Chapter 2

Submillimeter detection and instrumental requirements

2.1 Introduction/Overview

The submillimeter wavelength band ($\lambda = 100 \, \mu m$ to 1 mm) is a transition region between infrared and millimeter waves. It is a region in which radiometric techniques common to the longer wavelength regime are combined with optical techniques. In this thesis we primarily concern ourselves with advanced heterodyne receiver technology in the submillimeter and far-infrared portion of the electromagnetic spectrum.

At frequencies below 100 GHz coherent (heterodyne) detection is extensively used in for example; mobile phones, satellite communications, and radar applications. The advantage of coherent detection is that by utilizing an intermediate frequency (IF), a nearly arbitrary high spectroscopic resolution (defined as $R = \nu / \Delta \nu$) can be obtained. This is especially useful in the study of atomic and molecular line transitions (Chap. 1). For example, resolution requirements of $R \geq 10^6 - 10^7$ are not uncommon. The actual achievable resolution will depend on the backend (correlator) and phase jitter in the local oscillator. At a rest frequency of 1 THz, a resolving power of $R = 10^6 - 10^7$ corresponds to a spectral resolution bandwidth of 100 KHz – 1 MHz. When expressed in velocity terms (a more practical unit of measure for the astronomer) this equates to $v = c/R = 0.03 - 0.3 \, \text{kms}^{-1}$.

Despite the nearly infinitely high resolution advantage of coherent detection, there are some important drawbacks. Because both amplitude and phase are detected there is an intrinsic lower limit ($h\nu/k_B$), known as the quantum noise limit, on the achievable sensitivity. Here $h$ is Planck’s constant, $6.626 \times 10^{-34} \, \text{J}s$, and $k_B$ Boltzmann’s constant $1.381 \times 10^{-23} \, \text{J/K}$. In practice, depended on the frequency of operation, heterodyne receivers have sensitivities that range from 3 to 10 times the quantum noise temperature. They are also only sensitive to a single radiation mode, and to optimize coupling efficiency from the device to the telescope one polarization (Chaps. 5 – 9) is typically employed. Another very important issue with coherent detectors is the fractionally small detection (IF) bandwidth. Modern heterodyne superconducting-
insulator-superconducting (SIS) receivers (Chaps. 7 & 8) are constructed with a 4 – 8 GHz instantaneous IF bandwidth. However, due to instabilities in the LO pump signal, standing waves in the LO-mixer path, and intrinsic mixer instability not all of this IF bandwidth is available for continuum or spectral energy density (SED) flux detection. The situation is even worse for hot electron bolometer (HEB) mixers, which at best have an instantaneous IF gain bandwidth of 2 – 3 GHz, of which due to excessive $1/f$ noise only $\lesssim 500$ MHz is available for flux detection. If we take 500 MHz as the upper limit for the total power spectral bandwidth of an HEB mixer, then at 1.9 THz (the important C$^+$ cooling line) we find that this corresponds to a frequency resolution limit of $R_{\text{min}} \sim 4000$, or a maximum velocity coverage of $c/R = 75 \text{ kms}^{-1}$. As we shall see, this is barely enough for continuum observations toward the galactic center.

Thus we find that coherent detection, though excellent for high resolution spectroscopic astrophysics [1, 2, 3, 4], has limited value in the study of broadband emission from cold interstellar dust, unresolved ultraluminous infrared galaxies, and the cosmic background radiation. This is especially true in the context of low background space observations where the limited sensitivity and instability of coherent detectors is prohibitive in the detection of broadband emission. To address continuum or dust observations incoherent detectors are better suited. In the submillimeter and far-infrared wavelength regime bolometers (heat sensitivity detectors) have been found most sensitive. Though not the subject of this thesis we will outline in Sec. 2.2 the basics of this very important class of incoherent detectors.

Bolometers are insensitive to phase and as such do not have a quantum limit. They are however limited by photon statistics as discussed in Sec 2.2.3. And because the bolometer resolution is set by optical filters (for ground based observations optical filters are ordinarily designed to follow the atmospheric windows), the achievable frequency resolution is very low ($R \sim 10$). To understand why this is advantageous in the study of continuum emission (absorption), one only has to consider that background limited direct detection is best achieved with the instrumental bandwidth matched to the expected resolution of the astronomical source (no excess noise).

From the above discussion, it is evident that there is a large gap between the lowest possible resolving power of coherent ($R \gtrsim 4000$) and incoherent ($R \lesssim 10$) detectors. This gap is filled by a class of medium resolution direct detection instruments that place a Fourier Transform Spectrometer (FTS) in the optical path of a small bolometer array. Good examples of this are SPIFY ($R = 500 – 10,000$) [5], and SPIRE [6], a 3-band Imaging Fourier Transform Spectrometer on the Herschel space observatory [7]. SPIRE has a maximum resolution of $R = 1000$.

The spectral resolution of such instruments can be made variable by adjusting the travel length $L$ of the FTS ($\Delta \nu \approx c/4L$). This technique facilitates, in principle, observations of the galactic center ($c/R = 60 \text{ kms}^{-1}$ with $R = 5000$), nearby galaxies with velocities of 100 – 200 kms$^{-1}$ ($R = 1500 – 3000$), and distant galaxies with $c/R \gtrsim 500 \text{ kms}^{-1}$ ($R \lesssim 500$). However, by using a FTS a penalty in observation speed is paid by having to sample the source as a function of path length. Thus high resolution scans will take a considerable amount of time. To avoid this time penalty, a grating spectrometer with multiple detector elements may be used. This
2.2. INCOHERENT (DIRECT) DETECTION

technique considerably speeds up the detection process, however at a cost of a fixed frequency resolution. Nikola and Bradford et al. have successfully used this method with ZEUS [8], and Z-spec [9]. ZEUS is an echelle grating imaging spectrometer, designed for the JCMT [10], with a resolving power $R \sim 1000$. It operates in the 350 $\mu$m and 450 $\mu$m atmospheric windows. Z-spec complements ZEUS as it operates in the 190 – 310 GHz atmospheric window. It has a resolution of $R \sim 275$ and is operational at the Caltech Submillimeter Observatory (CSO) [11]. Both instruments are used to search for high redshift CO emission ($z = v/c$) from ultra luminous galaxies with velocities in excess of 1500 $\text{km s}^{-1}$. In the same category, PACS [12] is also a grating spectrometer with $R_{\text{max}} = 4000$. It is to be flown on Herschel space observatory.

As a final note, a new class of direct detectors has recently been introduced by Day and Zmuidzinas et al. [13]. These detectors are based on a change in the kinetic inductance component of the surface impedance when a superconducting thin-film material absorbs an optical photon. KIDs as they are known, have the advantage that inherent frequency multiplexing of the output channels greatly facilitates large format camera arrays. In principle this technique circumvents the complicated SQUID multiplexing of superconducting transition edge (TES) arrays [14]. The resolving power of kinetic inductance detectors is determined by an antenna or optical bandpass filter response. As such, KIDs are very low spectral resolution devices.

From the above discussion, it is clear that the optimal detector approach is very much dependent on the scientific goals of the experiment, e.g. spectral resolution, bandwidth, throughput, sensitivity, and instrument stability. With so many approaches to choose from, the instrument designer must carefully consider the scientific goals of the experiment [15].

2.2 Incoherent (direct) detection

As alluded to in the previous Section, heterodyne detection suffers from limited bandwidth and quantum noise. Though suitable for high resolution spectroscopy, the direct or incoherent detection mechanism allows for background limited observations over large bandwidth and low spectral resolution. This facilitates the study of broadband continuum emission from a large variety of galactic and extragalactic sources.

The reason that coherent detectors are subject to a quantum noise limit is that the preservation of signal phase results in an uncertainty in the photon occupation number. When expressed as a temperature, the radiation field that produces an equivalent noise equals $h\nu/2k_B$. It is in the submillimeter wavelength regime that the quantum noise limit becomes comparable, or even exceeds the fundamental photon fluctuation noise (Sec. 2.2.3). A submillimeter direct detector, especially under low background conditions, is therefore always more sensitive than a coherent detector.

