Coherent Receivers

• **Principles**
  ➢ **Downconversion**
  Heterodyne receivers mix signals of different frequency; if two such signals are added together, they beat against each other. The resulting signal contains frequencies only from the original two signals, but its *amplitude* is modulated at the difference, or beat, frequency; this downconverted beat signal is used for the detection. Unlike incoherent detectors, this signal encodes the spectrum of the incoming signal over a range of input frequencies and also retains information about the phase of the incoming wavefront. This “extra” information allows efficient spectral multiplexing (many spectral elements observed simultaneously with a single receiver) and very flexible use of arrays of telescopes and receivers for interferometry.

For heterodyne operation, the mixed field must be passed through a nonlinear circuit element or *mixer* that converts power from the original frequencies to the beat frequency. In the submillimeter- and millimeter-wave (and radio) region this element is a diode or other nonlinear electrical circuit component. For visible and infrared operation, the nonlinear element is a photon detector, sometimes also termed a photomixer.
If the mixer has a linear $I$-$V$ curve, then the conversion efficiency is zero. Similarly, any mixer having a characteristic curve that is an odd function of voltage around the origin will have zero conversion efficiency if operated at zero bias. Much greater efficiency is achieved, however, with a characteristic curve that is an even function of voltage; if $I \propto V^2$, the output current is proportional to the square of the signal amplitude. $I \propto V^2 \propto \xi^2 \propto P$, where $\xi$ is the strength of the electric field; hence, we prefer to use square law devices as fundamental mixers because their output is linear with input power.

Assume that we are using a mixer illuminated with a mixture of two sources of power, one a signal at frequency $\omega_S$ and the other at $\omega_{LO}$. Let the power at $\omega_{LO}$ originate from the local oscillator (LO) within the instrument. We also specify that $\omega_S > \omega_{LO}$. Then the mixed signal will be amplitude modulated at the intermediate frequency $\omega_{IF} = |\omega_S - \omega_{LO}|$. The signal at $\omega_{IF}$ contains spectral and phase information about the signal at $\omega_S$. The signal has been downconverted to a much lower frequency than $\omega_S$ or $\omega_{LO}$, in the frequency range where it can be processed by
There is no way of telling in the mixed signal whether $\omega_S > \omega_{LO}$ or $\omega_{LO} > \omega_S$. Because we have lost the initial information regarding the relative values of $\omega_S$ and $\omega_{LO}$, many of the derivations of receiver performance will assume that the input signal contains two components of equal strength, one above and the other below the LO frequency $\omega_{LO}$. Since the signal at $\omega_{IF}$ can arise from a combination of true inputs at $\omega_{LO} + \omega_{IF}$ and $\omega_{LO} - \omega_{IF}$, it is referred to as a double sideband signal. When observing continuum sources, the ambiguity in the frequency of the input signal is a minor inconvenience. When observing spectral lines, the *image* frequency signal at the off-line sideband results in complications.

- **Heterodyne Receivers**
  - **Mixer**
  
  At the highest frequencies at which heterodyne receivers are used, a continuous wave (CW) laser is used as the local oscillator. The laser light and the signal are combined by a beam splitter, sometimes called a diplexer. The output is mixed in a photon detector--a photoconductor, photodiode, photomultiplier, or bolometer. Because a photomixer responds to power, or field strength squared, it is a square law mixer by definition. However, such heterodyne receivers are uncommon. The technique comes into its own in the submm and at lower frequencies, where the mixer is a specialized electrical element onto which the energy...
is concentrated by antennae, waveguides, and other non-focussing optics.

The picture above is a SIS mixer block. The local oscillator and the input signal are coupled into the mixers by the twin slot antenna. The IF is taken out to the right, while the volume to the left helps tune the response for efficiency (but at a specific frequency). SIS stands for "superconductor-insulator-superconductor. A sandwich of these materials produces the "diode or other nonlinear device" described above as the heart of a mixer. Its operation is illustrated below.

