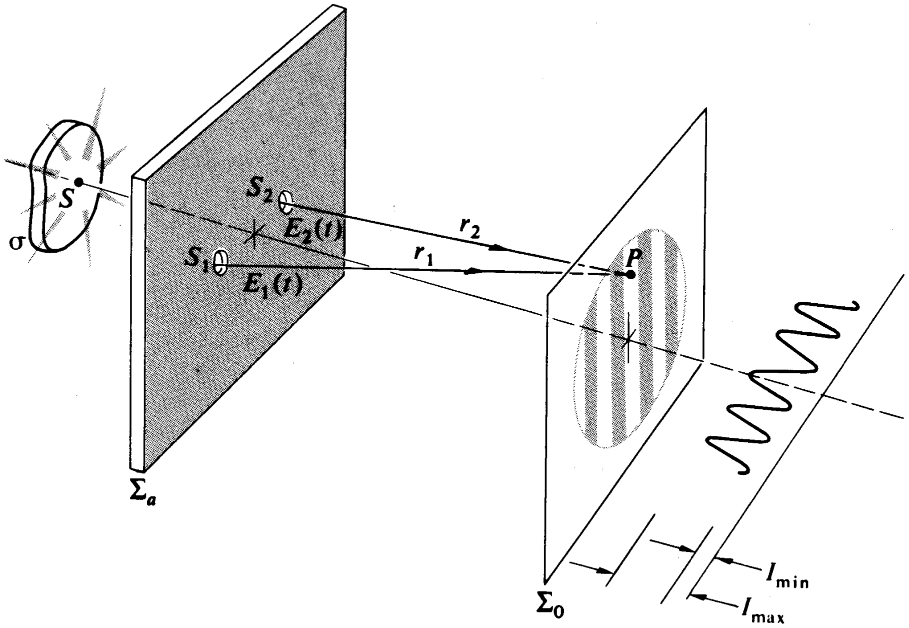

Figure 18: *Young's experiment using a quasi-monochromatic source S illuminating two pinholes S_1 and S_2. Figure taken from Hecht 1987.*

with $t_1 = r_1/c$ and $t_2 = r_2/c$, r_1 and r_2 representing the pathlengths to P as measured from S_1 and S_2, respectively. This expression tells us that the field at the space-time point (P, t) can be determined from the field that existed at S_1 and S_2 at $(t - t_1)$ and $(t - t_2)$, respectively, these being the the instants when the light, which is now overlapping at P, first emerged from the apertures. The quantities $\tilde{C}_1$ and $\tilde{C}_2$ are so-called *propagators*, they mathematically reflect the alterations in the field resulting from it having transversed either of the apertures. For example, the secondary wavelets issuing from the pinholes in the Young set-up are out of phase by $\pi/2$ radians with the primary wave incident on the aperture screen. In that case, $\tilde{C}_1$ and $\tilde{C}_2$ are purely imaginary numbers equal to $e^{i\pi/2}$.

5.3 The mutual coherence function

The resultant irradiance at P, averaged over some time interval which is long compared to the coherence time, is:

$$I = \mathbf{E}\left\{\tilde{E}_P(t)\tilde{E}_P^*(t)\right\} \tag{85}$$

Employing equation (84) this can be written as:

$$\begin{aligned} I &= \tilde{C}_1\tilde{C}_1^*\mathbf{E}\left\{\tilde{E}_1(t-t_1)\tilde{E}_1^*(t-t_1)\right\} \\ &+ \tilde{C}_2\tilde{C}_2^*\mathbf{E}\left\{\tilde{E}_2(t-t_2)\tilde{E}_2^*(t-t_2)\right\} \end{aligned}$$

$$
\begin{aligned}
&+ \tilde{C}_1\tilde{C}_2^*\mathbf{E}\left\{\tilde{E}_1(t-t_1)\tilde{E}_2^*(t-t_2)\right\} \\
&+ \tilde{C}_1^*\tilde{C}_2\mathbf{E}\left\{\tilde{E}_1^*(t-t_1)\tilde{E}_2(t-t_2)\right\} \qquad (86)
\end{aligned}
$$

It is now assumed that the wave field is *stationary*, as is almost universally the case, i.e. the statistical nature of the wave field does not alter with time and the time average is independent of whatever time origin we select (i.e. the wave field is designated as Wide Sense Stationary). Accordingly, the first two expectation values in equation (86) can be rewritten as:

$$I_{S_1} = \mathbf{E}\left\{\tilde{E}_1(t)\tilde{E}_1^*(t)\right\} \text{ and } I_{S_2} = \mathbf{E}\left\{\tilde{E}_2(t)\tilde{E}_2^*(t)\right\} \qquad (87)$$

where the time origin was displaced by amounts t_1 and t_2, respectively. The subscripts for the irradiances used here underscore the fact that they refer to the values at points S_1 and S_2. Furthermore, if we introduce the time difference $\tau = t_2 - t_1$, the time origin of the last two terms can be shifted by an amount t_2 yielding:

$$\tilde{C}_1\tilde{C}_2^*\mathbf{E}\left\{\tilde{E}_1(t+\tau)\tilde{E}_2^*(t)\right\} + \tilde{C}_1^*\tilde{C}_2\mathbf{E}\left\{\tilde{E}_1^*(t+\tau)\tilde{E}_2(t)\right\} \qquad (88)$$

This expression comprises a complex quantity plus its own complex conjugate and is therefore equal to twice its *Real* part:

$$2\ Re\left[\tilde{C}_1\tilde{C}_2^*\mathbf{E}\left\{\tilde{E}_1(t+\tau)\tilde{E}_2^*(t)\right\}\right] \qquad (89)$$

As noted before, the $\tilde{C}$-coefficients are purely imaginary, i.e. $\tilde{C}_1\tilde{C}_2^* = \tilde{C}_1^*\tilde{C}_2 = |\tilde{C}_1||\tilde{C}_2|$.

The expectation value contained in expression (89) is a cross-correlation function, which is denoted by:

$$\tilde{\Gamma}_{12}(\tau) = \mathbf{E}\left\{\tilde{E}_1(t+\tau)\tilde{E}_2^*(t)\right\} \qquad (90)$$

This equation is referred to as the **mutual coherence function** of the wave field at positions S_1 and S_2. Making use of the definitions above, equation (86) now takes the form:

$$I = |\tilde{C}_1|^2 I_{S_1} + |\tilde{C}_2|^2 I_{S_2} + 2|\tilde{C}_1||\tilde{C}_2|\ Re\,\tilde{\Gamma}_{12}(\tau) \qquad (91)$$

The terms $|\tilde{C}_1|^2 I_{S_1}$ and $|\tilde{C}_2|^2 I_{S_2}$ are the irradiance at P, arising when one or the other of the apertures is open alone: either $\tilde{C}_1 = 0$ or $\tilde{C}_2 = 0$. Denoting these irradiances as I_1 and I_2 we have:

$$I = I_1 + I_2 + 2|\tilde{C}_1||\tilde{C}_2|\ Re\,\tilde{\Gamma}_{12}(\tau) \qquad (92)$$

If S_1 and S_2 are made to coincide, the mutual coherence function becomes the autocorrelation function:

$$\tilde{\Gamma}_{11}(\tau) = \tilde{R}_1(\tau) = \mathbf{E}\left\{\tilde{E}_1(t+\tau)\tilde{E}_1^*(t)\right\} \qquad (93)$$

or:

$$\tilde{\Gamma}_{22}(\tau) = \tilde{R}_2(\tau) = \mathbf{E}\left\{\tilde{E}_2(t+\tau)\tilde{E}_2^*(t)\right\} \qquad (94)$$

One can imagine that two wavetrains emerge from these coalesced pinholes and somehow pick up a relative phase delay τ. In the situation at hand $\tau = 0$, since the optical path difference (*shorthand: OPD*) goes to zero. Hence:

$$
\begin{aligned}
I_{S_1} &= \mathbf{E}\left\{\tilde{E}_1(t)\tilde{E}_1^*(t)\right\} = \Gamma_{11}(0) = \mathbf{E}\left\{|\tilde{E}_1(t)|^2\right\} \text{ and} \\
I_{S_2} &= \mathbf{E}\left\{\tilde{E}_2(t)\tilde{E}_2^*(t)\right\} = \Gamma_{22}(0) = \mathbf{E}\left\{|\tilde{E}_2(t)|^2\right\} \qquad (95)
\end{aligned}
$$

These autocorrelation functions are also called *self-coherence functions.* For $\tau = 0$ they represent the (average) irradiance (power) of the radiation field at positions S_1 and S_2 respectively.

5.4 Interference law for a partially coherent radiation field: the complex degree of coherence

From equation (92) and the selfcoherence functions we can now write:

$$|\tilde{C}_1||\tilde{C}_2| \;=\; \frac{\sqrt{I_1}\sqrt{I_2}}{\sqrt{\Gamma_{11}(0)}\sqrt{\Gamma_{22}(0)}} \tag{96}$$

Hence, the normalized expression for the mutual coherence function can now be defined as:

$$\tilde{\gamma}_{12}(\tau) \;\equiv\; \frac{\tilde{\Gamma}_{12}(\tau)}{\sqrt{\Gamma_{11}(0)\Gamma_{22}(0)}} \;=\; \frac{\mathbf{E}\left\{\tilde{E}_1(t+\tau)\tilde{E}_2^*(t)\right\}}{\sqrt{\mathbf{E}\left\{|\tilde{E}_1(t)|^2\right\}\mathbf{E}\left\{|\tilde{E}_2(t)|^2\right\}}} \tag{97}$$

This quantity is referred to as the **complex degree of coherence**, equation (92) can now be recast into its final form:

$$I \;=\; I_1 \;+\; I_2 \;+\; 2\sqrt{I_1 I_2}\; Re\; \tilde{\gamma}_{12}(\tau) \tag{98}$$

which is the *general interference law for a partially coherent radiation field.*
The quantity $\tilde{\gamma}_{12}(\tau)$ *simultaneously* gives a measure of the *spatial coherence* by comparison of two locations in space (S_1 and S_2 in the above case) and the *coherence in the time domain* by accounting for a time lag τ between both signals.
$\tilde{\gamma}_{12}(\tau)$ is a complex variable and can be written as:

$$\tilde{\gamma}_{12}(\tau) \;=\; |\tilde{\gamma}_{12}(\tau)|e^{i\psi_{12}(\tau)} \tag{99}$$

From equation (97) and the Schwarz inequality it is clear that $0 \;\leq |\tilde{\gamma}_{12}(\tau)| \;\leq 1$. The phase angle $\psi_{12}(\tau)$ of $\tilde{\gamma}_{12}(\tau)$ relates to the phase angle between the fields at S_1 and S_2 and the phase angle difference concomitant with the OPD in P resulting in the time lag τ, as shown in equation (90). For quasi-monochromatic radiation at a mean wavelength $\overline{\lambda}$ and frequency $\overline{\nu}$, the phase difference ϕ due to the OPD can be expressed as:

$$\phi \;=\; \frac{2\pi}{\overline{\lambda}}(r_2 - r_1) \;=\; 2\pi\overline{\nu}\tau \tag{100}$$

If we designate a phase angle $\alpha_{12}(\tau)$ between the fields at S_1 and S_2, we have
$\psi_{12}(\tau) = [\alpha_{12}(\tau) - \phi)]$.
Hence:

$$Re\;\tilde{\gamma}_{12}(\tau) \;=\; |\tilde{\gamma}_{12}(\tau)|\cos\left[\alpha_{12}(\tau) - \phi\right] \tag{101}$$

Substitution of this expression in the interference law for partially coherent radiation given in equation (98) yields for the intensity observed at point P on the observation plane Σ_O:

$$I \;=\; I_1 \;+\; I_2 \;+\; 2\sqrt{I_1 I_2}\;|\tilde{\gamma}_{12}(\tau)|\cos\left[\alpha_{12}(\tau) - \phi\right] \tag{102}$$

The maximum and minimum values of I occur if the cosine term in equation (102) equals +1 and −1, respectively. The Visibility V (see definition (83)) at position P can therefore be expressed as:

$$V = \frac{2\sqrt{I_1}\sqrt{I_2}}{I_1 + I_2}|\tilde{\gamma}_{12}(\tau)| \tag{103}$$

In practice, frequently things are (or can be) adjusted in such a way that $I_1 = I_2$, giving rise to the following simplified expressions for the total irradiance I and Visibility V:

$$I \ = \ 2I_0 \left\{1 + |\tilde{\gamma}_{12}(\tau)| \cos\left[\alpha_{12}(\tau) - \phi\right]\right\} \text{ and } V \ = \ |\tilde{\gamma}_{12}(\tau)| \tag{104}$$

We note that in this case *the modulus of the complex degree of coherence is identical to the visibility of the fringes* ! This then provides an experimental means of obtaining $|\tilde{\gamma}_{12}(\tau)|$ from the resultant fringe pattern. Moreover, the off-axis shift in the location of the central fringe (no OPD → $\phi = 0$) is a measure of $\alpha_{12}(\tau)$, the relative retardation in phase of the fields at S_1 and S_2. Thus, measurements of the visibility and the fringe position yield both the amplitude and phase of the complex degree of coherence.

Dealing with the definition of the complex degree of coherence, it can be noted that the intensity fluctuations are also expected to be partially coherent, since the amplitude and phase fluctuations in a thermal signal tend to track each other. An impression of such a fluctuating wave signal is given in figure 19, both in absolute value and centered around the avarage value $\overline{I}$: $\triangle I = I - \overline{I}$. The random superposition of wave packets results in

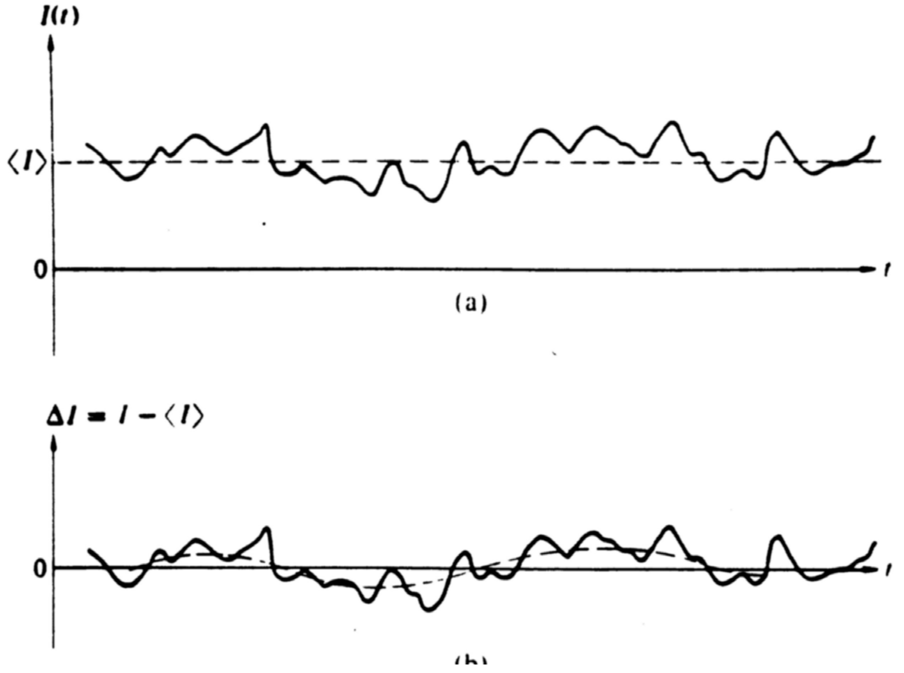

Figure 19: *Partially coherent intensity fluctuations in a thermal radiation source. Both absolute values $\overline{I}$ and relative values $\triangle I = I - \overline{I}$ are shown. Figure taken from Hecht 1987.*

a Gaussian distribution of the amplitude fluctuations for one direction of polarisation and, consequently, is sometimes referred to as *Gaussian light*. This fact directly follows from the *central limit theorem*. Taking the cross-correlation $\mathbf{E}\left\{\triangle I_1(t+\tau)\triangle I_2(t)\right\}$ between two different parts of the incoming radiation beam yields now in principle an interferometric tool, which does not involve the usage of phase information. This technique, labelled as *correlation* or *intensity*-interferometry, can indeed be succesfully applied for very bright stellar sources to obtain accurate estimations of stellar angular diameters. It was first tried in radio astronomy in the middle 1950's, and later on (1956) also succesfully used in an optical stellar interferometer by Hanbury Brown and Twiss. They employed search-light mirrors to collect starlight onto two apertures, which was then focussed on two photomultiplier devices. Although photomultipliers operate on the photo-electric effect, which is keyed to the quantum nature of the optical light, laboratory zline experiments showed that the correlation effect was indeed preserved in the process of photo-electric emission.

The star Sirius was the first to be examined, and it was found to possess an angular diameter of 6.9 milliseconds of arc. For certain stars angular diameters of as little as 0.5 milliseconds of arc can be measured in this way.

6 Indirect spectroscopy

6.1 Temporal coherence

Temporal coherence is characterised by the coherence time τ_c . The value of τ_c follows from the finite bandwidth of the radiation source under consideration. If we assume a quasi-monochromatic (QM) source, then we have $\tau_c \approx \frac{1}{\triangle\nu}$ with $\triangle\nu$ the line width (in radiation frequency) of the QM-source.
These effects can be assessed with the aid of the Wiener-Khinchine theorem:

$$S(\nu) = \int_{-\infty}^{+\infty} R(\tau) e^{-2\pi i\nu\tau} d\tau \tag{105}$$

$$R(\tau) = \int_{-\infty}^{+\infty} S(\nu) e^{2\pi i\nu\tau} d\nu \tag{106}$$

Take as an example a Gaussian shaped spectral line profile, i.e.

$$S(\nu) \sim e^{-\left(\frac{\nu}{\triangle\nu}\right)^2} \Longleftrightarrow R(\tau) \sim e^{-\left(\frac{\tau}{\tau_c}\right)^2} \tag{107}$$

As can be seen from the FT (indicated by $\Longleftrightarrow$ in expression(107), the wave packet corresponding to this line profile has an autocorrelation function that is also Gaussian with a characteristic width τ_c, moreover the *autocorrelation* $R(\tau)$ equals the *autocovariance* $C(\tau)$. This corresponds to a wave train with a Gaussian shaped envelope for the

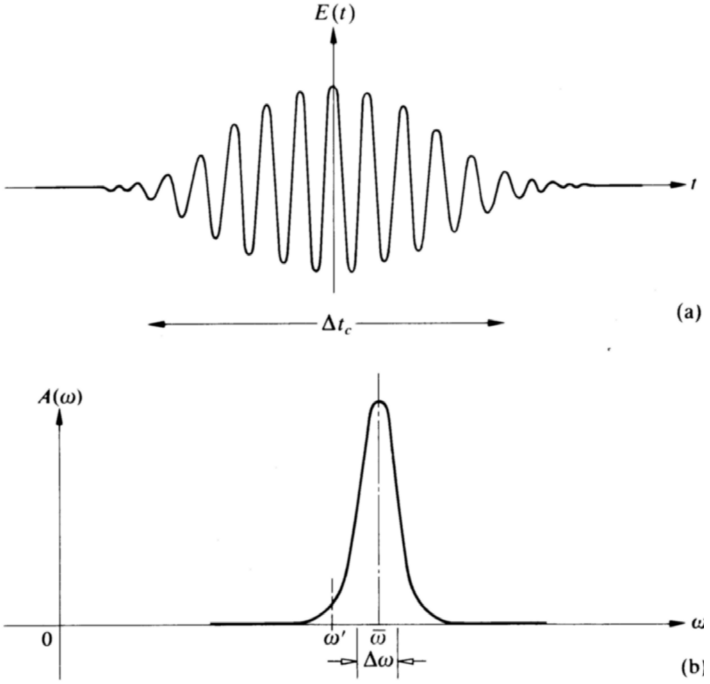

Figure 20: *A Gaussian shaped line profile of a quasi-monochromatic radiation source and the shape of the associated wavepacket. Figure taken from Hecht 1987.*

wave amplitude (see figure 20).
Try to memorize the following notions:

- A first order system shows an exponential autocorrelation function $R(\tau)$.
- A Gaussian line profile in the frequency domain shows an amplitude modulated wave train with a Gaussian envelope in the time domain (*see discussion above*).
- A Lorentz line profile in the frequency domain shows an exponentially damped oscillator profile in the time domain (*try this yourself*).

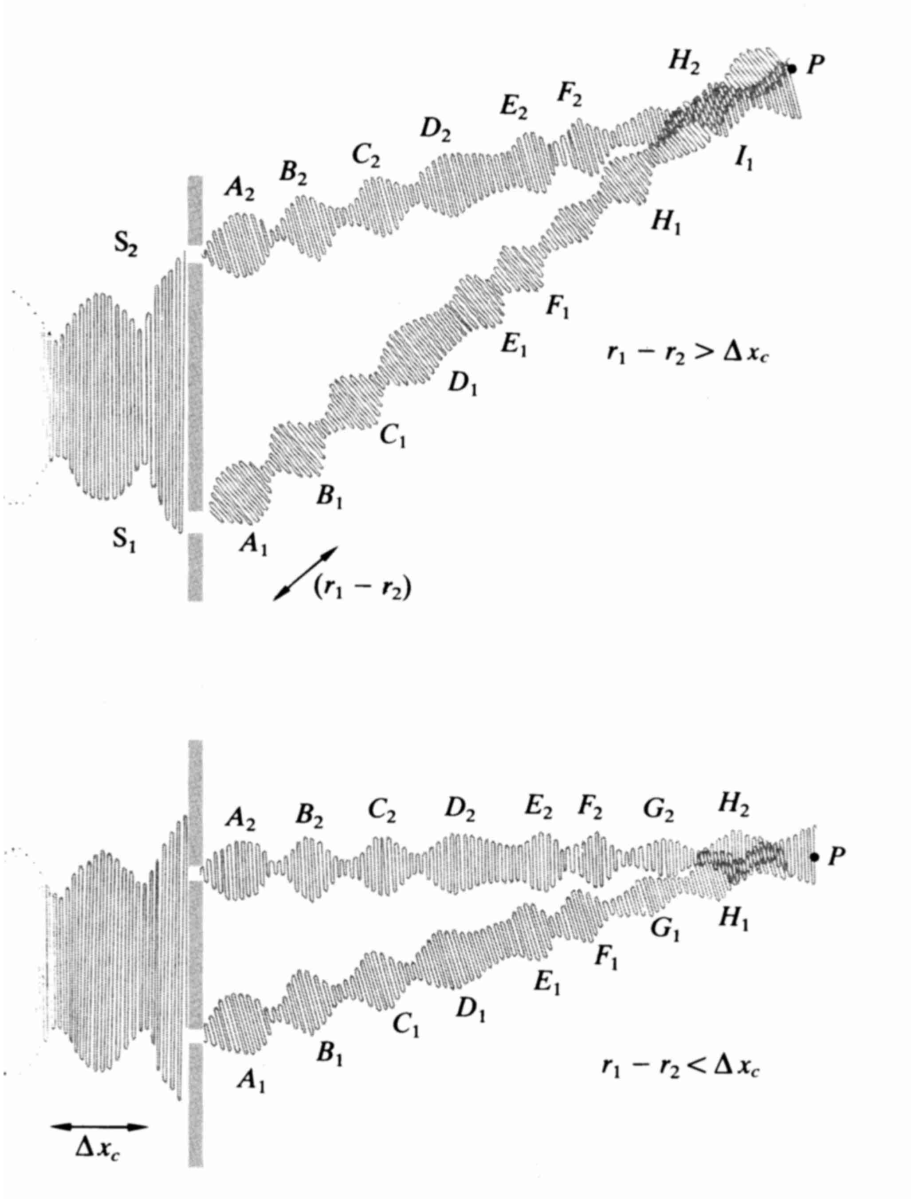

Figure 21: *The influence of the coherence length on the interference pattern of two diffracted coherent thermal radiation sources S_1 and S_2. Figure taken from Hecht 1987.*

For spectroscopic measurements at infrared and shorter wavelengths one can directly disperse the incoming radiation beam with the aid of a wavelength dispersive device, like for instance a transmission or a reflection grating, and measure the resulting intensity distribution (i.e. the spectrum). However for spectroscopy at radio and submillimeter wavelengths one employs an indirect method. The incoming wave signal is fed into a *correlator* that produces the temporal coherence function $R(\tau)$, a subsequent FT of this function yields the spectral distribution $S(\nu)$ by virtue of the Wiener-Khinchine relation.

6.2 Longitudinal correlation

Associated with the coherence time τ_c is the so-called coherence length $l_c = c\tau_c$.

Problem: Show that the coherence length can also be expressed as $l_c = \frac{\lambda^2}{\triangle\lambda}$ in which $\triangle\lambda$ refers to the equivalent of $\triangle\nu$ in the wavelength domain.

Now consider an EM-wave that propagates along a vector $\vec{r}$, and mark two positions P_1 and P_2 on this line of propagation at a mutual distance R_{12}. If $R_{12} \ll l_c$, there will be a strong correlation between the EM-fields at P_1 and P_2 and as a consequence interference effects will be possible. In the case of $R_{12} \gg l_c$, no interference effects are possible. This effect (i.e. potential interference *yes* or *no*) relates to the so-called *longitudinal correlation* or *longitudinal spatial coherence*.
This effect can be clearly demonstrated by considering the wave trains in Young's interference experiment (see figure 21). The diffracted beams emanating from S_1 and S_2, which are coherent radiation sources, cause an interference pattern. However, in the case of large path differences the interference contrast will diminish, since corresponding wave packets in the stochastic signal no longer overlap (see figure 21: packet H_1 and $H_2 \Longrightarrow$ packet I_1 and H_2).

7 Indirect Imaging

If we apply equation (104) to the situation of a quasi-monochromatic point source S, located on the central axis, the wavelets emanating from two infinitesimal pinholes S_1 and S_2 (fictitious case!) will be fully coherent and exactly in phase ($\alpha_{12}(\tau) = 0$) and constitute two coherent secondary sources. Interference will occur, provided the OPD between the interfering beams is less than the coherence length $l = c\tau_c$. With $V = |\tilde{\gamma}_{12}(\tau)| = 1$, the equation for the total irradiance in (104) reduces to:

$$I \;=\; 2I_0(1+\cos\phi) \;=\; 4I_0\cos^2\frac{\phi}{2} \tag{108}$$

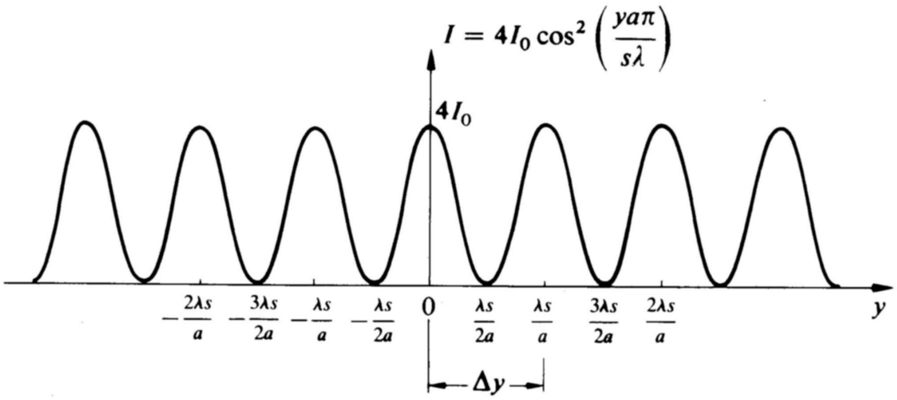

Figure 22: *Interferogram displaying the idealized irradiance as a function of the y-coordinate of the fringes in the observation plane Σ_O. Figure taken from Hecht 1987.*

Taking a distance a between the pinholes and assuming that the distance s to the observation plane Σ_O is very much larger than a, we can express the path difference $(r_2 - r_1)$ in equation (100) for ϕ in good approximation by:

$$r_2 - r_1 \;=\; a\theta \;=\; \frac{a}{s}\,y \tag{109}$$

Here y is the linear coordinate in the observation plane Σ_O, starting from the intersection of the central axis with this plane and running perpendicular to the fringes. Substituting ϕ in equation (108) by combining (100) and (109) we get an analytical expression for the interferogram in Σ_O:

$$I \;=\; 4I_0\cos^2\frac{\pi a y}{s\overline{\lambda}} \tag{110}$$

This (idealized) irradiance versus distance distribution is displayed in figure 22 and constitutes basically the response of an ideal two-element interferometer to a monochromatic point source, i.e. the PSF of an ideal two-element interferometer.

The derivation of the actual PSF for a non-ideal two-element interferometer, that also accounts for the finite size of the apertures, can be accomplished by utilizing the concept of the *pupil function* which was introduced for single apertures in section 3. This concept will now first be applied to a two-element interferometer, starting with the assumption of infinitesimal apertures (the ideal case discussed above) and subsequently by implementing the practical case with finite aperture sizes. Later on in this chapter, with the introduction of aperture synthesis, this will be extended to the derivation of the point source response function (PSF) and the optical transfer function (OTF) for *multi-aperture arrays.*

7.1 Quasi-monochromatic point source: spatial response function of a two-element interferometer

Consider two circular apertures with diameter d separated by a baseline direction vector $\vec{s}$. Take the origin of the pupil function $P(\vec{r})$ on the baseline vector $\vec{s}$ symmetrically positioned between the two apertures. The pupil function $P(\vec{r})$ can now be expressed as:

$$P(\vec{r}) \,=\, \Pi\left(\frac{\vec{r}-\vec{s}/2}{d}\right) \,+\, \Pi\left(\frac{\vec{r}+\vec{s}/2}{d}\right) \tag{111}$$

with Π the 2-dimensional circular window function.
Introducing the ***spatial frequency variable*** $\vec{\zeta} = \vec{r}/\lambda$, the pupil function can be rewritten as:

$$P(\vec{\zeta}) \,=\, \Pi\left(\frac{\vec{\zeta}-\vec{s}/2\lambda}{d/\lambda}\right) \,+\, \Pi\left(\frac{\vec{\zeta}+\vec{s}/2\lambda}{d/\lambda}\right) \tag{112}$$

Now if the diameter d of the apertures is assumed to be very much smaller than the length $|\vec{s}| = D$ of the baseline vector, i.e. $d \ll D$, the pupil function can be approximated by:

$$P(\vec{\zeta}) \,=\, \delta\left(\vec{\zeta}-\vec{s}/2\lambda\right) \,+\, \delta\left(\vec{\zeta}+\vec{s}/2\lambda\right) \tag{113}$$

The optical tranfer function (OTF) for this limiting case of infinitesimal apertures, follows from the *self-convolution* of the the function $(\lambda/R)P(\vec{\zeta})$. Since we have a symmetrical pupil function in this case, the self-convolution is identical to the *autocorrelation* of $(\lambda/R)P(\vec{\zeta})$:

$$\begin{aligned} OTF \,=\, H_\lambda(\vec{\zeta}) \,&=\, \left(\frac{\lambda}{R}\right)^2 \int\int_{\text{pupil plane}} P(\vec{\zeta}\,')P(\vec{\zeta}\,'-\vec{\zeta})d\vec{\zeta}\,' \\ &= \left(\frac{\lambda}{R}\right)^2 \int\int_{\text{pupil plane}} \left[\delta\left(\vec{\zeta}\,'-\vec{s}/2\lambda\right) \,+\, \delta\left(\vec{\zeta}\,'+\vec{s}/2\lambda\right)\right]\cdot \\ &\qquad \cdot\left[\delta\left(\vec{\zeta}\,'-\vec{\zeta}-\vec{s}/2\lambda\right) \,+\, \delta\left(\vec{\zeta}\,'-\vec{\zeta}+\vec{s}/2\lambda\right)\right]d\vec{\zeta}\,' \\ &= 2\left(\frac{\lambda}{R}\right)^2 \left[\delta(\vec{\zeta}) + \frac{1}{2}\delta\left(\vec{\zeta}-\vec{s}/\lambda\right) \,+\, \frac{1}{2}\delta\left(\vec{\zeta}+\vec{s}/\lambda\right)\right] \end{aligned} \tag{114}$$

This OTF shows that the pair of pinholes transmits three principal spatial frequencies: a DC-component $\delta(\vec{0})$ and two high frequencies related to the length of the baseline vector $\vec{s}$ at $\pm\vec{s}/\lambda$. These three spatial frequencies represent a three-point sampling of

the so-called uv-plane in 2-dimensional frequency(Fourier) space. **Note:** Full frequency sampling of the uv-plane will allow complete reconstruction of the brightness distribution of the celestial source being observed!
The PSF follows from FT$\{H_\lambda(\vec{\zeta})\}$:

$$\begin{aligned}\delta(\vec{\zeta}) \quad &\Leftrightarrow \quad \mathbf{1} \ \text{(2-dimensional sheet!)}\\ \delta\left(\vec{\zeta} - \vec{s}/\lambda\right) \quad &\Leftrightarrow \quad e^{i2\pi\vec{\theta}\cdot\vec{s}/\lambda}\\ \delta\left(\vec{\zeta} + \vec{s}/\lambda\right) \quad &\Leftrightarrow \quad e^{-i2\pi\vec{\theta}\cdot\vec{s}/\lambda}\end{aligned}$$

$$\Longrightarrow PSF \,=\, \left(\frac{\lambda}{R}\right)^2 \left[2(1+\cos 2\pi\vec{\theta}\cdot\vec{s}/\lambda)\right] \,=\, 4\left(\frac{\lambda}{R}\right)^2 \cos^2\pi\vec{\theta}\cdot\vec{s}/\lambda \tag{115}$$

in which $\vec{\theta}$ ***represents the 2-dimensional angular coordinate vector*** used to describe the angular distribution of the diffracted image. The attenuation factor $(\lambda/R)^2$ results from the spherical expansion of the diffracted field in the Fraunhofer limit.
If we assume a light flux I_0 emanating from each pinhole, we can express the brightness distribution for the diffracted beam as:

$$I \,=\, 4I_0\cos^2\pi\vec{\theta}\cdot\vec{s}/\lambda \tag{116}$$

This is the same equation we derived before with the aid of the interference law for partially coherent light, but now in a two-dimensional setting with $\vec{\theta}$ replacing y/s and $\vec{s}$ replacing the pinhole distance a in equation (110). We have full constructive interference:

$$I \,=\, 4I_0 \text{ for } \frac{\vec{\theta}\cdot\vec{s}}{\lambda} \,=\, n\,(=\,0,\pm1,\pm2,\ldots) \;\rightarrow\; |\vec{\theta}| \,=\, \frac{n\lambda}{|\vec{s}|\cos\phi} \tag{117}$$

and full destructive interference:

$$I \,=\, 0 \text{ for } \frac{\vec{\theta}\cdot\vec{s}}{\lambda} \,=\, (n+\frac{1}{2})\,(=\,\pm\frac{1}{2},\pm\frac{3}{2},\pm\frac{5}{2},\ldots) \;\rightarrow\; |\vec{\theta}| \,=\, \frac{(n+\frac{1}{2})\lambda}{|\vec{s}|\cos\phi} \tag{118}$$

with $\cos\phi$ the angle between $\vec{\theta}$ and the baseline vector $\vec{s}$.
The PSF represents a *corrugated sheet* with its modulation along the direction of the baseline vector $\vec{s}$ and a periodicity $(\Delta\theta)_s = \lambda/|\vec{s}|$, i.e. a pattern of alternating bright and dark *stripes* orthogonal to the direction of the baseline vector $\vec{s}$.
In actuality, the apertures have a finite size and the diffracted light by the apertures is localized, so we have $d < S$ but the approximation by δ-functions for infinitesimal apertures no longer holds! The PSF has now to be derived from the amplitude of the diffracted field by FT of the pupil function given in expression 112:

$$\tilde{a}(\vec{\theta}) \,\Leftrightarrow\, \frac{\lambda}{R}\left[\Pi\left\{\frac{\lambda}{d}\left(\vec{\zeta} - \frac{\vec{s}}{2\lambda}\right)\right\} + \Pi\left\{\frac{\lambda}{d}\left(\vec{\zeta} + \frac{\vec{s}}{2\lambda}\right)\right\}\right] \tag{119}$$