In the radio, or longer wavelength regime, the thermal energy ($KT$) dominates the photon energy ($h\nu$), e.g. $KT \gg h\nu$. In the submillimeter both quantities play an important role. For direct detectors at wavelength long wards of 200 $\mu$m the photon energy is however too low to use traditional infrared photoconductive detectors. At these frequencies bolometric detectors have proven very successful, with background
limited performance demonstrated in both single pixel and large format arrays [14].

2.2.1 Bolometers

Bolometers are thermal detectors that are used to sense broadband continuum radiation. These devices respond to the square of the signal input, in contrast to coherent detectors which respond to the product of the local oscillator and signal input (Sec. 2.3). Bolometers can however also be used under suitable conditions as heterodyne mixers. In Chap. 4 we discuss the fundamentals of “hot electron bolometers” (HEBs), and in Chap. 6 the IF bandwidth and mixer conversion gain.

Fig. 2.1 shows the model of a classic bolometer. An incoming photon raises the temperature in the absorber with heat capacity \( C \). In this case the absorbing film is weakly coupled via conductance \( G \) to a thermal bath \( T_0 \). In its most fundamental form the energy balance of the absorber can be described by

\[
C \frac{\partial (T_b - T_0)}{\partial t} + G \Delta T = P_s .
\]

(2.1)

\( \Delta T = (T_b - T_0) \) and \( P_s \) is the incoming power flow (photons). In steady state \( \partial \Delta T/\partial t = 0 \) and \( \Delta T = P_s/G \). Solving Eq. 2.1 for \( T_b \) yields

\[
T_b = T_0 + \frac{P_s}{G} e^{-t/\tau} ,
\]

(2.2)

with \( \tau = C/G \) the thermal time constant of the system. Sensitivity is maximized by reducing \( G \). However, in doing so \( \tau \) increases. This is generally undesirable. Thus to keep \( \tau \) small (or constant), the heat capacity of the absorber also needs to be reduced. This is typically achieved by utilizing smaller area devices. And since the resistance of the absorbing film depends on the temperature \( (T_b) \), a simple current

![Diagram of a bolometer](image)

Figure 2.1: Fundamental model of a bolometer. Sensitivity is optimized by reducing the thermal conductance \( G \), thereby increasing \( \Delta T \). In doing so the thermal response time \( \tau \) increases, unless the thermal capacitance \( C \) of the absorber is decreased. This is usually accomplished by reducing the area of the bolometer. The bolometer resistance is chosen to be a strong function of temperature at the operating condition.
source may be employed to sense the incoming radiation. In practice the situation is a bit more complicated as the bias is ordinarily modulated to stay away from $1/f$ noise in the detector and FET low noise amplifier (Fig. 2.1). It is generally desirable therefore to have the bolometer respond with maximum $\Delta T$, and short thermal time constant $\tau$. This allows for synchronous detection, the rapid differencing of source and background signals (chopping), away from the $1/f$ knee in the power spectrum of the device and atmospheric fluctuation noise. Aside from these issues it is also very important to operate the bolometer as close as possible to the theoretical (photon fluctuation dominated) background, while keeping the bolometer power saturation (linearity) in concurrence with its application.

The signal to noise ratio (SNR) of a direct detector system may be defined as

$$SNR = \frac{P_s}{NEP} \sqrt{2T_{int}},$$

(2.3)

where $T_{int}$ is the effective integration time, and the NEP the “Noise Equivalent Power”. The factor 2 is the result of the chopping efficiency (50 %). Since the NEP of a direct detection system is defined as the rms signal which produces a SNR = 1, we find in the Rayleigh-Jeans limit with $P_s = 2k_B T_s \Delta \nu$ (factor 2 for both polarizations) that the

$$NEP = 2k_B T_s \sqrt{2T_{int}} (W/\sqrt{Hz}).$$

(2.4)

Note that the NEP decreases linearly with decreasing bandwidth ($\Delta \nu$). Thus by increasing the resolving power of the instrument ($\Delta \nu = \nu/R$, Sec. 2.1), for instance by using a FTS or echelle grating, a lower detector NEP will be required to stay background limited.

The responsivity of a bolometer is defined as the change in voltage drop per Watt of absorbed power

$$S_v = \frac{V}{P_s} = I \frac{\partial R \cdot T}{\partial T} \cdot (V/W).$$

(2.5)

$S_v$ can be linked to an NEP via $V_n/S_v$, where $V_n$ equals the (measured) noise voltage.

Often however it is more convenient to measure the electrical responsivity of a bolometer ($S_e$) via the so called isothermal technique [16]. In this case the quantum efficiency of the bolometer is not included, e.g. $S_v = \eta_q S_e$. Note that for a high quality bolometer $\eta_q \sim 0.5$. In many cases it is found that the NEP of bolometers scales with the square root of the area. This is certainly the case with the photon fluctuation noise of Sec. 2.2.3. If $\tau$ is held constant, then $G \propto A$ and $S \propto A^{-1}$ so that the specific detectivity $D^* = \sqrt{A}/NEP$ is approximately constant, and thus a useful figure of merit [16].

Aside from fundamental photon fluctuation noise, there are many additional sources of excess noise that contribute to the overall NEP of a bolometer. Examples are the thermal fluctuation noise of the absorbing element (Johnson noise), phonon noise and $1/f$ noise. Of these three Johnson noise is the most dominating. Since these noise sources may be assumed uncorrelated, we can add the mean square fluctuations to estimate the sensitivity of the bolometer, e.g.
\[ \text{NEP}^2 = \text{NEP}_{ph}^2 + \text{NEP}_e^2, \tag{2.6} \]

where

\[ \text{NEP}_e^2 \approx 4k_B T_b G. \tag{2.7} \]

In an ideal detector with background limited performance (BLIP) the first term of Eq. 2.6 dominates. This is the subject of Sec. 2.2.4, however as a frame of reference neutron-transmutation-doped (NTD) germanium thermisters at 300 mK have achieved NEPs of \( \sim 1 \times 10^{-16} \text{W/} \sqrt{\text{Hz}} \) and \( \tau=11 \text{ ms} \) [17]. At 100 mK bath temperature the NTD Germanium response improves to \( 2 \times 10^{-17} \text{W/} \sqrt{\text{Hz}} \) with a \( \tau=30 \text{ ms} \) [18]. Richards et al. has reported a NEP as low as \( 7 \times 10^{-18} \text{W/} \sqrt{\text{Hz}} \) with a \( \tau=6 \text{ ms} \) [16]. And finally, the spider web bolometers developed for SPIRE [6] on Herschel [7] have a rated NEP of \( \sim 2 \times 10^{-18} \text{W/} \sqrt{\text{Hz}} \). The latter value is well below the thermal background provided by the space observatory. For high background ground observations a noise equivalent power on the order of a few times \( 10^{-15} \text{W/} \sqrt{\text{Hz}} \) or better, suffices for low resolving powers since atmospheric transmission and emissivity from the sky dominate the noise budget (Sec. 2.2.4).

### 2.2.2 TES

Classical bolometers as the one described above work very well. However, in the push towards large format cameras (> 300 pixels) there is an issue with the maximum number of signal wires that may be routed in/out of the cryostat. To circumvent this problem, superconducting transition edge sensors (TES) integrated with SQUID multiplexing readout arrays are being investigated. For example, SCUBA-2 is a 10,000 pixel submillimeter camera [14] based on TES technology. The superconducting multiplexing readout is necessary to manage the large number of signal wires coming from the 300 mK work surface.

The TES uses the superconducting film transition where the resistance is an extremely steep function of temperature. Actual sensors are open structure, and do not use a feedhorn or antenna to couple in the radiation. Thus they are sensitive to a field of view (FOV) of \( \sim \pi \text{ steradian} \). To define the beam with a top-hat like telescope illumination, Liot cold stops at various places in the optical chain are being used.

