Chapter 8: Submillimeter and Radio Astronomy

8.1. Introduction

The submillimeter and millimeter-wave regime – roughly $\lambda = 0.2$ mm to 3 mm - represents a transition between infrared and radio methods. Because of the infinitesimal energy associated with a photon, photo-detectors are no longer effective and we must turn to the alternative two types described in Section 1.4.3. Thermal detectors – bolometers - are useful at low spectral resolution. For high-resolution spectroscopy (and interferometers), coherent detectors are used. Coherent detectors – heterodyne receivers – dominate the radio regime ($\lambda > 3$ mm) both for low and high spectral resolution (Wilson et al. 2009).

As the wavelengths get longer, the requirements for optics also change. The designs of the components surrounding bolometers and submm- and mm-wave mixers must take account of the wave nature of the energy to optimize the absorption efficiency. “Pseudo-optics” are employed, combining standard lenses and mirrors with components that concentrate energy without necessarily bringing it to a traditional focus. In the radio region, energy can be conveyed long distances in waveguides, hollow conductors designed to carry signals through resonant reflection from their walls. At higher frequencies, strip lines or microstrips can be designed to have some of the characteristics of waveguides; they consist of circuit traces on insulators and between or over ground planes. These approaches typify the use of non-optical techniques to transport and concentrate the photon stream energy.

We first discuss the two general detector types and then derive the general observational regimes in the submm/mm range where each is predominant. We follow with the radio regime, where heterodyne techniques rule virtually completely. For bolometers the basic instrumentation – imagers, polarimeters, spectrometers - usually follows the design principles in Chapters 4 – 6. We

Figure 8.1. Thermal model of a bolometer. Important aspects of the design are the strength of the thermal link, $G$, the heat capacity of the detector, $C$, and the heat sink temperature, $T_0$. 


1
will concentrate in this chapter on describing bolometer operation. Because heterodyne receivers encode the spectrum of a source in their outputs, however, instrumentation for them takes fundamentally different forms, which will be discussed below.

8.2. Bolometers
8.2.1 Principles of Operation
Bolometers absorb the energy of a photon stream, but rather than freeing charge carriers, this energy is thermalized. The absorber is connected to a heat sink through a poorly conducting thermal link (Figure 8.1), so the deposited energy raises its temperature. This change is sensed by a thermometer mounted on the absorber, which produces an electronic signal. The signal is proportional to the total absorbed energy, not to the number of photons absorbed. Because of these differences in operation, the performance of bolometers needs to be described somewhat differently from that of the photo-detectors that have held center stage up to now.

The smallest photon stream power a bolometer can detect is the basic measure of its performance. This measure is characterized by the noise equivalent power, NEP, defined as follows. If the output of the bolometer is taken to electronics filtered to have an electronic bandwidth of 1Hz, then the NEP is the power onto the detector that produces a signal that is equal to the root mean square noise into the same electronics. A number of noise components are contained within the NEP. A fundamental noise source is associated with the storage of energy as a result of the heat capacity of the detector and the fluctuations of entropy across the thermal link; by the laws of thermodynamics, this energy is released (and restored) at a fluctuating rate (van der Ziel 1976, Mather 1982), yielding:

\[
NEP_T = \frac{(4kT^2G)^{1/2}}{\eta} \quad (8.1)
\]

where \( T = T_0 + T_1 \) in Figure 8.1) is the temperature of operation, \( G \) is the conductivity of the thermal link, in watts per degree (K), and \( \eta \) is defined below. In addition, the signal noise still has a component proportional to the square root of the number of absorbed photons:

\[
NEP_{p\phi} = \frac{hc}{\lambda} \left( \frac{2\phi}{\eta} \right)^{1/2} \quad (8.2)
\]
where $\varphi$ is the photons $s^{-1}$ and $\eta$ is the fraction of these photons that are absorbed (the quantum efficiency). The noise also includes Johnson noise, the thermal noise current for a resistor:

$$\langle I_f^2 \rangle = \frac{4kT \, df}{R} \quad (8.3)$$

where $df$ indicates the frequency bandwidth. The corresponding contribution to the NEP depends on construction details of the device (Rieke 2003).