Applying the shift and scaling theorems from Fourier theory, i.e. if $f(x) \Leftrightarrow F(s)$ then $f[a(x-b)] \Leftrightarrow \left(e^{-2\pi ibs}/a\right) F(s/a)$, we find for the amplitude of the diffracted field:

$$\tilde{a}(|\,\vec{\theta}\,|) \,=\, \left(\frac{\lambda}{R}\right)\left[\frac{1}{4}\pi\,(d/\lambda)^2\right]\left[\frac{2J_1(\pi|\,\vec{\theta}\,|d/\lambda)}{\pi|\,\vec{\theta}\,|d/\lambda}\right]\left(e^{-2\pi i\vec{\theta}\cdot\vec{s}/2\lambda} + e^{2\pi i\vec{\theta}\cdot\vec{s}/2\lambda}\right)$$

$$= \; 2\left(\frac{\lambda}{R}\right)\left[\frac{1}{4}\pi\,(d/\lambda)^2\right]\left[\frac{2J_1(|\vec{u}\,|)}{|\vec{u}\,|}\right]\cos\vec{u}\cdot\;\vec{s}/d \quad \text{with} \quad \vec{u}=\pi\,\vec{\theta}\,d/\lambda$$

$$\mathrm{PSF}=|\tilde{a}(|\vec{\theta}|)|^2 \;=\; 4\left(\frac{\lambda}{R}\right)^2\left[\frac{1}{4}\pi\,(d/\lambda)^2\right]^2\left[\frac{2J_1(|\vec{u}\,|)}{|\vec{u}\,|}\right]^2\cos^2\vec{u}\cdot\vec{s}/d \tag{120}$$

Again we have full constructive interference for:

$$\frac{\vec{\theta}\cdot\vec{s}}{\lambda} \;=\; n\,(=\;0,\pm1,\pm2,\ldots) \;\;\rightarrow\;\; |\vec{\theta}| \;=\; \frac{n\lambda}{|\vec{s}|\cos\phi} \tag{121}$$

and full destructive interference for:

$$\frac{\vec{\theta}\cdot\vec{s}}{\lambda} \;=\; (n+\frac{1}{2})\,(=\;\pm\frac{1}{2},\pm\frac{3}{2},\pm\frac{5}{2},\ldots) \;\;\rightarrow\;\; |\vec{\theta}| \;=\; \frac{(n+\frac{1}{2})\lambda}{|\vec{s}|\cos\phi} \tag{122}$$

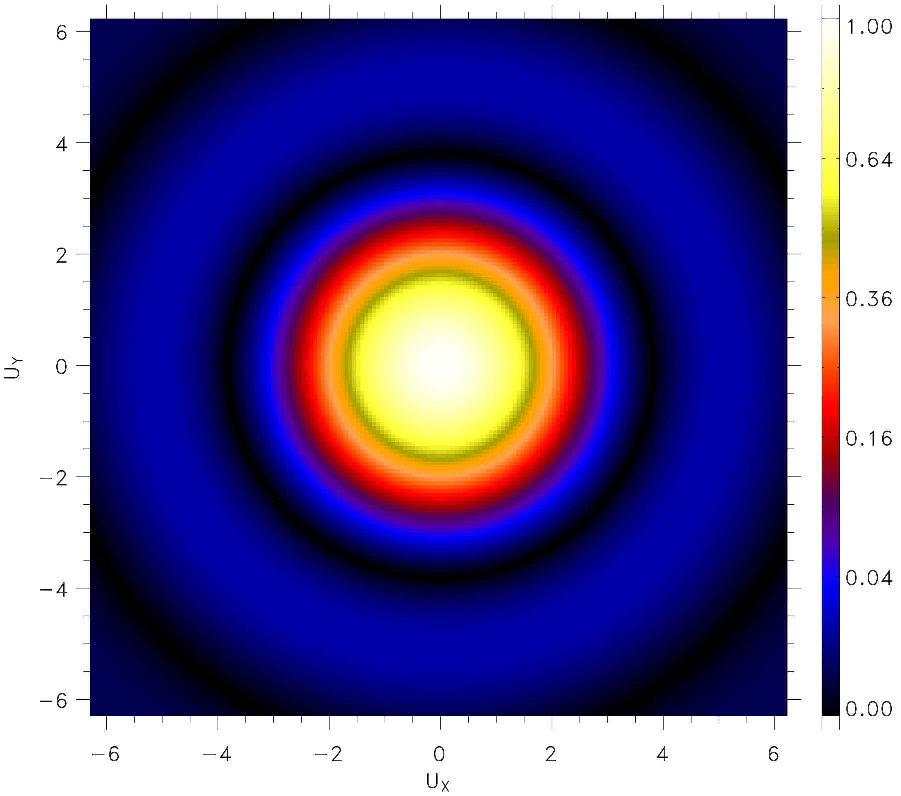

Figure 23: *PSF of a single circular aperture in pseudo-color as a function of the 2D-position vector $\vec{u}$ (λ-invariant display).*

with $\cos\phi$ the angle between $\vec{\theta}$ and the baseline vector $\vec{s}$.
The first two terms in the expression for the PSF give the normalisation for $|\vec{\theta}\,|=0$. The

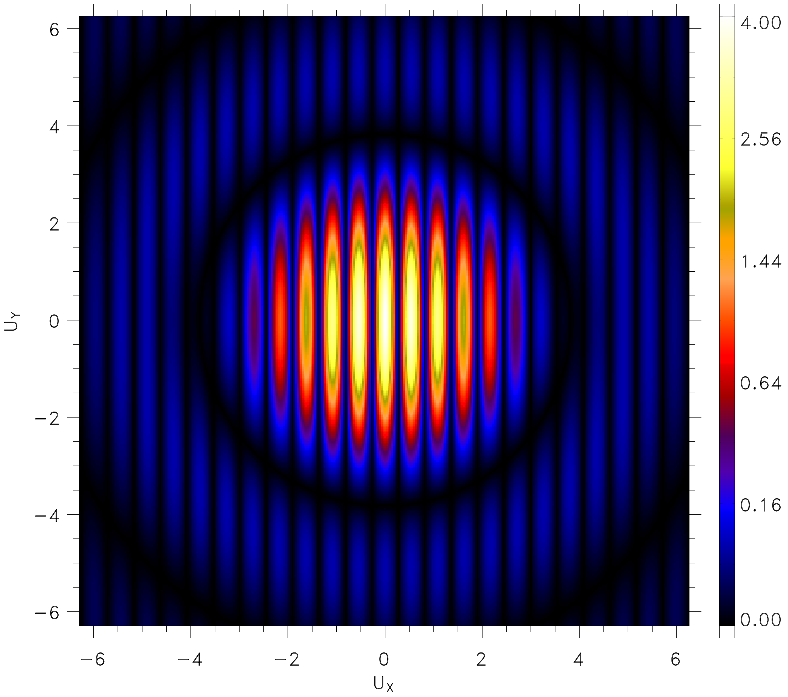

Figure 24: *PSF of a two-element interferometer in pseudo-color as a function of the 2D-position vector $\vec{u}$ (λ-invariant display). The aperture diameter d equals 25 meters, the length of the baseline vector $|\vec{s}|$ is chosen to be 144 meter.*

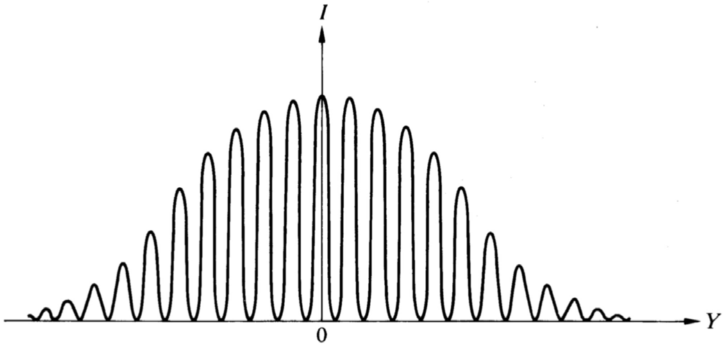

Figure 25: *Double beam interference fringes showing the modulation effect of diffraction by the aperture of a single pinhole. Figure taken from Hecht 1987.*

other terms represent a corrugated 2-dimensional Airy brightness distribution, intensity-modulated along the direction of the baseline vector $\vec{s}$ with periodicity $(\Delta\theta)_s = \lambda/|\vec{s}|$, i.e. a pattern of alternating bright and dark *annuli at a pitch determined by* $(\Delta\theta)_d = 1.220\lambda/d$, $2.233\lambda/d$, $3.238\lambda/d, ...$ *of the individual telescope mirrors as shown in figure 23 in pseudo-color* superimposed by periodic drop-outs in brightness orthogonal to the

direction of the baseline vector $\vec{s}$ *at a pitch determined by* $(\Delta\theta)_s = \lambda/|\vec{s}|$. This corrugated 2-dimensional Airy brightness distribution is also displayed (λ-invariant) in pseudo-color code as a function of the 2-D position vector $\vec{u}$ in figure 24.
A typical one-dimensional cross-section along $u_y = 0$ of the central part of the interferogram 24 is sketched in figure 25. Note that the visibilities in both figure 22 and in figure 25 are equal to one, because $I_{min} = 0$.
It can be shown that $|\tilde{\gamma}_{12}(\tau)|$ equals one for all values of τ and any pair of spatial points, if and only if the radiation field is *strictly monochromatic*, in practice such a situation is obviously *unattainable* ! Furthermore, a non-zero radiation field for which $|\tilde{\gamma}_{12}(\tau)| = 0$ for all values of τ and any pair of spatial points cannot exist in free space either.

7.2 Quasi-monochromatic extended source: spatial or lateral coherence

Spatial coherence (also: lateral coherence or lateral correlation) has to do with the spatial extent of the radiation source.

Problem: Prove that for $\tau \ll \tau_c$:

$$\tilde{\gamma}_{12}(\tau) = \tilde{\gamma}_{12}(0)e^{2\pi i\nu_0\tau} \tag{123}$$

with $|\tilde{\gamma}_{12}(\tau)| = |\tilde{\gamma}_{12}(0)|$ and a *fixed phase difference* $\alpha_{12}(\tau) = 2\pi\nu_0\tau$, ν_0 represents the average frequency of the wave carrier.
In the following treatment of spatial coherence, it is implicitly assumed that the frequency bandwidth of the radiation source is suffiently narrow that the comparison between two points with respect to spatial coherence occurs at times differing by $\triangle t \ll \tau_c$.

Query What is the quantitative relation between the *brightness distribution* of the spatially extended radiation source and the resulting *phase correlation* between two positions in the radiation field?

Approach Consider again Young's experiment for the case that the radiation source S is extended and illuminates the pinholes S_1 and S_2 (actually shown in figure 18). In the observers plane Σ, the interference is given by the expectation value of the product $\tilde{E}_1(t)\tilde{E}_2^*(t) = \mathbf{E}\{\tilde{E}_1(t)\tilde{E}_2^*(t)\} = \tilde{\Gamma}_{12}(0)$ with the subscripts 1 and 2 referring to the positions P_1 and P_2 in the Σ-plane. If $\tilde{E}_1$ and $\tilde{E}_2$ are uncorrelated, then $|\tilde{\Gamma}_{12}(0)| = 0$. In the case of full correlation $|\tilde{\gamma}_{12}| \left(= \frac{|\tilde{\Gamma}_{12}(0)|}{\sqrt{I_1 I_2}}\right) = 1$, for partial correlation we have $0 < |\tilde{\gamma}_{12}(0)| < 1$.
The extended source in Young's experiment is a collection of non-coherent infinitesimal radiators, this obviously reduces the contrast in the interferogram. This contrast can be observed and is described by the afore mentioned *Visibility function V*:

$$V = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} = |\tilde{\gamma}_{12}(0)| \tag{124}$$

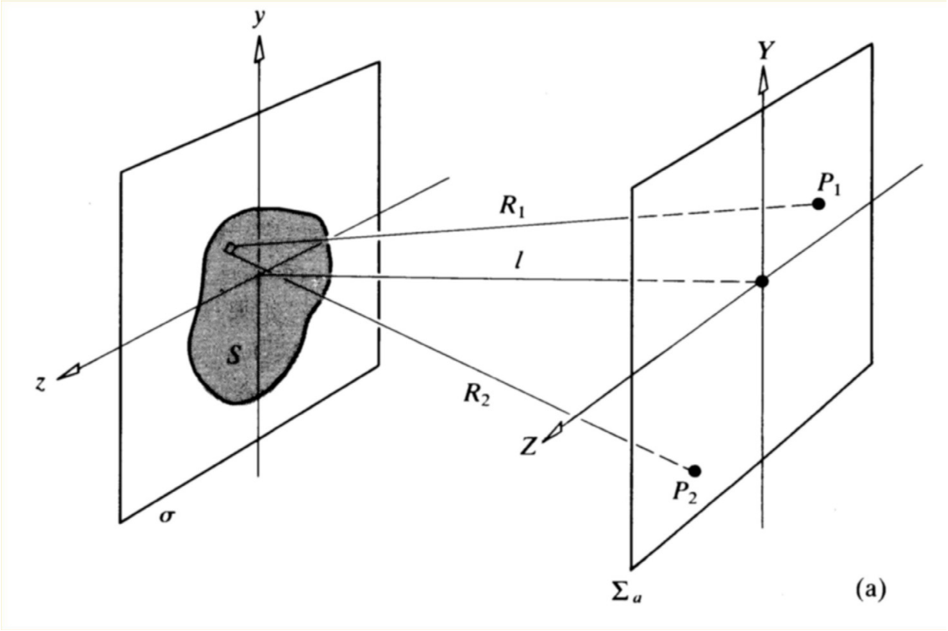

Figure 26: *Relating $\tilde{\gamma}_{12}(0)$ to the brightness distribution of an extended radiation source S: configuration for demonstrating the Van Cittert-Zernike theorem. Figure taken from Hecht 1987.*

7.3 The Van Cittert-Zernike theorem

So how can we now relate $\gamma_{12}(0)$ (or $\Gamma_{12}(0)$) to the brightness distribution of the extended radiation source S?
This can be done in the following way (see figure 26). Locate S, a QM-incoherent source, in a plane σ, with an intensity distribution $I(y, z)$. Consider next the observation plane Σ parallel to σ, l is perpendicular to both planes (coincident with the X-axis) connecting the centre of the extended source $(y = 0, z = 0)$ to the zero reference in Σ $(Y = 0, Z = 0)$. Select two positions P_1 and P_2. The objective is to describe the value of $\gamma_{12}(0)$ in this plane, i.e. the coherence of the radiation field in Σ. Consider furthermore a small (infinitesimal) radiation element dS in the source at distances R_1 and R_2 from P_1 and P_2 respectivily. Suppose now that S is *not a source* but *an aperture* of identical size and shape, and suppose that $I(y, z)$ is not a description of the irradiance (or intensity distribution) but, instead, its functional form corresponds to the *field distribution* across that aperture. In other words imagine that there is a transparancy at the aperture with amplitude transmission characteristics that correspond functionally to the irradiance distribution $I(y, z)$. Furthermore, imagine that the aperture is illuminated by a spherical wave converging towards the fixed point P_2, so that a diffraction pattern will result centered at P_2. This diffracted field distribution, normalised to unity at P_2, is everywhere (e.g. at P_1) equal to the value of $\gamma_{12}(0)$ at that point. This is the *Van Cittert-Zernike theorem.*
In the limit that R_1 and R_2 are much larger than the source diameter and the relevant part of the Σ-plane we have the equivalent of Fraunhofer diffraction, this condition is practically always satisfied for astronomical observations. In that case, the van Cittert-

Zernike theorem can be expressed mathematically as:

$$\tilde{\Gamma}(\vec{r}) = \int\int_{\text{source}} I(\vec{\Omega}) e^{\frac{2\pi i \vec{\Omega}.\vec{r}}{\lambda}} d\vec{\Omega} \tag{125}$$

$$I(\vec{\Omega}) = \lambda^{-2} \int\int_{\Sigma\text{-plane}} \tilde{\Gamma}(\vec{r}) e^{-\frac{2\pi i \vec{\Omega}.\vec{r}}{\lambda}} d\vec{r} \tag{126}$$

$I(\vec{\Omega})$ is the intensity distribution of the extended radiation source as a function of a unit direction vector $\vec{\Omega}$ as seen from the observation plane Σ. Taking the centre of the extended radiation source S as the zero-reference for $\vec{\Omega}$ (coincident with the central axis l in figure (26)) and assuming a relativily small angular extent of the source we can write $I(\vec{\Omega}) = I(\theta_y, \theta_z)$ and $d\vec{\Omega} = d\theta_y d\theta_z$, where θ_y and θ_z represent two orthogonal angular coordinate axes across the source starting from the zero-reference of $\vec{\Omega}$.
$\tilde{\Gamma}(\vec{r})$ is the coherence function in the Σ-plane, the vector $\vec{r}$ represents an arbitrary baseline $\vec{r}(X,Y)$ in this plane with $d\vec{r} = dYdZ$ (in the above example $\overline{P_1P_2} = \vec{r}_{P_1} - \vec{r}_{P_2}$). Expressions (125) and (126) for $\tilde{\Gamma}(\vec{r})$ and $I(\vec{\Omega})$ show that they are linked through a Fourier transform, except for the scaling with the wavelength λ. This scaling might be perceived as a "true" Fourier transform with the *conjugate variables* $\vec{\Omega}$ and $\vec{r}/\lambda$, i.e. by expressing $\vec{r}$ in units of the wavelength λ, *writing the van Cittert-Zernike theorem as the Fourier pair*:

$$I(\vec{\Omega}) \quad \Leftrightarrow \quad \tilde{\Gamma}(\vec{r}/\lambda) \tag{127}$$

The complex *spatial* degree of coherence, $\tilde{\gamma}(\vec{r})$, follows from:

$$\tilde{\gamma}(\vec{r}) = \frac{\tilde{\Gamma}(\vec{r})}{\int\int_{\text{source}} I(\vec{\Omega}) d\vec{\Omega}} \tag{128}$$

i.e. by normalising on the total source intensity.
Note: Although the extended source S is spatially incoherent, there still exists a partially correlated radiation field at e.g. positions P_1 and P_2, since all individual source elements contribute to a specific location P in the Σ-plane.

For a derivation of the Van Cittert-Zernike relations, consider the geometry given in figure 27.
The observation plane Σ contains the baseline vector $\vec{r}(Y,Z)$ and is perpendicular to the vector pointing at the centre of the radiation source. The angular coordinates θ_y and θ_z across the source (see above) correspond to the linear coordinates of the unit direction vector $\vec{\Omega}(\Omega_X, \Omega_Y, \Omega_Z)$, i.e. the direction cosines of $\vec{\Omega}$ relative to the X, Y, Z coordinate system ($\Omega_X^2 + \Omega_Y^2 + \Omega_Z^2 = 1$). The spatial coherence of the EM-field between the two positions 1 (for convenience chosen in the origin) and 2 is the outcome of a correlator device that produces the output $\mathbf{E}\left\{\tilde{E}_1(t)\tilde{E}_2^*(t)\right\}$.
In reality positions 1 and 2 are not point like, they represent *radio antennae* or *optical reflectors*, we shall come back to this later. From the geometry displayed in figure 27, regarding the Van Cittert-Zernike relations, we can note the following:
$\Rightarrow$ If $I(\vec{\Omega}) = I_0\delta(\vec{\Omega})$, i.e. a point source on the X-axis, the Van Cittert-Zernike relation yields $|\tilde{\Gamma}(\vec{r})| = I_0$ and $|\tilde{\gamma}(\vec{r})| = 1$: a plane wave hits the full YZ-plane everywhere at

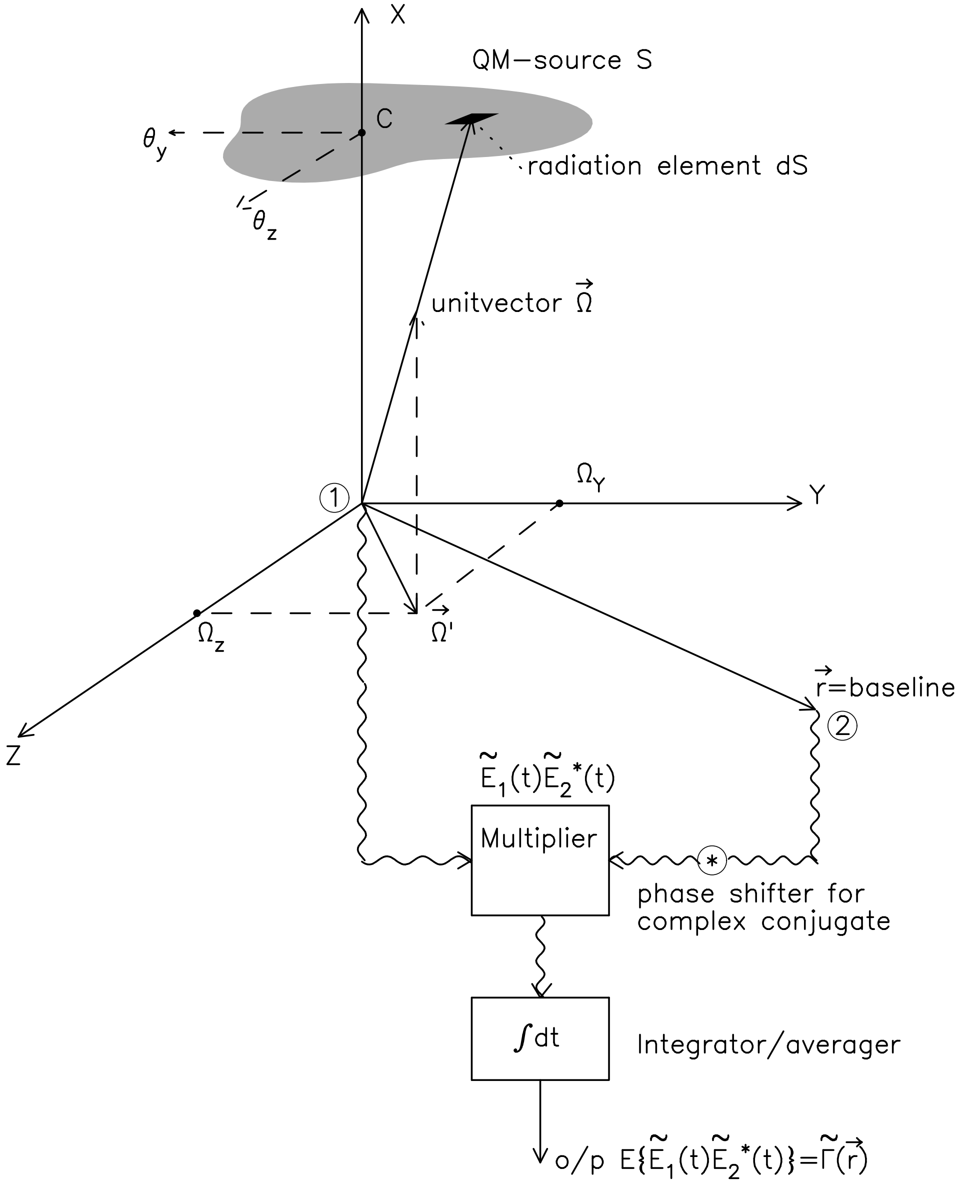

Figure 27: *Van Cittert Zernike relation: reference geometry.*

the same time, full coherence is preserved (by definition) on a plane wave front.
⇒ Next, let us consider an infinitesimal source element in the direction $\vec{\Omega}_0 \Longrightarrow I_0\delta(\vec{\Omega}-\vec{\Omega}_0)$. The projection of $\vec{\Omega}_0$ on the $\Sigma-$plane is $\vec{\Omega}'_0(\Omega_Y,\Omega_Z)$. There will now be a difference in path length between positions 1 and 2 given by the projection of $\vec{r}$ on $\vec{\Omega}_0$, i.e. $\vec{r}.\vec{\Omega}'_0 = \Omega_Y Y + \Omega_Z Z$. Then:

$$\tilde{E}_1(t) = \tilde{E}_0(t)e^{2\pi i\nu_0\left(t+\frac{\vec{\Omega}_0\cdot\vec{r}}{c}\right)} = \tilde{E}_0(t)e^{\left(2\pi i\nu_0 t+\frac{2\pi i\vec{\Omega}_0\cdot\vec{r}}{\lambda}\right)} \tag{129}$$

$$\tilde{E}_2^*(t) = \tilde{E}_0(t)e^{-2\pi i\nu_0 t} \tag{130}$$

Therefore:

$$\mathbf{E}\left\{\tilde{E}_1(t)\tilde{E}_2^*(t)\right\} = \mathbf{E}\left\{|\tilde{E}_0(t)|^2\right\}e^{\frac{2\pi i\vec{\Omega}_0\cdot\vec{r}}{\lambda}} = I_0(\vec{\Omega}_0)e^{\frac{2\pi i\vec{\Omega}_0\cdot\vec{r}}{\lambda}} \tag{131}$$

Integration over the full source extent(straight forward integration, since all source elements are spatially uncorrelated) yields:

$$\tilde{\Gamma}(\vec{r}) = \int\int_{\text{source}} I_0(\vec{\Omega}) e^{2\pi i \vec{\Omega}.\vec{r}/\lambda} d\vec{\Omega} \tag{132}$$

$$\tilde{\gamma}(\vec{r}) = \frac{\int\int_{\text{source}} I_0(\vec{\Omega}) e^{2\pi i \vec{\Omega}.\vec{r}/\lambda} d\vec{\Omega}}{\int\int_{\text{source}} I_0(\vec{\Omega}) d\vec{\Omega}} \tag{133}$$

The meaning of this relationship is, in physical terms, that $\tilde{\Gamma}(\vec{r})$ at a certain point represents a *single* Fourier component (with baseline $\vec{r}$) of the intensity distribution of the source with strength $\tilde{\Gamma}(\vec{r})d\vec{r}$. A short baseline (small $|\vec{r}|$) corresponds to a low frequency (*spatial frequency!*) component in the brightness distribution $I(\theta_y, \theta_z)$, i.e. *coarse* structure, large values of $|\vec{r}|$ correspond to *fine* structure in $I(\theta_y, \theta_z)$. The *diffraction limited* resolution in *aperture synthesis* corresponds to:

$$|\vec{r}_{max}| = L_{max} \Longrightarrow \theta_{min} = \frac{\lambda}{2L_{max}} \tag{134}$$

The factor 2 in the denominator of the expression for θ_{min} follows from the rotation symmetry in aperture synthesis.

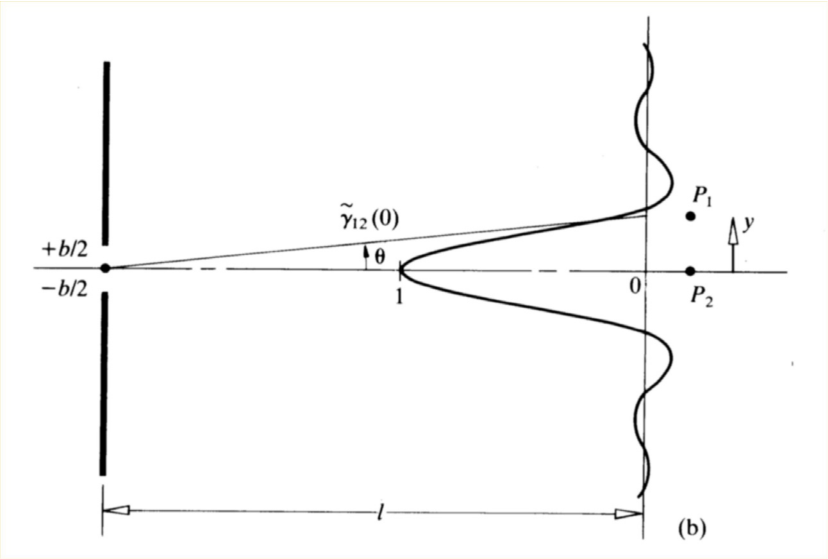

Figure 28: *The coherence function $\tilde{\gamma}_{12}(0)$ for a uniform slit source. Figure taken from Hecht 1987.*

Example Consider a one dimensional case. This can be done by taking the slit source of uniform intensity shown in figure 28 , slit width b and running coordinate ξ, the observation plane Σ, running coordinate y, is located at large distance l from the slit source (i.e. the Fraunhofer limit is applicable). The source function can be expressed as the window function $\Pi\left(\frac{\xi}{b}\right)$, in angular equivalent $\Pi\left(\frac{\beta}{\beta_0}\right)$, with $\beta_0 = b/l$.

Application of the Van Cittert-Zernike theorem $I(\vec{\Omega}) \Leftrightarrow \tilde{\Gamma}(\vec{r}/\lambda)$ yields:

$$\Pi\left(\frac{\beta}{\beta_0}\right) \Leftrightarrow \beta_0 \text{sinc}\left(\frac{y\beta_0}{\lambda}\right) = \beta_0 \text{sinc}\left(\frac{yb}{\lambda l}\right) \tag{135}$$

with $\text{sinc}(x) = \left(\frac{\sin \pi x}{\pi x}\right)$. The modulus of the normalised complex coherence function becomes:

$$|\tilde{\gamma}(y)| = \left|\frac{\beta_0 \text{sinc}\frac{yb}{\lambda l}}{\beta_0}\right| = \left|\text{sinc}\frac{yb}{\lambda l}\right| = V \Rightarrow \text{Visibility} \tag{136}$$

Note that:

- *Enlarging* b with a factor 2, *shrinks* the coherence function with the same factor. This is of course a direct consequence of the scaling law under Fourier transform.

- The width of the coherence function follows from: $\left(\frac{y\beta_0}{\lambda}\right) \approx 1 \Rightarrow y = \left(\frac{\lambda}{\beta_0}\right)$. If the radiation source exhibitis a smooth brightness distribution over the angle $\beta_0 = \Delta$ radians, as is the case with the slit source, then $\gamma(y)$ also displays a smooth distribution over a distance of λ/Δ meters.

- If the brightness structure of a radiation source covers a wide range of angular scales, say from a largest angular scale Δ to a smallest angular scale δ (in radians), then the spatial coherence function shows a finest detail of λ/Δ and a maximum extent of $\approx \lambda/\delta$ in meters.

7.4 Etendue of coherence

Consider the two-dimensional case of a circular source of uniform intensity with an angular diameter θ_s, the source brightness distribution can then be described as a circular two-dimensional window function: $I(\vec{\Omega}) = \Pi\left(\frac{\theta}{\theta_s}\right)$. To compute the complex degree of coherence in the observation plane Σ take again two positions, position 1 in the centre (origin, as before) and position 2 at a distance ρ from this centre point. Applying the van Cittert-Zernike theorem, we find for $\tilde{\Gamma}(\rho)$:

$$\Pi\left(\frac{\theta}{\theta_s}\right) \Leftrightarrow \tilde{\Gamma}(\rho/\lambda) = \frac{(\theta_s/2)J_1(\pi\theta_s\rho/\lambda)}{\rho/\lambda} \tag{137}$$

where J_1 represents the Bessel function of the first kind. Normalisation to the source brightness, through division by $(\pi\theta_s^2)/4$, yields the expression for the complex degree of coherence:

$$\tilde{\gamma}(\rho) = \frac{2J_1(\pi\theta_s\rho/\lambda)}{\pi\theta_s\rho/\lambda} \tag{138}$$

The modulus of the complex degree of coherence is therefore:

$$|\tilde{\gamma}(\rho)| = \left|\frac{2J_1(u)}{u}\right| \tag{139}$$

with the argument of the Bessel function $u = \pi\theta_s\rho/\lambda$. We can derive the extent of the coherence in the observation plane Σ by evaluating $J_1(u)$. If we take $u = 2$, $|\tilde{\gamma}(\rho)| = J_1(2) = 0.577$, i.e. the coherence in Σ remains significant for $u \leq 2$, or $\rho \leq 2\lambda/(\pi\theta_s)$. The area S in Σ over which the coherence remains significant equals $\pi\rho^2 = 4\lambda^2/(\pi\theta_s^2)$. In this expression, $\pi\theta_s^2/4$ equals the solid angle Ω_{source} subtended by the radiation source. Significant coherence will thus exist if the following condition is satisfied:

$$\epsilon = S\Omega_{\text{source}} \leq \lambda^2 \tag{140}$$

The condition $\epsilon = S\Omega_{\text{source}} = \lambda^2$ is called the *Etendue of Coherence*, to be fulfilled if coherent detection is required!

Example Consider a red giant star, of radius $r_0 = 1.5$ x 10^{11} meter, at a distance of 10 parsec. For this object $\theta_s = 10^{-6}$ radians. If this object is observed at $\lambda = 0.5\mu$m, the value of the coherence radius ρ, on earth, on a screen normal to the incident beam is $\rho = 2\lambda/(\pi\theta_s) = 32$ cm. In the infrared, at $\lambda = 25\mu$m, the radius ρ is increased fifty fold to ≈ 15 meter. In the radio domain, say at $\lambda = 6$ cm, $\rho \approx 35$ km.

In general, *good coherence* means a Visibility of 0.88 or better. For a uniform circular source this occurs for $u = 1$, that is when $\rho = 0.32\lambda/\theta$. If we consider a narrow-bandwidth uniform radiation source at a distance R away, we have

$$\rho = 0.32(\lambda R)/D \tag{141}$$

This expression is very convenient to quickly estimate the required physical parameters in an interference or diffraction experiment. For example, if we put a red filter over a 1-mm-diameter disk-shaped flashlight source and stand back 20 meters from it, then $\rho = 3.8$ mm, where the mean wavelength is taken at 600 nm. This means that a set of apertures spaced at about 4 mm or less should produce clear fringes. Evidently the area of coherence increases with the distance R, this is why one can always find a distant bright street light to use as a convenient source.

Important: Remember that throughout the treatment of spatial coherence it was assumed that the comparison between the two points occurs at times differing by a $\triangle t \ll \tau_c$. If this condition is not fulfilled, for example because the frequency bandwidth of the radiation source is too large, interferometric measurements will not be possible (see section on temporal coherence). Frequency filtering will then be required to reduce the bandwidth of the source signal, i.e. make it more quasi-monochromatic.

8 Aperture synthesis

8.1 Telescope elements of finite size

As already stated, the positions 1 and 2 in the observation plane Σ are in practise not pointlike, but encompass a telescope element of finite size, say a radio dish of 25 meter diameter. This dish has then a diffraction sized beam of $\lambda/D(= 25 \text{ meter})$. In that case the Van Cittert-Zernike relation needs to be "weighted" with the telescope element (single dish) transfer function $H(\vec{\Omega})$. For a circular dish antenna $H(\vec{\Omega})$ is almost the *Airy brightness function*, well known from the diffraction of a circular aperture. The Van Cittert-Zernike relations now become:

$$\tilde{\Gamma}'(\vec{r}) = \int\int_{\text{source}} I(\vec{\Omega})H(\vec{\Omega})e^{\frac{2\pi i\vec{\Omega}.\vec{r}}{\lambda}}d\vec{\Omega} \tag{142}$$

$$I(\vec{\Omega})H(\vec{\Omega}) = \lambda^{-2}\int\int_{\Sigma\text{-plane}} \tilde{\Gamma}'(\vec{r})e^{-\frac{2\pi i\vec{\Omega}.\vec{r}}{\lambda}}d\vec{r} \tag{143}$$

The field of view scales with λ/D, e.g. if λ decreases the synthesis resolution improves but the field of view reduces proportionally!
So, in aperture synthesis the incoming beams from antenna dish 1 and antenna dish 2 are fed into a *correlator (multiplier)* that produces as output the product $\tilde{E}_1(t)\tilde{E}_2^*(t)$. This output is subsequently fed into an *integrator/averager* that produces the expectation value $\mathbf{E}\left\{\tilde{E}_1(t)\tilde{E}_2^*(t)\right\} = \tilde{\Gamma}'(\vec{r})$. By applying the Fourier transform given in (143), and correcting for the beam profile of the single dish $H(\vec{\Omega})$, the source brightness distribution $I(\vec{\Omega})$ can be reconstructed.
Important: Indirect imaging with an aperture synthesis system is limited to measuring image details within the *single pixel* defined by the beam profile of an individual telescope element, i.e. a single dish!