### 2.2.3 Photon fluctuation noise

As alluded to in the previous Section, fundamental noise in bolometers is limited by fluctuations in the photon arrival rate. This statistical fluctuation with time has to do with the quantum nature of photons, and is known as the background limit. For a single mode, the photon occupation number is given by the Bose-Einstein formula

\[ n_0 = \frac{1}{e^{\hbar \nu/k_B T} - 1}, \tag{2.8} \]

where \( T \) is the effective blackbody radiation temperature. The uncertainty in optical power has been discussed by a great number of people [19, 20, 21] and can for a single mode be expressed as
\[ \sigma_{ph}^2 = \frac{(h\nu)^2 \Delta \nu}{\eta_q T_{int}} \eta_q n_0 (1 + \eta_q n_0) . \]  
(2.9)

This expression is identical to Lammarre et al. [19] with \( P \), the degree of polarization, equal to zero and \( A\Omega/\lambda^2 = 1 \) (single mode diffraction limited beam). In the radio regime with \( k_B T \gg h\nu \) (\( n_0 \gg 1 \)), Eq. 2.9 reduces to the Dicke Radiometer equation (Eq. 3.1). In the optical regime where Poisson statistics dominate, we have \( h\nu \gg k_B T \) (\( n_0 \ll 1 \)) and Eq. 2.9 reduces to [20]

\[ \sigma_{ph}^2 = \frac{(h\nu)^2 n_0 \Delta \nu}{\eta_q T_{int}} . \]  
(2.10)

In the submillimeter and terahertz, we operate in a regime where \( n_0 \sim 1 \), and both terms in Eq. 2.9 need to be considered.

To estimate the photon background limit we need to convert \( \sigma_{ph} \) to a noise equivalent power. This can be done by considering how observations are made. To circumvent system instabilities (Appendix A) synchronous detection, e.g. chopping between source and background via a nutating mirror, is usually applied. This reduces the observing efficiency by a factor of two in time. And because the signal for astronomical observations is deeply embedded in the noise, subtraction of the two essentially white noise signals (Gaussian distribution) increases the noise by \( \sqrt{2} \). The SNR is thus given by

\[ SNR = \frac{P_s}{\sqrt{2} \sqrt{2} \sigma_{ph}} . \]  
(2.11)

Via Eq. 2.3 we obtain the photon noise dominated NEP (W/\( \sqrt{Hz} \)), defined in 1 s integration time, as

\[ NEP_{ph} = 2 \sqrt{2} \sigma_{ph} . \]  
(2.12)

Note that this quantity is referred to the input of the detector.

### 2.2.4 Emissivity, optical efficiency, and sky noise

Use of a telescope for observations will give additional sources of photon noise, of which the dominant ones are thermal noise due to a finite telescope temperature, and Ohmic loss in the optics chain. There is also the issue of optical coupling efficiency, and for ground based observations the non-negligible thermal noise of a partially transparent sky (Sec. 2.4). To include these effects, we incorporate in Eq. 2.9 the emissivity \( \varepsilon \) of an equivalent blackbody with radiation temperature \( T \), and transmittance \( \tau \). It is important to quantize these numbers if background limited performance (detector noise \(< \) photon noise) observations are to be achieved. We thus obtain the generalized variance of the photon noise as

\[ \sigma_{ph}^2 = \frac{(h\nu)^2 N_m \Delta \nu (\varepsilon \tau n_0)}{\eta_q T_{int}} (1 + \varepsilon \tau n_0 \eta_q) . \]  
(2.13)
Chapter 2: Submillimeter detection and instrumental requirements

\( N_m \) are the number of modes (2 for both polarizations). For a perfect blackbody the emissivity equals the absorbivity or \((1-\Gamma^2)\), where \(\Gamma^2\) is the optical voltage reflection coefficient. For metallic mirrors the emissivity can be approximated as \(\varepsilon_m=0.1\lambda^{-0.5}\) [19]. Rewriting Eq. 2.13 in terms of the resolving power \(R\) and \(\text{NEP}_{ph}\) yields

\[
\text{NEP}^2_{ph} = \frac{8(h\nu)^2}{\eta_q R} N_m \nu (\varepsilon \tau n_0) (1 + \varepsilon \tau n_0 \eta_q) \left( W^2/Hz \right). \tag{2.14}
\]

Thus the higher the resolving power the lower the photon noise, and the more sensitive the detector has to be. Note that the \(\text{NEP}_{ph}\) is referred to the detector input. To refer it to the telescope input, or the top of the atmosphere as the case may be, one has to divide by the instrument optical efficiency \((\eta_{opt})\), telescope spillover efficiency \((\eta_s)\), and atmospheric transmission \((\eta_{atm})\).

So far we have ignored the astronomical background. In low background observations (space) galactic and Zodiacal foreground emission can not be ignored. From Lamarre et al. [19] we estimate a \(\text{NEP}_{ph,\text{space}}\) of \(2.25 \times 10^{-18}/\sqrt{R}\) in the \(0.3 – 3\) THz frequency range. The \(\text{NEP}_{ph,\text{space}}\) may be added in a statistical manner since the noise is uncorrelated from the other noise sources.

For terrestrial observations the atmospheric transmission, \(\eta_{atm}=e^{-\tau_{atm}}\), is incorporated in the general emissivity via

\[
\varepsilon = \varepsilon_m' + (1 - \varepsilon_m')[(1 - \eta_s) + \eta_s(1 - \eta_{atm})]. \tag{2.15}
\]

In Eq. 2.15 \(\varepsilon_m'\) is the overall emissivity of the telescope mirrors, and \(\eta_s\) the telescope spillover, or forward efficiency \((\sim 0.9)\). Defined in this way, \((1-\eta_s)\) is that part of the beam which terminates somewhere in the (warm) telescope. The sky temperature, for use in the mode occupation number \(n_0\), can be estimated from \(T_{sky}=0.95T_{\text{amb}}\).

In case of a space telescope \(T\) equals the equivalent telescope temperature.

### 2.2.5 Background limited detection

The next three examples demonstrate the effect that emissive telescope structures and the atmosphere have on the photon background limit. This limit is a useful figure of merit as it sets a sensitivity requirement on the described incoherent detectors.

#### 2.2.5.1 Example 1: A 80 K space telescope

SPIRE [6], the imaging Fourier transform spectrometer instrument on Herschel [7], has a resolving power of \(R \sim 1000\). Assuming a telescope emissivity of \(\varepsilon=0.04\), telescope temperature of 80 K, 50 % detector quantum efficiency, 50 % instrumental optical efficiency \((\eta_{opt})\), frequency of operation 850 GHz (350 \(\mu\)m), sensitivity to both polarizations, and a 90 % spillover efficiency, we find that \(n_0 = 0.601\), and that to be background limited the SPIRE detector NEP has to be less than \(1.4 \times 10^{-17} W/\sqrt{Hz}\).

In fact the SPIRE detector NEP \(~ 2 \times 10^{-18}\), which ensures that the instrument is indeed limited by the photon background.

To estimate the NEP referred to the output of the telescope we have
\[ NEP_{tel}^{ph} = \sqrt{\left( \frac{NEP_{ph}}{\eta_q\eta_{opt}\eta_s\eta_{atm}} \right)^2 + NEP_{ph,\text{space}}^2}, \]  
(2.16)

or \( 6.4 \times 10^{-17} \text{ W/} \sqrt{\text{Hz}} \) (\( \eta_{atm} = 1 \)).

Note that \( n_0 = 0.6 \) indicates that we are indeed midway between the optical and radio regimes. In this case, some of the photons are bunched and both terms of Eq. 2.14 need to be used to properly describe the photon statistics.

### 2.2.5.2 Example 2: A 4 K (cooled) space telescope

Now let's assume the same specifications as above, but with the telescope cooled to 4 K. In this case the background limit is \( 1.1 \times 10^{-19} \text{ W/} \sqrt{\text{Hz}} \), a significantly more difficult requirement. \( NEP_{tel}^{ph} = 5.1 \times 10^{-19} \text{ W/} \sqrt{\text{Hz}} \).