(a) shows the device without a bias voltage. The superconductors are shown in the band diagram approximation; the band gap is a few meV. A bias has been applied in (b). When it becomes big enough to align the "valence" band on the left with the "conduction" one on the right, suddenly Cooper pairs tunnel through the insulator efficiently, causing a current. See the plot in (c) for the overall behavior. Because the inflection in the bias curve is so sharp at $2\Delta/q$, relatively little local oscillator power is needed to get a good IF signal. However, the tuning required for good signal coupling generally restricts the operation to a relatively narrow spectral range (10-20%).

The full receiver is shown below:
Using the photodetector case as an example, the mixer need not respond fast enough to track the frequency of the two input signals, so those signals just produce constant photocurrents. In addition, the photocurrent contains a component oscillating at the intermediate frequency, \( \omega_{IF} = |\omega_S - \omega_L| \). The IF current is the heterodyne signal and has a mean-square-amplitude of

\[
\langle I_{IF}^2 \rangle = 2 I_L I_S, \quad (1)
\]

where \( I_L \) is the current in the detector from the LO signal and \( I_S \) is that from the source.

It is important to note that the signal strength in equation (1) depends on the LO power. As a result, many forms of noise can be overcome by increasing the output of the local oscillator. The ability to provide an increase in power while downconverting the input signal frequency is characteristic of quantum mixers, such as the photomixers discussed here. The conversion gain is defined as the IF output power that can be delivered by the mixer to the next stage of electronics divided by the input signal power. Classical mixers do the downconversion without gain - but the use of very low noise electronics in the GHz range of the IF signal makes it useful to carry out this operation even without gain.

**Post mixer Electronics**

The heterodyne signal derived above is a low level, high frequency AC current; it needs to be amplified and converted into a slowly varying
voltage that is proportional to the time-averaged input signal power. The first step in this process is amplification. We want to preserve as high an IF frequency range as possible, since the spectral band over which the receiver works is just $2\Delta f_{\text{IF}}$ (assuming we use it double sideband).

The best performance for the IF amplifier is obtained with high electron mobility transistors (HEMTs) built on GaAs. The HEMT is based on the metal-semiconductor field effect transistor (MESFET). It consists of a substrate of GaAs with an n-doped layer grown on it to form the channel, with contacts for the source and drain and a gate formed as a Schottky diode between them on this layer. The electron flow between source and drain in this channel is regulated by the reverse bias on the gate; as with the JFET, with an adequately large reverse bias the depletion region grows to the semi-insulating layer and pinches off the current. Because this structure is very simple, MESFETs can be made extremely small, which reduces the electron transit time between the source and drain and increases the response speed.
In the HEMT, the MESFET performance is further improved by using a heterojunction (junction between two different semiconductors with different bandgaps) so the electrons flow in undoped GaAs.

To achieve this result, the MESFET is grown on heavily doped GaAlAs, whose Fermi level lies above the bottom of the conduction band in the undoped GaAs layer. Thus, the conduction electrons collect in the GaAs and flow through it under the influence of the MESFET fields. The very high mobility in undoped GaAs makes for very fast response, to \( \sim 10^{11} \) Hz.
**Detector Stage**

The conversion to a slowly varying output can be done by a detector stage that rectifies the signal and sends it through a low-pass filter. We would like the circuit to act as a square law detector because \( \langle I_{IF}^2 \rangle \) is proportional to \( I_S \) (see equation (1)), which in turn is proportional to the power in the incoming signal.

![Detector Stage Circuit Diagram](image)

To demonstrate how the detector stage achieves this goal, we solve the diode equation for voltage and expand in \( I/I_0 \):

\[
V = \frac{kT}{q} \ln \left( 1 + \frac{I}{I_0} \right) \approx \frac{kT}{q} \left[ I \left( I_0 \right) - \frac{1}{2} \left( \frac{I}{I_0} \right)^2 + \frac{1}{3} \left( \frac{I}{I_0} \right)^3 - \frac{1}{4} \left( \frac{I}{I_0} \right)^4 + \ldots \right]. \tag{2}
\]

The first and third terms in the expansion will have zero or small conversion efficiency, and the 4th and higher terms will be small if \( I \ll I_0 \). Thus, the detector stage does act as a square law device.