8.2 Quasi-monochromatic point source: spatial response function (PSF) and optical tranfer function (OTF) of a multi-element interferometer

The pupil function of a linear array comprising N circular apertures with diameter d, aligned along a baseline direction $\vec{b}$ (unit vector) and equally spaced at a distance $|\vec{s}| = \vec{b}\cdot\vec{s}$ can be written as:

$$\begin{aligned} P(\vec{\zeta}) = &\left[\Pi\left(\frac{\lambda\vec{\zeta}}{d}\right) + \Pi\left\{\frac{\lambda}{d}\left(\vec{\zeta} - \frac{\vec{s}}{\lambda}\right)\right\} + \Pi\left\{\frac{\lambda}{d}\left(\vec{\zeta} - 2\cdot\frac{\vec{s}}{\lambda}\right)\right\} + \ldots\right. \\ &\left. \ldots + \Pi\left\{\frac{\lambda}{d}\left(\vec{\zeta} - (N-1)\cdot\frac{\vec{s}}{\lambda}\right)\right\} = \sum_{n=0}^{N-1}\Pi\left\{\frac{\lambda}{d}\left(\vec{\zeta} - n\cdot\frac{\vec{s}}{\lambda}\right)\right\}\right] \end{aligned} \tag{144}$$

Applying the shift and scaling theorems from Fourier theory, i.e. if $f(x) \Leftrightarrow F(s)$ then $f[a(x-b)] \Leftrightarrow \left(e^{-2\pi ibs}/a\right)F(s/a)$, we find for the amplitude of the diffracted field:

$$\begin{aligned}\tilde{a}(|\vec{\theta}\,|) \;=\; & \left(\frac{\lambda}{R}\right)\left[\frac{1}{4}\pi(d/\lambda)^2\right]\left[\frac{2J_1(\pi|\,\vec{\theta}\,|d/\lambda)}{\pi|\,\vec{\theta}\,|d/\lambda}\right]\cdot\Big[1 \,+\, e^{-i(2\pi\vec{\theta}\cdot\vec{s}/\lambda)} \,+\, \left(e^{-i(2\pi\vec{\theta}\cdot\vec{s}/\lambda)}\right)^2 \,+\, \ldots \\ & \qquad\qquad\qquad\qquad\qquad\qquad \ldots + \,\left(e^{-i(2\pi\vec{\theta}\cdot\vec{s}/\lambda)}\right)^{N-1}\Big] \\ =\; & \left(\frac{\lambda}{R}\right)\left[\frac{1}{4}\pi(d/\lambda)^2\right]\left[\frac{2J_1(\pi|\,\vec{\theta}\,|d/\lambda)}{\pi|\,\vec{\theta}\,|d/\lambda}\right]\sum_{n=0}^{N-1}\left(e^{-i(2\pi\vec{\theta}\cdot\vec{s}/\lambda)}\right)^n \qquad (145)\end{aligned}$$

The sum of the *geometric series* of N complex exponentials, accomodating the accumulating phase shifts introduced by the addition of each subsequent telescope element equals:

$$\sum_{n=0}^{N-1}\left(e^{-i(2\pi\vec{\theta}\cdot\vec{s}/\lambda)}\right)^n \;=\; \frac{e^{-iN(2\pi\vec{\theta}\cdot\vec{s}/\lambda)}-1}{e^{-i(2\pi\vec{\theta}\cdot\vec{s}/\lambda)}-1}$$

$$=\; \frac{e^{-iN(2\pi\vec{\theta}\cdot\vec{s}/2\lambda)}\left(e^{-iN(2\pi\vec{\theta}\cdot\vec{s}/2\lambda)} \,-\, e^{iN(2\pi\vec{\theta}\cdot\vec{s}/2\lambda)}\right)}{e^{-i(2\pi\vec{\theta}\cdot\vec{s}/2\lambda)}\left(e^{-i(2\pi\vec{\theta}\cdot\vec{s}/2\lambda)} \,-\, e^{i(2\pi\vec{\theta}\cdot\vec{s}/2\lambda)}\right)} \;=\; e^{-i(N-1)\pi\vec{\theta}\cdot\vec{s}/\lambda}\left[\frac{\sin N(\pi\vec{\theta}\cdot\vec{s}/\lambda)}{\sin(\pi\vec{\theta}\cdot\vec{s}/\lambda)}\right] (146)$$

The PSF for the diffracted intensity distribution follows from the relation PSF $=\,\tilde{a}(|\vec{\theta}|)\cdot\tilde{a}^*(|\vec{\theta}|)$:

$$\mathrm{PSF} \;=\; \left(\frac{\lambda}{R}\right)^2\left[\frac{1}{4}\pi\,(d/\lambda)^2\right]^2\left[\frac{2J_1(|\vec{u}\,|)}{|\vec{u}\,|}\right]^2\frac{\sin^2 N(\vec{u}\cdot\vec{s}/d)}{\sin^2(\vec{u}\cdot\vec{s}/d)} \quad \text{with} \quad \vec{u}=\pi\vec{\theta}\,d/\lambda \qquad (147)$$

For $N\!=\!\mathbf{1}$, we recover the Airy brightness function for a single circular aperture, for $N\!=\!\mathbf{2}$ (Michelson) we have $\sin^2 N(\vec{u}\cdot\vec{s}/d)/\sin^2(\vec{u}\cdot\vec{s}/d) \;=\; [2\sin(\vec{u}\cdot\vec{s}/d)\cos(\vec{u}\cdot\vec{s}/d)]^2\,/\,\sin^2(\vec{u}\cdot\vec{s}/d) \;=\; 4\cos^2(\vec{u}\cdot\vec{s}/d)$, commensurate with expression (120).
For N apertures, **maximum** constructive interference occurs for $\sin N(\pi\vec{\theta}\cdot\vec{s}/\lambda)/\sin(\pi\vec{\theta}\cdot\vec{s}/\lambda) \;=\; N$, that is when:

$$\frac{\vec{\theta}\cdot\vec{s}}{\lambda} \;=\; n\,(=\;0,\pm1,\pm2,\ldots) \;\;\rightarrow\;\; |\vec{\theta}| \;=\; \frac{n\lambda}{|\vec{s}|\cos\phi} \qquad (148)$$

These so-called *principal maxima* are apparently found at the same $|\vec{\theta}|$-locations, regardless of the value of $N\geq 2$.
Minima, of **zero** flux density, exist whenever $\sin N(\pi\vec{\theta}\cdot\vec{s}/\lambda)/\sin(\pi\vec{\theta}\cdot\vec{s}/\lambda) \;=\; 0$, i.e. for:

$$\begin{aligned}\frac{\vec{\theta}\cdot\vec{s}}{\lambda} \;&=\; \pm\frac{1}{N},\pm\frac{2}{N},\pm\frac{3}{N},\ldots,\pm\frac{N-1}{N},\pm\frac{N+1}{N},\ldots \\ \Rightarrow\;\; |\vec{\theta}| \;&=\; \frac{n\lambda}{N|\vec{s}|\cos\phi}, \quad \text{for} \quad n=\pm1,\pm2,\ldots \quad \text{but} \quad n\neq kN\ (k=0,\pm1,\pm2,\ldots \quad (149)\end{aligned}$$

with $\cos\phi$ the angle between $\vec{\theta}$ and the baseline vector $\vec{s}$.
Between consecutive principal maxima there will therefore be **N-1 minima**. Since between each pair of minima there will have to be a **subsidiary maximum**, i.e. a

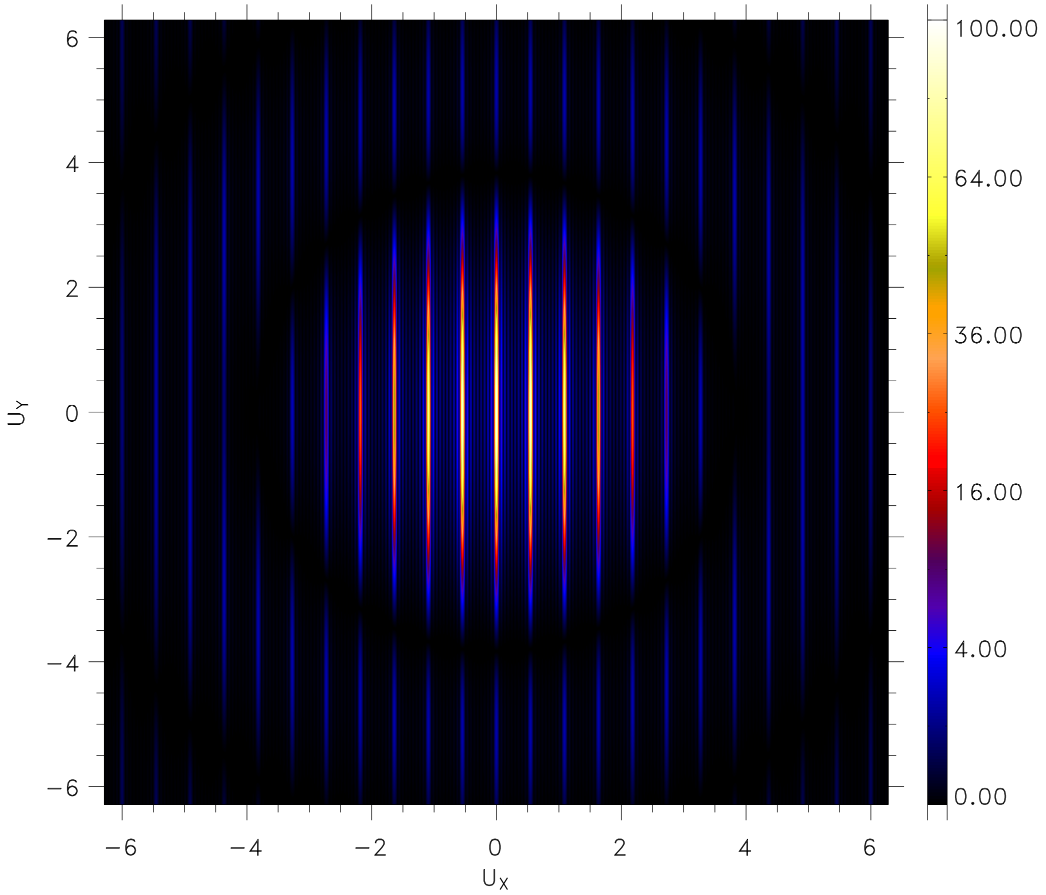

Figure 29: *PSF of a 10-element interferometer with circular apertures in pseudo-color as a function of the 2D-position vector $\vec{u}$ (λ-invariant display).*

total of **N-2 subsidiary maxima** between consecutive principal maxima. The first two terms in the expression for the PSF give the normalisation for $|\vec{\theta}\,| = 0$. The other terms represent a corrugated 2-dimensional Airy brightness distribution, intensity-modulated along the direction of the baseline vector $\vec{s}$ with a periodicity $(\Delta\theta)_s = \lambda/|\vec{s}|$ of narrow bright principal maxima and with a periodicity $(\Delta\theta)_{Ns} = \lambda/(N|\vec{s}|)$ of narrow weak subsidiary maxima, interleaved with zero-intensity minima.

The corrugated 2-dimensional Airy brightness distribution for $N = 10$, yet again for $d = 25$ meters and $|\vec{s}| = 144$ meters, is displayed (λ-invariant) in pseudo-color code as a function of the 2D-position vector $\vec{u}$ in figure 29. Also shown, in figure 30, is a magnification of the central part of the PSF that more clearly shows the interleaved subsidiary maxima and minima. A cross-section for $u_y = 0$ of the central part of the interferogram, delineating the profiles of the sharply peaked principal maxima and the adjacent series of subsidiary maxima and minima, is presented in figure 31.

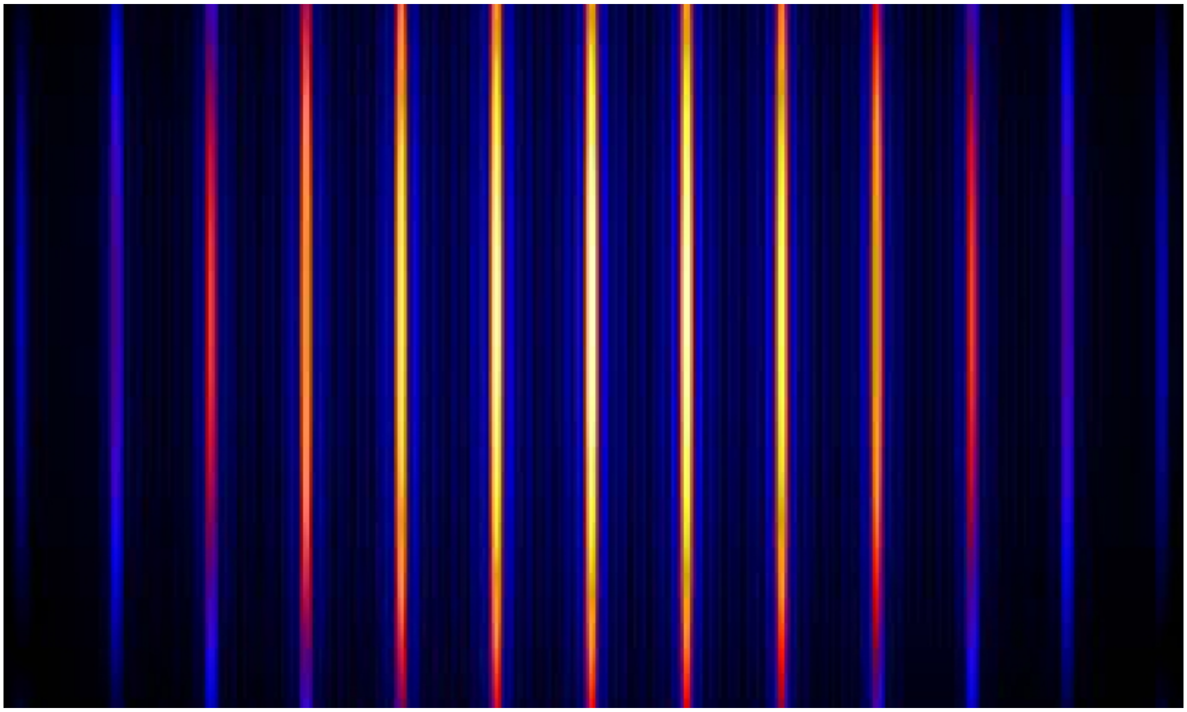

Figure 30: *Magnification of the central part of figure 29, clearly showing the locations of the subsidiary maxima and minima.*

The OTF for an array of N circular apertures can be obtained from the autocorrelation of the pupil function given in expression (144):

$$\begin{aligned} H_\lambda(\vec{\zeta}, N\vec{s}/\lambda) \, &= \, \left(\frac{\lambda}{R}\right)^2 \left[\sum_{n=0}^{N-1} \Pi\left\{\frac{\lambda}{d}\left(\vec{\zeta} - n \cdot \frac{\vec{s}}{\lambda}\right)\right\}\right] \, * \, \left[\sum_{m=0}^{N-1} \Pi\left\{\frac{\lambda}{d}\left(\vec{\zeta} - m \cdot \frac{\vec{s}}{\lambda}\right)\right\}\right] \\ &= \, \left(\frac{\lambda}{R}\right)^2 \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} A_{nm}(\vec{\zeta}, \, \vec{s}/\lambda) \end{aligned} \tag{150}$$

in which A_{nm} represents an element of the NxN autocorrelation matrix $\mathbf{A}$, that is given by:

$$A_{nm}(\vec{\zeta}, \vec{s}/\lambda) \, = \, \int\!\!\int_{pupil\,plane} \Pi\left\{\frac{\lambda}{d}\left(\vec{\zeta}' - n \cdot \frac{\vec{s}}{\lambda}\right)\right\} \Pi\left\{\frac{\lambda}{d}\left(\vec{\zeta}' - \vec{\zeta} - m \cdot \frac{\vec{s}}{\lambda}\right)\right\} d\vec{\zeta}' \tag{151}$$

Values $A_{nm} \neq 0$ are given by Chinese-hat functions as derived for a single circular aperture in section **3**. However in the multi aperture case here, we have a series of **principal maxima** in the $H_\lambda(\vec{\zeta}, \vec{s}/\lambda)$ plane (= the uv-plane representing 2-dimensional frequency(Fourier) space) at spatial frequency values:

$$\vec{\zeta}_{max} \, = \, \vec{\zeta} - k \cdot \frac{\vec{s}}{\lambda} \;\text{ with }\; k = n - m = 0, \pm 1, \pm 2, \ldots, \pm(N-2), \pm(N-1) \tag{152}$$

Hence we can replace A_{nm} by A_k, where the single index k refers to the *diagonals* of the autocorrelation matrix $\mathbf{A}$: $k = 0$ refers to the main diagonal, $k = \pm 1$ to the two

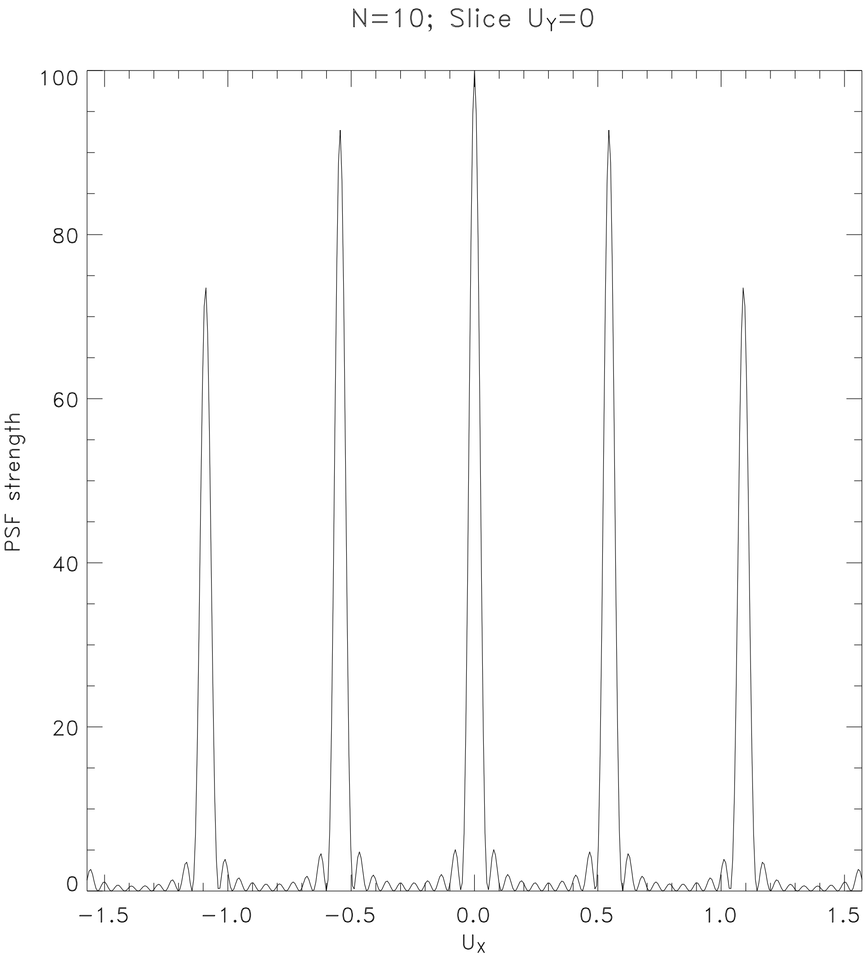

Figure 31: *Cross-section of the central part of the PSF for a 10-element interferometer, delineating the brightness profiles of the principal maxima and the subsidiary maxima and minima.*

diagonals contiguous to the main diagonal and so on. The analytical expression for the diagonal terms A_k can now be computed in the same way as for a single circular aperture, however in this case we need vector notation:

$$A_k = \frac{1}{2}\left(\frac{d}{\lambda}\right)^2 \cdot$$

$$\cdot \left[\arccos\left(\frac{\lambda}{d}\left|\vec{\zeta} - k\cdot\frac{\vec{s}}{\lambda}\right|\right) - \left(\frac{\lambda}{d}\left|\vec{\zeta} - k\cdot\frac{\vec{s}}{\lambda}\right|\right)\left(1 - \left(\frac{\lambda}{d}\left|\vec{\zeta} - k\cdot\frac{\vec{s}}{\lambda}\right|\right)^2\right)^{\frac{1}{2}}\right]\Pi\left(\frac{\lambda}{2d}\left|\vec{\zeta} - k\cdot\frac{\vec{s}}{\lambda}\right|\right) \quad (153)$$

which we rewrite in terms of Chinese-hat functions $\hat{C}_k(\vec{\zeta} - k\cdot\vec{s}/\lambda)$ normalised to unit aperture area:

$$A_k = \frac{1}{4}\pi\left(\frac{d}{\lambda}\right)^2 \hat{C}_k(\vec{\zeta} - k\cdot\vec{s}/\lambda) \quad (154)$$

The sum over all elements of matrix $\mathbf{A}$ can now be obtained from:

$$\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} A_{nm} \equiv A_{sa} \sum_{k=-(N-1)}^{N-1} (N-|k|)\hat{C}_k(\vec{\zeta}-k\cdot\vec{s}/\lambda) \tag{155}$$

$$\text{Sum check:} \quad \sum_{k=-(N-1)}^{N-1} (N-|k|) = N + 2\sum_{k=1}^{N-1}(N-k)$$

$$= N + \frac{1}{2}(N-1)\left[2(2N-2)-2(N-2)\right] = N + N^2 - N = N^2,$$

compliant with the required total number of matrix elements! The quantity $A_{sa} = \frac{1}{4}\pi(d/\lambda)^2$ represents the geometrical area of a single telescope element.
Hence, we arrive at the following expression for the OTF of the array:

$$H_\lambda(\vec{\zeta}, N\vec{s}/\lambda) = \left(\frac{\lambda}{R}\right)^2 NA_{sa} \sum_{k=-(N-1)}^{N-1} \frac{(N-|k|)}{N}\,\hat{C}_k(\vec{\zeta}-k\cdot\vec{s}/\lambda) \quad \text{with} \tag{156}$$

$$\hat{C}_k(\vec{\zeta}-k\cdot\vec{s}/\lambda) =$$

$$= \frac{2}{\pi}\left[\arccos\left(\frac{\lambda}{d}\left|\vec{\zeta}-k\cdot\frac{\vec{s}}{\lambda}\right|\right) - \left(\frac{\lambda}{d}\left|\vec{\zeta}-k\cdot\frac{\vec{s}}{\lambda}\right|\right)\left(1-\left(\frac{\lambda}{d}\left|\vec{\zeta}-k\cdot\frac{\vec{s}}{\lambda}\right|\right)^2\right)^{\frac{1}{2}}\right]\Pi\left(\frac{\lambda}{2d}\left|\vec{\zeta}-k\cdot\frac{\vec{s}}{\lambda}\right|\right) \tag{157}$$

The geometrical term NA_{sa} reflects the total geometrical collecting area of the telescope array (N times the area of a single aperture A_{sa}) and the term $(\lambda/R)^2$ accomodates the attenuation due to the spherical expansion of the wave field.
The OTF for the spatial frequency throughput of the aperture array is described by a linear array of discrete spatial frequency domains in the uv-plane, characterised by Chinese-hat functions centered at zero frequency and at multiples of the baseline vector $\vec{s}/\lambda$, that specifies the equidistance (magnitude and direction) between adjacent apertures. The peak transfer declines linearly with increasing mutual separation between aperture elements relative to the zero-frequency response, i.e. proportional to $(N-|k|)/N$ for spatial frequencies centered around $k\cdot\vec{s}/\lambda$. This can be easily explained by considering the monotonously decreasing number of aperture elements, and hence the associated array collecting power, involved at larger baseline values. For the maximum available baseline $(N-1)\vec{s}/\lambda$ this throughput reduction amounts to $1/N$.

Intermezzo: Verify that $H_\lambda(\vec{\zeta}, N\vec{s}/\lambda) \Leftrightarrow$ PSF(N-apertures).

Proof: Application of the *shift theorem* yields for the Fourier transform of the single-aperture-weighted ($A_{sa} = \frac{1}{4}\pi\,(d/\lambda)^2$) Chinese-hat function at baseline position $k\cdot\vec{s}/\lambda$ the following expression for its FT:

$$\frac{1}{4}\pi\,(d/\lambda)^2\,C_k(\vec{\zeta}-k\cdot\vec{s}/\lambda) \Leftrightarrow \left[\frac{1}{4}\pi\,(d/\lambda)^2\right]^2\left[\frac{2J_1(|\vec{u}\,|)}{|\vec{u}\,|}\right]^2 e^{-ik(2\pi\vec{\theta}\cdot\vec{s}/\lambda)} \tag{158}$$

Putting $x \equiv 2\pi\vec{\theta}\cdot\vec{s}/\lambda$, the Fourier transform of (156) can be expressed as:

$$\text{FT}\left[H_\lambda(\vec{\zeta}, N\vec{s}/\lambda)\right] = \left(\frac{\lambda}{R}\right)^2\left[\frac{1}{4}\pi\,(d/\lambda)^2\right]^2\left[\frac{2J_1(|\vec{u}\,|)}{|\vec{u}\,|}\right]^2 \cdot$$

$$\cdot \left[N\left(1+\sum_{k=1}^{N-1} e^{ikx} + \sum_{k=1}^{N-1} e^{-ikx}\right) - \left(\sum_{k=0}^{N-1} ke^{ikx} + \sum_{k=0}^{N-1} ke^{-ikx}\right)\right] \tag{159}$$

The first two sums of (159)involve simple *geometric* series and can be easily calculated, i.e.:

$$\begin{aligned} \sum_{k=1}^{N-1} e^{ikx} &= \frac{e^{ix}\left(e^{i(N-1)x}-1\right)}{e^{ix}-1} = \frac{e^{i(N-\frac{1}{2})x}-e^{ix/2}}{2i\sin\left(x/2\right)} \qquad \text{and:} \\ \sum_{k=1}^{N-1} e^{-ikx} &= \frac{e^{-ix}\left(e^{-i(N-1)x}-1\right)}{e^{-ix}-1} = -\left(\frac{e^{-i(N-\frac{1}{2})x}-e^{-ix/2}}{2i\sin\left(x/2\right)}\right) \end{aligned}$$

Hence, after some rearrangement of the complex exponentials, this results in a real function:

$$N\left(1+\sum_{k=1}^{N-1} e^{ikx} + \sum_{k=1}^{N-1} e^{-ikx}\right) = N\,\frac{\sin\left[(2N-1)(x/2)\right]}{\sin\left(x/2\right)} \tag{160}$$

The third and the fourth sum in equation (159) involve a combination of an *arithmetic* and a *geometric* series (a so-called *arithmetico-geometric* series). These sums can be simply derived by differentation of the expression for the sum of the geometric series, resulting in:

$$\begin{aligned} \sum_{k=0}^{N-1} ke^{ikx} &= \frac{Ne^{\,i(N-1)x}-(N-1)\,e^{\,iNx}-1}{4\sin^2(x/2)} \qquad \text{and:} \\ \sum_{k=0}^{N-1} ke^{-ikx} &= \frac{Ne^{\,-i(N-1)x}-(N-1)\,e^{\,-iNx}-1}{4\sin^2(x/2)} \end{aligned}$$

Straightforward recombination of the complex exponentials again results in a (trigonometric) real function:

$$\sum_{k=0}^{N-1} ke^{ikx} + \sum_{k=0}^{N-1} ke^{-ikx} = \frac{N\cos(N-1)x-(N-1)\cos Nx-1}{2\sin^2(x/2)} \tag{161}$$

Combining expressions (160) and (161), we arrive at :

$$\begin{aligned} &\left[N\left(1+\sum_{k=1}^{N-1} e^{ikx} + \sum_{k=1}^{N-1} e^{-ikx}\right) - \left(\sum_{k=0}^{N-1} ke^{ikx} + \sum_{k=0}^{N-1} ke^{-ikx}\right)\right] = \\ &= \frac{2N\sin(x/2)\sin(N-\frac{1}{2})x - N\cos(N-1)x+(N-1)\cos Nx+1}{2\sin^2(x/2)} = \\ &= \frac{N\sin x\sin Nx \,+\, N\cos x\cos Nx \,-\, \cos Nx \,-\, N\cos x\cos Nx \,-\, N\sin x\sin Nx \,+\, 1}{2\sin^2(x/2)} = \\ &= \frac{1-\cos Nx}{2\sin^2(x/2)} = \frac{2\sin^2(Nx/2)}{2\sin^2(x/2)} \end{aligned} \tag{162}$$

Substituting $x \equiv 2\pi\vec{\theta}\cdot\vec{s}/\lambda \equiv 2(\vec{u}\cdot\vec{s}/d)$, yields:

$$\mathrm{FT}\left[H_\lambda(\vec{\zeta},N\vec{s}/\lambda)\right] = \left(\frac{\lambda}{R}\right)^2 \left[\frac{1}{4}\pi\left(d/\lambda\right)^2\right]^2 \left[\frac{2J_1(|\vec{u}\,|)}{|\vec{u}\,|}\right]^2 \frac{\sin^2 N(\vec{u}\cdot\vec{s}/d)}{\sin^2(\vec{u}\cdot\vec{s}/d)} \qquad \text{Q.E.D.} \tag{163}$$

Note: The modulation term $\sin^2 N(\vec{u}\cdot\vec{s}/d)/\sin^2(\vec{u}\cdot\vec{s}/d)$ can be simplified for low N-values:

$$\begin{aligned} &= 1 && \text{for N=1} \\ &= 4\cos^2(\vec{u}\cdot\vec{s}/d) && \text{for N=2} \\ &= [2\cos 2(\vec{u}\cdot\vec{s}/d)+1]^2 && \text{for N=3} \\ &= 8\cos^2 2(\vec{u}\cdot\vec{s}/d)\cdot[1+\cos 2(\vec{u}\cdot\vec{s}/d)] && \text{for N=4} \end{aligned} \tag{164}$$

End Intermezzo

8.3 Earth Rotation Aperture Synthesis (ERAS)

By employing the rotation of the earth, the baseline vectors $k\cdot\vec{s}/\lambda$ of the linear N-element interferometer array, as defined in the previous section, will scan the YZ-plane displayed in figure (27) in case the X-axis is lined up with the North polar axis, i.e. in this particular geometry the X-coordinate of the baseline vectors is zero. The principal maxima or 'grating lobes' in the PSF of a multiple aperture array, as computed in the preceding section, will now manifest themselves as concentric annuli around a central source peak at angular distances $k\cdot\lambda/|\vec{s}|$. If the circular scans in the YZ-plane are too widely spaced, i.e. if $|\vec{s}|$ is larger than the single dish diameter, the (2-dimensional) Nyquist criterion is not respected and undersampling of the spatial frequency uv-plane (=YZ-plane) occurs. Consequently, the grating lobes will show up within the field of view defined by the single-dish beam profile. This can be avoided by decreasing the sampling distance $|\vec{s}|$. In the following section we shall now demonstrate these notions with the concrete configuration of the Westerbork Radio Synthesis Telescope.

8.4 The Westerbork Radio Synthesis Telescope (WSRT)

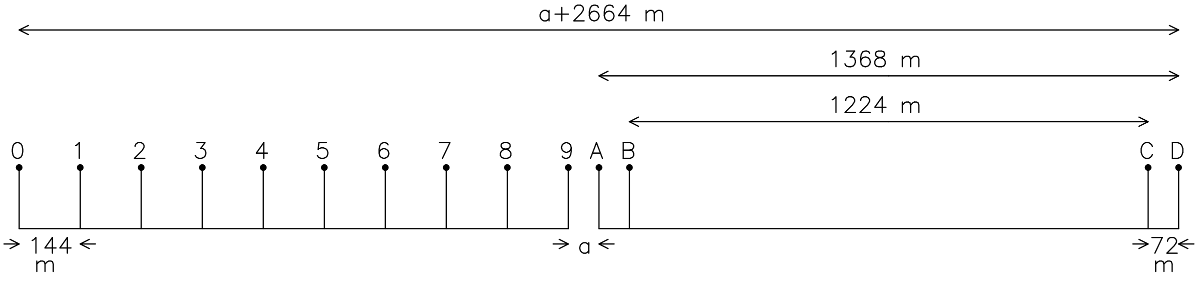

Figure 32: *Configuration of the Westerbork Radio Synthesis Telescope.*

The WSRT consists of 14 parabolic antenna's, with single dish diameters D of 25 meter. They are accurately lined up along the East-West direction with an overall length of $\approx$ 2750 meter. Ten antenna's (numbers 0 thru 9) are fixed with a mutual distance of 144 meter. The other four antenna's (A thru D) can be moved collectively with respect to the fixed array, without altering their mutual distance, i.e. the variable distance between antenna's 9 and A can be preselected as a suitable base line increment $\triangle L$

(Figure 32).
These 14 antenna's comprise 40 simultaneously operating interferometers. By employing the rotation of the earth, the full antenna array is rotated in a plane containing Westerbork perpendicular to the rotation axis of the earth. This is the YZ-plane introduced in figure 27 that contains all the base line vectors $\vec{r}$. In this particular geometry, the X component of the base line vectors is zero. This implies that in this case we are limited to sources near the North polar axis, all single dishes are thus pointed in that direction. In practise, the standard distance between antenna's 9 and A equals 72 meters. Consequently, after half a rotation of the earth the YZ-plane is covered with 38 concentric semi-circles with radii ranging from $L_{min} = 72$ meters to $L_{max} = 2736$ meters, with increments of $\triangle L = 72$ meters. The *WSRT-correlators* integrate and average over 10 seconds, this implies sampling of the concentric semi-circles every 1/24 degrees. After 12 hours, half the YZ-plane has been covered. The other half need not be covered, it can be found by mirroring the first half since $I(\vec{\Omega})$ is a real function.
The brightness distribution $I(\vec{\Omega})$ can now be reconstructed by Fourier inversion according to expression (126). Since we have only obtained samples of the spatial coherence function $\tilde{\Gamma}(\vec{r})$, the integral of expression (126) is replaced by a sum. Moreover, one normally applies in addition a *weighting function* to get a considerable reduction of the side lobes, this goes slightly at the expense of the ultimate angular resolution. This is expressed in terms of a degradation factor $\alpha \geq 1$. The simplest form of such a weighting function is a triangular shaped, circular symmetric, function (i.e. a cone), the attenuation effect on the side lobes is called *apodisation*. Leaving any constants aside for the moment , we obtain $\hat{I}(\vec{\Omega})$:

$$\hat{I}(\vec{\Omega}) = \sum_k w(\vec{r}_k)\tilde{\Gamma}(\vec{r}_k)e^{-\frac{2\pi i\vec{\Omega}.\vec{r}_k}{\lambda}} \tag{165}$$

with $w(\vec{r}_k)$ the weighting (apodisation) function.
Expression (165) yields a radio map on a discrete grid in $\vec{\Omega}$-space. The cell size of this grid (pitch) is chosen in such a way that oversampling of the spatial resolution of the array is achieved, so that *contour plots* can be constructed.
Note: The reconstructed $\hat{I}(\vec{\Omega})$ needs to be corrected for the single dish response function $H(\vec{\Omega})$, see equation (143).
If we consider half an earth rotation, the sum in (165) involves $\approx$ 165.000 numbers (i.e. 38 semi-circles, 12x60x6 correlator samples/semi-circle). This sum will have to be taken for roughly the same number of image points in $\vec{\Omega}$-space. This task is accomplished by using the Fast Fourier Transform algorithm.