### 2.2.5.3 Example 3: A warm ground based telescope

As a third example let's assume a direct detector with similar specifications as in the example 1, but now located at Chajnantor, Chile. The instrument is again operational in the 350 \( \mu \text{m} \) atmospheric window under 0.5 mm (25 percentile) precipitable water vapor condition with an ambient temperature of 273 K and airmass \( A = 1.0 \). \( \eta_s \) is again 90\%, however the telescope emissivity has increased to 0.1 (larger telescope, higher surface rms and more blockage). From Fig. 2.8 we have an atmospheric transmission of 50\%. \( \varepsilon \) is therefore 0.64, and \( n_0=0.861 \). To achieve BLIP conditions, the bolometer NEP should be \( \leq 5.2 \times 10^{-17} \text{ W/} \sqrt{\text{Hz}} \). This is readily achieved with today's technology. Thus to gain more sensitivity one is left with increasing the telescope aperture for point source observations (diffraction limited beam size = 1.22 D/\( \lambda \)) and/or increasing the number of pixels. The NEP referred to the top of the atmosphere (Eq. 2.15) is \( 4.6 \times 10^{-16} \text{ W/} \sqrt{\text{Hz}} \). Excess voltage noise has been ignored in the discussion, thus the above given NEP is somewhat optimistic. This result is for a resolving power of 1000. For \( R = 10 \), the NEP referred to the top of the atmosphere becomes \( 4.6 \times 10^{-15} \text{ W/} \sqrt{\text{Hz}} \).

Some final notes: Johnson (electronics) noise is minimized by lowering the bath temperature \( T_b \) (Eq. 2.7). Low background bolometers are therefore operated below 100 mK. For ground based bolometers this requirement may be relaxed to 300 mK as shown by the required NEPs in this example.

Lamarre et al. (1995) [19] showed that for a wavelength < 400 \( \mu \text{m} \) the telescope temperature is the dominating factor in meeting the background limited BLIP condition. For \( \lambda > 400 \mu \text{m} \) both emissivity and temperature are important and thus low emissive telescope/mirror designs must be utilized.

### 2.3 Coherent (heterodyne) detection

In this section we present a variety of coherent detection concepts. These form the basis for the heterodyne detection techniques discussed in this thesis.
A heterodyne receiver down-converts the RF signal to a lower intermediate frequency (IF) without loss of phase. This stands in contrast to the incoherent detection described in the previous Section. The purpose of down-converting the signal to a lower frequency is that commercial low noise electronics can than be used in the signal post detection process.

A characteristic of heterodyne detection is the preservation of signal phase. This feature of the heterodyne down-conversion process is well suited for arraying multiple antennas. This technique is known as interferometry (Chap. 3) and effectively synthesizes a larger telescope area, thereby enhancing the instrument sensitivity. Good examples are ALMA [1], the SMA [4], IRAM [3], and OVRO/CARMA [2]. The IF output signal is usually processed by a high resolution backend spectrometer or correlator. In the case of a single dish antenna an acousto-optical spectrometer (AOS) [22, 23] or fast Fourier transform spectrometer (FFTS) [24] is typically employed. Interferometers (Chap. 3) have their IF outputs signals cross-correlated by either a commercial [25] or custom build digital correlator. The preservation of phase in a heterodyne system is also the foundation of modern day communication systems. Examples include phase modulation in mobile phones, radio communication, and satellite telemetry.

To understand the mechanism of heterodyne detection we depict in Fig. 2.2 the layout of a fundamental double sideband (DSB) heterodyne receiver. The block diagram is very similar to the schematic layout of the single-ended receiver in Fig. 3.2. Refering to Fig. 2.2, the input signal $g_s(t)$ is multiplied by a local oscillator signal $g_{lo}(t)$ as:

$$g_s(t) \times g_{lo}(t) = V_s \cos(\omega_s t + \varphi_s) V_{lo} \cos(\omega_{lo} t + \varphi_{lo})$$
$$= \frac{V_s V_{lo}}{2} \left[ \cos[(\omega_s + \omega_{lo}) + (\varphi_s + \varphi_{lo})] + \cos[(\omega_s - \omega_{lo}) - (\varphi_s + \varphi_{lo})] \right].$$

(2.17)

For an ideal mixer with unity mixer gain and $V_s = V_{lo} = 1$, the mixer IF output, after passing through an appropriate bandpass filter, thus becomes
2.3. COHERENT (HETERODYNE) DETECTION

Figure 2.3: Schematic representation of heterodyne detection. The signal of interest is combined with a strong local oscillator signal and down-converted to an intermediate frequency $\nu_{IF}$. Top) 2SB mixer. At the mixer IF output a mathematically negative frequency will be positive, with a $180^\circ$ phase shift. Bottom) DSB mixer. See text for details.

$$g_{IF}(t) = \cos (|\omega_s - \omega_{lo}| t + \phi_{IF}).$$

$\phi_{IF}$ is the down converted phase, $(\phi_s + \phi_{lo})$. The sign of $\phi_{IF}$ depends on whether the IF frequency is mathematically positive or negative. From the above expression it is clear that phase jitter, or frequency modulation noise (FM) on the local oscillator signal will place a limit on the maximum achievable frequency resolution of the down-converted IF signal. In telecommunication systems stringent requirements on the phase stability of local oscillators is therefore a common requirement. From Eq. 2.18 it is evident that if the signal frequency is greater than the LO frequency ($\omega_s > \omega_{lo}$), that the corresponding intermediate frequency is positive. Such a signal is said to be in the upper sideband (USB). If however the signal frequency is less than the LO frequency ($\omega_s < \omega_{lo}$) the intermediate frequency is mathematically negative, and the signal will reside in the lower sideband (LSB). This process is graphically shown in the top panel of Fig. 2.3. Of course negative IF frequencies do not exist (time only moves forward), and the down-converted lower sideband will be present as a positive frequency in the 2SB receiver IF output, albeit with a $180^\circ$ phase shift. For a double-sideband receiver, the upper and lower sidebands fold in the IF. This is undesirable for many applications, and significant efforts have been expanded to design mixers that separate the upper and lower sidebands at the IF output. In Chap. 3 we take a close look at these mixers configurations, and in Chap. 8 we present the design, fabrication, and measurement of a $600 – 720$ GHz [26] sideband separating receiver.

In practice the signal, or astronomical line, will be present in either the upper or lower sideband. The sideband the received signal resides in is often referred to as
the “signal” sideband, with the opposite sideband than being the “image” sideband. These nomenclatures will be used quite extensively in the remaining chapters of the thesis. It should also be noted that an astronomical line placed in the upper sideband will move in a negative frequency direction with increasing LO frequency, whereas an astronomical line in the lower sideband will move in a positive frequency direction with increasing LO frequency. In other words, a change in LO frequency results in either a positive or negative translation, depending which sideband the astronomical line resides in. This feature is extensively used to verify which sideband the received signal resides in. Fig. 2.3 demonstrates this principle.

As a final note, the down-conversion process may also be understood in Fourier space by the principle of frequency shifting [27]. In this case, let \( g_s(t) \) again be the RF signal, than

\[
\mathcal{F}[g_s(t)e^{i\omega_s t}] = \int_{-\infty}^{\infty} g_s(t) e^{i\omega_s t} e^{-i\omega t} dt \\
= \int_{-\infty}^{\infty} g_s(t) e^{-i(\omega - \omega_s) t} dt \\
= G(\omega - \omega_s) .
\] (2.19)

If now \( g_{lo}(t) = \cos(\omega_{lo} + \varphi_{lo}) \) we obtain, using Euler’s identity, the spectrum in phasor notation as

\[
\mathcal{F}[g_s(t)\cos(\omega_{lo} + \varphi_{lo})] = \frac{1}{2} \left[ G(\omega_s + \omega_{lo}) e^{-i\varphi_{lo}} + G(\omega_s - \omega_{lo}) e^{i\varphi_{lo}} \right] 
\] (2.20)

Again, by means of an appropriate bandpass filter the difference frequency (IF) can be selected. It is however important to note that there are many occasions where up-conversion, filtering, and down-conversion is more appropriate if interference of harmonic content (intermodulation) is to be avoided. A more detailed description is beyond the scope of this thesis as in the submillimeter and terahertz harmonic content, thanks to device parasitics, can usually be ignored.