Sometimes it is desirable to carry out a variety of operations with the IF signal itself before smoothing it. For example, imagine that the heterodyne signal is sent to a bank of narrow bandpass electronic filters with a smoothing circuit on the output of each filter. The frequencies present at the input to the filter bank are limited to the bandwidth of the IF amplifier or of the photomixer (typically \( \sim 1 \text{GHz} \)). The filters can be tuned to divide this IF signal into components at a range of frequencies over the IF stage bandpass. Because the intermediate frequency goes as \( |\omega_S - \omega_L| \), the frequencies in the heterodyne signal correspond to a similar range of frequencies in the source centered on the signal and image frequencies. The filter bank therefore provides the spectrum of the source under study.

➤ **Local Oscillator**
At the high frequencies of the infrared and visible regions, the only local oscillator with reasonably high power output is a continuous wave laser. In the submm and lower frequencies, tunable LOs are available - usually by starting with a lower frequency oscillator and multiplying it up to the operating frequency. In any case LO power is a critical asset.

- **Fundamental limits**
  - **Bandwidth**
    The spectral bandwidth of a heterodyne receiver is determined by the achievable bandwidth at the intermediate frequency.
  - **Throughput**
    The signal photons cannot be concentrated onto the mixer in a parallel beam; even for a point source, they will strike it over a range of angles. The requirement that interference occurs at the mixer between the laser and the signal photons sets a requirement on the useful range of acceptance angle for the heterodyne receiver. This condition can be expressed as
    \[
    \Phi \approx \frac{\lambda}{D}, \quad (3)
    \]
    where \( D \) is the diameter of the telescope aperture and \( \Phi \) is the angular diameter of the field of view on the sky. Thus, a coherent receiver must operate at the diffraction limit of the telescope.
A second restriction is that the interference that produces a heterodyne signal only occurs for components of the source photon electric field vector that are parallel to the electric field vector of the laser power; i.e., only a single polarization of the source emission produces any signal. The relation between wavelength and etendue and the constraints on polarization are manifestations of the antenna theorem, applicable to all heterodyne receivers.

- **Signal to Noise and Detection Limits**

  The detection limits of a heterodyne receiver are considered differently than for incoherent detectors. We need to distinguish noise: (1) that is independent of the LO-generated current, $I_L$, and (2) that depends on $I_L$. In principle, the first category can be eliminated by using a local oscillator with sufficient power to raise the signal strength out of the noise. Thus, the second category alone contains the fundamental noise limits. Two types of fundamental noise are: (1) quantum noise in the mixer from the generation of charge carriers by the LO power; and (2) noise from thermal background detected by the system. The division between the two regimes can often be simplified by stating that for $h\nu >> kT_B$, the quantum limit holds, while for $h\nu << kT_B$ we get the thermal limit. Here, $T_B$ is the temperature associated with the background power.

- **Noise temperature**

  To describe the performance on continuum sources, a thermal source is introduced through a noise temperature, $T_N$, defined such that a matched blackbody at the receiver input at a temperature $T_N$ produces $S/N = 1$. The lower $T_N$, the fainter a source gives $S/N = 1$, and the better is the performance of the receiver (so the behavior is again like NEP – the smaller, the better).