The Point Spread Function (PSF):

Taking a point source S as the quasi monochromatic radiation source, the spatial frequency response function of the rotated interferometer array in the uv-plane can be obtained explicitly from expression (156) for the OTF by implementing the geometry of concentric scans. In that case, owing to the circular symmetry, the vectorial expression $H_\lambda(\vec{\zeta}\,, N\vec{s}/\lambda)$ can be replaced by a, radially symmetric, scalar expression $H_\lambda(p\,, N\triangle L/\lambda)$, with the scalar p the (λ - normalised) spatial frequency variable ($= |\,\vec{\zeta}\,|$) and $\triangle L$ the baseline increment of the array ($= |\,\vec{s}\,|$). Although an exact expression for the scalar function $H_\lambda(p\,, N\triangle L/\lambda)$ would involve integration along one coordinate of the

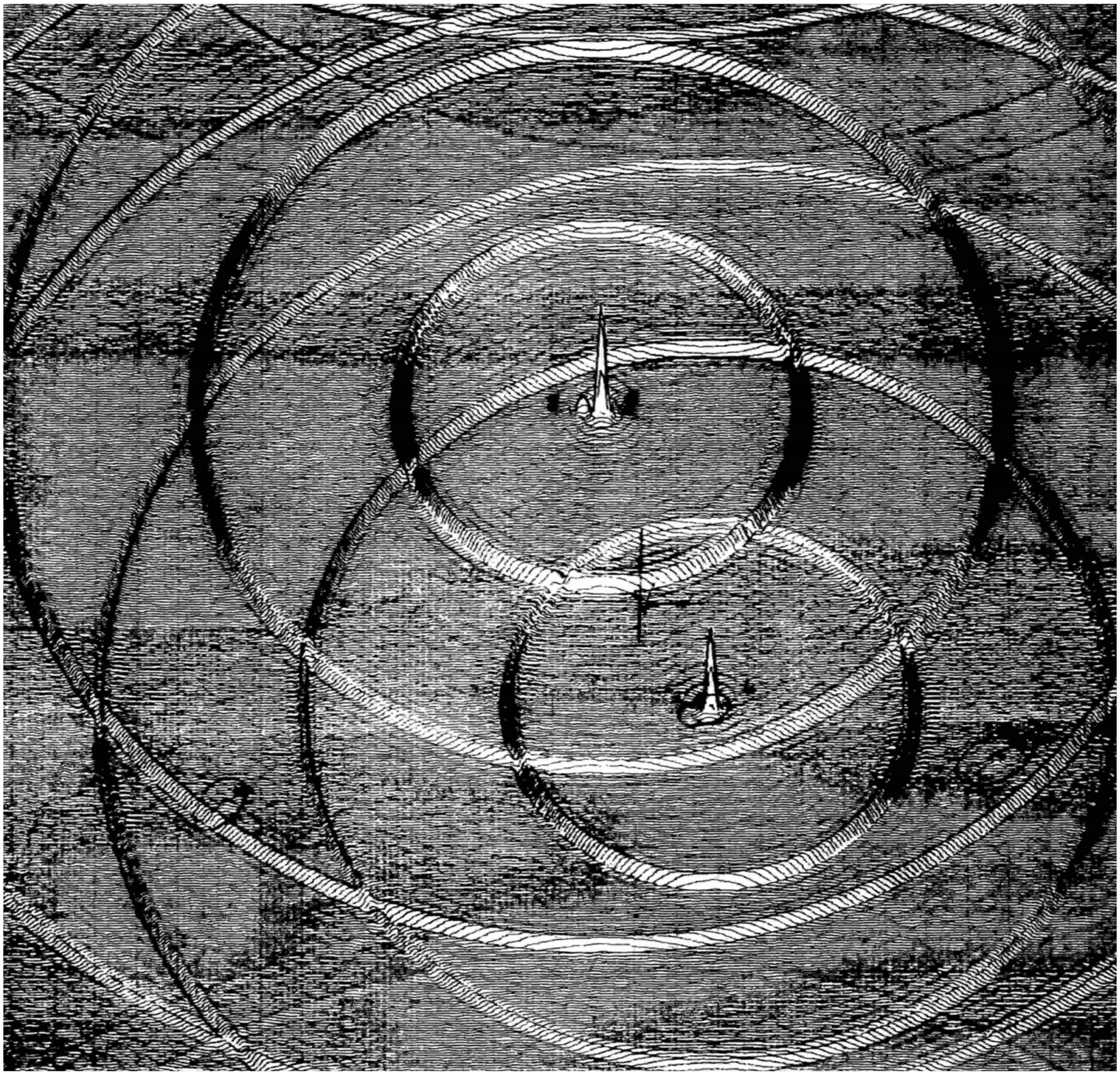

Figure 33: *A reconstructed "dirty" radio map showing central source peaks and grating lobes of increasing order.*

two-dimensional Chinese-hat function, a straightforward one-dimensional cross-section constitutes an adequate approximation. Hence we can write:

$$H_\lambda(p\,, N\triangle L/\lambda) \,=\, \left(\frac{\lambda}{R}\right)^2 NA_{sa} \sum_{k=-(N-1)}^{N-1} \frac{(N-|k|)}{N}\, \hat{C}_k(p - k\cdot\triangle L/\lambda) \qquad (166)$$

The PSF derives from the Fourier transform of $H_\lambda(p\,, N\triangle L/\lambda)$. Following the approach outlined in the *intermezzo* above, we arrive at:

$$PSF_{ERAS} \,=\, \left(\frac{\lambda}{R}\right)^2 \left[\frac{1}{4}\pi\,(d/\lambda)^2\right]^2 \left[\frac{2J_1(u)}{u}\right]^2 \frac{\sin^2 N(u\triangle L/D)}{\sin^2(u\triangle L/D)} \qquad (167)$$

where we have again utilised the reduced angular variable $u = \pi\theta D/\lambda$, θ represents the, radially symmetric, diffraction angle.

Evaluation of expression (167) reveals two main image components:
A Central Peak:
This distribution is rather similar to the *Airy brightness pattern*, with a typical breadth, i.e. spatial resolution:

$$\triangle\theta = \frac{\lambda}{2L_{max}} radians \tag{168}$$

with $2L_{max}$ the maximum diameter of the array in the YZ-plane. In the derivation of (167), no weighting function was applied. If a weighting function is applied for efficient apodisation, expression (168) needs to be multiplied by a degradation factor $\alpha = 1 - 1.5$ to accomodate the loss in ultimate angular resolution. Moreover, moving outward, the sidelobes in (167) will progressively be reduced depending on the particular choice of the weighting function.
Concentric Grating Lobes:
The angular distances of these annuli from the central peak follow from the location of the *principal maxima* given by the modulation term $\sin^2 N(u\triangle L/D)/\sin^2(u\triangle L/D)$ in expression (167). For an N-element array with increment $\triangle L$, these angular positions are given by:

$$\theta_{grating} = \frac{\lambda}{\triangle L}, 2\frac{\lambda}{\triangle L}, ..., (N-1)\frac{\lambda}{\triangle L} \tag{169}$$

A typical source field containing these grating lobes is shown in figure (33).
It is clear from this figure that severe undersampling of the YZ-plane has occurred since the grating lobes are well within the field of view. For the WSRT, one way to remedy these imperfections in the PSF is by decreasing the distance between antenna's 9 and A during the second half rotation of the earth. Combined with the first half rotation, a 36 meter increment coverage is achieved at the expense of doubling the observation time from 12 to 24 hours. In the same way, combining four half rotations in 48 hours, we can increase the coverage to 18 meter increments. Since the single dish diameter D equals 25 meter, no empty space is now left in the YZ-plane, the undersampling is corrected and the grating lobes have been moved outside the field of view defined by the single dish beam profile. This is summarized in the following tables that show exposure times, maximum baselines and increments, and at four different radio wavelengths single dish fields of view, central peak angular resolutions for $\alpha = 1.5$, and angular distances of the grating lobes.

Exposure(hours)	a (meter)	L_{min} (meter)	$\triangle L$ (meter)	L_{max} (meter)
1x12	72	72	72	2736
2x12	36,72	36	36	2736
4x12	36,54,72,90	36	18	2754

λ (cm)	6	21	49	92
f (MHz)	5000	1420	612	326
Single dish pixel λ/D (arcsec)	500	1800	4200	7600
Resolution $\alpha\lambda/(2L_{max})$ (arcsec)	3	11.5	27	50
Grating lobes $\lambda/(\triangle L = 72$ m) (arcsec)	172	7600	1405	2640
Grating lobes $\lambda/(\triangle L = 36$ m) (arcsec)	344	1205	2810	5270
Grating lobes $\lambda/(\triangle L = 18$ m) (arcsec)	688	2410	5620	10540

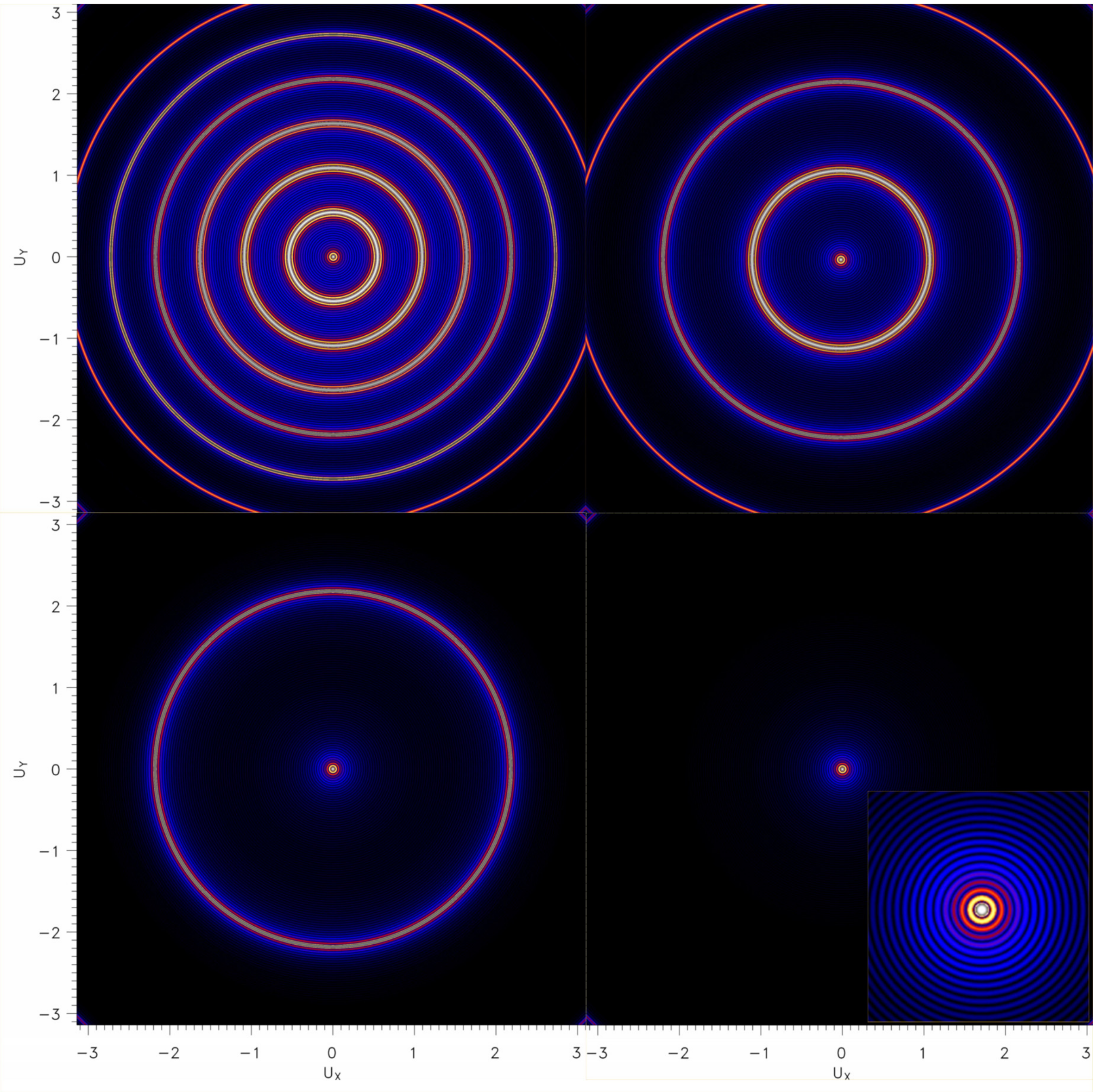

Figure 34: *Position of the concentric grating rings in the PSF for various sampling densities of the uv-plane. FOV:* $-\lambda/D \leq \theta_{x,y} \leq \lambda/D$ $(-\pi \leq u_{x,y} \leq \pi)$*. Upper left:* $\triangle L/D = 144/25$*, upper right:* $\triangle L/D = 72/25$*, lower left:* $\triangle L/D = 36/25$*, lower right:* $\triangle L/D = 18/25$*. The insert in the lower right panel shows a blow-up of the central peak image, displaying the circular subsidiary maxima and minima.*

As is clear from these tables, the grating lobes can only be moved outside the single dish pixel if the empty spaces between the samplings are filled in, the table shows that this is obviously the case for $\triangle L = 18$ m: for example the first order grating lobe at 6 centimeters is situated at an angular distance of 688 arcseconds from the centre, whereas a single dish pixel λ/D equals 500 arcseconds.
Figure (34) stipulates the outward movement of the grating lobes for a WSRT-type configuration by taking $D = 25$ meter and by reducing the baseline increment values

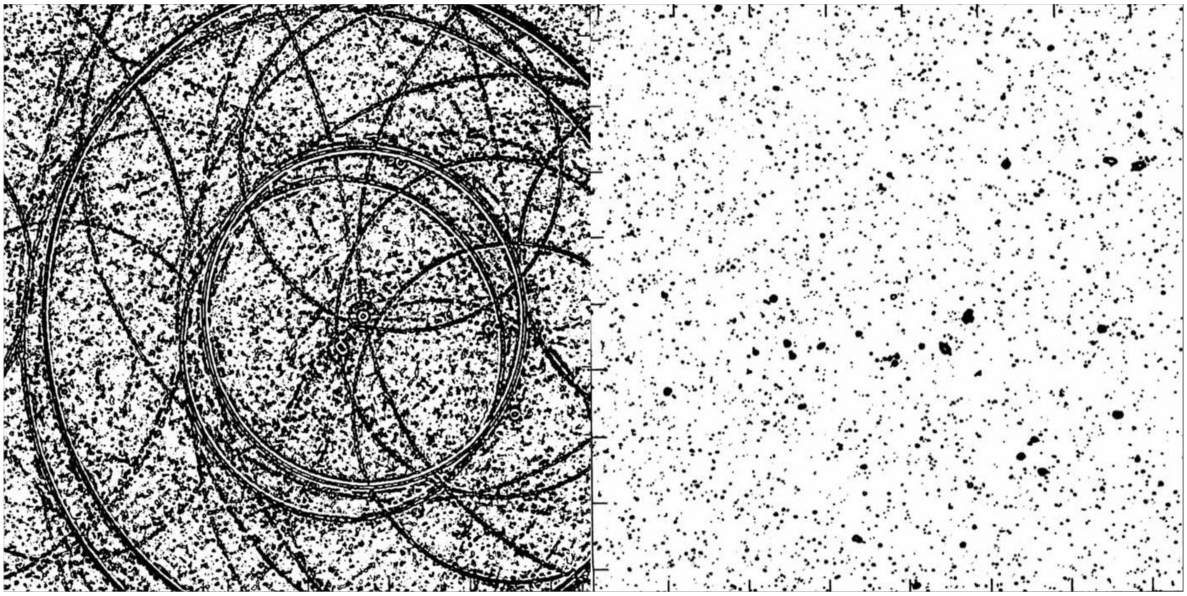

Figure 35: *Applying the CLEAN method for improving dirty radio maps. This is a radio map taken at $\lambda = 50$ cm, with a $\triangle L = 36$ meter increment. Left panel before CLEAN, right panel after CLEAN.*

from $\triangle L$=144 to 72, 36 and 18 meters respectively.
The problem of incomplete coverage of the YZ-plane is that the values of the coherence function $\tilde{\Gamma}(\vec{r})$ are set to zero in the empty spaces, which will certainly give an erroneous result. This can be circumvented by employing *interpolation* between the sampling circles based on certain assumptions regarding the complexity (or rather simplicity) of the brightness distribution of the sky region observed. Requirement is that the reconstructed (contour) map always remains consistent with the measurements at the grid points! Two mathematical techniques are often applied to get rid of the grating lobes: the CLEAN and MEM (Maximum Entropy Method) techniques. Obviously this is potentially much more efficient than elongating the observation times by a factor two or four, as was done in the above table. We shall not discuss these particular algorithms in any detail here, however the MEM technique will, in a more general fashion, be treated later on in this course. The improvement that can be achieved by applying the CLEAN method to a dirty radio map is displayed in figure (35), the left panel shows a dirty radio map at a wavelength of 50 cm, the right panel shows the "cleaned" map after applying the CLEAN process.

8.5 Maximum allowable bandwidth of a quasi-monochromatic source

As was already stated, the coherence length of the radiation source needs to be larger than the maximum path length difference at the longest baseline present in the interferometer array. This imposes a maximum frequency bandwidth for the observation of the radiation source, which is disadvantageous since the noise in the image decreases with $\sqrt{\triangle\nu T_{obs}}$. The largest angle of incidence equals half the field of view, i.e. $\lambda/2D$, with D the single dish diameter. At this angle, the coherence length compliant with the largest baseline needs to obey:

$$L_{coh} \gg \frac{\lambda}{2D} L_{max} \tag{170}$$

This translates in the following condition for the allowable frequency bandwidth:

$$\frac{\triangle\nu}{\nu_0} \ll \frac{2D}{L_{max}} \tag{171}$$

In the case of the WRST the ratio $2D/L_{max} \approx 1/50$. At a wavelength of 21 centimeters ($\approx$ 1400 MHz), this yields $\triangle\nu \ll 28$ MHz. This corresponds to a coherence length of more than 10 meters. In practise a value of $\triangle\nu \approx 10$ MHz is chosen, which corresponds to a coherence length of about 30 meters.

If one wishes to increase the bandwidth in order to improve sensitivity, than division in frequency subbands is required. For instance, if one observes at 21 centimeters and $\triangle\nu = 80$ MHz is required, this bandwidth is subdivided into eight 10 MHz subbands by filters. The individual subband maps are subsequently scaled (with λ) and added. This then yields the required signal to noise ratio.

8.6 The general case: arbitrary baselines and a field of view in arbitrary direction

In deriving the Van Cittert-Zernike relations, we took the baseline vector $\vec{r}(Y, Z)$ in a plane perpendicular to the vector pointing at the centre of the radiation source. In the WSRT case we were, therefore, limited to consider a field of view near the north polar direction.

Consider now an extended source in a plane σ in an arbitrary direction, an observation plane Σ parallel to σ, the centre of the radiation source is located on the X-axis, which is perpendicular to both planes. We designate the intersection point of the X-axis with the Σ-plane as the position of antenna 1 (for convenience, as we did before) and position antenna 2 at arbitrary coordinates (X, Y, Z). This defines an arbitrary baseline vector $\vec{r}(X, Y, Z)$. During the earth rotation the antenna beams are kept pointed at the source direction $\vec{\Omega}$, i.e. this vector does not move, however the tip of the baseline vector $\vec{r}(X, Y, Z)$ describes a trajectory $X(t), Y(t), Z(t)$ in space. Consider again radiation incident on the Σ-plane parallel to the X-axis, like in the case of the original Van Cittert-Zernike derivation. In that case the path difference at positions 1 and 2 was zero, since the baseline was located in the Y, Z-plane, now the path difference is given by $X(t)$. To restore maximum coherence between positions 1 and 2, this path difference can be compensated by delaying the signal of antenna 2 by $X(t)/c$, in other words radiation from direction X arrives at the same time at both inputs of the correlator. Taking the centre of the source as the zero reference point for the source direction vector in the X, Y, Z reference frame, we can select an infinitesimal source element in the direction $\vec{\Omega}_0(\Omega_{X0}, \Omega_{Y0}, \Omega_{Z0})$ (unit vector) $\Longrightarrow I_0\delta(\vec{\Omega} - \vec{\Omega}_0)$. The path difference at the input of the correlator is the projection of the baseline on that unit vector under subtraction of the compensation $X(t)$:

$$\vec{r}.\vec{\Omega}_0 - X = [\Omega_{Y0}Y + \Omega_{Z0}Z + (\Omega_{X0} - 1)X] \tag{172}$$

If the extent of the radiation source (or the field of view defined by a single antenna beam) is sufficiently small, the value of the direction cosine $\Omega_{X0} \approx 1$, i.e. $(\Omega_{X0} - 1)X \ll \lambda$. In that case the difference in path length, like in the case of the original VanCittert-Zernike derivation, equals the scalar product of two vectors in the YZ-plane: the path difference equals $\approx \vec{r}_p.\vec{\Omega}'_0 = \Omega_{Y0}Y + \Omega_{Z0}Z$, with $\vec{r}_p$ the projection of the baseline vector on the YZ-plane and $\vec{\Omega}'_0$ the projection of the unit direction vector on the YZ-plane. If $I(\vec{\Omega})$ represents the brightness distribution over the sky, the VanCittert-Zernike relation for obtaining the coherence function is now given by:

$$\tilde{\Gamma}(\vec{r}_p) = \int\int_{\text{source}} I_0(\vec{\Omega}) e^{2\pi i \vec{\Omega}.\vec{r}_p/\lambda} d\vec{\Omega} \tag{173}$$

with $\vec{r}_p$ the *projected* baseline!
Conclusion: Arbitrary baselines do not change the aperture synthesis technique, except that the 3D baselines are projected on the observation plane Σ and become 2D.
Note: the pathlength difference $X(t)$ changes continually during the rotation of the earth and, hence, needs continues electronic compensation with an accuracy of a small fraction of the wavelength λ. In practise, for the WSRT, this is accomplished in two steps: *coarse compensation* with a delay line and subsequent *fine tuning* with the aid of an electronic phase rotator.
At the beginning of the description of earth rotation aperture synthesis, we considered the radiation source to be located in the direction of the rotation axis. Suppose now that the extended radiation source (or the single telescope field of view) is located along a direction vector that makes an angle ϕ_0 with the rotation axis of the earth. This is then the direction of the X-axis, perpendicular to this axis is the YZ-plane. The East-West oriented WSRT baselines physically rotate in a plane perpendicular to the earth axis, producing concentric circles as described earlier. These concentric circles now need to be projected on the YZ-plane, i.e. the observation plane Σ perpendicular to the viewing direction of the centre of the source (or the centre of the field of view). The circles change into ellipses and the coherence function is now sampled on ellipses, rather then on circles. The major axes of these ellipses remain equal to the physical length of the WSRT baselines, the minor axes are shortened by $\cos\phi_0$. This causes the *point spread function (PSF)* to become elliptical as well, the angular resolution therefore reduces with the lowering of the declination $[(\pi/2) - \phi_0]$. As a result, celestial directions along the declination are subject to a broadening of the central peak of the PSF according to:

$$PSF = \frac{\alpha\lambda}{2L_{max}\cos\phi_0} \tag{174}$$

Apart from the central peak, the *grating lobes* are scaled accordingly, this can be clearly seen in figures (33) and (35). For the WSRT the most extreme case of baseline shortening would occur with a source in the equatorial plane (declination 0), no resolution would be left in one direction. To circumvent this problem, baselines need to contain always North-South components. This is for example the case with the US VLA (Very Large Array) aperture synthesis telescope.

9 High energy imaging

9.1 Grazing incidence telescopes

Wavelengths shortward of ≈ 50 nanometer and longer than ≈ 0.1 nanometer, i.e. the regime of extreme-ultraviolet (EUV) and X-radiation, are absorbed by metallic surfaces at normal incidence, the only way to reflect and focus this radiation is to employ *total reflection*, since the refractive index of metals at these wavelenghts is slightly less than one. We can write the refractive index as:

$$n \;=\; 1 - \delta - i\beta \quad \text{with} \quad \delta,\, \beta \ll 1 \tag{175}$$

δ being a measure for the reflection and β a measure for the absorption. If we only consider the real part of n, referring to reflection, we have $n \;=\; 1 - \delta$. Employing Snell's law, the critical angle for total reflection follows from:

$$\cos\theta_{cr} \;=\; 1 - \delta \qquad \Longrightarrow \qquad \theta_{cr} \;=\; \sqrt{2\delta} \tag{176}$$

where the cosine term can been approximated by the first terms of the Taylor series, since $\delta \;\ll\; 1$. Typical values for the critical angle, depending on wavelength and specific metal(coating), range from 1 to 3 degrees. These small angles imply that focussing optics for this wavelength range requires very high incidence angles. It is therefore commonly referred to as *grazing incidence optics*, the grazing angle constitutes the angle between the incoming ray and the metal reflecting surface (i.e. the complement of the angle of incidence). Figure 36 shows the reflection efficiencies ϵ, for several grazing angles, as a function of wavelength in Å-units for gold as the reflection coating

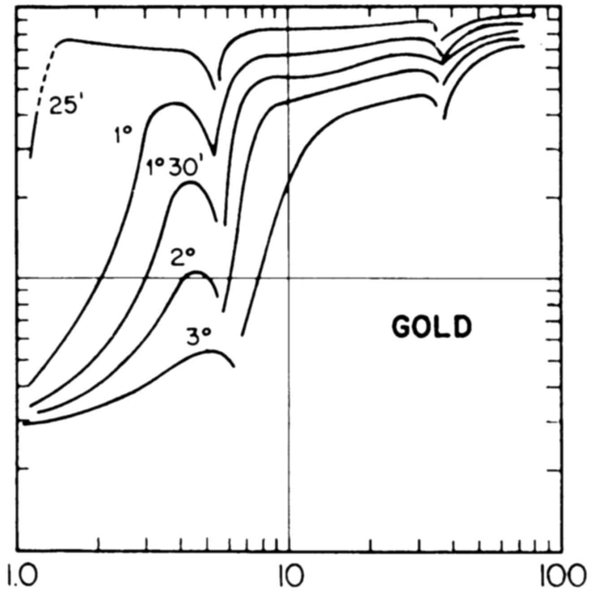

Figure 36: *Reflection efficiencies of gold for several grazing angles as a function of wavelength (in Å) of the incoming radiation beam. The ordinate scale for the efficiency ϵ ranges from 0.01 to 1. Figure taken from Giacconi and Gursky 1974.*

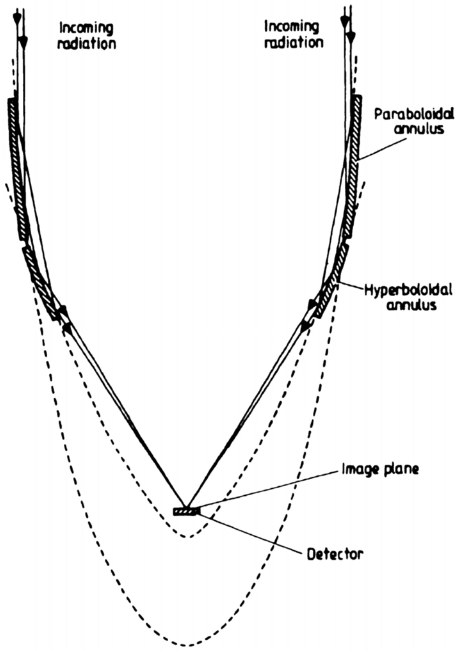

Figure 37: *Grazing incidence optics: Wolter-I configuration. Figure taken from Kitchin 1998*

deposited on the mirror substrate.

Several grazing incidence telescope configurations have been devised, the most practical use has been a design by Wolter comprising adjacent annular sections of a paraboloidal and a hyperboloidal surface irradiated at grazing incidence. This so-called Wolter-I telescope is displayed in figure 37, the focus of the paraboloid coincides with the focus of the "virtual" part of the hyperboloid and the telescope focus is the focal point of the actual hyperboloidal annulus. Obviously, like in the case of the Cassegrain system, the hyperboloidal surface acts as a correction element for the large coma error of a single paraboloid for off-axis radiation. In practise one can obtain resolutions of the order of 1-10 arcseconds over fields of view (FOVs) of several arcminutes. Another design, also due to Wolter, is shown in figure 38. In this case the grazing incidence rays reflect from the inner surface of the paraboloidal mirror and, consecutively, from the outer surface of the hyperboloid. This so-called Wolter-II configuration is occasionally used for soft X-ray and Extreme Ultraviolet imaging, the typical grazing angles are larger than in the case of the Wolter-I configuration, however the design can be much more compact. The location of the foci of both the individual annuli and of the full telescope is indicated in figure 38. A drawback of grazing incidence optics is the fact that the aperture of such telescopes is only a thin ring in projection, since only the radiation incident onto the paraboloidal annulus is transmitted to the focus. In order to increase the effective area, and hence the sensitivity of the telescope system, several systems with different radii are nested inside each other in a confocal fashion. This is schematically shown in figure 39.

It is important to realise that at these short wavelengths it is normally no longer the diffraction that limits the resolution of the telescope (λ/D very small), but rather

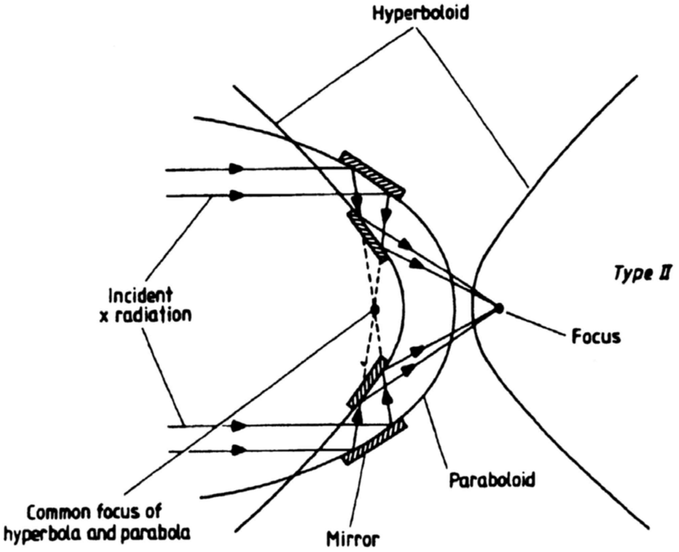

Figure 38: *Grazing incidence optics: Wolter-II configuration. Figure taken from Kitchin 1998.*

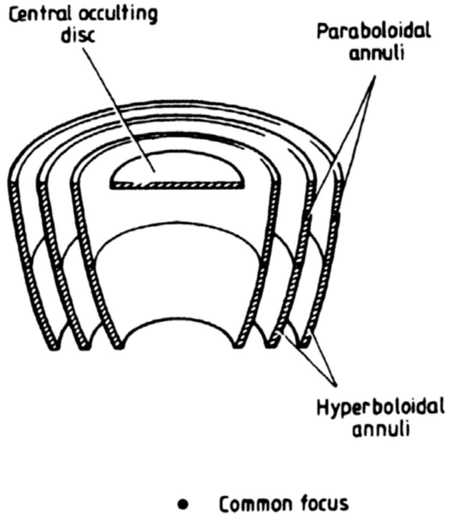

Figure 39: *Confocal nesting of Wolter-I telescopes to enlarge the aperture effective area. Figure taken from Kitchin 1998.*

the accuracy of the surface *figure* such as *roundness* and *profile*. Moreover, the micro-finish of the reflecting surface (i.e. the microscopic surface roughness) plays a dominant role in the potential quality of the image: a surface with irregularities scatters rather than reflects the incoming beam, this will show up as *"wings"* in the PSF that might contain a substantial fraction of the beam energy. Surface smoothness of the order of 0.1 nanometer has be achieved, providing very high quality X-ray mirrors.

Grazing incidence mirrors can be produced in a variety of ways, either by direct machining and polishing or by a so-called replication technique. In the latter case a thin metal reflective layer is deposited on a highly polished (i.e. X-ray quality) mandrel of the inverse shape to that required for the mirror. Subsequently electro-deposition of e.g. nickel onto the mandrel is applied to built up a self-supporting mirror shell. When the appropriate thickness for the required mechanical integrity of the shell has been reached, the shell can be separated from the mandrel by thermal shock (cooling).

Grazing incidence mirrors with intrinsically much lower, but nevertheless still very useful, angular resolution of the order of 1-3 arcminutes can be fabricated from nesting foil mirrors that approximate the paraboloidal and hyperboloidal surfaces by truncated conical surfaces. Thin aluminum foil with a lacquer coating to provide the required smoothness and reflection efficiency has been successfully employed for this purpose, the conical approximation much reduces the complexity of the fabrication process and therefore the costs. Obviously, given the mandatory short length of the mirror cones to retain, at least geometrically, the required resolution, many hundreds of foil shells need

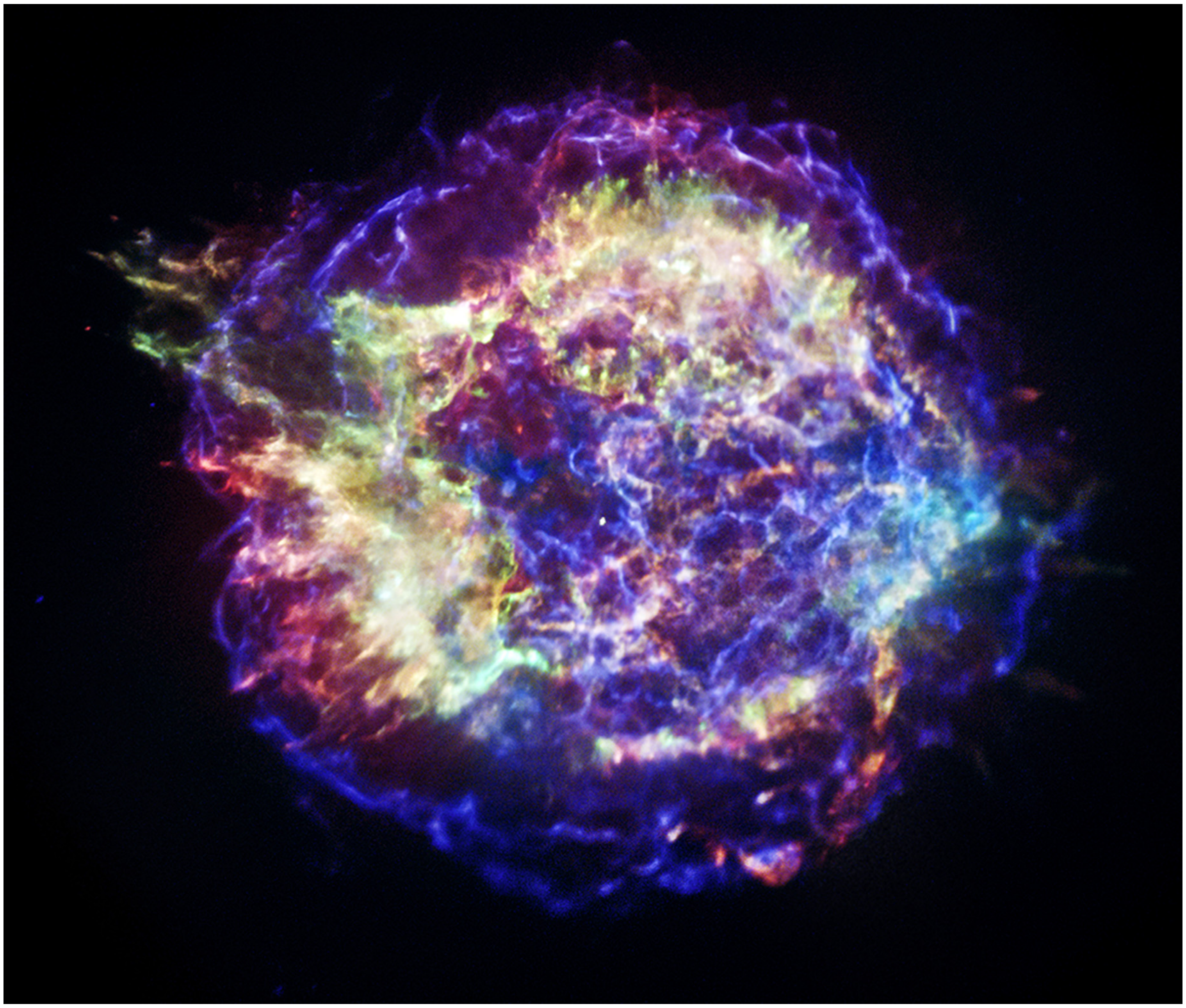

Figure 40: *High resolution X-ray image of the Cas A SNR (SN type II) obtained with the Chandra grazing incidence telescope. The telescope comprises a nest of 4 confocal Iridium coated Zerodur shells. A Chromium binding layer is applied between the high-Z Ir-coating and the glass. Credit NASA/Chandra Science Data Centre.*

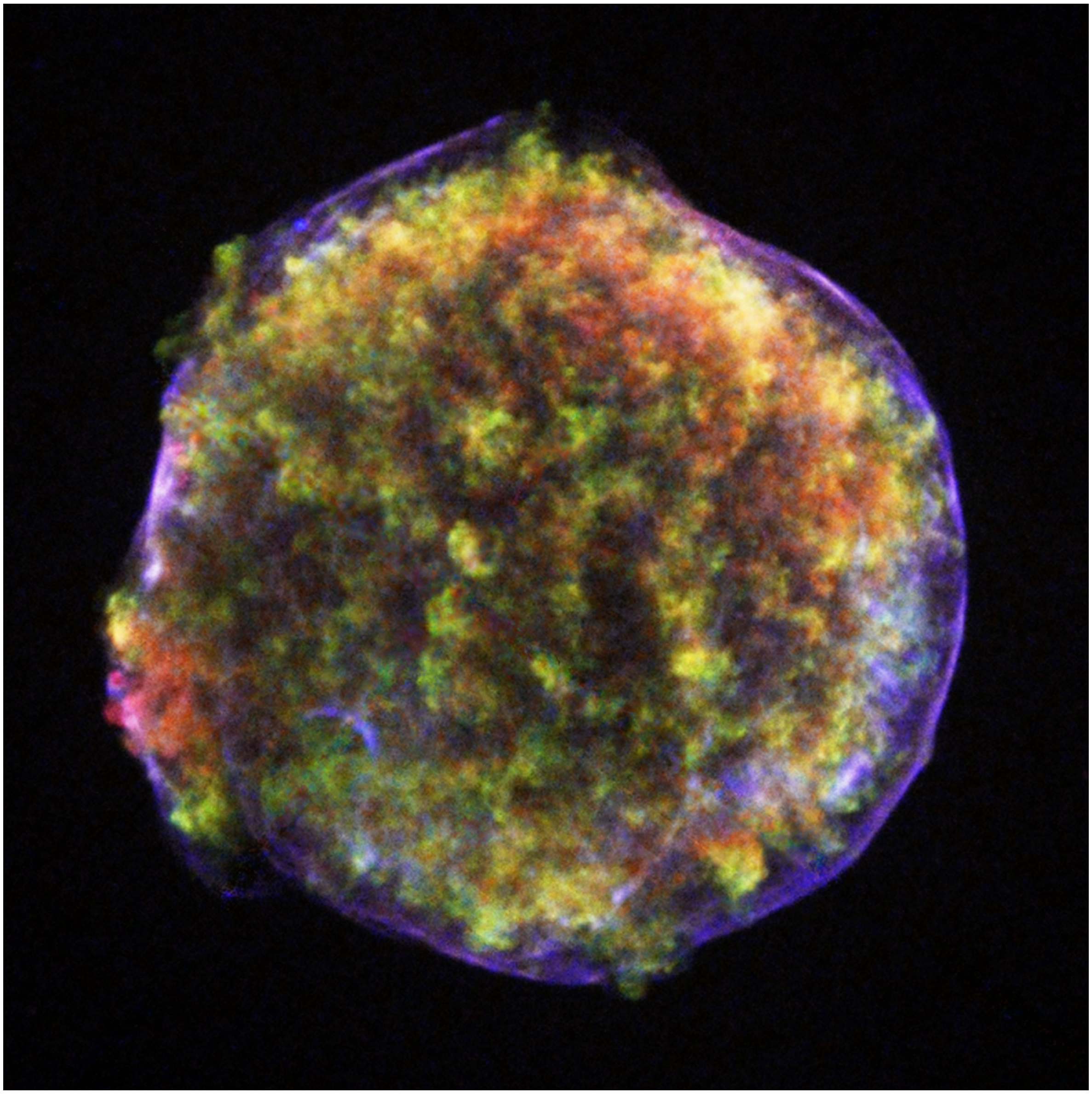

Figure 41: *High resolution image of the historical SNR Tycho (SN type Ia, observed in 1572) obtained with the Chandra X-ray telescope. Credit NASA/Chandra Science Data Centre.*

to be nested to reach a sufficiently large effective aperture!