### 2.3.1 Sensitivity of a heterodyne receiver

An important characteristic of heterodyne receivers is the added noise, and efficiency of the down-conversion process. A practical mixer (or amplifier) adds noise in the form of thermal, quantum, flicker (1/\( f \)), and/or shot noise. In addition, a mixer has conversion efficiency, commonly known as mixer gain. This process is somewhat similar to amplifiers, with the exception that amplifiers have by definition a gain greater than unity, whereas mixers typically exhibit a conversion gain less than unity and have an output frequency that is by definition different from the input frequency. (In Chap. 4 we discuss SIS mixers which due to their quantum nature can in principle have a mixer conversion gain). It is often convenient to consider a mixer, or amplifier, as being lossless with an equivalent noise source at the input. In Fig. 2.4 we demonstrate this mechanism. Mathematically we can describe this situation as
2.3. COHERENT (HETERODYNE) DETECTION

\[ P_{\text{out}} = G(P_{\text{in}} + P_n) . \]  
\[ (2.21) \]

\[ G \] is the gain, \( P_n \) the noise presented to the input, and \( P_n \) the equivalent noise temperature of the device. And because radio-astronomy and microwave instrumentation operates in the Rayleigh-Jeans limit of the Planck blackbody power spectrum (a 77 K blackbody has a peak radiation at 3.9 \( \mu \)m, or 76 THz) noise powers are typically expressed in terms of equivalent temperature via \( P = k_B T \). We can therefore re-write Eq. 2.21 as

\[ T_{\text{out}} = G(T_{\text{in}} + T_n) , \]  
\[ (2.22) \]

and thus the noise referred to the input of the mixer becomes

\[ T_n = \frac{T_{\text{out}}}{G} - T_{\text{in}} . \]  
\[ (2.23) \]

In electrical engineering applications, it is common practice to express the noise temperature of a device as a noise figure (\( F \)) via

\[ F_{\text{db}} = 10 \log_{10}(1 + T_n/T_0) . \]  
\[ (2.24) \]

\( T_0 \) is referenced to room temperature, 290 K by definition. For an ideal instrument with no added noise (\( T_n = 0 \)) and unit conversion gain we have the situation that \( T_{\text{in}} = T_{\text{out}} \). In practice having a heterodyne receiver with a zero Kelvin noise temperature is physically impossible by virtue of the Heisenberg uncertainty principle. This can be understood from the phase coherent down-conversion process where uncertainty in the photon occupation number results in at least half a quantum of noise \( (h\nu/2k) \) being present at the input of the receiver. Whether this half a quantum of noise is attributed to the blackbody input load (the mixer is noiseless in this case), or to the mixer itself is a matter of convention. In actual superconducting heterodyne receivers the sensitivity ranges from a few to 10 times the quantum noise.

Figure 2.4: Top) Definition of equivalent noise power referred to the input of a mixer or amplifier. Bottom) Circuit description of an amplifier with noise temperature \( T_n \).
The Planck formulation of a true blackbody assigns the half a quantum of noise to the mixer. The Callen & Welton definition \([28, 29]\) assigns it to the input load. For reference we give here both definitions. In a unit bandwidth

\[
P_{\text{Planck}}(T, \nu) = \frac{h\nu}{e^{h\nu/k_B T} - 1} = n_0 h\nu,
\]

and

\[
P_{\text{CW}}(T, \nu) = \frac{h\nu}{e^{h\nu/k_B T} - 1} + \frac{h\nu}{2} = (n_0 + \frac{1}{2}) h\nu.
\]

In the Rayleigh-Jeans limit \((h\nu \ll k_B T)\) \(P_{\text{CW}} \sim P_{\text{RJ}}\), and \(P_{\text{Planck}}\) is always half a photon below \(P_{\text{CW}}\).

If we now cascade \(m\) devices, each with a noise temperature and gain referred to the input, it is easy to show that the equivalent noise temperature of the cascaded chain will be given by

\[
T_{\text{eq}} = T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1 G_2} + \cdots + \frac{T_n}{G_1 G_2 \cdots G_{n-1}}
\]

This formulism of the equivalent noise temperature of a cascaded chain is useful in determining the overall noise temperature of a receiving system (Eq. 7.6).

To quantify the equivalent noise of a receiver or amplifier the “Y-factor” method is usually employed. Here a blackbody radiator with known (calibration) temperature is connected to the device under test (DUT). Under the assumption that the output of the receiver or amplifier responds linearly to a change in input noise temperature (Eq. 2.22), the equivalent noise temperature is then readily obtained via

\[
T_{\text{eq}} = \frac{T_{\text{hot}} - Y T_{\text{cold}}}{Y - 1}.
\]

\(Y\) is the ratio of the corresponding output powers \((P_{\text{hot}}/P_{\text{cold}})\). It is common to use the Callen & Welton definition for the equivalent noise temperatures of the blackbody input load. If there is (optical) loss between the blackbody radiator and the device under test, then this may be taken into account via the procedure outlined in Sec. 5.1.2. In receiving systems \(T_{\text{eq}}\) is usually referred to as \(T_{\text{DSB}}\) with noise being contributed by both sidebands. For an ideal SSB receiver, \(T_{\text{SSB}}^{\text{rec}} = 2T_{\text{DSB}}^{\text{rec}}\). A procedure to derive the DSB receiver noise temperature of a SIS receiver using the Tucker theory (Chap. 4) is provided in Sec. 4.1.7.3.

### 2.3.2 Conversion of NEP to system noise temperature

It is sometimes convenient to express the NEP \((W/\sqrt{Hz})\) in temperature units \((K)\). This also facilitates comparison between coherent and incoherent systems. Caution is warranted however when comparing direct detection systems with heterodyne receivers as the former is primarily designed to measure broadband continuum and the latter vibration and rotational atomic and molecular fine line transitions. Nevertheless, for medium resolution direct detection systems such as PACS, SPIRE, Z-Spec, and ZEUS it can be useful.
The noise equivalent flux density, referred to the top of the atmosphere is given by

$$NEFD = \frac{NEP_{ph}}{A_e \Delta \nu} \left( \text{Jy/}\sqrt{\text{Hz}} \right).$$  \hspace{1cm} (2.29)

$A_e$ is the effective antenna aperture ($\eta_A A_{tel}$), $\Delta \nu$ the resolution bandwidth ($\nu/R$), and $\eta_A$ the telescope aperture efficiency. 1 Jy is the unit of flux density, defined as $10^{-26}$ W m$^{-2}$ Hz$^{-1}$. The flux density sensitivity in units of W m$^{-2}$ Hz$^{-1}$ of a direct detector is

$$S^{dd}(\text{rms}) = \frac{NEFD}{\sqrt{n2T_{int}}},$$  \hspace{1cm} (2.30)

where $n$ is the number of samples taken in case of a FTS direct detector. The factor of two accounts for the on-off source chopping efficiency (50%).

For a double sideband heterodyne receiver noise temperature $T^{\text{DSB}}_{\text{rec}}$ the single sideband (SSB) system temperature referred to the top of the atmosphere is given by Eq. 2.41. The antenna temperature, inclusive of a factor 2 due to chopping efficiency (Sec. 2.3.5) can be obtained from the Dicke or radiometer equation (Eq. 3.1)

$$T_A^* = \frac{2T^{\text{SSB}}_{\text{sys}}}{\sqrt{\Delta \nu T_{int}}} \left( \text{K} \right),$$  \hspace{1cm} (2.31)

with $T_A^*$ related to the flux density via [30]

$$S^{\text{het}}(\text{rms}) = \frac{2k_B T_A^*}{A_e \eta_{mb}}.$$  \hspace{1cm} (2.32)

Note that for a single mode (Gaussian) telescope illumination $A_e \Omega / \lambda^2 = 1$. Equating $S^{dd}$ to $S^{\text{het}}$ and solving for $T_{\text{sys}}^{\text{SSB}}$ yields

$$T_{\text{sys}}^{\text{SSB}} = \frac{\eta_{mb} NEP}{4\sqrt{2nk_B \Delta \nu}}.$$  \hspace{1cm} (2.33)

For example, the photon background limited direct detector in example 3 (Sec. 2.2.5.3) has, with $n = 1$, and $\eta_{mb} = 0.7$, an equivalent $T_{\text{sys}}^{\text{SSB}} = 142$ K. This can be compared to the record NbTiN twin-slot DSB receiver of Chap. 5 ($T_{\text{rec}}^{\text{DSB}} = 205$ K) via Eq. 2.41, by using the same atmospheric and telescope conditions. In this case $T_{\text{sys}}^{\text{SSB}} = 1580$ K. In other words, we see that the direct detector of example 3 is eleven times more sensitive. Caution is warranted however since we have taken $n = 1$, and have ignored electronic noise in the detector. On a point source, the sensitivity of a FTS based direct detector will decrease by $\sqrt{n}$. This explains the popularity of multi-pixel grating spectrometers such as ZEUS, Z-spec, and PACS [8, 9, 12].