  In the thermal limit, if the effective source emissivity $\varepsilon = 1$, then by definition $T_N = T_B$. For the ideal double sideband case, the quantum limit is

  $$T_N \approx \frac{h\nu}{k}.$$  \hspace{1cm} (4)

  The quantum limit expressed in equation (4) can be justified in terms of the Heisenberg uncertainty principle, which states that the uncertainty, $\Delta P$, in a measurement of power will be
\[ \Delta P = \frac{h \nu}{\Delta t}, \quad (5) \]

where \( \Delta t \) is the observation time. Starting with the equation for Johnson noise in a resistor within a frequency bandwidth \( df \) and converting to power noise within a time interval \( \Delta t \),

\[ < P > \Delta t = k T_N. \quad (6) \]

Setting \( \Delta P \approx <P> \), we obtain

\[ T_N = \frac{h \nu}{k}. \]

It is often convenient to express the flux from a source as an antenna temperature, \( T_S \), in analogy with the noise temperatures. This concept is particularly useful in the millimeter and submillimeter (and radio) regions, where the observations are virtually always at frequencies that are in the Rayleigh Jeans regime \((h \nu << kT)\). In this case, the antenna temperature is linearly related to the input flux density:

\[ \frac{P}{\Delta \nu} = 2kT_S, \quad (7) \]

where \( \Delta \nu \) is the frequency bandwidth. To maintain the simple formalism in terms of noise and antenna temperatures, it is conventional to use a Rayleigh Jeans equivalent temperature such that equation (7) holds by definition whether the Rayleigh Jeans approximation is valid or not.

The achievable signal-to-noise ratio for a coherent receiver is given in terms of antenna and system noise temperatures by the Dicke radiometer equation:

\[ \left( \frac{S}{N} \right)_c = K \frac{T_S}{T_N} (\Delta f_{IF} \Delta t)^{1/2}, \quad (8) \]

where \( \Delta t \) is the integration time of the observation and \( K \) is a constant of order one.
Comparisons with incoherent detectors

Equations (7) and (8) give us the means to compare the performance of coherent and incoherent detectors, as long as we also keep in mind the bandwidth and single mode detection restrictions that we have already discussed. From equation (7) and the definition of NEP, the signal-to-noise ratio with an incoherent detector system operating at the diffraction limit is

\[
\left( \frac{S}{N} \right)_i = \frac{2kT_S \Delta \nu (\Delta t)^{1/2}}{\text{NEP}} \quad \ldots \ (9)
\]

Therefore, using equation (8), we obtain the ratios of signal to noise achievable with the two types of system under the same measurement conditions:

\[
\frac{(S/N)_c}{(S/N)_i} = \frac{\text{NEP} (\Delta f_{IF})^{1/2}}{2kT_N \Delta \nu}. \quad \ldots \ (10)
\]

Suppose a bolometer is operating background limited and we compare its signal to noise on a continuum source with a heterodyne receiver operating at the quantum limit. We set the bolometer field of view at the diffraction limit, \( A_\Omega = \lambda^2 \), and assume that the background is in the Rayleigh Jeans regime (e.g., thermal background at 270K observed near 1mm). The background limited NEP is:

\[
\text{NEP} = \frac{hc(2\varphi/\eta)^{1/2}}{\lambda}. \quad \ldots \ (11)
\]

The photon incidence rate, \( \varphi \), can be shown to be

\[
\varphi = \frac{2\eta kT_B \Delta \nu}{h\nu}. \quad \ldots \ (12)
\]

If we assume the bolometer is operated at 25% spectral bandwidth, \( \Delta \nu = 0.25\nu \), then

\[
\frac{(S/N)_c}{(S/N)_i} \sim \frac{4 \times 10^6}{\nu} (4\Delta f_{IF})^{1/2} \sim \frac{2.6 \times 10^{11}}{\nu}. \quad \ldots \ (13)
\]

We have taken the IF bandwidth to be \( 3 \times 10^9 \) Hz, a typical value.
Thus, the incoherent detector becomes more sensitive near $2.6 \times 10^{11}$Hz, or at a wavelength just longer than 1mm. Actually, this comparison is slightly unfair to it (since, for example, it does not have to work at the diffraction limit), so it is the detector of choice for continuum detections to wavelengths of 2 to 3mm. Hence, the development of large scale bolometer cameras for mm-wave and submm telescopes.

Of course, the coherent detectors are preferred for high resolution spectroscopy and for interferometry.