9.2 Non-focussing optics: beam modulation

If the wavelength of the incident radiation beam becomes smaller than ≈ 0.1 nanometer, it becomes increasingly difficult to use reflection optics. Employing multi-layer coated mirrors still allows focussing of radiation in selected wavebands down to wavelengths as short as 0.02 nanometer, but not much further. Therefore, images will have to be obtained by other means. In the photon energy range where the photo-electric effect is still dominant, imaging can be accomplished by applying the *coded mask* technique. In

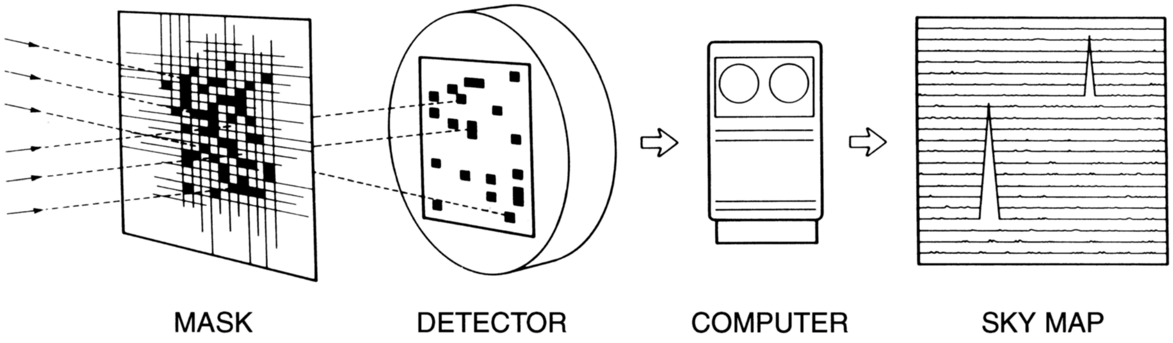

Figure 42: *Elements of a coded mask telescope. Figure taken from Bleeker in BeppoSAX 2003.*

the photon energy range where Compton scattering and pair creation are the dominant interaction processes, the *kinematics* of the interaction processes can be used to extract directional information to built up images. In that case telescope and detecting device have become one. We shall treat these latter two techniques under Gamma-ray imaging, and shall first briefly describe the principle of coded mask imaging.

Coded mask imaging is based on an old principle, the *Camera Obscura*, where a hole in a mask allows an image to form on a screen. The smaller the hole , the better the angular resolution, but unavoidably at the expense of the irradiance. Moreover, if the hole becomes too small, diffraction effects will degrade the resolution. In the X-ray domain diffraction effects become negligible considering the small wavelengths involved, and the irradiance level can be enhanced by employing a large number of holes, so that the transmission of the mask may reach a value of ≈ 50 percent. The resulting image from such a multi-hole mask is a linear superposition of the images produced by each individual hole. The distribution of the holes needs to be done in such a way that the brightness distribution of the illuminating celestial source can be reconstructed in an unambiguous fashion from the measured intensity distribution in the registered image. The appropriate pattern of *holes and stops* constitutes a so-called coded mask, this surface is described by a two-dimensional transmission function $\mathbf{M}(x,y)$. If we assume an extended radiation source $\mathbf{S}(x,y)$, the intensity distribution $\mathbf{D}(x,y)$ on the detector can be written as a convolution:

$$\mathbf{D}(x,y) \;=\; \mathbf{M}(x,y) \,*\, \mathbf{S}(x,y) \,+\, \mathbf{N} \tag{177}$$

with $\mathbf{N}$ representing the detector noise.

The working principle of the coded mask camera is displayed in figure 42. To reconstruct the brightness distribution of the extended celestial radiation source, the image $\mathbf{D}(x,y)$ needs to be deconvolved with the transmission function $\mathbf{M}(x,y)$ of the mask. The simplest operation to perform this deconvolution is a *two-dimensional cross-correlation* of the mask transmission function with the detector image. Formally, an estimate of the source distribution, $\mathbf{\Sigma}(x,y)$, can be obtained by applying a decoding function $\mathbf{A}(x,y)$:

$$\mathbf{\Sigma}(x,y) \;=\; \mathbf{A}(x,y) * \mathbf{D}(x,y) \;=\; \mathbf{A}(x,y) * \mathbf{M}(x,y) * \mathbf{S}(x,y) + \mathbf{A}(x,y) * \mathbf{N} \tag{178}$$

If the image detector were to be *noise free*, there should exist a one-to-one correspondence between $\mathbf{S}$ and $\mathbf{\Sigma}$, hence we require $\mathbf{A} * \mathbf{M} = \delta$ (Kronecker delta). Secondly, the noise distribution should preferably be uniform over the deconvolved image. Mask patterns $\mathbf{M}(x, y)$ that satisfy these two requirements are designated *optimal.*

Note: *It is important to realize that the noise level and distribution in the detector image depends on the brightness distribution of the sky within the field of view of the coded mask telescope. Each source in the sky field observed does not only contribute to its specific location (i.e. pixel) in the detector image but also to all the other image pixels. This means that all the radiation sources in the observed sky field contribute to the noise level of the detector image.*

The angular resolution of a coded mask telescope is directly related to the size m_e of an individual mask element and the distance l between the mask plane and the detector plane:

$$\delta\theta_{res} \; = \; \arctan(m_e/l) \tag{179}$$

This resolution implies a proper sampling by the image detector of the mask element m_e, the Nyquist criterion is amply satisfied for a spatial resolution of the detector $d_e \approx m_e/3$. Usually the size of the mask pattern covers the same area as the image detector and the *sensitivity is not uniform across the whole field of view*, being dependent on the coded fraction of the detector that is illuminated by the mask pattern. Hence, the characterisation of the field of view of a coded mask telescope can be done along three coding limits: the fully coded limit, the semi-coded limit and the zero-coded limit. If the mask has the same size as the image detector and does not comprise repetition of a (smaller size) basic coding pattern, the field that is fully coded equals zero. The field of view at half the sensitivity will be equal to the solid angle subtended by the mask as seen from the image detector. Several coded mask telescopes have been developed and built for hard X-ray and low-energy gamma-ray detection, e.g. the Dutch-Soviet coded mask X-ray camera on the Soviet MIR Space Station, the French coded mask telescope SIGMA onboard the Soviet GRANAT spacecraft, the Dutch-Italian Wide Field coded mask Cameras on the BeppoSAX satellite and the low-energy Gamma-ray Imager on the European INTEGRAL Space Observatory. The Wide-Field coded mask Cameras (WFC) onboard BeppoSAX have been particularly successful during the six year mission life time of the satellite and have made a crucial contribution to unveiling the origin of gamma-ray burst sources. The two, anti-parallel, WFCs in BeppoSAX had a 40 x 40 degrees^2 field of view (FWZR) and an angular resolution $\delta\theta_{res} = 5$ arcminutes with location accuracy (i.e. position resolution, see Chapter 5) of 0.7 arcminutes. The coded mask comprised a matrix of 256 x 256 elements with m_e =1 mm. The open fraction of the mask was chosen from numerous sky simulations to be 0.33 for optimum signal to noise ratio. The resolution of the position sensitive X-ray image detector was $\approx$ 0.4 mm, i.e. commensurate with the Nyquist requirement on proper sampling.

Figure 43 shows raw image data (left panel) from a sky region close to the Galactic Centre in terms of a photon intensity map. It is evident from this raw image that the sky field observed contains two bright point sources, which are directly visible as two light rectangles in the image arising from mechanical collimation by the tube connecting the mask with the image detector. No other sources can yet be discerned. The right-hand panel shows the deconvolved detector image, displaying a host of point sources in

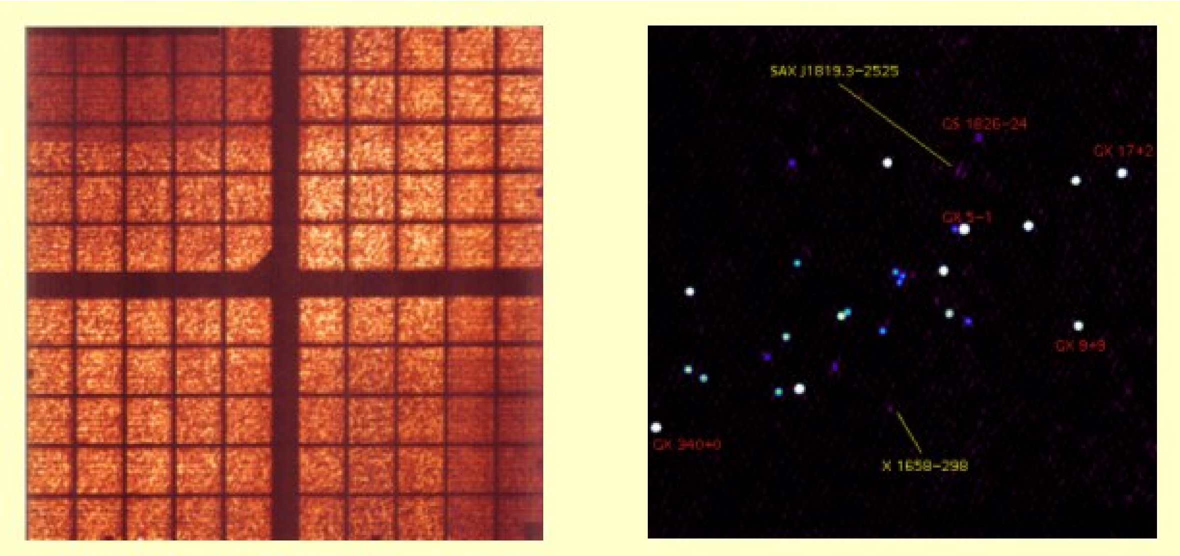

Figure 43: *Left: A BeppoSAX WFC coded mask picture of a sky region near the Galactic Centre in terms of a raw photon intensity map. The dark bars and stripes reflect the detector window support structure. Right: the deconvolved detector image showing the wealth of compact X-ray sources in the Galactic Bulge area. Figure taken from Bleeker BeppoSAX 2003.*

a wide brightness range. All these sources can be simultaneously monitored regarding their time variability (e.g. transients, bursts, flashes, periodicities) and their spectral morphology and variability.

9.3 Compton imaging

In the energy range from several hundred keV to several MeV, the interaction of γ-ray photons with matter is dominated by Compton scattering. The cross-section for this process is governed by the Klein-Nishina equation. The atomic cross section for a target material with atomic number Z simply follows from multiplication of the KN-cross section with the atomic charge Z, where the atomic electrons are considered to be essentially free. The Compton process is therefore proportional to Z. The energy of the photon can only be measured after a cascade process of consecutive scattering interactions until the down-scattered, i.e. in energy degraded, photon is eventually absorbed by photoelectric absorption. This requires a thick absorber, that suffers from a great deal of noise due to the interaction of cosmic particle radiation with the detector material. This noise can be effectively eliminated by splitting the absorber in two separate parts, located at a certain distance l from each other. In this way directional sensitivity can be introduced by employing the kinematics of the Compton effect. The upper part of this so-called *Compton telescope* then comprises a position sensitive sensor array of a suitable low-Z material with a thickness maximised for a single Compton scattering to occur. In contrast, the lower detector in the telescope entails a position sensitive array of high-Z material of sufficient thickness to ensure total absorption of the scattered photon energy.

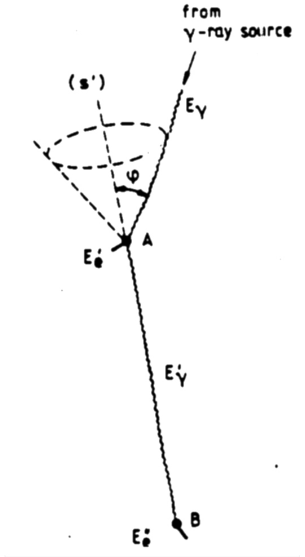

Figure 44: *Kinematics of the Compton effect. The sky location of the incident photon can be traced back to a circular contour by measuring the scattering angle $\phi = \theta_{sc}$ and the orientation of the line element AB.*

The sky location of the incident photon can now be traced back to a circular sky contour by measuring the scattering angle and the direction of the scattered photon leaving the upper telescope sensor. This is schematically shown in figure 44. By measuring many source photons, the intersecting sky contours will indicate the most probable sky position of the gamma-ray source.

The scattering angle θ_{sc} between the incident photon and the scattered photon, leaving the upper telescope element, follows from the kinematics of the Compton effect:

$$\cos\theta_{sc} \;=\; 1 \;-\; mc^2\left(\frac{1}{\epsilon_{\gamma'}} \;-\; \frac{1}{\epsilon_\gamma}\right) \quad \text{with} \quad \epsilon_\gamma \;=\; \epsilon_{\gamma'} \;+\; E_e \tag{180}$$

in which ϵ_γ represents the energy of the incoming photon, $\epsilon_{\gamma'}$ the energy of the first scattered photon and E_e the energy of the scattered Compton electron. The direction of the first Compton scattered photon, the line element AB in figure 44, can be reconstructed from measuring the interaction positions in the upper and lower telescope elements respectively.

Figure 45 depicts such a telescope configuration for the case of the Comptel instrument, launched in 1991 aboard the Compton Gamma-Ray Observatory (Compton GRO) by NASA. The energy and position of the scattered Compton electron is measured in the upper detector, comprising a liquid scintillator array in which each element is viewed by a number of photomultipliers to obtain position resolution within a single array module. The energy and position of the first scattered photon is measured in the lower detector, constituting an array of optically thick NaI(Tl) scintillator crystals for total absorption of the scattered gamma-ray photon, again a cluster of photomultipliers is used to determine the position of the absorbed scattered photon inside a single NaI-module. By requiring simultaneous triggers from the upper and lower detector arrays within a small

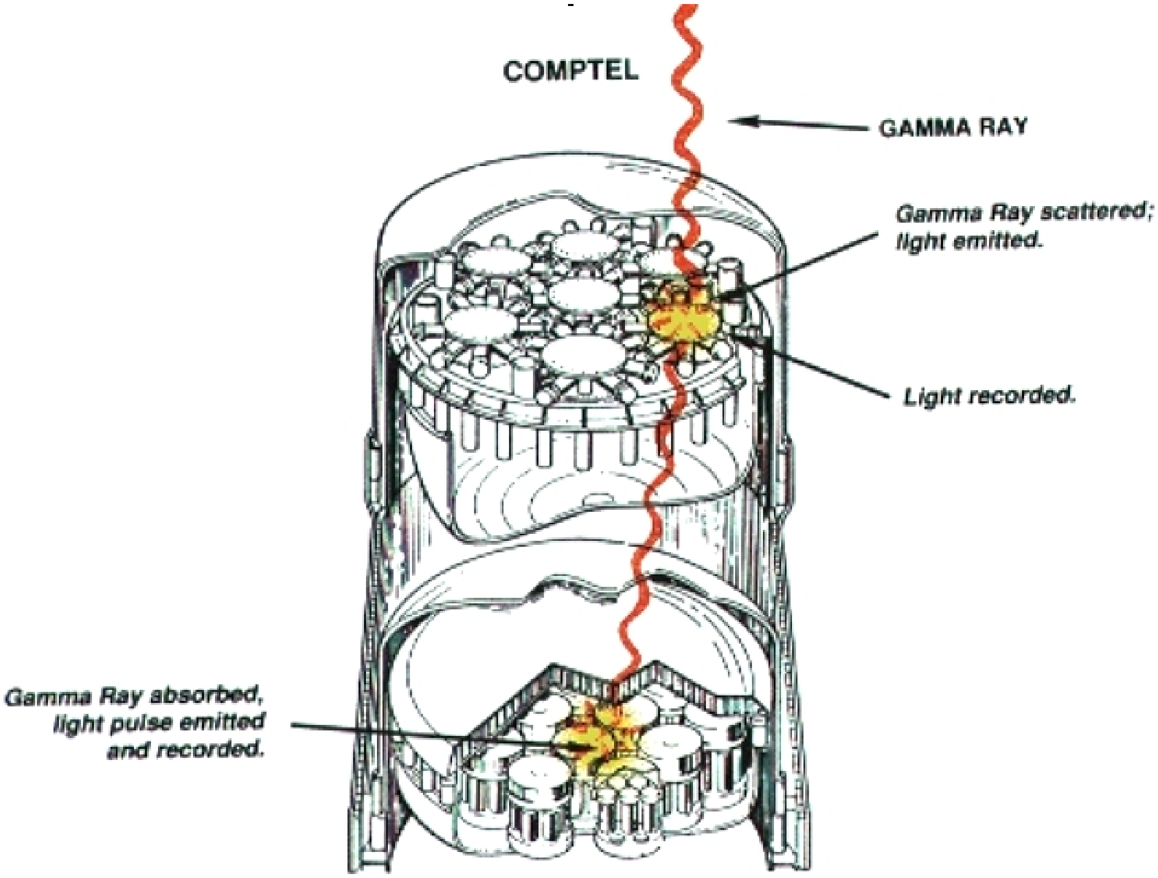
COMPTEL
GAMMA RAY
Gamma Ray scattered; light emitted.
Light recorded.
Gamma Ray absorbed, light pulse emitted and recorded.

Figure 45: *Configuration of the Compton telescope as flown on the NASA Compton Gamma-Ray Observatory (Compton-GRO). The upper detector array (the scatterer) consists of 7 cylindrical modules filled with liquid scintillator (NE 213A, geometric area $\approx$ 0.42 m^2), each viewed by several photomultiplier tubes to obtain position resolution. The lower detector (the absorber) consists of 14 cylindrical modules of an anorganic scintillator (NaI(Tl), geometric area $\approx$ 0.86 m^2, each of them also viewed by a multitude of photomultipliers. The distance between both detector assemblies is 1.5 meter. Credit NASA/Comptel Consortium.*

time window commensurate with the time-of-flight $t_{TOF} \approx l/c$ of the first scattered photon from the upper to the lower detector array, almost all background events can be eliminated. This can be understood by remembering that these background events either trigger only one detector array or they produce the wrong time sequence (practically all background events are produced in the high-mass high-Z absorber array). This particular Compton telescope is the most sensitive instrument in the medium-energy gamma-ray range (0.5-30 MeV) developed and operated thus far. The angular resolution depends on the measurement accuracy of the orientation of the line element AB and the scattering angle θ_{sc} and ranges from 1.7 to 4.4 degrees (FWHM) depending on the gamma-ray energy. The accuracy of source position determination, i.e. the position resolution, is 5-30 arcminutes dependent on photon energy. The field of view (FOV) of such an instrument can be quite large, depending on l and the typical diameter of the

upper and lower detector-arrays, the FOV of the Comptel instrument subtends ∼ one steradian on the sky, similar to that of a coded mask telescope.
Figure 46 shows a gamma-ray map of the Orion region, the nearest birthplace of massive stars in our Galaxy. The orange-white contours depict the areas of enhancement of the measured gamma-ray flux in the 3-7 MeV region. Most of this flux appears to be concentrated near specific energies and, although the energy resolution of Comptel is rather limited (5-8 percent FWHM), this is strongly suggestive for the presence of gamma-ray line radiation at 4.4 and 6.1 MeV respectivily. The blue contours display the areas of high-density interstellar clouds. The spatial coincidence of the detected gamma-ray enhancements with these dense clouds points to interaction between en-

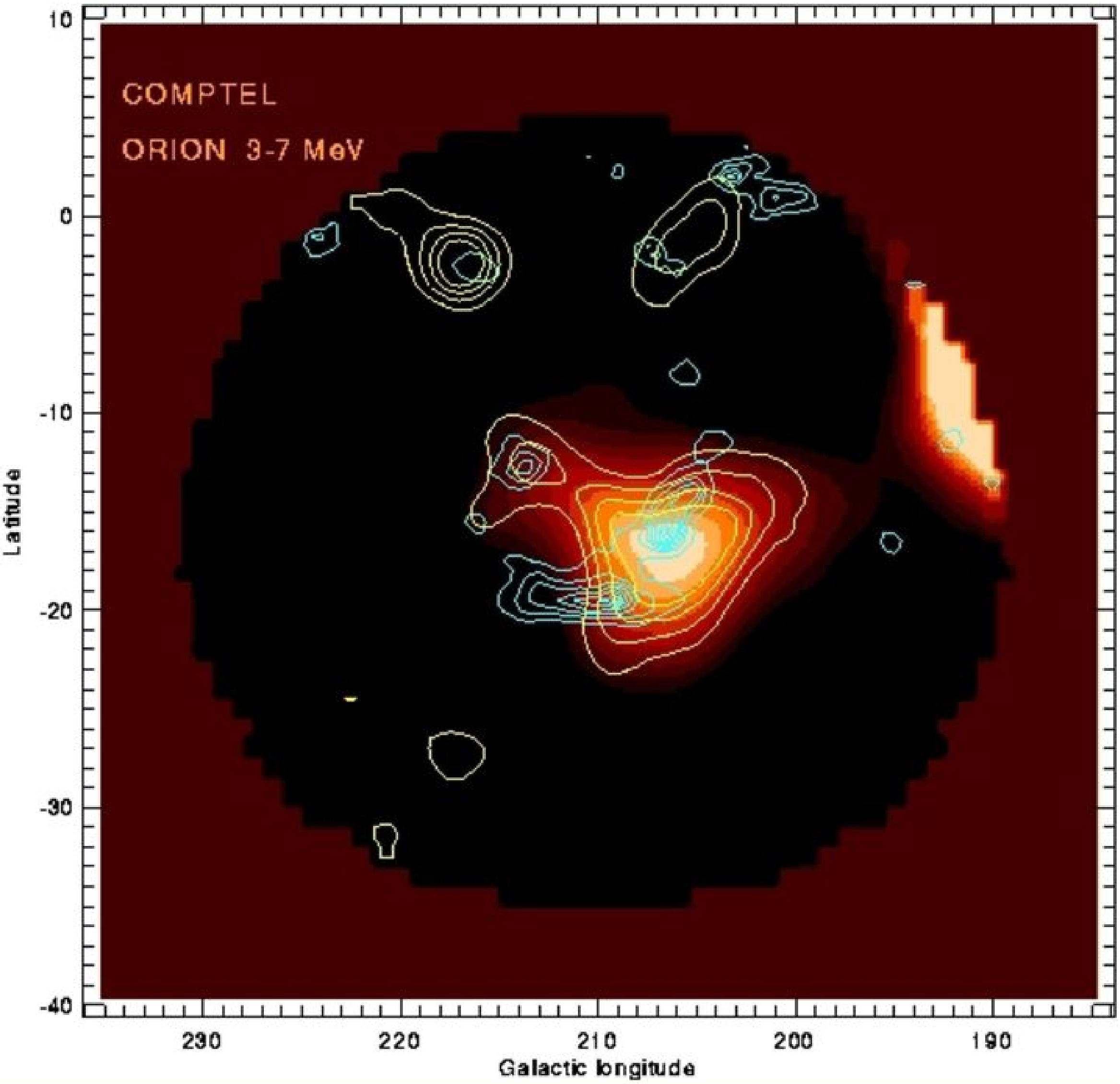

Figure 46: *Gamma-ray map of the Orion region in the 3-7 MeV energy range as measured by Comptel. The orange-white contours indicate the locations of gamma-ray enhancements, the blue contours represent the locations of the densest interstellar gas clouds. Credit Comptel Consortium.*

ergetic cosmic-ray nuclei and the gas nuclei constituting the dense interstellar clouds. The lines could then originate from the radioactive decay of energetic cosmic-ray nuclei of carbon and oxygen.

9.4 Imaging through pair formation

At gamma-ray energies $\geq$ 20 MeV the main interaction process with matter is pair formation, causing the disappearance of the gamma-ray photon (destructive interaction) which is converted into an electron-positron pair, both having positive kinetic energy. The interaction can only occur in the presence of a strong electric field, as encountered in the vicinity of an atomic nucleus or near an orbital electron. The nucleus takes care of conserving the momentum balance (*recoil-nucleus*), but gains negligible energy. Consequently, the energy of the incident gamma-ray photon is practically entirely converted to electron/positron mass and kinetic energy. In the *centre-of-mass system*, the positron and the electron are emitted in opposite directions, however in the *laboratory-frame* they are emitted in the direction of the initiating photon, and the tracks of the electron and positron reveal the direction of the incident gamma-ray photon. *Directionality* can thus be obtained by measuring (”visualising”) the 3D-orientations of the electron and positron trajectories that originate at the location of the recoil-nucleus. From these orientations the sky position of the incident gamma-ray photon can be reconstructed.

The angular distribution $N(\phi)d\phi$ of the electron(positron) relative to the direction of the incident photon is, in good approximation, given by:

$$N(\phi)\,d\,\phi \;\approx\; \frac{\phi\,d\,\phi}{(\phi_{\pm}^{2}\,+\,\phi^{2})^{2}} \quad \text{with} \quad \phi_{\pm} \;=\; \frac{mc^{2}}{E_{\pm}} \tag{181}$$

where $E_{\pm}$ represents the total energy of the pair-electron(positron). The most probable angle ϕ_p for electron(positron) emission follows from differentiation of expression (181), this results in $\phi_p \;=\; \phi_{\pm}/\sqrt{3}$.

The *angle of bipartition* is given in close approximation by:

$$\phi_b \;\approx\; \phi_{\pm} \;=\; \frac{mc^{2}}{E_{\pm}} \tag{182}$$

and is thus *inversely proportional* to the total energy of the pair-electron(positron). The angle ψ between the partners of a pair is geometrically fixed for a given combination of ϕ_{-}, ϕ_{+} and the *dihedral* angle Φ between the photon-electron and the photon-positron planes. To a first approximation, if $\Phi = \pi$ (radian), and the positron and electron are emitted symmetrically with respect to the initiating photon, we may put $\phi_{+} \;\approx\; \phi_{-} \;\approx\; \phi$, consequently $\psi \;=\; 2\phi$. Hence, the opening angle of the pair also diminishes inversely proportional to the energy of the initiating gamma-ray photon.

The characteristic interaction length for pair formation is given by the *radiation length* (expressed in mass per unit area). This interaction length is proportional to $[Z(Z+1)]^{-1}$, i.e. $\approx Z^{-2}$ for the higher Z-elements, furthermore it depends on the material properties under consideration but is *independent* of the energy of the incident photon.

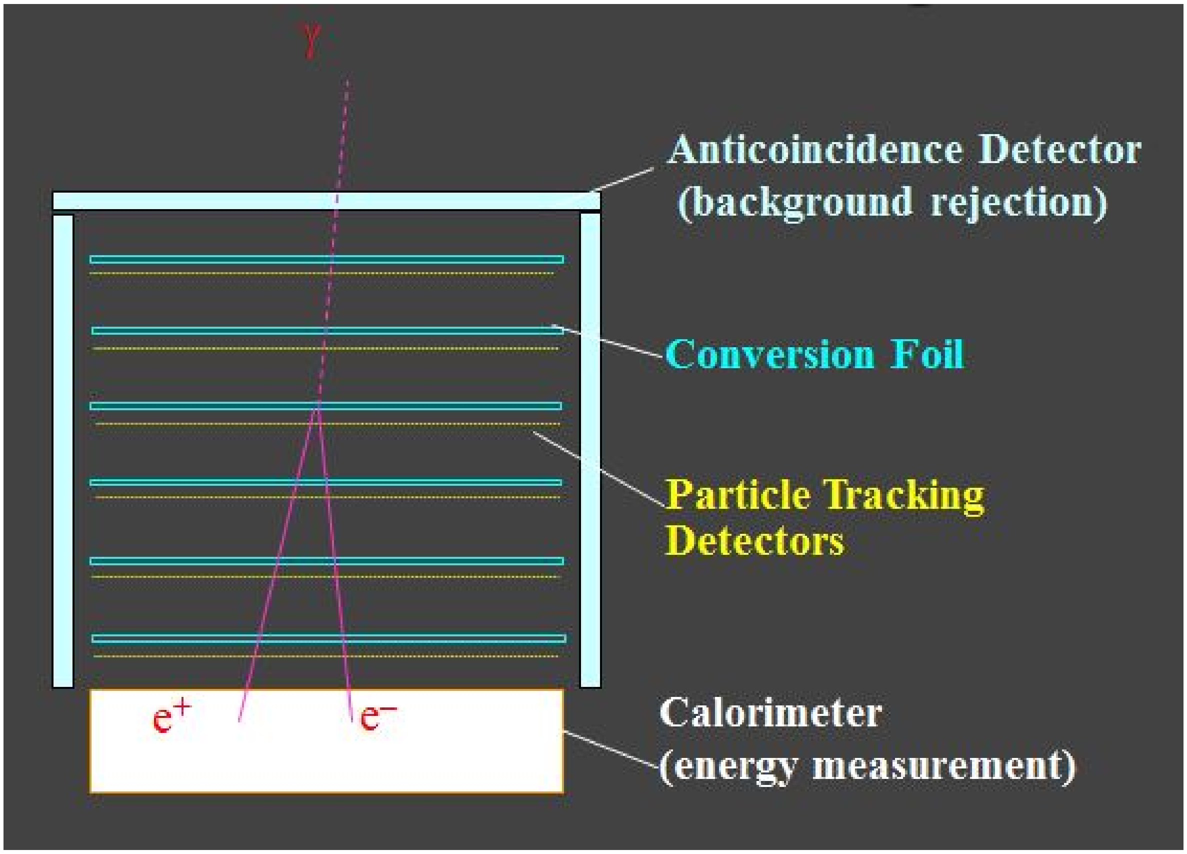

Figure 47: *Principle of a pair conversion telescope. Thin metal converter plates are interleaved with tracking layers, comprising gas layers or thin 2D-position-sensitive semiconductor detector planes. Credit NASA/EGRET website.*

A pair-conversion telescope can now be constructed by stacking several thin high-Z-metal converter plates at an appropriate mutual distance, each with a thickness of a fraction of a radiation length. The full stack is built up to a total thickness of a few radiation lengths, suitable high-Z converters comprise lead or tungsten plates. The gaps between the plates in the stack, the so-called *tracking layers*, are either filled with gas or do contain thin, position sensitive, semiconductor strip detectors, this is schematically shown in figure 47. Once the pair has formed, the electron/positron either ionizes the gas filling along their tracks or produce an electric charge in the semiconductor strips at the particular spots of their passage.

In the case of a gas filled chamber, the registered presence of a particle-pair triggers a high voltage pulse that discharges along the particle tracks between the plates, delineating the orientation of their trajectories. Such a detector is called a *spark chamber*, the principle of which was already developed during the late 1950's and the 1960's for diagnostics in high-energy physics at particle accelerators and in cosmic ray physics. This technique was also used in the first space-borne gamma-ray telescopes on SAS-2 (NASA) and COS-B (ESA), that resulted in the first (partial) maps of the high-energy gamma-ray sky. Although more sophisticated read-out techniques, employing thin-wire grids for detection of the charges produced along the ionisation tracks, were utilized

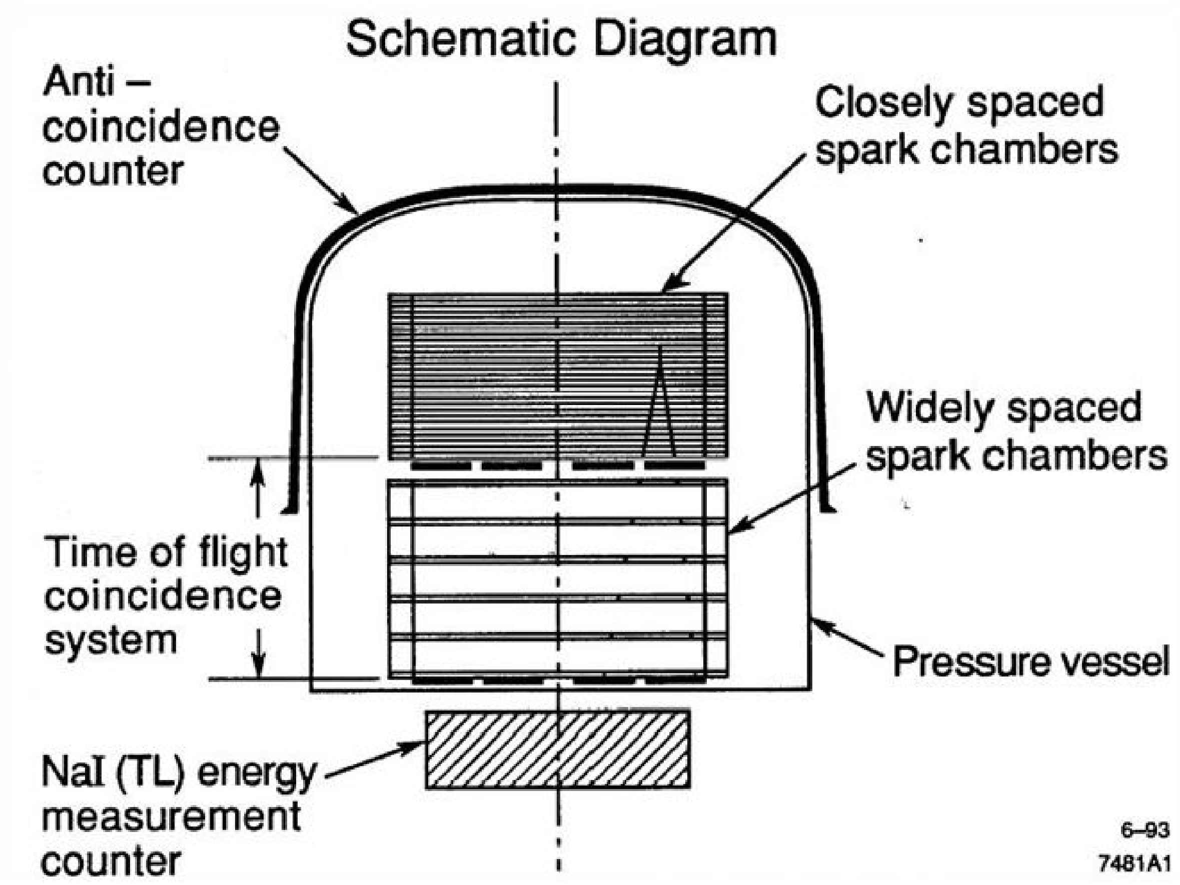
Schematic Diagram
Anti – coincidence counter
Closely spaced spark chambers
Widely spaced spark chambers
Time of flight coincidence system
Pressure vessel
NaI (TL) energy measurement counter
6–93
7481A1

Figure 48: *Schematic of the EGRET pair telescope aboard the Compton Gamma Ray Observatory. The principal telescope elements are indicated, for explanation see text. Credit NASA/EGRET Consortium.*

in the second generation EGRET gamma-ray telecope on the Compton Gamma-Ray-Observatory, the basic detection principle remained unaltered. Figure 48 shows the main constitutive elements of this pair telescope, that was designed to cover the energy range from 20 MeV to 30 GeV.