As a final note, the flux density to temperature conversion of a given telescope may be obtained by dividing $S^{\text{het}}/T_A^*$. In units of Jy/K this gives

$$\frac{S^{\text{het}}}{T_A^*} = \frac{2.76 \times 10^3 \text{Jy}}{\eta_A A_{tel} \eta_{mb}}.$$  \hspace{1cm} (2.34)
For the CSO, with a telescope diameter of 10.4 m, a low frequency main beam efficiency of 70 %, and an aperture efficiency of \( \sim 90 \% \), \( S^{\text{het}}_{\nu}/T^{*}_{A} = 51 \text{ Jy/K} \). Thus a 150 \( \mu \text{K} \) line equates to 7.6 mJy of flux.

### 2.3.3 IF Bandwidth

Another important mixer parameter is the IF bandwidth. It has direct bearing on the time efficiency of spectral line surveys as demonstrated in Fig. 2.5. It follows that an increased IF bandwidth allows for the simultaneous observation of several (or many) molecular lines. A large IF bandwidth is also important in the measurement of broad emission lines from external galaxies.

Not until quite recent (1995), the IF bandwidth of heterodyne SIS receivers was limited by the availability of low noise amplifiers to approximately 500 MHz. This put a significant constraint on the maximum observable linewidth (\( \sim 400 \text{ kms}^{-1} \) at 345 GHz), and even less at higher frequencies. It meant that only molecular line transitions of galactic sources and some nearby galaxies could be observed. Though considered small these days, it was enormous in comparison to the IF bandwidth afforded by the thermal relaxation time of the InSb hot electron bolometer [31]. These offered IF bandwidth of just a few MHz, and to build up a spectrum the LO had to be scanned!

Today’s technology has changed all that. IF bandwidth of 4 GHz, and higher are possible thanks to advances in SIS junction and low noise InP MMIC technology.
To demonstrate the significance we show an Orion-KL spectrum obtained with Trex, the technology development receiver described in Chap. 7 [32]. What used to take 8 observations (or 4 with 1 GHz IF bandwidth [33, 34]) now takes just one.

Significant, IF bandwidths of 4 or even 8 GHz allow for observations of distant galaxies with $\sim 3000 \text{ kms}^{-1}$ linewidth. With the increased IF bandwidth, instrument stability as discussed in Sec. 2.3.5, Chap. 7, and Appendix A becomes of utmost importance.

As a final note, HEB mixers operating in the terahertz regime have a limited IF bandwidth due to the thermal time constants of the electron-phonon interaction time and phonon escape time (Chap. 6). The IF gain bandwidth of NbN based hot electron bolometers is limited to approximately $3 – 5 \text{ GHz}$. In practice essentially all HEB mixers operate at IF frequencies below $4.8 \text{ GHz}$ with practical IF passbands of $1 – 2.4 \text{ GHz}$ [37, 38]. This puts a similar velocity coverage constraint on the HEB mixers as the SIS mixers experience in the early days. For example, a $\Delta \nu = 1 \text{ GHz}$ at $1.900 \text{ THz} (\text{C}^+)$ gives maximum velocity resolution of $160 \text{ kms}^{-1}$, just barely enough for observations toward the galactic center. For the HIFI instrument, with an IF bandwidth of $2.4 \text{ GHz}$, a maximum velocity coverage of $380 \text{ kms}^{-1}$ at $1.900 \text{ THz}$ is possible.

### 2.3.4 RF Bandwidth

The first submillimeter heterodyne SIS receivers had RF instantaneous bandwidths of just a few GHz [39]. Integrated SIS junction RF matching networks (Sec. 4.1.8) coupled to wide bandwidth (fixed tuned) waveguide probes, and antenna coupled quasi-optical mixers (Chap. 5) changed all that. Today with the advent of high current density AlN barrier junctions and modern electromagnetic simulations tools such as HFSS [40], Sonnet [41], and Supermix [42] instantaneous RF bandwidth of $140 \text{ GHz}$ or more are possible (Chap. 7).

HEB mixers, having very little parasitic capacitance, are sensitive to a large range of photon energies. They can be coupled to with either waveguide or open structure antennas (Sec. 4.3). The very large instantaneous RF bandwidth of these devices is liable to result in direct detection (see Sec. 4.3.2), due to a change in heat loading by the different calibration loads. Heat loading on the HEB mixer can however be minimized by either reducing the RF bandwidth (via optical or electrical means) or by reducing the brightness temperature of the calibration loads.

### 2.3.5 Instrument stability and baseline quality

Throughout the thesis instrument stability is discussed as an important system parameter in establishing time efficient observations. Considering the generally large expense, demand on telescope time, and desired quality of the data products, system stability has become an important design parameter for modern heterodyne instrumentation. In the bottom panel of Fig. 2.6 we depict two simulated spectra from actual HIFI [38] data, as obtained in HEB mixer band 7 during instrument level tests (ILT).
The spectra are shown for four different velocity binning resolutions; native (0.086 km/s\(^{-1}\)), 1 km/s\(^{-1}\), 2 km/s\(^{-1}\), and 5 km/s\(^{-1}\). To convert the velocity resolution to spectral resolution we use the Doppler relationship \(v = c/R\) where \(R = \nu/\Delta\nu\). In the example of Fig. 2.6, the LO frequency equals 1.8980 THz (C\(^{+}\)). The corresponding spectral resolution \(\Delta\nu\) may thus be obtained as: native 0.5462 MHz [22], 6.33 MHz, 12.67 MHz, and 31.67 MHz. In the top panel of Fig. 2.6 we depict the total power and spectroscopic Allan variance for the total (full) and individual spectrometer subbands. Details on the Allan variance method can be found in Appendix A.

For standard position switch observations the source is observed for a time \(t_{on}\), after which an off-source (\(t_{off}\)) reference measurement is taken. The duration of the reference measurement is ordinarily \(\sqrt{t_{on}}\). To remove the sky, telescope and instrumental baselines the “on” source signal is subtracted from the “off” source.

Figure 2.6: a) Total power stability of HIFI HEB mixer band 7 at 1.8970 THz. The Allan time, in a fluctuation noise bandwidth of \(\sim 1.8\) MHz, is \(\leq 8\) s. b) Spectroscopic stability (text) with a measured stability time of \(\sim 80\) s. c) Synthesized position switched spectrum. Total “on-source” integration time in one 600 s cycle is \(\sim 464\) s, clearly much too long given the stability of the instrument. Severe baseline distortion is the result. d) Synthesized double beam switched (DBS) spectrum for different velocity resolutions. Each phase of the chop cycle is 4 s, well below the stability time of the instrument. For a native resolution the theoretical and synthesized 1\(\sigma\) rms noise levels are virtually identical (64 mK vs. 68 mK). Total “on-source” integration time is 2829 s.
signal as part of the calibration routine (Sec. 2.3.6).

For a “stable” receiver (instrument) position switching is the most efficient method of observing, since a relatively large percentage of the time is spent integrating on the source. Unfortunately for HEB based heterodyne receivers, “stable” appears to be a bit of an oxymoron, as evidenced by the total power Allan variance stability measurement of Fig. 2.6a. Typical total power Allan variance times, defined by a $\geq \sqrt{2}$ deviation from the ideal radiometer response, are commonly less than 8 s. The spectroscopic stability is obtained by subtraction of the common mode (Appendix A, Chap. 10), and may be as large as $\sim 80$ s in a 1.8 MHz noise fluctuation bandwidth ($\Delta \nu$) as evidenced from Fig. 2.6. The data presented here was obtained from measurements on the HIFI instrument [38] during instrument level tests at SRON, the Netherlands. Note that a balanced HEB configuration, by virtue of the in Chap. 8 discussed amplitude noise cancellation properties, is liable to have a significantly enhanced ($\sim 10$ dB) immunity to environmental and synthesizer induced local oscillator instability (Chap. 10). Recent results by Pantaleev et al. appear to bear this out [43].