The instrument uses a multiple-layer spark chamber with thin metal pair-conversion plates. The total energy of the gamma-ray photon is measured by a NaI(Tl) scintillation counter beneath the spark chambers to provide good energy resolution over a wide dynamic range in energy. The absorption proces involves a cascade of non-thermal Bremsstrahlung losses by the pair-particles and renewed pair formation by the Bremsstrahlung photons, followed by Compton scattering and finally photo-electric absorption until all the initial photon energy has been depleted. The instrument is covered by a plastic scintillator anticoincidence dome to discriminate against charged particle radiation incident on the telescope. Moreover, to further separate background events from true celestial gamma-rays, a time-of-flight system is incorporated similar to the one incorporated in the Compton telescope, to ensure that the radiation is arriving in the proper time sequence.

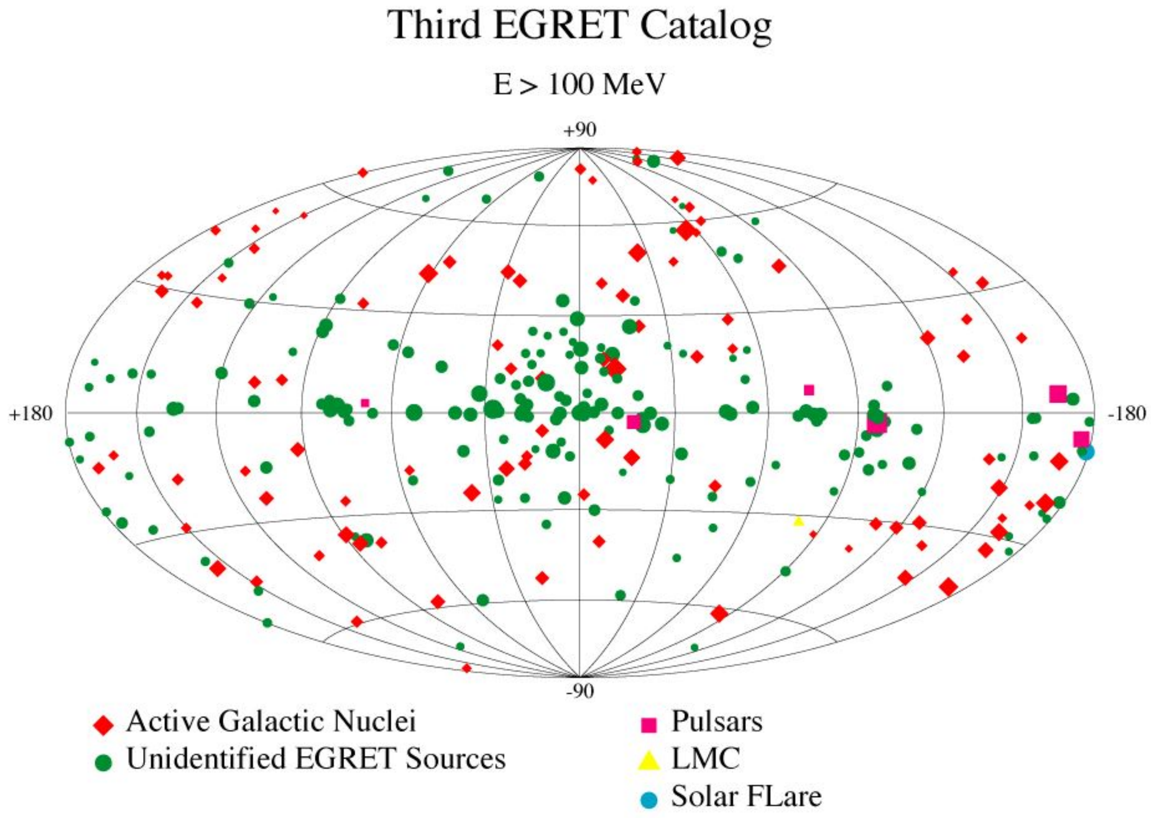

Figure 49: *The third EGRET all sky map of gamma-ray point sources above 100 MeV. The majority of these sources remains as yet unidentified. Credit NASA EGRET Consortium.*

The EGRET telescope instantaneously covers a wide field of view of ~ 0.6 steradians, the position resolution is dependent on photon energy and source strength. For a strong, hard spectrum, gamma-ray source this position resolution amounts to ~ 5 - 10 arcminutes. As outlined above, at lower photon energies the pair-angle is large, however the kinetic energy of the pair particles is relatively low and they suffer from scattering in the metal converter plates of the spark chamber. This introduces stochastic fluctuations in the particle-track determination due to the straggling behaviour of the electron/positron pairs. At high photon energies, the statistical fluctuations due to straggling are less severe, since the pair particles possess a much higher kinetic energy. However, as we have seen, the separation angle ψ diminishes with $1/E_{\pm}$, which makes accurate determination of the bipartition angle less straightforward.
Figure 49 shows the third catalogue of EGRET detected gamma-ray point sources above 100 MeV, the large majority of these sources has not yet been identified. This is either due to the limited positional resolution of the EGRET instrument or to an intrinsic lack of counterparts at other wavelengths.

The next generation gamma-ray telescope GLAST (Gamma-ray Large Area Space Tele-

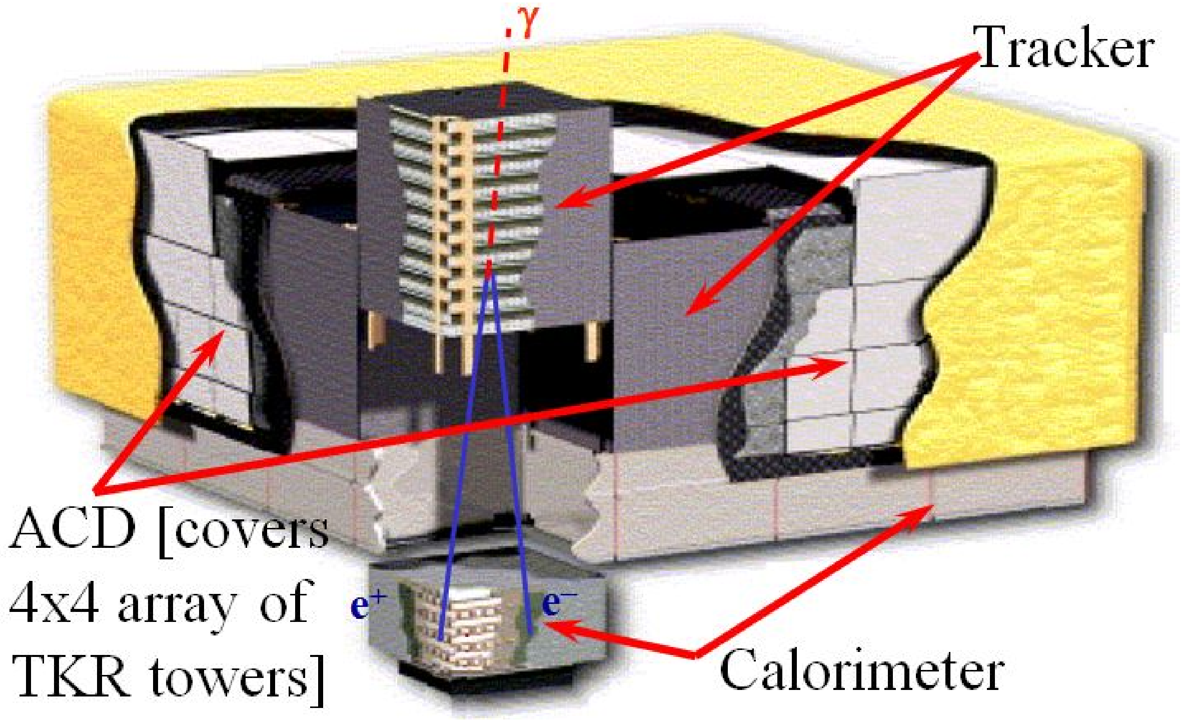

Figure 50: *GLAST: concept of the tracker modules. Credit NASA/GLAST Consortium.*

scope),launched by NASA in 2008, exploits a new technology in track detection: the gas-filling of the chamber has been replaced by solid state detectors as the tracking material. This allows for improved energy and spatial resolution. An energy resolution as good as 5 percent(FWHM) is aimed for and a position resolution as good as a few arcminutes for a single gamma-ray photon and of the order of $\leq$ 30 arcseconds for a strong gamma-ray source. In addition, a major advantage is the fact that a replenishable supply of chamber gas is no longer needed, which will make a much longer mission life time potentially feasible.

The baseline design for the GLAST pair forming telescope involves a modular tracker array comprising four-by-four *tower modules* (total of 16), of which each module consists of interleaved planes of thin lead converter sheets and silicon-strip solid state detectors. The silicon strips are arranged in a stack of 19 pairs-of-planes, one plane of each pair for read-out in the x-direction, the other plane of the pair for read-out in the perpendicular y-direction. When the electron/positron interacts with such a plane-pair, an accurate position can be determined in two dimensions. The third dimension of the track is derived by analysing the signals from consecutive adjacent planes as the particle travels downward through the stack towards the energy calorimeter. A multi-fold (at least three) coincidence between adjacent plane-pairs triggers the read-out sequence of the particle tracks.

Figure 50 shows the concept of the tower modules and the stack of tracking detectors. The energy bandwidth of GLAST ranges from $\sim$ 15 MeV to $\geq$ 100 GeV, with a field of view in excess of 2 steradians. The sensitivity of GLAST for point source detection will be $\sim$ 50 times that of EGRET at 100 MeV, with an enhanced positional resolving

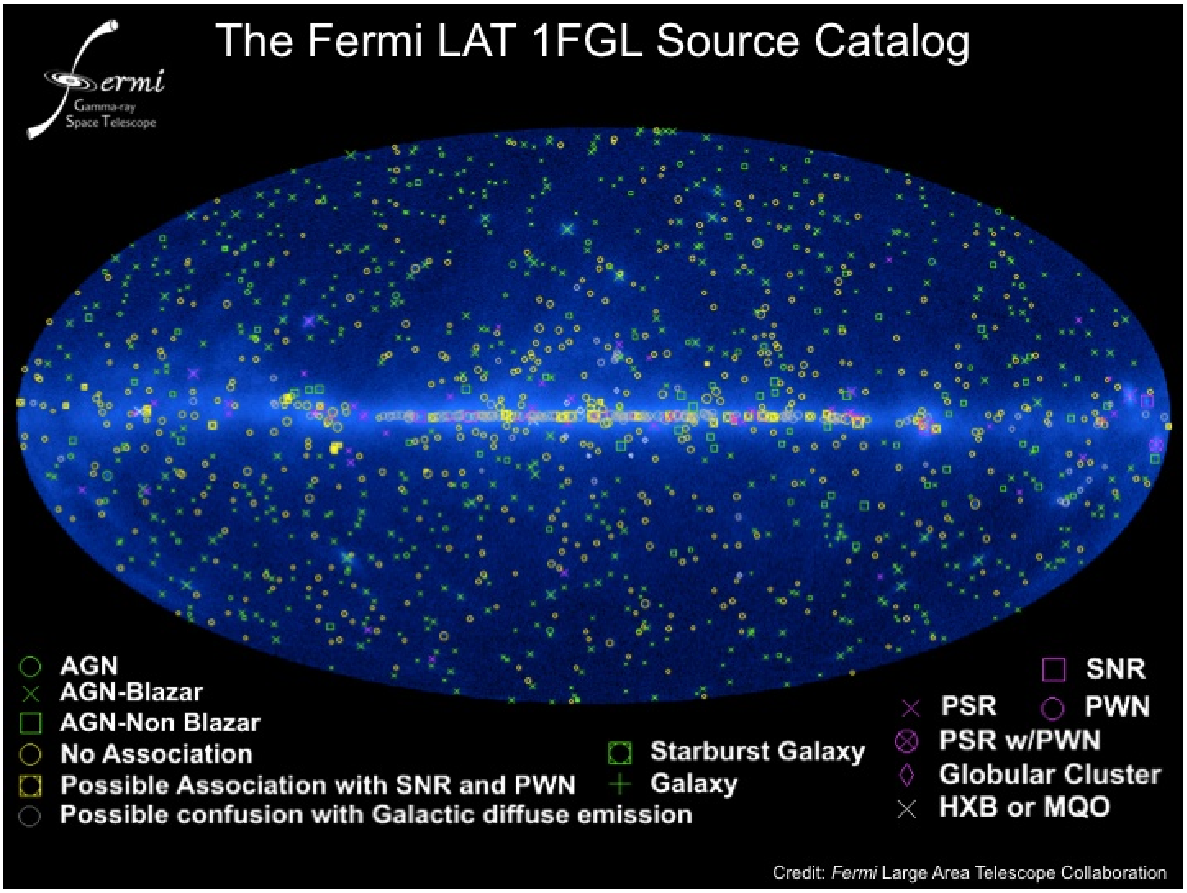
The Fermi LAT 1FGL Source Catalog
fermi
Gamma-ray Space Telescope
○ AGN
× AGN-Blazar
□ AGN-Non Blazar
○ No Association
□ Possible Association with SNR and PWN
○ Possible confusion with Galactic diffuse emission
□ Starburst Galaxy
+ Galaxy
□ SNR
× PSR
○ PWN
⊗ PSR w/PWN
◊ Globular Cluster
× HXB or MQO
Credit: Fermi Large Area Telescope Collaboration

Figure 51: *The first point source sky map obtained with the Fermi Gamma-ray Space Observatory. Credit Fermi-LAT Consortium.*

power of 0.5 - 5 arcminutes.

After launch on June 16 2008, GLAST was renamed the Fermi Gamma-ray Space Observatory. The first point source catalogue produced by Fermi is shown in Figure 51. The large number of detected gamma-ray point sources, of which most are still unidentified (designated 'No Association'), demonstrates the huge leap in sensitivity as compared to EGRET.

10 Signal to Noise Ratio

10.1 General

The feasibility of detecting a signal from a cosmic source depends on the level of noise (and disturbance) in which the signal is embedded during the measurement. Contributions to the noise signal potentially comprise other sources of radiation in the field of view of the observing instrument (sky-noise), background radiation from the operational environment of the telescope (e.g. atmosphere, radiation belts, earth) and noise arising from the constituting elements of the measurement chain (transducers, transmission lines, wave tubes, amplifiers, mixers, digitisers, etc.). Consequently, the quality of a particular observation is determined by the magnitude of the so-called signal-to-noise ratio (*SNR*). This *SNR* is generally a function of integration time (T_{obs}) of the observation and depends on the bandwidth ($\lambda\lambda$, $\nu\nu$, $\epsilon\epsilon$) of the measurement. From this a limiting sensitivity can be derived, i.e. the weakest source signal that can still be detected significantly.
For electromagnetic radiation three types of noise characterisations dependent on wavelength, are in use.

1. In the radio-band coherent detection is employed, i.e. the phase information is preserved, with $h\nu \ll kT$. Noise and *SNR* are normally represented by *characteristic temperatures* and *temperature ratios*.

2. In the (far) infrared band ($h\nu \approx kT$) incoherent detection is used. A transducer (Latin: *transducere*) converts the radiation field into a voltage or current and noise is commonly expressed in terms of an *equivalent radiation power* (*power characterisation*).

3. At shorter wavelengths (Optical, UV, X-rays, γ-rays), the incoming photons can be registered individually: so-called single photon counting. The signal and noise contributions can then be evaluated by a statistical treatment of the accumulated quanta, i.e. *quantum characterisation*. The latter treatment of noise and *SNR* applies by definition for corpuscular radiation and neutrinos.

The three notions introduced above are discussed in some detail in the following sections and the parameters which are commonly used for *SNR* characterisation in these cases will be shortly reviewed.

10.2 Temperature characterisation

10.2.1 Brightness and antenna temperature

Consider a thermal signal source radiating at radio wavelengths. If this source is optically thick, the specific intensity $B(\nu)$ is given by the Rayleigh-Jeans approximation ($h\nu \ll kT$) of a blackbody radiator:

$$B(\nu) = \frac{c}{4\pi}\rho(\nu) = \frac{2kT_b\nu^2}{c^2} = \frac{2kT_b}{\lambda^2} \tag{183}$$

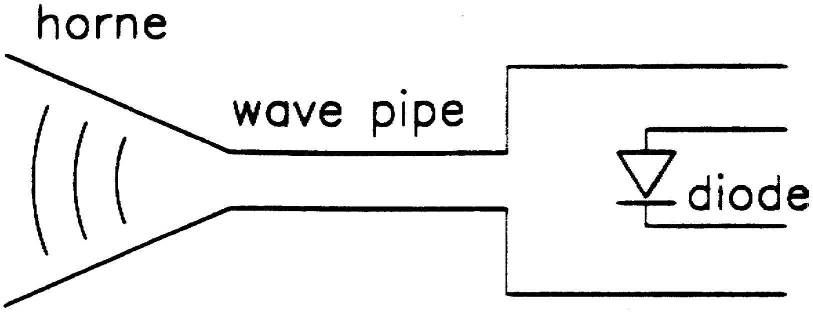

Figure 52: *Schematic view of a radio telescope, with the receiving horne, the wave tube and the detector diode.*

in which T_b is the *brightness temperature*. (Note that although the last expression contains λ, the units remain however per unit frequency, e.g. Watt$\cdot$m$^{-2}\cdot$sr$^{-1}\cdot$Hz^{-1}. The convenience of using λ becomes apparent when integration over the beamsize is performed, see below.) In this way a radio brightness distribution $B(\nu, \vec{\Omega})$ on the sky can be described by an equivalent brightness temperature distribution $T_b(\vec{\Omega})$.
Radio waves arriving at the focus of a telescope enter the receiver input by a horn (or "feed") which matches the impedance of the vacuum to that of a wave-tube, that selects one degree of polarisation. The wave is subsequently guided through the wave-tube into a resonance cavity, which defines by its selectivity a bandwidth $\Delta\nu$ centered around a frequency ν_0 of the incoming radiation. A (always) non-linear detection element (transducer) in this resonance cavity converts the wave field (i.e. the electric vector) into a current. This is the simplest detection configuration for a radio telescope, nevertheless it suffices to describe the essential characteristics of the measurement process, see figure 52. The power reaching the non-linear element is in good approximation given by:

$$P(\nu) = \frac{1}{2}\,\eta(\nu)\,A(\nu)\,\frac{2k}{\lambda^2}\int\limits_{\Omega_{beam}} T_b(\vec{\Omega})\,d\vec{\Omega} \equiv \eta(\nu)kT_{ant} \tag{184}$$

in which the factor $\frac{1}{2}$ refers to a single polarisation component, $A(\nu)$ represents the telescope effective area at frequency ν, Ω_{beam} is the diffraction limited beamsize and $\eta(\nu)$ takes account of the frequency dependent transmission losses prior to detection by the non-linear element. $P(\nu)$ is given in units of Watt$\cdot$Hz^{-1}. Expression 184 can be set equal to the product of the transmission losses $\eta(\nu)$ and a thermal power kT_{ant} (per unit frequency) available at the receiver input. T_{ant} is called the *antenna temperature* and is therefore independent of the transmission in the receiver system. It should be noted that T_{ant} is only equal to the physical temperature T_b of the radio source if the following conditions are met:

- The source is optically thick at the frequency considered.
- The source is sufficiently extended to fill the diffraction limited beamsize Ω_{beam} of the telescope, i.e. Ω_{beam} satisfies $A(\nu)\ \Omega_{beam} = \lambda^2$, the 'etendue' of coherence introduced in section 7.

If the radiation source is not a blackbody but has an arbitrary spectrum (e.g. an optically thin thermal source or a non-thermal source), the antenna temperature T_{ant} becomes frequency dependent and does not relate to a physical temperature. The power received at the detection element is now expressed as

$$P(\nu) = \eta(\nu) k T_{ant}(\nu) \tag{185}$$

in which $T_{ant}(\nu)$ is the antenna temperature at the specific frequency ν. In this picture the antenna may be thought to be replaced by its characteristic resistance R_a at a fictitious temperature $T_{ant}(\nu)$ producing a thermal noise power $kT_{ant}(\nu)$ per unit frequency at frequency ν or kT_{ant} if the power received is white (frequency independent) over the bandwidth $\Delta\nu$ considered. In turn, this replacement is equivalent to a fictitious noise free resistor R_a in series with a voltage source which generates a power spectral density, in units of Volt^2·Hz^{-1}:

$$S_V(\nu) = 4kT_{ant}(\nu) R_a \tag{186}$$

It is important to realize that $T_{ant}(\nu)$ has generally nothing to do with the physical temperature of the resistor R_a. To demonstrate this, consider an amplifier (bandwidth $\Delta\nu$ = 1 MHz), with an output impedance R = 50 Ω, generating a rms-white noise voltage σ_V = 1 mV. The available power per Hz at the output of this amplifier follows from

$$P = \frac{\sigma_V^2}{4\ R\ \Delta\nu} = 5 \cdot 10^{-15} \quad Watt \cdot Hz^{-1} \tag{187}$$

The corresponding noise temperature is

$$T_{ant} = \frac{P}{k} \approx 3.5 \cdot 10^8 \quad K \tag{188}$$

whereas the physical temperature is about 300 Kelvin.
Only in the case of passive elements (i.e. which do not need energy), like wave guides, resistor networks and transmission cables embedded in the earth, the noise temperature and the physical temperature are about equal.

10.2.2 Noise sources at radio wavelengths

In practise, if the radio antenna is pointed at a sky region devoid of sources, a non-zero output will be measured at the output of the receiver. This arises from various noise sources:

- Residual thermal emission from the atmosphere and from the telescope itself. A simple approximation is available when the atmosphere is optically thin, i.e. if the optical depth $\tau(\nu)$ along the zenith is small compared to unity. In this

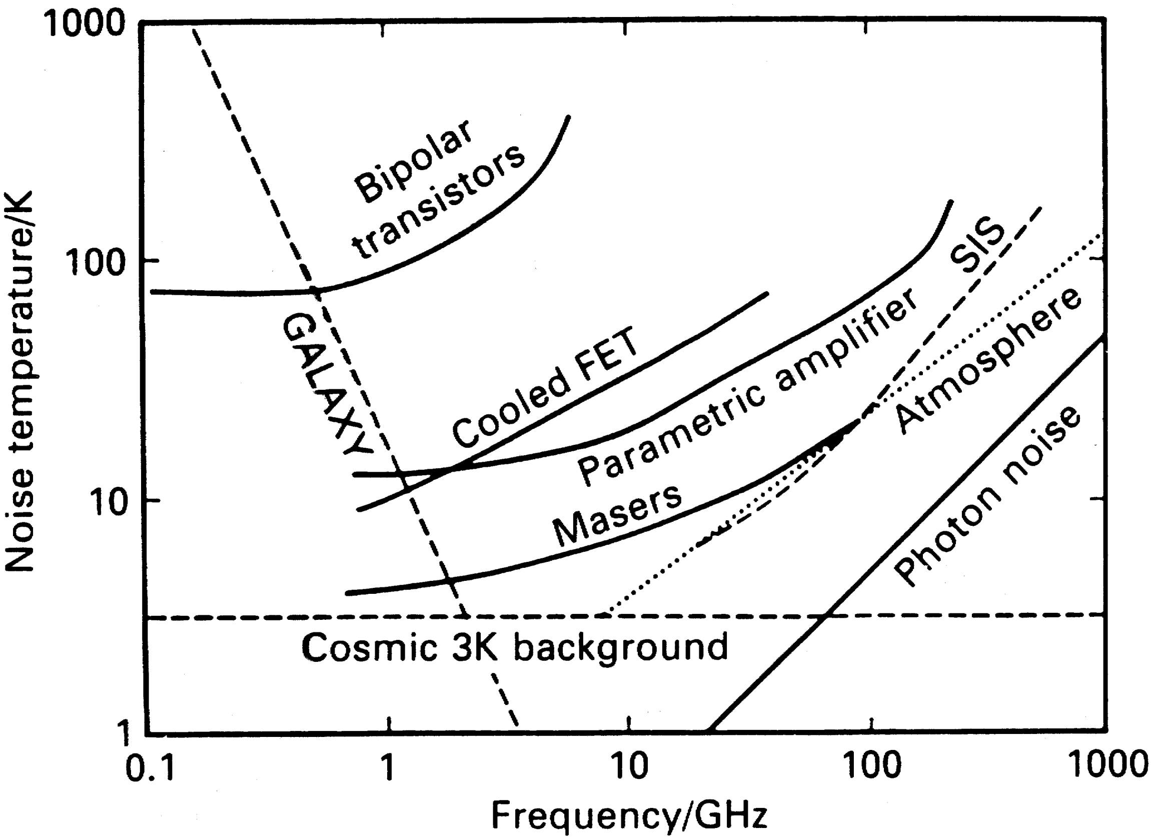

Figure 53: *Frequency dependence of the noise temperature for several sources of background radiation and receiver components. Figure taken from Longair 1992.*

case, the specific intensity $I(\nu) \approx \tau(\nu) B(\nu, \bar{T}_{atm})$, with $B(\nu, \bar{T}_{atm})$ the blackbody function at the average temperature $\bar{T}_{atm}$ of the atmosphere. Since $\tau(\nu)$ is a strong function of wavelength, the contribution of the atmospheric thermal emission to the noise is also a strong function of wavelength and so is the associated noise temperature $T_{atm}(\nu)$. For example at $\nu = 100$ GHz ($\lambda \approx 3$ mm), $\tau(\nu) \approx 0.2$, i.e. the atmospheric emission corresponds to an antenna temperature of about 50 Kelvin. Figure 53 shows, among others, the frequency dependence of the noise temperature associated with atmospheric emission.

- Thermal emission from the telescope environment and the earth surface, detected in the side lobes of the diffraction pattern of the telescope's angular response function (diffraction limit). This contribution to the antenna temperature is very site dependent. Symbol: $T_{lobe}(\nu)$.

- Thermal noise from the Galaxy and the microwave background. The Galactic contribution is again strongly frequency dependent, the microwave background is white. Both components are also indicated in figure 53. Symbol: $T_{Gal}(\nu)$ and T_{cb}.

- Contributions from the receiver chain, i.e. feed, wave guides, detection element, amplifiers, etc. This contribution to the total noise is normally expressed as an *effective noise temperature* $T_{eff}(\nu)$. The value of $T_{eff}(\nu)$ is derived in the following way. Suppose that the total noise power available at the antenna $P_{ant}(\nu)$, is subject to a power amplification $G(\nu)$ in the receiver chain. At the output stage of the receiver, the total power $P_{out}(\nu)$ is then given by

$$P_{out}(\nu) = G(\nu) \cdot P_{ant}(\nu) + P_{rec}(\nu) = G(\nu) \cdot \left[P_{ant}(\nu) + \frac{P_{rec}(\nu)}{G(\nu)} \right] \tag{189}$$

in which $P_{rec}(\nu)$ represents the noise power introduced by the receiver chain. Division by Boltzmann's constant k yields:

$$T_{out}(\nu) = G(\nu) \cdot [T_{ant}(\nu) + T_{eff}(\nu)] = G(\nu) \cdot T_{op}(\nu) \tag{190}$$

$T_{out}(\nu)$ equals the noise temperature at the output of the whole telescope system (antenna + receiver chain), $T_{eff}(\nu)$ is the effective noise temperature introduced above. $T_{eff}(\nu)$ can be regarded as the temperature which has to be assigned to the characteristic resistance R_a of the antenna in order to have a fictitious *noise-free* system produce the same noise temperature as the noisy system for $T_{ant} =$ 0 Kelvin. $T_{op}(\nu)$ represents in a similar way the temperature to be assigned to R_a to have a noise free system produce the same noise temperature as the noisy system at the actual value of the antenna temperature T_{ant}:

$$T_{op}(\nu) = T_{ant}(\nu) + T_{eff}(\nu) \tag{191}$$

$T_{op}(\nu)$ is the so-called *operational temperature* of the telescope system. This relation effectively contains all essential noise contributions: $T_{ant}(\nu)$ includes all components due to radiation noise, $T_{eff}(\nu)$ includes all components referring to the receiver chain. Figure 53 shows some typical values of $T_{eff}(\nu)$ for masers, cooled field-effect transistors (FET) used in the first amplifier stage, parametric amplifiers and superconducting mixer elements (SIS) which are employed in case of a heterodyne detection chain.

Note 1: If the receiver chain consists of a series of n noisy elements, each with a power gain $G_i(\nu)$ and an associated effective temperature $T_{eff_i}(\nu)$, the operational temperature $T_{op}(\nu)$ derives from:

$$T_{op}(\nu) = T_{ant}(\nu) \; + \; T_{eff_1}(\nu) \; + \; \frac{T_{eff_2}(\nu)}{G_1(\nu)} \; + \; \cdots \; + \; \frac{T_{eff_n}(\nu)}{\prod_{i=1}^{n-1} G_i(\nu)} \tag{192}$$

This is the cascade rule according to Friis (verify yourself). Expression 192 shows that the first stages in the receiver chain are mainly determining the noise behavior of the system, once the amplification is sufficiently high, the noise contributions of additional stages become small. In particular the input-stage is crucial with respect to the achievable noise performance.

Note 2: The power gains $G_i(\nu)$ may partly be smaller than unity, this represents transmission losses in passive components like wave tubes, transmission lines and resistor networks: $G_i(\nu) = \eta_i(\nu)$. These losses can seriously degrade the noise characteristics of the system since the physical temperature of the attenuating components enters the determination of $T_{op}(\nu)$. According to Kirchhoff's *reciprocity law*, in equilibrium the *absorbed power* by the passive components equals the *radiated power*. Hence, the contribution of a single loss-element at physical temperature T_L to the noise can be expressed as $(1 - \eta(\nu))T_L$ or $\frac{1}{L(\nu)}[L(\nu) - 1]T_L$, with $L(\nu) = \eta^{-1}(\nu)$ the frequency dependent loss-factor (by definition larger than unity). Employing this single loss-element behind an antenna yields a system output noise given by the expression:

$$T_{out}(\nu) = \frac{1}{L(\nu)} \; (T_{ant}(\nu) + [L(\nu) - 1]T_L) \tag{193}$$

i.e. $T_{eff}(\nu) = [L(\nu) - 1]\, T_L$, and

$$T_{op}(\nu) = T_{ant}(\nu) + [L(\nu) - 1]T_L \tag{194}$$

To show that this potentially strongly influences the value of $T_{op}(\nu)$, consider the following example.
Suppose the telescope is pointed at a cloudless sky region devoid of sources with $T_{ant}(\nu) = 20$ K. If a small piece of wave tube between the antenna and the input of the amplifier introduces a loss of 10 %, i.e. $L(\nu) = 1.1$, and this wave tube is at room temperature ($T \approx 290$ K), the operational temperature becomes $T_{op}(\nu) = 20 + 0.1 \cdot 290 \approx 50$ K; a deterioration of a factor 2.5. If the wave tube is cooled with liquid Nitrogen ($T \approx 70$ K), $T_{op}(\nu)$ becomes about 27 K (35 % deterioration) and at liquid Helium temperature ($T \approx 4$ K), $T_{op}(\nu) = 20$ K, i.e. practically no deterioration. The importance of cooling of passive components at the receiver chain front-end is therefore amply demonstrated.

10.2.3 The *SNR*-degradation factor

With the above assessment of the various noise contributions, the total operational noise temperature is expressed as:

$$T_{op}(\nu) = T_{Gal}(\nu) + T_{cb} + T_{atm}(\nu) + T_{lobes}(\nu) + T_{eff}(\nu) \tag{195}$$

This represents the background power against which a potential source signal will have to be detected. If the signal power (i.e. radiation flux) is given by $\Phi_s(\nu) = kT_s(\nu)$, the *momentaneous SNR* at the input of the receiver chain is given by the ratio between the available source power and the power contained in the background radiation fields:

$$SNR_{in}(\nu) = \frac{T_s(\nu)}{T_{Gal}(\nu) + T_{cb} + T_{atm(\nu)} + T_{lobes}(\nu)} \tag{196}$$

The noise generated by the receiver chain degrades the *SNR* at the output to:

$$SNR_{out}(\nu) = \frac{G(\nu) \cdot T_s(\nu)}{G(\nu) \cdot [T_{Gal}(\nu) + T_{cb} + T_{atm}(\nu) + T_{lobes}(\nu) + T_{eff}(\nu)]} \tag{197}$$

The *SNR*-degradation factor $D(\nu)$ (≥ 1) is now simply expressed as

$$D(\nu) = \frac{SNR_{in}(\nu)}{SNR_{out}(\nu)} = \frac{T_{op}(\nu)}{T_{ant}(\nu)} \tag{198}$$

Referring to the example in Note 2 in the previous subsection, the deterioration of the *SNR* due to the loss in the wave tube would be specified by $D(\nu) \approx 2.5$ (room temperature), $D(\nu) \approx 1.35$ (liquid Nitrogen) and $D(\nu) \approx 1$ (liquid Helium). The *momentaneous SNR* $= \frac{T_s(\nu)}{T_{op}(\nu)}$ refers to the value obtained considering the radiation energy during one second (e.g. Watt) and in a bandwidth of 1 Hz. In reality this is of course greatly improved by integration over the selected bandwidth $\Delta\nu$ and over an exposure period T_{obs}. The derivation of this *frequency (band) limited SNR* is the subject of the next paragraph.

10.2.4 Derivation of the band limited noise in the thermal limit: signal to noise ratio and limiting sensitivity

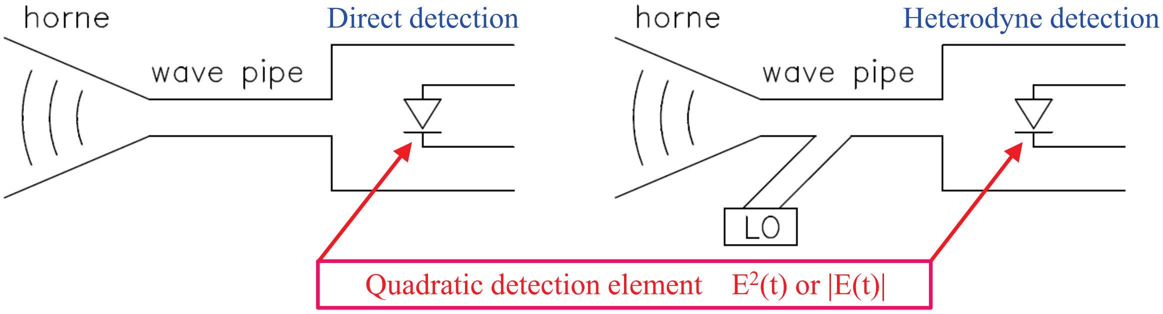

Figure 54: *Schematic view of a radio receiver with the receiving horne, which selects the frequency ν_s and the frequency bandwidth $\Delta\nu_s$, the wave pipe, and the 'quadratic' detection element (left). By mixing the radio signal with that of a local oscillator (LO), the carrier frequency can be shifted to much lower frequencies without any information loss (heterodyne detection), suitable for electronic processing (right).*

The average spectral noise power in the thermal limit ($h\nu \ll kT$) for one degree of polarization, equals kT. This is the situation which prevails in radio-, microwave- and submillimeter receivers and hence gives rise to a description of signals and noise with the aid of characteristic temperatures, e.g. source temperature T_s and noise temperature T_n. The one-sided power spectral density $S(\nu) \equiv \overline{P}(\nu) = kT$ Watt Hz^{-1} is constant as a function of frequency and, as a consequence, is termed white noise.

In practice, due to the finite frequency response of any receiver system, this will be frequency limited. Hence we can express the double-sided power spectral density for a thermal source for one degree of polarization as:

$$S_d(\nu) = \frac{1}{2}kT\Pi\left(\frac{\nu}{2\nu_c}\right) \quad \text{with} \quad \nu_c \ll \frac{kT}{h} \tag{199}$$

where ν_c constitutes the cut-off frequency of the receiver system under consideration, i.e. the total energy contained in the power spectrum remains finite, as it should be for any physical system.

Continuing now with the detection of radio signals, we wish to consider the signal-to-noise ratio. A linearly polarized signal can mathematically be expressed by the real function:

$$E(t) = E_0(t)\cos(2\pi\bar{\nu}t + \phi(t)) \tag{200}$$

The amplitude $w(t) \equiv E_0(t)$ of the quasi-monochromatic wave is a *wide-sense stationary Gaussian* random time function of zero mean. Moreover the stochastic process is assumed to be *mean- and correlation-ergodic*, i.e. for an arbitrary real stochastic variable $w(t)$ its *expectation value* at time t, $\mathbf{E}\left\{w(t)\right\}$, can be interchanged with its time average.

Detection of such a radio signal requires, like in the case of *mixing*, a non-linear operation like $\psi(t) \equiv |w(t)|$ or $\psi(t) = w^2(t)$, to extract the power present in the signal. We shall consider here the case of quadratic detection, i.e. $\psi(t) = w^2(t)$, since this is mathematically straightforward in contrast to absolute-value transformations.

If $w(t)$ can be described as a stationary random process with a normally distributed amplitude around zero mean, the probability density function is given by:

$$f(w) = \frac{1}{\sigma_w\sqrt{2\pi}}e^{-w^2/2\sigma_w^2} \qquad (\mu_w = 0) \tag{201}$$

and the measuring process is schematically indicated by

$$w(t) \rightarrow \text{transformation} \rightarrow \psi(t) \equiv w^2(t) \tag{202}$$

For the transformation $\psi(t) = w^2(t)$, with $\sigma_w^2 = R_w(0)$, we can write the probability density of ψ as

$$f(\psi) = \frac{1}{2[2\pi R_w(0)\psi]^{1/2}}\exp\left[\frac{-\psi}{2R_w(0)}\right]U(\psi) \tag{203}$$

with $U(\psi)$ the Heaviside step function. (In understanding expression (203), don't forget the transformation $dw \Rightarrow d\psi$!). Thus, the stochastic process $\psi(t)$ is apparently *not* normally distributed, and of course also $\mu_\psi \neq 0$.

To derive the power spectral density of $\psi(t)$ we need to find an expression for the autocorrelation function $R_\psi(\tau)$ of $\psi(t)$. One can show that for a normally distributed $w(t)$ the autocorrelation of $\psi(t) \equiv w^2(t)$ follows from:

$$\begin{aligned} R_\psi(\tau) &= \mathbf{E}\left\{\psi(t)\psi(t+\tau)\right\} = \mathbf{E}\left\{w^2(t)w^2(t+\tau)\right\} = \\ &= \mathbf{E}\left\{w^2(t)\right\}\mathbf{E}\left\{w^2(t+\tau)\right\} + 2\mathbf{E}^2\left\{w(t)w(t+\tau)\right\} \end{aligned} \tag{204}$$

The derivation of this relation applying the theory of stochastic processes makes use of the so-called moment-generating functions of $w(t)$. Hence:

$$R_\psi(\tau) = R_w^2(0) + 2R_w^2(\tau) = \mu_\psi^2 + C_\psi(\tau) \tag{205}$$

The average μ_ψ of $\psi(t)$ equals the variance of $w(t)$, the autocovariance of $\psi(t)$ equals twice the square of the autocovariance of $w(t)$, and the variance of $\psi(t)$ is $\sigma_\psi^2 = 2\sigma_w^4$. The *double-sided* power spectral density of $\psi(t)$ follows from the Wiener-Khinchin theorem:

$$S_{d_\psi}(\nu) = R_w^2(0)\delta(\nu) + 2S_{d_w}(\nu) * S_{d_w}(\nu) \tag{206}$$

It is important to realize that in the case of quadratic detection a number of frequency components is introduced in the case of an amplitude-modulated signal like $w(t)$. We shall demonstrate this, as an example, for a deterministic signal, i.e. an amplitude-modulated high frequency carrier of the form:

$$x(t) = A(1 + m\cos\eta t)\cos\omega t \tag{207}$$

This represents a high frequency carrier (ω) the amplitude (A) of which is modulated by a much lower frequency (η). m is called the modulation index.
The signal $x(t)$ contains three discrete frequencies, $\omega - \eta$, ω, and $\omega + \eta$. This is easily seen by using $\cos\alpha\cos\beta = \frac{1}{2}[\cos(\alpha+\beta) + \cos(\alpha-\beta)]$.

The instantaneous intensity of this signal can be expressed as $I = x^2(t)$, and for the average intensity we get

$$\overline{I} = \overline{x^2(t)} = \frac{1}{2}A^2(1 + \frac{1}{2}m^2) \tag{208}$$

The equivalent of the term $\mathbf{E}\left\{a^2(t)\right\}\mathbf{E}\left\{a^2(t+\tau)\right\}$ in expression (204) amounts in this case to a DC-term representing the square of the average intensity:

$$\overline{I}^2 = \frac{1}{4}A^4\left(1 + \frac{1}{2}m^2\right)^2 \tag{209}$$

To arrive at the equivalent of the autocovariance term $2\mathbf{E}^2\left\{a(t)a(t+\tau)\right\}$ in equation (204), we first have to compute the autocovariance $C_x(\tau)$ of $x(t)$:

$$C_x(\tau) = \overline{x(t)x(t+\tau)} = \frac{1}{2}A^2[\cos\omega\tau + \frac{1}{4}m^2\cos(\omega+\eta)\tau + \frac{1}{4}m^2\cos(\omega-\eta)\tau] \tag{210}$$

Subsequently:

$$\overline{x(t)x(t+\tau)}^2 = \frac{1}{4}A^4(1 + \frac{1}{2}m^2\cos\eta\tau)^2\cos^2\omega\tau \tag{211}$$

Using the relation $\cos^2\alpha = \frac{1}{2}(\cos 2\alpha + 1)$ this can be disentangled in various components:

$$\begin{aligned}
\overline{x(t)x(t+\tau)}^2 &= \frac{A^4}{8}(1 + \cos 2\omega\tau) + \frac{m^2A^4}{8}[\cos\eta\tau + \frac{1}{2}\cos(2\omega-\eta)\tau + \\
&+ \frac{1}{2}\cos(2\omega+\eta)\tau] + \frac{m^4A^4}{64}[1 + \cos 2\eta\tau + \frac{1}{2}\cos(2\omega-2\eta)\tau + \\
&+ \cos 2\omega\tau + \frac{1}{2}\cos(2\omega+2\eta)\tau]
\end{aligned} \tag{212}$$

This expression shows that the original frequencies in the power spectral density of $x(t)$, ω and $\omega \pm \eta$ have been transformed in the squaring process to frequency components at DC, η, 2η, 2ω, $2\omega \pm \eta$ and $2\omega \pm 2\eta$. Displayed on a double-sided frequency diagram this shows twice the original frequency bandwidth centered at DC and $\pm 2\omega$ (see figure 56). In the case of the stochastic signal $w(t)$ we shall see the same characteristics in the power spectral density of the autocovariance $C_\psi(\tau)$.

Consider now the front-end of a receiver behind a radio antenna (telescope), see (Figure (54). This front-end generally contains a resonance cavity tuned at a central frequency $\overline{\nu} = \nu_s$ with a bandwidth $\Delta\nu_s$. If the noise entering the receiver can be

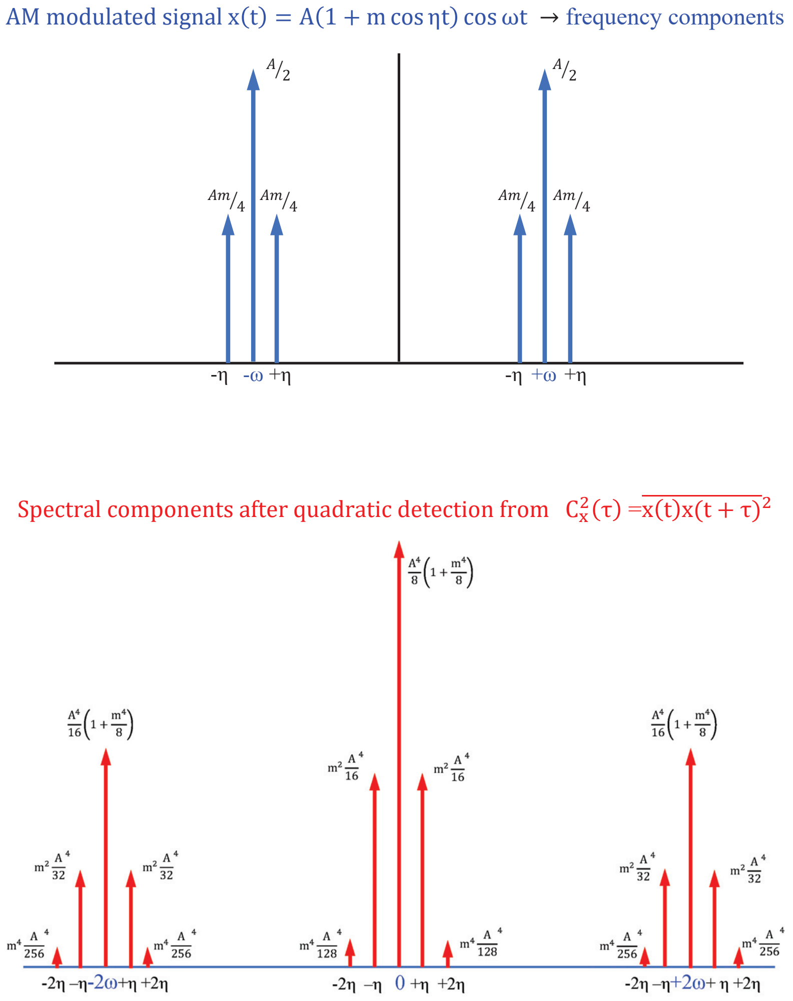

Figure 55: *Frequency components of an amplitude modulated (frequency η) carrier (frequency ω) with modulation index m (upper panel). The spectral components that emerge after quadratic detection present in the autocovariance function are shown in the lower panel. Evidently the sidebands appear at twice the carrier frequency ω, moreover the signal bandwidth 4η amounts to twice the original bandwidth 2η.*

characterized by a noise temperature T_n the double-sided power spectral density of $w(t)$ for one degree of polarization is given by:

$$S_{d_w}(\nu) = \frac{1}{2}kT_n\left[\Pi\left(\frac{\nu-\nu_s}{\Delta\nu_s}\right) + \Pi\left(\frac{\nu+\nu_s}{\Delta\nu_s}\right)\right] \tag{213}$$

This signal is then fed to a non-linear detection element, like a Schottky-diode or an induction coil, which introduces the transformation $\psi(t) = w^2(t)$. Consequently, we have:

$$R_w^2(0) = (\sigma_w^2)^2 = \left(\frac{1}{2}kT_n \int\limits_{-\infty}^{+\infty} \left[\Pi\left(\frac{\nu-\nu_s}{\Delta\nu_s}\right) + \Pi\left(\frac{\nu+\nu_s}{\Delta\nu_s}\right)\right] d\nu\right)^2 = (kT_n\Delta\nu_s)^2 \qquad (214)$$

and

$$\begin{aligned} 2\left[S_{d_w}(\nu) * S_{d_w}(\nu)\right] &= \frac{1}{2}(kT_n)^2 \left[\Pi\left(\frac{\nu-\nu_s}{\Delta\nu_s}\right) + \Pi\left(\frac{\nu+\nu_s}{\Delta\nu_s}\right)\right] * \left[\Pi\left(\frac{\nu-\nu_s}{\Delta\nu_s}\right) + \Pi\left(\frac{\nu+\nu_s}{\Delta\nu_s}\right)\right] \\ &= (kT_n)^2\Delta\nu_s \left[\Lambda\left(\frac{\nu}{\Delta\nu_s}\right) + \frac{1}{2}\Lambda\left(\frac{\nu-2\nu_s}{\Delta\nu_s}\right) + \frac{1}{2}\Lambda\left(\frac{\nu+2\nu_s}{\Delta\nu_s}\right)\right] \qquad (215) \end{aligned}$$

$S_{d_\psi}(\nu)$ consists therefore of a component $(kT_n\Delta\nu_s)^2\delta(\nu)$, a *time independent average value at zero frequency* (stationary Gaussian random amplitudes) and three 'triangle-functions' centered at zero frequency and at frequencies $-2\nu_s$ and $+2\nu_s$ with a basewidth of $2\Delta\nu_s$, as illustrated in figure 56. Note the correspondence in frequency shift and bandwidth with the example involving the deterministic amplitude modulated signal above! In practice one always has $\nu_s \gg \Delta\nu_s$, in the centimeter range for example one typically has $\nu_s \simeq 10^{10}$ Hz and $\Delta\nu_s \simeq 10^7$ Hz. The detection of a *potential radio source*

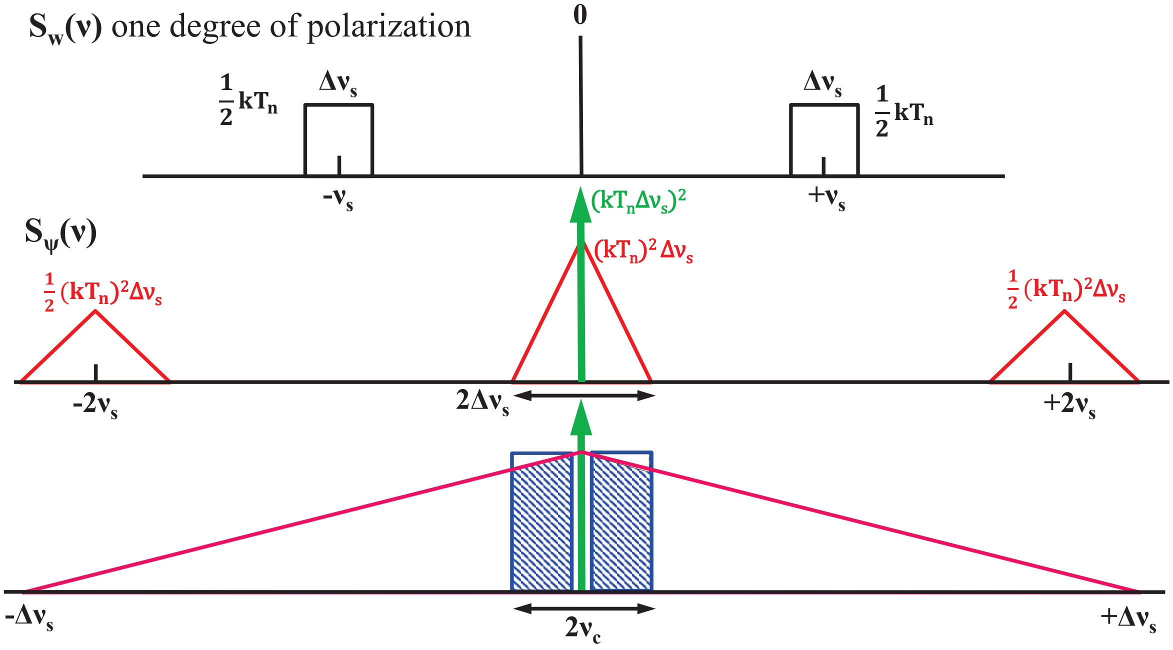

Figure 56: *Upper panel: Double-sided power spectral density of thermal radiation for one degree of polarization ($\bar{P} = kT$ Watt Hz^{-1}), over a channel bandwidth $\Delta\nu_s$ centered at frequency ν_s, incident on a non-linear detection element. Middle panel: double-sided power spectral density at the output of the detection element. Lower panel: low frequency filtering (cut-off ν_c) providing signal averaging over a time interval $\Delta T_{av} = 1/2\nu_c$. This time averaging process obviously does not influence the DC-component, this is schematically indicated in the figure by showing an exclusion of the green arrow.*

signal will then have to be assessed in the context of the *autocovariance of the noise signal*, i.e. we need an expression for $C_\psi(\tau)$. This follows from the Fourier transform of the term $2\left[S_{d_w}(\nu) * S_{d_w}(\nu)\right]$ in equation (215). Applying the shift theorem and the $\Lambda \Leftrightarrow \mathrm{sinc}^2$ transform from Fourier analysis, we get:

$$\begin{aligned} C_\psi(\tau) &= (kT_n\Delta\nu_s)^2\left[\frac{e^{-2\pi i(2\nu_s\tau)}+e^{2\pi i(2\nu_s\tau)}}{2}\right]\mathrm{sinc}^2\tau\Delta\nu_s + (kT_n\Delta\nu_s)^2\mathrm{sinc}^2\tau\Delta\nu_s \\ &= (kT_n\Delta\nu_s)^2(1+\cos 2\pi(2\nu_s\tau))\mathrm{sinc}^2\tau\Delta\nu_s \end{aligned} \tag{216}$$

The $\cos 2\pi(2\nu_s\tau)$ term refers to the high frequency carrier which will be filtered off in any low frequency averaging process. This averaging process can be taken over an arbitrary time intervalΔT_{av}. This is equivalent to filtering in the frequency domain with a filter $\Pi(\nu/2\nu_c)$ with $\nu_c = 1/(2\Delta T_{av})$ commensurate with the Nyquist sampling theorem.

The averaged value of the autocovariance is then obtained in the τ domain by convolution of $C_\psi(\tau)$ with the Fourier transform $\Pi(\nu/2\nu_c) \Leftrightarrow 2\nu_c\mathrm{sinc}2\nu_c\tau$.

Assuming $\nu_c \ll \Delta\nu_s \ll \nu_s$, the $\cos 2\pi(2\nu_s\tau)$ term in expression (216) averages to zero. Consequently we have:

$$\begin{aligned} \left[C_\psi(\tau)\right]_{\Delta T} &= (kT_n\Delta\nu_s)^2\mathrm{sinc}^2\tau\Delta\nu_s * 2\nu_c\mathrm{sinc}2\nu_c\tau \\ &= (kT_n\Delta\nu_s)^2\cdot 2\nu_c\int\limits_{-\infty}^{+\infty}\mathrm{sinc}^2\tau'\Delta\nu_s\mathrm{sinc}2\nu_c(\tau-\tau')d\tau' \end{aligned} \tag{217}$$

and with a change of variables $u' \equiv \tau'\Delta\nu_s$ this becomes:

$$\left[C_\psi(u)\right]_{\Delta T} = (kT_n)^2\Delta\nu_s\cdot 2\nu_c\int\limits_{-\infty}^{+\infty}\mathrm{sinc}^2u'\mathrm{sinc}\frac{2\nu_c}{\Delta\nu_s}(u-u')du' \tag{218}$$

Since $\nu_c/\Delta\nu_s \ll 1$, $\mathrm{sinc}(2\nu_c/\Delta\nu_s)u'$ varies very slowly compared to sinc^2u'. We may therefore regard sinc^2u' as a δ-function in comparison to $\mathrm{sinc}(2\nu_c/\Delta\nu_s)u'$. Moreover we also have the proper normalization, since $\int\limits_{-\infty}^{\infty}\mathrm{sinc}^2u'du' = \int\limits_{-\infty}^{+\infty}\delta(u')du' = 1$. Applying this approximation we get:

$$\begin{aligned} \left[C_\psi(u)\right]_{\Delta T} &= (kT_n)^2\Delta\nu_s\cdot 2\nu_c\int\limits_{-\infty}^{+\infty}\delta(u')\mathrm{sinc}\frac{2\nu_c}{\Delta\nu_s}(u-u')du' \\ &= (kT_n)^2\Delta\nu_s\cdot 2\nu_c\mathrm{sinc}\frac{2\nu_c}{\Delta\nu_s}u \end{aligned} \tag{219}$$

Substituting $\tau = u/\Delta\nu_s$ we arrive at the final expression for the ΔT-averaged value of the noise autocovariance:

$$\left[C_\psi(\tau)\right]_{\Delta T} = (kT_n)^2\Delta\nu_s\cdot 2\nu_c\mathrm{sinc}2\nu_c\tau \tag{220}$$

The noise variance $\left[C_\psi(0)\right]_{\Delta T} = (kT_n)^2\Delta\nu_s\cdot(2\nu_c)$ should be compared to the strength of a radio source signal characterized by a source temperature T_s. The average value of this source signal after quadratic detection follows from:

$$(\mu_\psi)_s = R_{w_s}(0) = \sigma^2_{w_s} = kT_s\Delta\nu_s \tag{221}$$

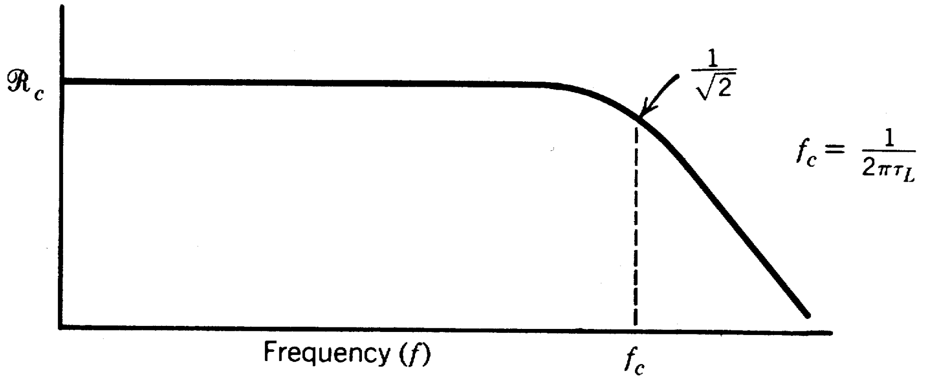

Figure 57: *Frequency spectrum of the noise of a transducer. Noise power is along the vertical axis. Figure taken from Dereniak & Crowe 1984.*

and the signal-to-noise ratio thus becomes:

$$\mathrm{S/N} = \frac{(\mu_\psi)_s}{[C_\psi(0)]^{1/2}_{\Delta T}} = \frac{T_s}{T_n}\left(\frac{\Delta\nu_s}{2\nu_c}\right)^{1/2} \tag{222}$$

i.e. the signal to noise is proportional to the square root of the receiver bandwidth $\Delta\nu_s$ and inversely proportional to the double-sided bandwidth of the integrating (averaging) low-pass filter ν_c. Introducing $\Delta T_{av} = 1/(2\nu_c)$ (Nyquist sampling) we get:

$$\mathrm{S/N} = \frac{T_s}{T_n}\left(\Delta\nu_s \Delta T_{av}\right)^{1/2} \tag{223}$$

i.e. the S/N ratio improves with the square root of the product of the radio-channel bandwidth and the integration time ΔT_{av}. In practice the noise temperature of a radio-wave receiving system is designated as the system or operational temperature that includes the contributions to the noise of the sky, the antenna and the receiver system. The S/N =1 sensitivity for detecting a radio source against the system thermal noise temperature is sometimes referred to as the *radiometer equation*:

$$\mathrm{S/N} = \frac{T_{source}}{T_{system}}\left(\Delta\nu_s \Delta T_{av}\right)^{1/2} \quad \text{radiometer equation!} \tag{224}$$

The minimum detectable *source power for one degree of polarization* with a signal to noise ratio of one is given by:

$$(P_s)_{min} = kT_n\left(\Delta\nu_s \Delta T_{av}\right)^{-1/2} \tag{225}$$

10.3 Power characterisation

10.3.1 Typical set-up of observation

Consider the case of incoherent detection of a radiation field by a transducer which converts the incident radiant power into an electrical output (usually a current or a voltage). Two types of flux will be used to characterise the radiation source:

- Monochromatic flux $\Phi(\lambda_0)$ at a particular wavelength λ_0, defined as:

$$\Phi(\lambda_0) = \int\limits_0^\infty \Phi(\lambda) \cdot \delta(\lambda - \lambda_0)\, d\lambda \tag{226}$$

- Blackbody flux $\Phi(T)$, which is the specific flux integrated over a blackbody profile:

$$\Phi(T) = \int\limits_0^\infty \Phi_{bb}(\lambda, T)\, d\lambda \tag{227}$$

The incoming signal is modulated (e.g. by a chopper) with a fixed frequency f_{chop}. This signal is now the input for the transducer. This transducer has a spectral bandwidth $\lambda\lambda$ over which its integrates the incoming flux. It produces a time-dependent output voltage V_{out} or current I_{out}, which also contains the noise of the transducer. If thermal noise (so-called Johnson noise) dominates, this noise is "white" in the temporal frequency domain: kT Watt·Hz^{-1} up to a cut-off frequency f_c, see figure 57. The *frequency bandwidth* of the transducer is in this case $\Delta f = f_c$. Since the source signal is periodic with known period $T = \frac{1}{f_{chop}}$, the outcoming signal can be folded *modulo* this period. In this process all transducer noise at other temporal frequencies can be filtered out.

$$\boxed{\text{Source}} \overset{\Phi}{\Longrightarrow} \boxed{\text{Chopper}} \Longrightarrow \boxed{\text{Transducer}}$$

10.3.2 Responsivity

For the case of incoherent detection a basic figure of merit is *responsivity*. This is the ratio of the electrical output (in Amperes or Volts) to the radiant input, i.e. total radiation flux (in Watts).

The *spectral voltage responsivity* of a detection system at a particular output wavelength λ_0 is the measured voltage output $V_{out}(f)$ divided by the monochromatic radiation flux $\Phi(\lambda_0, f)$ incident on the transducer:

$$R_V(\lambda_0, f) = \frac{V_{out}(f)}{\Phi(\lambda_0, f)} \qquad (in\ Volt \cdot Watt^{-1}) \tag{228}$$

Since the transducer generally has a limited frequency response or time resolving power (in figure 57 it is flat up to a cut-off frequency f_c) the output V_{out} will depend on the temporal behaviour of $\Phi(\lambda_0)$, which can be modulated with a chopping frequency f. For instance, if $f \gg f_c$, the response of the transducer will be zero. Simularly, the *spectral current responsivity* is defined as:

$$R_I(\lambda_0, f) = \frac{I_{out}(f)}{\Phi(\lambda_0, f)} \qquad (in\ Ampere \cdot Watt^{-1}) \tag{229}$$

Alternatively, a *blackbody responsivity* $R(T, f)$ can be defined, which represents the detector output signal divided by the incident radiation flux from a blackbody source $\Phi(T)$ modulated by frequency f:

$$R_V(T, f) = \frac{V_{out}(f)}{\Phi(T, f)} \tag{230}$$

The relation of $R_V(T,f)$ with $R_V(\lambda_0,f)$ can be easily seen by computing the voltage at the output of the transducer through integration of the spectral responsivity over a blackbody spectrum:

$$V_{out} = \int\limits_0^\infty R_V(\lambda_0,f)\ \Phi_{bb}(\lambda_0,T)\ d\lambda_0 \equiv R_V(T,f)\ \int\limits_0^\infty \Phi_{bb}(\lambda_0,T)\ d\lambda_0 \qquad (231)$$

hence,

$$R_V(T,f) = \frac{\int\limits_0^\infty R_V(\lambda_0,f)\ \Phi_{bb}(\lambda_0,T)\ d\lambda_0}{\int\limits_0^\infty \Phi_{bb}(\lambda_0,T)\ d\lambda_0} = \overline{R_V(\lambda_0,f)} \qquad (232)$$

which shows that $R_V(T,f)$ is obtained by averaging $R_V(\lambda_0,f)$ over a blackbody spectral distribution. Note that the blackbody responsivity is a measure of the detector response to incident radiation integrated over all wavelengths even though the transducer is only sensitive to a finite wavelength interval.

Example: consider a steady blackbody source with area A_s irradiating a sensor (transducer) area A_{tr}, both areas are normal to the optical axis connecting them. The power emitted by the blackbody source per unit solid angle equals $A_s\sigma_{SB}T^4/\pi$ (in Watt·sr^{-1}) with σ_{SB} = 5.67·10^{-8} Watt·m^{-2}·K^{-4} Stefan-Boltzmann's constant. The transducer area A_{tr} subtends a solid angle as seen by the blackbody source of A_{tr}/R^2 if the source is at a distance R. Hence, the blackbody voltage responsivity follows from:

$$R_V(T) = \frac{\pi R^2 V_{out}}{A_s A_{tr}\sigma_{SB}T^4} \qquad \textit{(in Volt· Watt}^{-1}\textit{)} \qquad (233)$$

This equation also holds if the transducer is preceded by a loss-less optical system that images all of the blackbody source area onto the detector area, since the ratio V_{out}/A_{tr} remains constant.

10.3.3 The Noise Equivalent Power (*NEP*)

The noise equivalent power (*NEP*) of a detector is the required power incident on the detector to produce a signal output equal to the rms-noise voltage at the output. Stating this in a different way, the *NEP* is the signal power that is required to produce a *SNR* equal to one. This signal power is given by

$$V_{out} = R_V \cdot \Phi \qquad (234)$$

This yields a *SNR* of

$$SNR = \frac{R_V \cdot \Phi}{\sqrt{\overline{V^2_{noise}}}} = \frac{R_I \cdot \Phi}{\sqrt{\overline{I^2_{noise}}}} \qquad (235)$$

and consequently for a *SNR* = 1:

$$\Phi \equiv NEP = \frac{\sqrt{\overline{V_{noise}^2}}}{R_V} = \frac{\sqrt{\overline{I_{noise}^2}}}{R_I} \tag{236}$$

Either the spectral responsivity $R(\lambda_0, f)$ or the blackbody responsivity $R(T, f)$ may be inserted in equation 236 to define two different noise equivalent powers. The spectral *NEP*(λ_0, f) is the monochromatic radiant flux $\Phi(\lambda_0)$ required to produce a *SNR* of one at a frequency f. The blackbody *NEP*(T, f) represents the blackbody radiant flux required to produce a *SNR* of one.

The noise equivalent power is useful for comparing similar detectors that operate under identical conditions. It should however not be used as a general measure of detector performance for comparing dissimilar detectors. Firstly, the larger the temporal frequency bandwidth Δf the larger the noise that is present. Also, increasing the detector area A_{tr} will in general decrease the responsivity if all other factors are held constant.

A more useful figure of merit is therefore the normalized noise equivalent power, either per unit bandwidth:

$$NEP^\star = \frac{NEP}{\sqrt{\Delta f}} \qquad (in\ Watt \cdot Hz^{-\frac{1}{2}}) \tag{237}$$

or per unit bandwidth and per unit area:

$$NEP^* = \frac{NEP}{\sqrt{A_{tr}\ \Delta f}} \qquad (in\ Watt \cdot Hz^{-\frac{1}{2}} \cdot m^{-1}) \tag{238}$$

Normalisation entails proportionality to the square root of bandwidth and collecting area A_{tr}.

The reciprocal values of the *NEP* and the *NEP** are the so-called *Detectivities*, D and D^*, and are often used, i.e.:

$$D = \frac{1}{NEP} \tag{239}$$

$$D^* = \frac{1}{NEP^*} = D \cdot \sqrt{A_{tr}\ \Delta f} \tag{240}$$

$$\tag{241}$$

implying the notion that *"larger is better"*. Again, either the spectral or blackbody *NEP* may be used to define spectral or blackbody detectivity, $D^*(\lambda_0, f)$ and $D^*(T, f)$ respectively.

An alternative equivalent expression for D^* is:

$$D^* = \frac{\sqrt{A_{tr}\ \Delta f}}{\Phi} \cdot SNR \tag{242}$$

with Φ the radiant power incident on the detector. Expression 242 can be interpreted as D^* to be equal to the *SNR* at the output of the transducer when 1 Watt of radiant power is incident on a detector area of 1 m^2 with a bandwidth of 1 Hz. This is of course only meant as a mental concept because most transducers are much smaller than 1 m^2 and they reach their limiting sensitivity output well below 1 Watt of incident power.

The figure of merit D^* may be used to compare directly the merit of transducers (sensors, detectors) of different physical size, whose performance was measured using different bandwidths.

10.4 Quantum characterisation

10.4.1 The unfiltered Poisson process

In the quantum limit no coherence effects occur and the radiation field fluctuations can be described by photon statistics only. Consider an incident radiation beam (wide-sense stationary, ergodic) with an average flux of λ photons (or particles or neutrinos) per second. The generation of photons at random times t_i can be described by a staircase function, with discontinuities at time locations t_i (see figure 58):

$$\begin{aligned} Z(t) &= \sum_i U(t-t_i),\ U(t) = \text{unit-step function} \qquad (243) \\ U(t) &= \begin{cases} 1 & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases} \end{aligned}$$

The photon *flow rate* [number of photons per second] follows from time differentiation

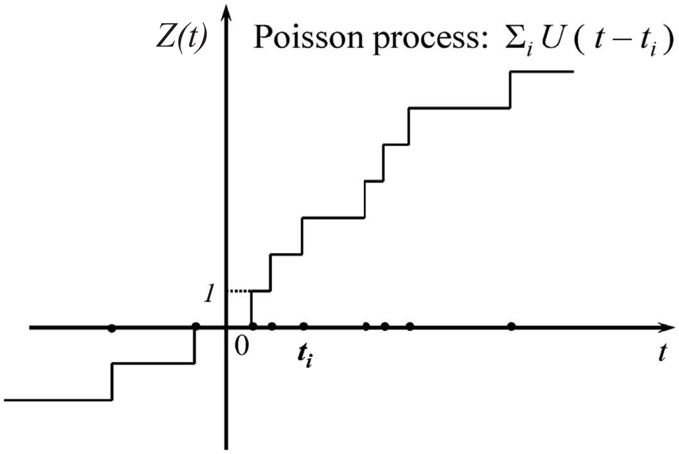

Figure 58: *Staircase function describing a Poisson process.*

of the stochastic variable $Z(t)$:

$$X(t) = \frac{dZ(t)}{dt} = \sum_i \delta(t-t_i) \qquad (244)$$

and represents a train of Dirac impulses at random time locations t_i.
At a constant photon rate, $X(t)$ is a wide sense stationary (WSS) stochastic signal with a *time independent* average $\overline{X(t)} = \lambda$ photons per second, λ is the *rate parameter* characteristic for the process under consideration.
We can now express the stochastic process $Z(t)$, displayed in figure (58), in the following way:

$$Z(t) = \int_0^t X(t')dt' = \int_0^t \sum_i \delta(t'-t_i)dt' = k(0,t) \qquad (245)$$

in which $k(t_1, t_2)$ represents the *number* of photons in a time period (t_1, t_2) of length $t = t_2 - t_1$. This number $k(t_1, t_2)$ is a Poisson distributed random variable (RV) with parameter λt, i.e. $Z(t)$ expresses an unfiltered Poisson process:

$$\mathbf{p}\{k, \lambda t\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad \text{with } \lambda \text{ the rate parameter (see above)} \tag{246}$$

Note: For Poisson distributed RVs hold that if two time periods (t_1, t_2) and (t_3, t_4) are considered that are non-overlapping, then the RVs $k(t_1, t_2)$ and $k(t_3, t_4)$ are independent.
From expression (246) we can construct a Poissonian probability density function featuring a continuous random variable (κ):

$$\mathbf{p}(\kappa, \lambda t) = \sum_{k=0}^{\infty} \mathbf{p}(k, \lambda t)\delta(\kappa - k) \tag{247}$$

The average value of κ and of κ^2 for assessment of the fluctuation magnitude follow from:

$$\begin{aligned} \mathbf{E}\{\kappa\} &= \int_{-\infty}^{+\infty} \kappa \, \mathbf{p}(\kappa, \lambda t) d\kappa = \lambda t & (248) \\ \mathbf{E}\{\kappa^2\} &= \int_{-\infty}^{+\infty} \kappa^2 \, \mathbf{p}(\kappa, \lambda t) d\kappa = (\lambda t)^2 + \lambda t & (249) \end{aligned}$$

The average value for $\kappa = \lambda t$ in equation (248) is of course as expected; the first term of equation (249) is the square of the average and its second term represents the variance. Since the variance of the fluctuations associated with the flow of the photons equals λt, the standard deviation becomes $\sqrt{\lambda t}$, i.e. the 'strength' of the noise in the photon flow. The relative fluctuation or signal to noise ratio (SNR) is then:

$$SNR = \frac{\lambda t}{\sqrt{\lambda t}} = \sqrt{\lambda t} \tag{250}$$

Consequently, the larger λt, the smaller the relative shot noise in the photon flow. With very small λ we apparently need a long filter time to suppress this shot noise.
To determine the autocorrelation function $R_Z(t_1, t_2)$ of the Poisson process $Z(t)$ let us first consider $t_2 \geq t_1$. The variables $k(0, t_1)$ and $k(t_1, t_2)$, referring to adjacent but non-overlapping time periods, are then independent Poisson variables with parameters λt_1 and $\lambda(t_2 - t_1)$ respectively. Thus we have:

$$\begin{aligned} \mathbf{E}\{k(0,t_1)k(t_1,t_2)\} &= \mathbf{E}\{k(0,t_1)\}\mathbf{E}\{k(t_1,t_2)\} = \lambda^2 t_1(t_2 - t_1), \text{ also} \Rightarrow & (251) \\ k(t_1,t_2) &= k(0,t_2) - k(0,t_1) = Z(t_2) - Z(t_1), \Rightarrow \text{ in (251)} \Rightarrow & \\ \mathbf{E}\left\{Z(t_1)\left[Z(t_2) - Z(t_1)\right]\right\} &= R_Z(t_1,t_2) - \mathbf{E}\{Z^2(t_1)\} \Rightarrow & \\ R_Z(t_1,t_2) &= \lambda^2 t_1(t_2 - t_1) + \lambda^2 t_1^2 + \lambda t_1 = \lambda^2 t_1 t_2 + \lambda t_1 & (252) \\ \text{If } t_2 < t_1 \Rightarrow R_Z(t_1,t_2) &= \lambda^2 t_1 t_2 + \lambda t_2 & (253) \end{aligned}$$

Introducing the autocovariance $C_Z(t_1, t_2)$ of $Z(t)$ we can write:

$$R_Z(t_1, t_2) = \lambda^2 t_1 t_2 + C_Z(t_1, t_2) = \lambda^2 t_1 t_2 + \lambda t_1 U(t_2 - t_1) + \lambda t_2 U(t_1 - t_2) \tag{254}$$

Regarding the stochastic variable $X(t)$, the time derivative of $Z(t)$ and representing the train of Dirac impulses at random time locations, we have the time independent average value $\mathbf{E}\{X(t)\} = \lambda =$ the rate parameter.
The autocorrelation function follows from successive partial differentiation of the autocorrelation of $Z(t)$ with respect to t_1 and t_2, thus:

$$R_X(t_1, t_2) = \frac{\partial^2 R_Z(t_1, t_2)}{\partial t_1 \partial t_2} = \lambda + \delta(t_2 - t_1) \tag{255}$$

Designating the time difference $(t_2 - t_1) = \tau$, we arrive at the general expression for the autocorrelation of a train of unit-value Dirac impulses at random time positions (WSS ergodic signal):

$$R_X(\tau) = \lambda^2 + \lambda\delta(\tau) \tag{256}$$

The second term in equation (256) represents the covariance $C(\tau)$ of $X(t)$, which equals in this case the variance $C(0)$ since it is zero for every value of τ except for $\tau = 0$. This is of course evident, since the Dirac impulses are randomly distributed in time and are thus mutually completely uncorrelated.