In panel c of Fig. 2.6 we show a synthesized position switched spectrum with an “on-source, slew time, off-source, and again slew time” cycle of 600 s. Total on-source integration time of the entire data set is 2789 s. For the Herschel space observatory the roundtrip slew time is assumed 80 s. In position switch mode, each 600 s cycle thus spends $\sim 464$ s integrating on the source with the remainder in slew time (80 s) and off-source integration ($\sim 56$ s). Clearly the “on-off” switching time is way beyond the spectroscopic stability time of the system. This is obvious from the extremely poor baseline quality of Fig. 2.6c.

A far better, though less efficient approach, is to symmetrically beam switch at a rate less than the spectroscopic Allan stability time of the system, by means of nutating mirror. In Fig. 2.6d we show the result of a double-beam switch (DBS) 0.25 Hz “off-on-on-off” chopping pattern. Again each cycle is 600 s and includes one position switch cycle as described above. Including the telescope chopping efficiency and position switch overhead, the total integration time is 2829 s. To estimate $T_A^*$ and compare it to the $1\sigma$ noise obtained from the SSB spectrum of Fig. 2.6d we have included in Eq. 2.31 the chopping efficiency (factor 2 in time), and a factor $\sqrt{2}$ due to the subtraction of two statistically uncorrelated noise levels.

$\Delta \nu$ is the noise fluctuation bandwidth of the spectrometer (1.8 MHz) which is larger than the spectrometer native resolution of 0.5462 MHz. Given a measured DSB HEB receiver noise temperature (in the lower region of the IF band) of $\sim 1600$ K and assuming equal sideband ratios, we calculate a theoretical $1\sigma$ rms noise level of 64 mK. This compares favorably with the, from the spectrum of Fig. 2.6d, obtained $1\sigma$ rms noise level of 68 mK (native resolution). Thus we find that symmetric beam switching on time scales less then the spectroscopic Allan time provides proper baseline quality with rms noise levels in agreement with theory. The penalty of differential beam switch measurements over (ideal) position switch measurements is a factor 2 increase in the rms noise level. The overwhelming benefit is of course the quality of the obtained baseline (spectra). The popular use of DBS techniques (analogous to synchronous detection) comes therefore as no surprise.
2.3.6 Calibration

Proper calibration of heterodyne radio-astronomical data is of utmost importance for correct scientific interpretation. To this extent a distinction can be made between space based observations (the HIFI instrument on the Herschel space observatory [44]) and ground based observations [45]. The latter is primarily concerned with the calibration of the Earth atmospheric emission and absorption properties. Due to the short nature of the atmospheric fluctuations (1/f power spectrum), ground based intensity calibrations are naturally exposed to significant errors. This situation allows for some important generalizations and simplifications of the calibration routine. For space based missions the lack of atmosphere puts a more stringent requirement on the amplitude calibration of the instrument. Significant efforts have been expanded by a variety of authors to tackle this problem [44, 45]. For the purpose of this thesis we provide an overview of the problem.

In general, intensity calibration of the astronomical source can be achieved with either a one- or two load chopper wheel. In the first scenario the brightness temperature of an off-source measurement (this includes the atmosphere in case of ground based observations) and a single load with known emissivity is used to derive the proper antenna temperature of the observed astronomical source. In the second scenario the radiation temperatures of both loads are used. The two chopper wheel approach provides a more robust solution since the actual atmospheric temperature, in case of the one load measurement, is difficult to estimate accurately.

One of the difficulties in absolute amplitude calibration are instrumental fluctuations over longer time scales. For SIS receivers this is generally not a problem. The ALMA specification is $10^{-4}$ in 1 s, which for a properly designed system (no ground loops, minimal thermal fluctuations, optical standing waves...) is not too difficult to achieve. Hot electron bolometer (HEB) mixers on the other hand are intrinsically very sensitive to external influences (Sec. 4.3) and, as may be seen from Fig. 2.6, a total power stability of a few seconds in a 1.8 MHz noise fluctuation bandwidth is unfortunately not unrealistic. In this case the normalized rms noise, $\sigma/\langle s(t) \rangle = 1/\sqrt{\Delta f T_{int}}$, is $\sim 5 \times 10^{-4}$. For a 1 GHz total power bandwidth this quantity may be degraded by approximately the square of the bandwidth, e.g. $1.2 \times 10^{-2}$ or about 1 % in 1 s. This is approximately a factor 100 less stable than the ALMA specification. In practice short time scale differential “on-off” source measurements such as the symmetric double beam switching of Fig. 2.6d are needed with HEB based receiver systems.

2.3.6.1 Two-load calibration

For the load calibration procedure to work, the radiation temperature ($J_{load}$), and coupling efficiency of both sidebands to the calibration load ($\eta_{load}$) have to be known. The radiation temperature of a load with physical temperature $T$ may be computed from the Planck formula as

$$J_b = B_\nu(T) = \frac{2h\nu^3}{c^2} \left[ e^{h\nu/k_B T} - 1 \right]^{-1}. \quad (2.35)$$
Generally speaking, the observed signal will only be in one of the sidebands. Under this condition the equation for $T_A^*$ observed at an airmass $A$ becomes

$$T_A^* = \frac{(V_{\text{source}} - V_{\text{sky}})}{G_s K \eta e^{-A \tau_{\text{atm}}}}. \tag{2.36}$$

$V_{\text{source}}$ is the observed signal, $V_{\text{sky}}$ the off-source position, and $G_s$ the fractional signal sideband gain ($G_s + G_i = 1$). For modern large instantaneous RF bandwidth receivers (Sec. 2.3.4) the sideband ratio ($G_s/G_i$) may, to an error of $\approx 5\%$, be approximated as unity. $K$ is a calibration factor. It links the backend spectrometer IF output to the antenna temperature. Following [45], the calibration factor $K$ in case of a two-load chopper wheel can be derived as

$$K = \frac{V_{\text{load}1} - V_{\text{load}2}}{\eta_{\text{load}1} G_s J_{\text{load}1}^s + G_i J_{\text{load}1}^i - \eta_{\text{load}2} G_s J_{\text{load}2}^s + G_i J_{\text{load}2}^i}, \tag{2.37}$$

with

$$T_A^*(2\text{load}) = \frac{(V_{\text{source}} - V_{\text{sky}})}{(V_{\text{load}1} - V_{\text{load}2})} \left( \frac{\eta_{\text{load}1} G_s J_{\text{load}1}^s + G_i J_{\text{load}1}^i - \eta_{\text{load}2} G_s J_{\text{load}2}^s + G_i J_{\text{load}2}^i}{\eta_s e^{-A \tau_{\text{atm}}}} \right). \tag{2.38}$$

In the above equation, $\eta_{\text{load}1}$ and $\eta_{\text{load}2}$ are the respective coupling efficiencies to the load. The superscripts $s$ and $i$ stand for the signal and image sideband. It is apparent from Eq. 2.38 that the atmospheric transmission in zenith, $\eta_{\text{atm}}$, has to

![Figure 2.7: Differential stability measured on the HIFI internal loads (physical temperatures of 4 K and 80 K). The result is for the same mixer (B7b, 1.890 THz) as in Fig. 2.6. The results show that the internal load calibration loop is stable to at least 900 s. Thus load intensity and standing wave calibrations with the HIFI HEB mixers may occur on a time scale of at least 15 minutes.](image-url)
be known accurately. Mangum et al. [45] found that for a two-load calibration, an atmospheric opacity accuracy of 1% corresponds to an ~2% error in the calibration budget. Referring back to Sec. 2.4, this is actually a difficult feat in the upper submillimeter and terahertz frequency bands [46]. For observations from space, the lack of a dominating atmosphere allows a two load calibration to correct for optical standing waves in the obtained spectra (Ossenkopf et al. [44]).

To characterize the two-load calibration stability of HIFI [38], differential internal hot (80 K) and cold (4 K) load measurements for all mixer bands, at a variety of LO frequencies have been obtained. In Fig. 2.7 we show the differential internal load stability of HEB mixer band 7 at an LO frequency of 1.890 THz. The result gives a sense for the secondary loop calibration timing. 15 minutes between load calibration appears to be a reasonable value.

2.3.6.2 One-load calibration

The one-load antenna temperature can be derived from Eq. 2.38 by substitution of 

\[ V_{load} \rightarrow V_{load1}, \quad V_{sky} \rightarrow V_{load2}, \quad T_{load} \rightarrow T_{load1}, \quad T_{sky} \rightarrow T_{load2}. \]

Defining the atmospheric transmission \( t_{atm} \) as 

\[ e^{-A \tau_{atm}} \]

with \( T_{sky} \) from Eq. 2.42 and \( G_s/G_i = 1 \), yields

\[ T_A^*(1\text{load}) = \frac{V_{source} - V_{Sky}}{\eta_s t_{atm}(V_{load} - V_{sky})} \left( (J_{load}^s - \eta_s J_{sky}^s - (1 - \eta_s)J_{spill}^s) + (J_{load}^i - \eta_s J_{sky}^i - (1 - \eta_s)J_{spill}^i) \right). \tag{2.39} \]

Because only one load is used, the sky emission \( J_{sky}^s \) must be determined. This is often difficult as the exact brightness temperature of the sky is not known with great precision. We further assume in Eq. 2.39 an equal sideband ratio (true DSB receiver) with 100% coupling efficiency to the (oversized) warm blackbody calibration load, and ignore the cosmic background radiation temperature (2.725 K). If we further make the simplifying assumption that \( T_{spill}, T_{sky}, T_{load} \) have more or less the same physical temperature (±5% error), and that \( \tau_{atm}^s \sim \tau_{atm}^i \) then Eq. 2.39 simplifies to

\[ T_A^*(1\text{load}') = 2 \frac{(V_{source} - V_{sky})}{(V_{load} - V_{sky})}. \tag{2.40} \]

Eq. 2.40 is independent of the atmospheric opacity. Of course this solution is a first order approximation and is liable to lead to significant calibration errors, especially above 200 GHz (Fig. 2.8).

From the discussion it is clear that both systematic errors (sideband ratio, load calibration parameters...) and statistical errors (atmospheric opacity, temperature of the atmosphere, integration times on the astronomical source, the loads and the off-source calibration) will be present in the calibration error budget. Calibration uncertainties below ~5%, especially in the upper submillimeter and terahertz frequency bands, will therefore be very difficult to achieve in practice, even with the two-load calibration system. Of all the possible sources of error, atmospheric opacity in the line of site of the observation is the most significant cause of error for ground
based observations. For space based observations the dominant cause of error will likely be the sideband ratio uncertainty.

2.4 Atmospheric transmission

The Earth atmosphere has a large influence on submillimeter and far-infrared observations. For example, due to pressure broadened rotational lines such as H$_2$O, O$_2$ (the main culprits) and some minor constituents, observations above $\sim 230$ GHz suffer from significant atmospheric absorption.

In order to illustrate the atmospheric influence and instrumental system degradation we model the atmosphere above three astronomically interesting sites: Mauna Kea (HI), Chajnantor (Chile), and Sofia. Our calculations use the radiative transfer model “ATM” of Pardo et al. [47] which, aside from the above mentioned molecules, also includes the dry air continuum. What is not included in the model are ground (or surface) layer temperature and humidity changes. These typically occur during the daytime by solar heating of the soil. Since most low opacity observations are presumably taken at night, we are allowed to ignore ground effects.

As input to the model we provide the precipitable water vapor (pwv) column height, atmospheric pressure, and temperature. The radiative transfer model output provides the atmospheric opacity, or optical depth, in neperes ($\tau_{atm}$) for a number of molecules. The sum of these can be linked to the atmospheric transmission in zenith via $\eta_{atm} = e^{-\tau_{atm}}$. We shown in Figs. 2.8 & 2.9 the atmospheric transmission in zenith as a function of precipitable water vapor for three high elevation sites. In general the single sideband system noise temperature, assuming unity sideband ratio, referred to the top of the atmosphere is obtained from

$$T_{SSBsys} = 2 \frac{T_{DSBrec} + T_{atm}(1 - \eta_{s}e^{-A\tau_{atm}})}{\eta_{s}e^{-A\tau_{atm}}}.$$  \hspace{1cm} (2.41)

The sky brightness temperature is thus seen to be

$$T_{sky} = T_{atm}(1 - \eta_{s}e^{-A\tau_{atm}}),$$  \hspace{1cm} (2.42)

with the source brightness, as seen from the Earth

$$T_{b} = T_{s}\eta_{s}e^{-A\tau_{atm}} + T_{atm}(1 - \eta_{s}e^{-A\tau_{atm}}).$$  \hspace{1cm} (2.43)

In Eq. 2.43, $T_{s}$ is the source brightness temperature referred to the top of the atmosphere, and $A$ the airmass (see also Sec. 8.4.3) as measured from zenith. From the above it is evident that both $\eta_{atm}$ and $T_{SSBsys}$ are a strong function of the precipitable water vapor column directly above the site. Thus knowing the exact column height at the time of the observation is very important since it effects the calibration accuracy of the obtained astronomical data. Though being the subject of the next Sec. 2.3.6, it is important to note that in the submillimeter wavelength regime the sky is usually assumed isothermal and uniform throughout. This allows the atmospheric opacity to be derived from sky dip measurements or by scaling 225 GHz radiometer tipping data [48]. In the higher submillimeter and terahertz frequency regimes (Fig. 2.9)
Figure 2.8: Top) Submillimeter (200 – 1000 GHz) atmospheric transmission on Mauna Kea (HI) for 25, 50, & 80 observing percentiles. Elevation is 4.2 km, ground temperature 268.15 K, and atmospheric pressure 623 mbar. Middle) Atmospheric transmission on Chajnantor (Chile) for 25, 50, & 80 percentiles. Elevation is 5 km, ground temperature is 270 K, and the atmospheric pressure 560 mbar. Bottom) Submillimeter atmospheric transmission on the Stratospheric Observatory for Infrared Astronomy (SOFIA) at an altitude of 13.1 km and 7 µm of precipitable water vapor. The atmospheric transmission calculations are based on a model (ATM) by J. Pardo et al. [47].
Figure 2.9: Top) Terahertz (1 – 2 THz) atmospheric transmission on Mauna Kea (HI) for 10 & 25 observing percentiles. Clearly, the atmosphere prohibits observations in this frequency range. Middle) Terahertz atmospheric transmission on Chajnantor (Chile) for 10 & 50 percentiles. The 10% is averaged over the entire year with the southern hemisphere winters being considerably better (15%) than the summer time (7%) [48, 49]. Bottom) The terahertz atmosphere from SOFIA for a 7µm column of precipitable water vapor. Above the moist troposphere the atmosphere is essentially dry, and it is in this regime that observations from SOFIA will be most competitive. The atmospheric transmission calculations are based on a model (ATM) by J. Pardo et al. [47].
the notion of a stable isobaric atmosphere is not correct. Here real time “FTS”
measurements fit to an atmospheric model have proven very successful in obtaining
the actual opacity at the frequency of the observation [46]. Of course for space based
observations with $\eta_{\text{atm}} = 100\%$ this discussion does not apply.

2.5 Summary

In this Chapter we took a look at a variety of important submillimeter detection
and instrumental requirements. From the discussion it is clear that the design and
implementation of far-infrared instrumentation demands a careful case-by-case sci-
cific motivation and outlook. Thus it is the scientific case that motivates a particular
instrument design, and not vice versa.
Bibliography


[7] [Online]. Available: Herschel; http://sci.esa.int/


[34] [Online]. Available: http://www.submm.caltech.edu/cso/receivers/lna.html/


[40] Ansoft Corporation Four Station Square, Suite 200, Pittsburgh, PA 15219-1119, USA