10.4.2 Frequency limited shot noise

By applying the Wiener Khinchin theorem to $R_X(\tau)$ we can compute the power spectral density:

$$R_X(\tau) \Leftrightarrow S_{d_X}(\nu) = \int_{-\infty}^{+\infty} R_X(\tau) e^{-2\pi j \nu \tau} d\tau = \lambda^2 \delta(\nu) + \lambda \tag{257}$$

which is inconsistent with physical reality since it implies an infinitely high power signal. In practice there is always a frequency cut-off at say ν_c, owing to some (high frequency) filtering process. We might perceive this as follows. The photon detection process involves conversion to charge carriers that are subsequently fed into a filter network, e.g. a first order RC filter. The RC-network acts on each individual charge impuls q (δ-function) with a current response function $h(t)$. If we now assume for convenience that each single photon generates a charge carrier (detection efficiency=1), implying also an average charge carrier rate λ, the resulting photo-current follows from a convolution of the Dirac δ-function train $X(t)$ with $h(t)$:

$$\frac{I(t)}{q} \; = \; X(t) \to h(t) \to Y(t) \tag{258}$$

with $h(t)$ the filter circuit *impulse* current response function (*Note*: $h(t) = 0$ for $t < 0$ and is a normalized function: $\int_0^\infty h(t)dt = 1$).
Hence we have:

$$Y(t) = h(t) * X(t) = \int_0^\infty \sum_k \delta(t' - t_k) h(t - t') dt' = \sum_k h(t - t_k) = \overline{Y} + \Delta Y(t) \tag{259}$$

Owing to the high carrier density in the charge flow, there will be a large degree of overlap between subsequent responses. This will result in a total current $I(t)$ that shows a Gaussian (normal) distribution around a mean value $\overline{I}$. For the expectation value of $Y(t)$ we thus find

$$\begin{aligned}\overline{Y} &= \mathbf{E}\{Y(t)\} = \mathbf{E}\{\int_0^\infty X(t-t')h(t')dt'\} \\ &= \int_0^\infty \mathbf{E}\{X(t-t')\}\, h(t')dt' = \lambda \int_0^\infty h(t')dt' = \lambda\, H(0) \qquad (260)\end{aligned}$$

where for the last transition we have used:

$$H(2\pi j\nu) \equiv \int_0^\infty h(t')e^{-2\pi j\nu t'}dt' \Rightarrow H(0) = \int_0^\infty h(t')dt' \qquad (261)$$

In the Fourier domain we write for the *current* power spectral density:

$$\begin{aligned}S_{d_Y}(\nu) &= |H(2\pi j\nu)|^2\; S_{d_X}(\nu) \\ &= \lambda^2\,|H(2\pi j\nu)|^2\,\delta(\nu) + \lambda\,|H(2\pi j\nu)|^2 = \lambda^2\; H^2(0) + \lambda\,|H(2\pi j\nu)|^2 \quad (262)\end{aligned}$$

Evidently the power is now finite, as it should be. We obtain the autocorrelation by taking the Fourier transform of the *current* power spectral density $S_{d_Y}(\nu)$:

$$R_Y(\tau) = \lambda^2\; H^2(0) + \lambda\,[h(\tau) * h(\tau)] \qquad (263)$$

where the first term on the right hand side gives the *mean* charge response of the linear dynamic system, and the second term represents the noise. Taking the autocovariance at $\tau = 0$ we obtain the variance of the noise signal:

$$C_Y(0) = \lambda \int_{-\infty}^{+\infty} h^2(t)dt = \lambda \int_{-\infty}^{+\infty} |H(2\pi j\nu)|^2 d\nu = 2\lambda \int_0^{+\infty} |H(2\pi j\nu)|^2 d\nu \qquad (264)$$

in taking the last steps we have applied Parseval's theorem and changed from a double sided $S_{d_I}(-\infty < \nu < +\infty)$ to a one-sided S_I: twice the integral from $0 < \nu < \infty$ to accommodate physically real frequencies.

We shall now apply the above analysis to the specific case of a photoconductive semiconductor device, where the shot noise is associated with the random generation (G) and recombination (R) of charge carriers and where the frequency filtering arises intrinsically from the finite life time τ_ℓ of the generated charge carriers.

10.4.3 Photoconducters: shot noise limited sensitivity

Generation/Recombination(GR) shot noise in photoconductors, intrinsic frequency filtering Generation-Recombination noise (GR noise) originates in thermally or optically-stimulated electronic transitions between valance and conduction

band or transitions between impurity levels, traps, or recombination centers and one of these bands. Associated with these transitions are fluctuations in the numbers of free carriers and in their lifetimes, thus giving rise to the GR noise. The detailed mathematical treatment of GR noise depends on many specific parameters like the number of energy levels, the energies corresponding to these levels, the electron population, and the occupancy of states.
As a simple example, the common case of an extrinsic semiconductor such as Germanium or Silicon containing both donors and acceptors, one being predominant and exceeding in number the number of free carriers, will be treated here.
For a simple GR two-level system we assume a generation rate (number per unit time) $g(N)$ and a recombination rate $r(N)$ which describe the transition from the impurity level to the conduction band and the reverse (recombination) process. N is a random variable that represents the number of free carriers (predominantly electrons) in the conduction band at time t. We further assume that both rates g and r depend explicitly *only* on the momentaneous number of free carriers in the conduction band, $N(t)$. In general g is a decreasing function of N (i.e. negative slope), whereas r is an increasing function of N (i.e. positive slope). In the equilibrium (steady state) situation we have balance between the generation and recombination rates, say at a free carrier average number value $\overline{N} = N_e$ at time t. Hence, for $g_e = g(N_e)$ and $r_e = r(N_e)$ we have

$$\begin{aligned} g_e &= r_e \quad \text{and } N \text{ normally distributed around } N_e \quad \Rightarrow & (265) \\ \mathbf{p}(N) &= p(N_e) \exp -\frac{1}{2}\left[\frac{(N-N_e)^2}{\overline{\Delta N^2}}\right] \text{ with variance } \overline{\Delta N^2} = N_e & (266) \end{aligned}$$

Taking the derivatives $g'_e = (dg/dN)_{N=N_e}$ and $r'_e = (dr/dN)_{N=N_e}$ as the generation rate and the recombination rate at the equilibrium number value N_e respectively (dimension [sec^{-1}]), we can assign specific time scales to the GR process by defining $1/\tau_g = -g'_e$ and $1/\tau_r = r'_e$ leading to a free carrier life time:

$$\begin{aligned} \frac{1}{\tau_\ell} &= -\frac{d}{dN}(g-r)_{N=N_e} \quad \Rightarrow \quad \frac{1}{\tau_\ell} = -(g'_e - r'_e) = \frac{1}{\tau_g} + \frac{1}{\tau_r} \text{ and:} & (267) \\ \overline{\Delta N^2} &= N_e = g_e \cdot \tau_\ell & (268) \end{aligned}$$

The carrier life time τ_ℓ dictates the dynamical response of the semiconductor on changes in free carrier generation, this response can be quantified by solving the time dependent continuity equation for change $dN(t)/dt$:

$$\frac{dN(t)}{dt} = g - \frac{N(t)}{\tau_\ell} \quad \Rightarrow \quad N(t) = g\tau_\ell\left(1 - e^{\frac{-t}{\tau_\ell}}\right) \tag{269}$$

The dynamical behavior expressed in equation (269) is characterized by a first order system with transfer function $H(2\pi j\nu) = 1/(1 + 2\pi j\nu\tau_\ell)$, its frequency response $|H(2\pi j\nu)|$ trails off at high frequencies with a tipping point at $\nu_\ell = 1/(2\pi\tau_\ell)$. This is shown, one-sided, in the left panel of figure (59). The associated autocovariance function is shown in the right panel of figure (59). It constitutes a double-sided exponential

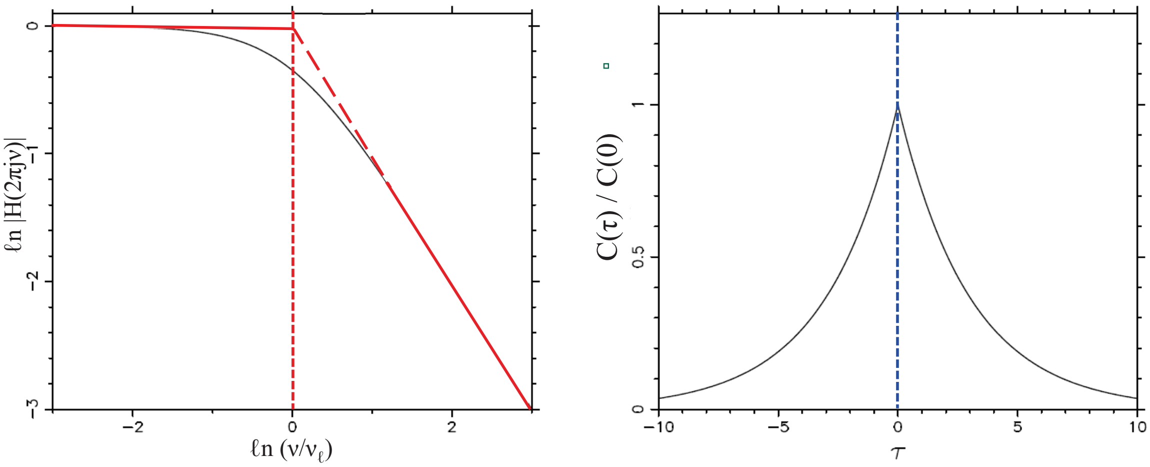

Figure 59: *For a first order transfer function with tipping point $\nu = \nu_\ell$ (left), the autocovariance drops exponentially with $|\tau|$ (right).*

function centered on $\tau = 0$ with a decay constant τ_ℓ that follows from the Fourier transform of the *double sided* transfer function:

$$\begin{aligned} H(2\pi j\nu) + H(-2\pi j\nu) &= \left(\frac{1}{(1+2\pi j\nu\tau_\ell)} + \frac{1}{(1-2\pi j\nu\tau_\ell)}\right) \Rightarrow \text{FT} \Rightarrow \\ &\Leftrightarrow \frac{e^{-\frac{\tau}{\tau_\ell}}}{2\tau_\ell}U(\tau) + \frac{e^{\frac{\tau}{\tau_\ell}}}{2\tau_\ell}U(-\tau) = \\ &= \frac{e^{-\frac{|\tau|}{\tau_\ell}}}{2\tau_\ell} = C(\tau) \end{aligned} \tag{270}$$

where $U(\tau)$ is the Heaviside step function: $U(\tau) = 0$, $\tau < 0$; $U(\tau) = 1$, $\tau \geq 0$. Hence we have:

$$C(\tau) = C(0)\, e^{-\frac{|\tau|}{\tau_\ell}} \quad \text{with} \quad \tau_\ell = \frac{1}{2\pi\nu_\ell} \tag{271}$$

For the GR-process, featuring the random count variable N, we can thus express the variance $\overline{\Delta N^2}$ for some time delay τ following excitation *or* decay as an exponential autocovariance according to:

$$C_N(\tau) = \overline{\Delta N^2}\, e^{-\frac{|\tau|}{\tau_\ell}} \tag{272}$$

The associated power spectral density $S_N(\nu)$ can now be obtained by applying the Wiener-Khinchin theorem. Since we are dealing here with two *independent* random processes, i.e. an excitation process followed by a decay process, in computing the spectral noise power we incorporate a factor 2 in taking the integral of the autocovariance over all physical delays τ. Subsequently we need to convert this *one-sided* spectral density to a *double-sided* spectral density (S_{d_N}) to accomodate the *negative* frequencies and time delays used in Wiener-Khinchin theorem. Thus we have:

$$S_N(\nu) = 2\int_0^\infty \overline{\Delta N^2}\, e^{-\frac{|\tau|}{\tau_\ell}}\, e^{-2\pi j\nu\tau} d\tau \quad \Rightarrow \quad S_{d_N}(\nu) = \int_{-\infty}^\infty \overline{\Delta N^2}\, e^{-\frac{|\tau|}{\tau_\ell}}\, e^{-2\pi j\nu\tau} d\tau \tag{273}$$

Performing the Fourier transform in equation (273) results in:

$$S_{d_N}(\nu) = \frac{2\tau_\ell \overline{\Delta N^2}}{1+(2\pi\nu\tau_\ell)^2} = \frac{2g_e\tau_\ell^2}{1+(2\pi\nu\tau_\ell)^2} \tag{274}$$

As shown before, the value of the average current density $|\vec{j}|$ in the semiconductor equals $nq|\vec{v}_d|$ with $n = N/V$ the charge carrier volume density, q the elementary charge and $\vec{v}_d$ the drift velocity in the applied electric field. With a cross sectional area A we have a total average current $I_e = A\cdot|\vec{j}| = A\cdot(N_e/V)q(d/\tau_{tr}) = qN_e/\tau_{tr} = qg_e(\tau_\ell/\tau_{tr})$ in which d represents the distance between the electrodes of the semiconductor and $\tau_{tr} = d^2/(\mu V)$ the charge carrier transit time between the electrodes (with μ the carrier mobility and V the bias voltage). Substituting in (274) and multiplying $S_{d_N}(\nu)$ by $(q/\tau_{tr})^2$ yields an expression for the current spectral density of the GR noise:

$$S_{d_I} = \left(\frac{q}{\tau_{tr}}\right)^2 S_{d_N}(\nu) = 2qI_e\left(\frac{\tau_\ell}{\tau_{tr}}\right)\left(\frac{1}{1+(2\pi\nu\tau_\ell)^2}\right) \tag{275}$$

The mean square GR current noise $\overline{\Delta I^2}$ follows from integration of S_{d_I} over all frequencies:

$$\begin{aligned} \overline{\Delta I^2} &= 2qI_e\left(\frac{\tau_\ell}{\tau_{tr}}\right)\int_{-\infty}^{+\infty}\frac{d\nu}{1+(2\pi\nu\tau_\ell)^2} \Rightarrow \quad \overline{\Delta I^2} = 4qI_e\left(\frac{\tau_\ell}{\tau_{tr}}\right)\Delta\nu_c \text{ with} & (276)\\ \Delta\nu_c &= \int_0^{+\infty}\frac{d\nu}{1+(2\pi\nu\tau_\ell)^2} \text{ the noise equivalent bandwidth within } 0<\nu<\infty \end{aligned}$$

For low frequencies the (*one-sided*) current spectral power can be expressed as:

$$S_I(0) = \frac{\overline{\Delta I^2}}{\Delta\nu_c} = 4qG_nI_e \quad [\text{Ampere }^2\text{ Hz }^{-1}] \quad \text{with} \quad G_n = \left(\frac{\tau_\ell}{\tau_{tr}}\right) = \frac{\tau_\ell\mu E}{d} \tag{277}$$

G_n is the so-called *noise gain.* In case the semiconductor has uniform resistance and an uniform electric field, the noise gain is proportional to this applied electric field. The GR-current noise can then be expressed as $\left(\sqrt{\overline{\Delta I^2}}\right)_{GRn} = \sqrt{4qG_nI_e\Delta\nu_c}$, with $\left(\sqrt{\overline{\Delta I^2}}\right)_{GRn}$ the rms-noise current, I_e the average total current and $\Delta\nu_c$ the noise equivalent bandwidth.

Shot noise in the signal-photon limit The average number of charge carriers generated by a radiation beam with an average monochromatic photon flux density $F(\lambda)$ (average spectral photon irradiance) at wavelength λ equals $G\eta(\lambda)F(\lambda)A_{pc}$, in which

A_{pc} represents the active area of the photoconductor, $\eta(\lambda)$ the quantum efficiency for photo-absorption and G the photoconductive gain. The average photocurrent $\mathbf{E}\{I_{ph}(t)\}$ can therefore be expressed as:

$$\mathbf{E}\{I_{ph}(t)\} = \overline{I_{ph}(t)} = qG\eta(\lambda)F(\lambda)A_{pc} \tag{278}$$

with q the elementary charge. The frequency response can be expressed as (see equation (269)):

$$H_{pc}(2\pi j\nu) = \frac{1}{1 + 2\pi j\nu\tau_\ell} \tag{279}$$

where τ_ℓ is the charge carrier life time.

$$\left(\sqrt{\overline{\Delta I^2}}\right)_{GR_{ph}} = \sqrt{4qG\overline{I_{ph}(t)}\Delta\nu_c} = 2qG\sqrt{\eta(\lambda)F(\lambda)A_{pc}\Delta\nu_c} \tag{280}$$

with $\left(\sqrt{\overline{\Delta I^2}}\right)_{GR_{ph}}$ the rms-noise current, $\overline{I_{ph}(t)}$ the average total photo-current, $F(\lambda)$ the average radiant signal photon flux and $\Delta\nu_c = \int\limits_0^{+\infty} d\nu/(1 + [2\pi\nu\tau_\ell]^2)$ the one-sided noise equivalent bandwidth within the frequency range $0 < \nu < \infty$. Substituting $\Delta\nu_c$ by performing the integration over frequency we get:

$$\left(\sqrt{\overline{\Delta I^2}}\right)_{GR_{ph}} = qG\left(\frac{\eta(\lambda)F(\lambda)A_{pc}}{\tau_\ell}\right)^{\frac{1}{2}} \tag{281}$$

Finally, for the signal to noise ratio *in the signal photon limit* we obtain

$$\mathrm{SNR} = \left(\frac{\overline{I_{ph}(t)}}{(\sqrt{\overline{\Delta I^2}})_{GR_{ph}}}\right) = (\eta(\lambda)F(\lambda)A_{pc}\tau_\ell)^{1/2} \tag{282}$$

This last equation tells us that a high value for the frequency cutoff $\nu_\ell = 1/(2\pi\tau_\ell)$ leads to a lower signal to noise for the photo-current; the reason for this is that the intrinsic system noise is less filtered.

Shot noise in the background-photon limit The rms-noise in the photocurrent, derived in the previous paragraph, dominates over thermal noise if the photoconductor is sufficiently cooled.
From expression (281) for the rms-noise in the photocurrent $\left(\sqrt{\overline{\Delta I^2}}\right)_{GR_{ph}}$ we can also assess the spectral Noise Equivalent Power ($NEP(\lambda,\nu)$) if we assume that this photocurrent arises from an *average monochromatic background-photon flux density* $B(\lambda)$ instead of the average signal-photon flux density $F(\lambda)$ considered above. This $NEP(\lambda,\nu)$ represents the so-called *radiation Background Limited Performance (BLIP)*. Let us use the relations:

$$NEP(\lambda,\nu) = \frac{\left(\sqrt{\overline{\Delta I^2}}\right)_{GR_{ph}}}{R^I_{pc}(\lambda,\nu)} = \frac{hc}{\lambda}\left(\frac{B(\lambda)A_{pc}}{\eta(\lambda)\tau_\ell}\right)^{\frac{1}{2}} \tag{283}$$

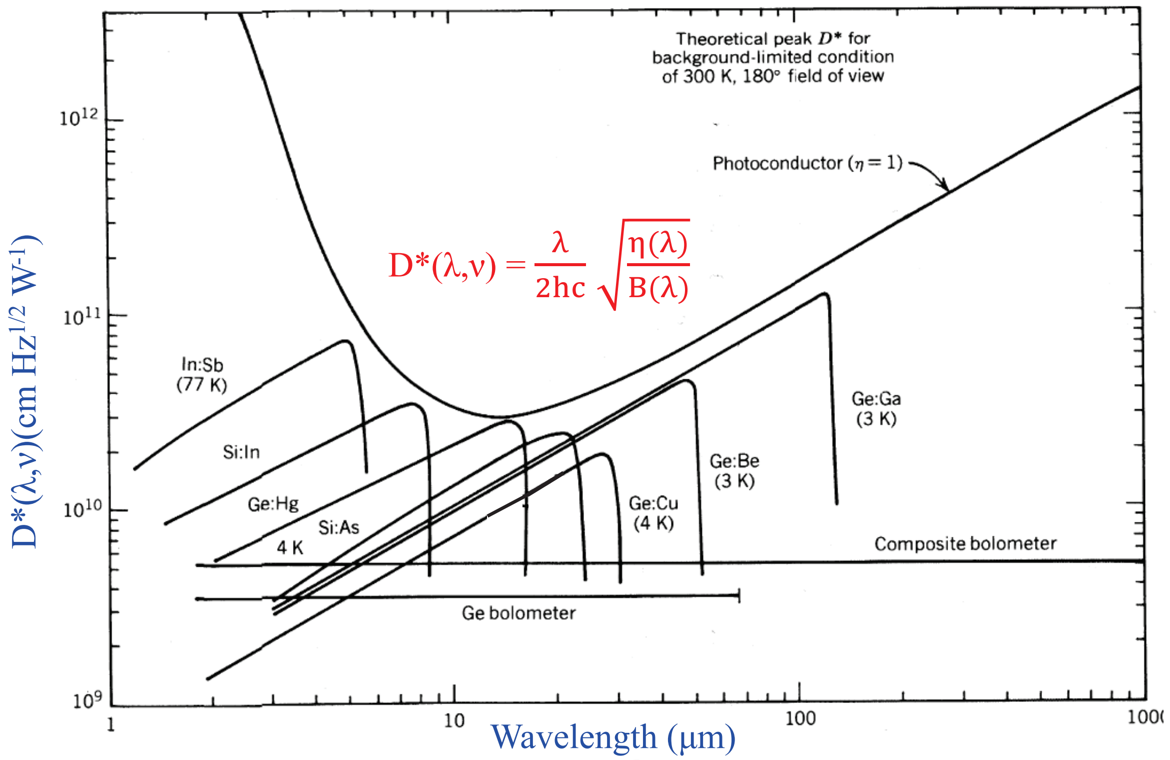

Figure 60: *Typical detectivities for several photo-conductors. Figure taken from Dereniak and Crowe 1984.*

and the BLIP normalised detectivity $D^*(\lambda, \nu)$:

$$D^*(\lambda, \nu) = \frac{\sqrt{A_{pc}\Delta\nu_c}}{NEP(\lambda, \nu)} = \frac{\lambda}{2hc}\left(\frac{\eta(\lambda)}{B(\lambda)}\right)^{\frac{1}{2}} \quad \text{by substituting } \Delta\nu_c = 1/(4\tau_\ell) \tag{284}$$

Figure 60 shows the theoretical *peak detectivity* for $\eta(\lambda) = 1$ as a function of wavelength, assuming radiation Background Limited Performance(BLIP) arising from an omnidirectional blackbody radiation field at 300 K integrated over the upper hemisphere of 180°. Also included in figure (60) are a number of common extrinsic photoconductors and their corresponding operating temperatures. Detectivities of two types of thermal detectors (bolometers) are displayed for comparison. The thermal detectors lack the cut-off feature in wavelength due to the different detection principle, they show a considerably lower value of $D^*(\lambda, \nu)$ but cover a wider spectral band than the photoconductive devices.

10.4.4 Single photon detection: noise analysis

In observations where the individual information carriers are registered (photons, cosmic-ray particles or neutrinos), the signal to noise ratio considerations are based on a statistical treatment of the data. This leads to expressions for the limiting sensitivity for source detection depending on collecting area, exposure time and noise level. In what

follows, statistical independence is assumed between subsequent events; the process possesses no internal coherence and the stochastic nature of the data can therefore be described by Poissonian statistics.
Consider a radiation beam, originating from a distant point source, with a photon flux density n_s (e.g. photons·m^{-2}·s^{-1}). Suppose this point source is embedded in uniformly distributed background radiation noise with a photon intensity n_{bg} (photons·m^{-2}·s^{-1}·sr^{-1}). Furthermore, the quantum noise of the detector is given by n_{det} counts per unit detector area per unit time. If the telescope effective area equals $\bar{A}_{eff}$, where the energy dependent collecting area has been averaged over the energy bandwidth of the observation, the number of registered counts over an integration period T_{obs} in one image pixel equals:

$$N_1 = ((n_s + n_{bg}\Delta\Omega)\bar{A}_{eff} + n_{det}A_{pix})T_{obs} = (n_s + n_{bg}\Delta\Omega + n_{det}\delta)\bar{A}_{eff}T_{obs} \qquad (285)$$

in which $\Delta\Omega$ represents the solid angle subtended on the sky by the angular resolution of the telescope and δ the ratio between the area of a single pixel on the face of the image detector A_{pix} and $\bar{A}_{eff}$. An adjacent pixel without the point source accumulates in the same observing time:

$$N_2 = (n_{bg}\Delta\Omega + n_{det}\delta)\bar{A}_{eff}T_{obs} \qquad (286)$$

According to Poissonian statistics the fluctuations in N_1 and N_2 equal $\sqrt{N_1}$ and $\sqrt{N_2}$ respectively. The *SNR* can now be defined as

$$SNR = \frac{N_1 - N_2}{\sqrt{N_1 + N_2}} \qquad (287)$$

in which the signal strength is evaluated in terms of the magnitude of the statistical fluctuation in the noise component (beware: **not** in relation to the *absolute* magnitude of the noise component).
Consider two extreme cases:

- *The source signal strongly dominates the noise, i.e.* $N_2 \ll N_1$.

  In this case the *SNR* equals

  $$SNR = \sqrt{N_1} = \sqrt{n_s\bar{A}_{eff}T_{obs}} \qquad (288)$$

  In terms of the limiting sensitivity, a minimum number of photons N_{min} is required in order to be able to speak of a detection, for example 10 or 25. In those cases the *SNR* equals 3 or 5.

  The limiting sensitivity follows from

  $$n_{s_{min}} = \frac{N_{min}}{\bar{A}_{eff}T_{obs}} \sim (\bar{A}_{eff}T_{obs})^{-1} \qquad (289)$$

  This is the best possible case. The detection is so-called *signal-photon-noise limited.* The limiting sensitivity improves linearly with the effective collecting area and the integration time.

- *The source signal is drowned in the noise, i.e.* $N_1 \approx N_2 = N$.

  The criterium for source detection can be formulated on the basis of a certain level of confidence (minimal *SNR*) k. The limiting sensitivity now follows from

$$n_{s_{min}} = k\frac{\sqrt{2N}}{\bar{A}_{eff}T_{obs}} \tag{290}$$

  Substituting $N = (n_s + n_{bg}\Delta\Omega + n_{det}\delta)\bar{A}_{eff}T_{obs}$:

$$n_{s_{min}} = k\sqrt{\frac{2(n_{bg}\Delta\Omega + n_{det}\delta)}{\bar{A}_{eff}T_{obs}}} \sim (\bar{A}_{eff}T_{obs})^{-\frac{1}{2}} \tag{291}$$

  In this case the limiting sensitivity only improves with the square root of the telescope collecting area and the integration time.

  In the case $n_{det}\delta \ll n_{bg}\Delta\Omega$ (detector noise negligible), the detection is said to be *background-photon-noise limited.*

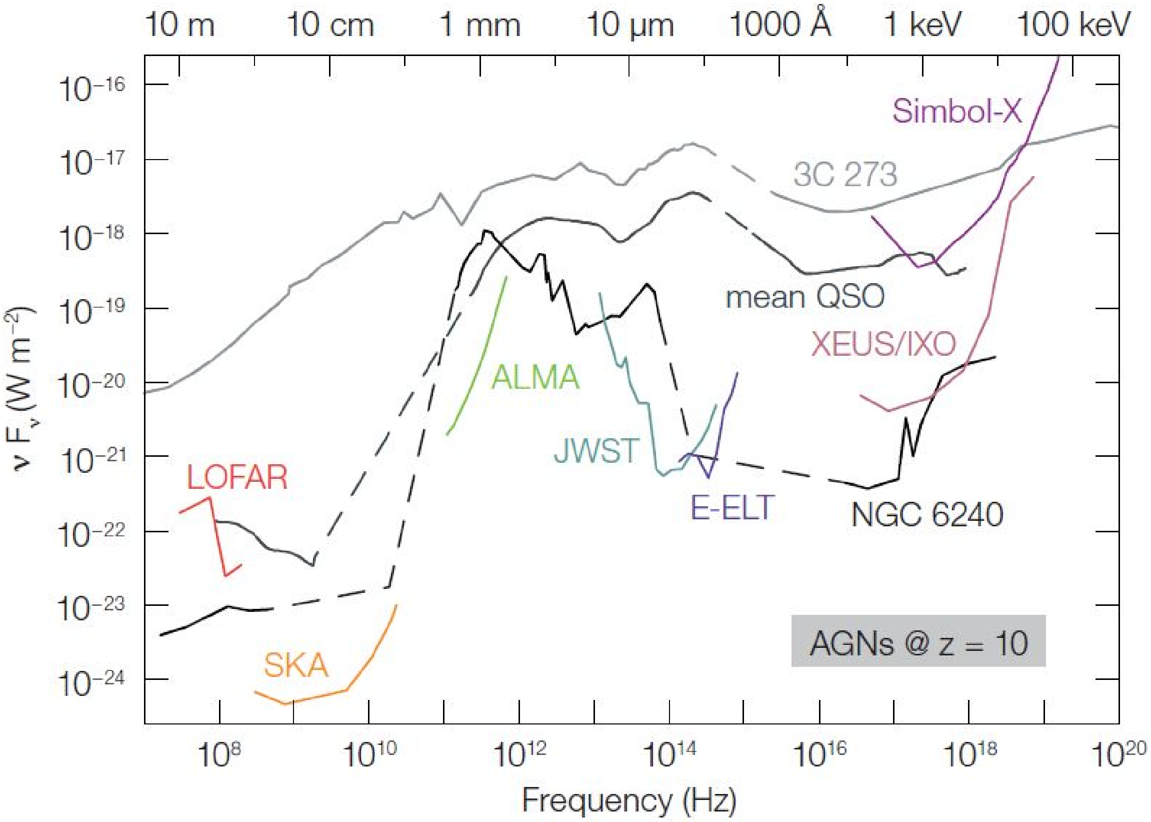

Figure 61: *Comparison of sensitivities of major (future) astronomical observing facilities. Spectral energy distributions for 3C273, for an average QSO template, and for the obscured star-forming galactic merger NGC 6240 are shown at a redshift of z=10. Sensitivities assume 12-hour 1-σ detections (SNR=1) for all instruments except for the X-ray observatories XEUS/IXO and Simbol-X where an equivalent 5-σ detection (SNR=5) in a 1 Megasecond exposure is assumed. Credit Astronet Report 2008.*

Note: implicitly the bandwidth of the observation also comes in when evaluating the *SNR*. The values n_s, n_{bg} and n_{det} were defined as integral values over a certain predetermined energy bandwidth. For example, if an energy bandwidth $\epsilon_2 - \epsilon_1 = \Delta\epsilon$ is considered, $n_s = \int\limits_{\Delta\epsilon} n_s(\epsilon)\, d\epsilon = \bar{n}_s(\epsilon)\Delta\epsilon$. Similarly, $n_{bg} = \bar{n}_{bg}(\epsilon)\Delta\epsilon$ and $n_{det} = \bar{n}_{det}(\epsilon)\Delta\epsilon$.

Now the following relations hold:

- signal-photon-noise limited: $n_{s_{min}}(\epsilon) \sim (\bar{A}_{eff} T_{obs} \Delta\epsilon)^{-1}$
- background-photon-noise limited: $n_{s_{min}}(\epsilon) \sim (\bar{A}_{eff} T_{obs} \Delta\epsilon)^{-\frac{1}{2}}$

Figure 61 shows a comparison of limiting sensitivities, assuming realistic observing times, for a number of current and planned major astronomical observing facilities, both ground-based and space-based. Typical spectral energy distributions for AGNs and a star-forming galactic merger, positioned at a redshift z=10, are included for reference.

11 References

- M.S. Longair, ***High Energy Astrophysics***, 1992, Cambridge University Press.

- E.L. Dereniak and Crowe, ***Optical Radiation Detectors***, 1984, John Wiley and Sons Inc.

- C.R. Kitchen, ***Astrophysical Techniques***, 1998, Institute of Physics Publishing.

- R. Giacconi and H. Gursky, ***X-ray Astronomy***, 1974, Reidel Publishing Company.

- P. Lena, F. Lebrun and F. Mignard, ***Observational Astrophysics***, 1998, Springer-Verlag.

- E. Hecht, ***Optics***, 1987, Addison-Wesley Publishing Company Inc.

- A. Papoulis, ***Probability, Random Variables, and Stochastic Processes***, 1991, McGraw-Hill Book Company International Editions.

- W.H. Press et al, ***Numerical Recipes***, 1992, Cambridge University Press.

- R.N. Bracewell, ***The Fourier Transform and its Applications***, 1986, McGraw-Hill Book Company International Editions.

- J.A.M. Bleeker, ***BeppoSAX: the Wide Field Camera prospect***, in BeppoSAX (Eds: E. van den Heuvel, R. Wijers and J. in 't Zand), 2004, Nuclear Physics B, Proc. Supplements.

- ASTRONET Infrastructure Roadmap, ***A Strategic Plan for European Astronomy***, 2008.