The fundamental speed of response of a bolometer is set by the ratio of its heat capacity, $C$, to the strength of the thermal link, which determines the thermal time constant $\tau_T$:

$$\tau_T = \frac{C}{G} \quad (8.4)$$

Making $G$ small increases the temperature excursion for a given power and tends to reduce the NEP as in equation (8.1) (which is good). It also makes the detector slower as in equation (8.4) (which is bad); high heat capacity also slows the detector. Even though the realized speed of the detector can be increased through feedback from the bias circuit (discussed in Section 8.2.2), high performance with adequate time response demands low heat capacity. The specific heats of the crystalline materials in a bolometer decrease as $T^3$; the metallic component specific heats decrease as $T$. Bolometers are therefore operated at very low temperature to reduce thermal and Johnson noise and heat capacities. These temperatures also enable use of superconducting readout electronics.

A bolometer operates best when the combination of heat dissipated in the thermometer plus the heat from the infrared background raises its temperature to about 1.5 times that of its heat sink. For linear response, the power dissipated in the thermometer must exceed the infrared power. These constraints set optimum values to $G$ and the operating temperature even when the time response is not an issue. Empirically, it is found that the achievable $NEP$ scales approximately as $T^{2.5}$ to $T^{2.9}$ (Rieke 2003). To achieve photon-noise-limited performance requires temperatures of $\sim 0.3K$ on the ground and $\sim 0.1K$ when using cold optics in space (or moderately high spectral resolution on the ground).
Very high performance bolometers have been built into small arrays for some time, but until recently these devices were based on parallel operation of single pixels. The obstacle to true array-type construction was that the very small signals required use of junction field effect transistor (JFET) amplifiers that needed to operate above about 50K, far above the operating temperature of 0.3K or below for the bolometers themselves. It is difficult to implement the simple integration of detector and amplifier that is the heart of array construction with this temperature difference.

8.2.2. The PACS Bolometer Array

With the development of adequately low-noise readouts that can operate near the bolometer temperature, high-performance bolometer arrays for the far infrared and submillimeter spectral ranges are now available. To illustrate their design, we describe two of them specifically. One channel of the Herschel/PACS instrument uses a 2048 pixel array of bolometers (Billot et al. 2006). The architecture of this array is vaguely similar to the direct hybrid arrays for the near- and mid-infrared. One silicon wafer is patterned with

Figure 8.2. A single pixel in the Herschel/PACS bolometer array, pixel size about 750μm.
bolometers, each in the form of a silicon mesh, as shown in Figure 8.2. A second silicon wafer is used to fabricate the MOSFET-based readouts, and the two are joined by indium bump bonding.

The development of "silicon micromachining" has enabled substantial advances in bolometer construction generally and is central to making large-scale arrays. The delicate construction of the PACS detectors, as shown in Figure 8.2, depends on the ability to etch exquisitely complex miniature structures in silicon. In this instance, the silicon mechanical structure around the mesh region provides the heat sink; the mesh is isolated from it with thin and long silicon rods. The rods and mesh both need to be designed to achieve appropriate response and time constant characteristics. The mesh is blackened with a thin layer of titanium nitride. Quarter-wave resonant structures can tune the absorption to higher values over limited spectral bands. For each bolometer, a silicon-based thermometer doped by ion implantation to have appropriate temperature-sensitive resistance lies at the center of the mesh. Large resistance values are used so the fundamental noise is large enough to utilize MOSFET readout amplifiers. When far infrared photons impinge on the array, they are absorbed by the grids and raise the temperatures of the thermometers. The resulting resistance changes are sensed by the readouts, amplified, and conveyed to the external electronics. To minimize thermal noise and optimize the material properties, the bolometer array is operated at 0.3K. Further details are in Billot et al. (2006).

8.2.3 The SCUBA-2 Bolometer Array
Another approach is taken in transition edge sensor (TES) arrays such as the ones used in the submillimeter camera SCUBA-2 (Holland et al. 2006; Craig et al. 2010). The name of these devices is derived from their thermometers, which are based on thin superconducting films. In superconductivity, pairs of electrons with opposite spin form weakly (order of a few meV) bound “Cooper pairs.” The pairs are not subject to the Pauli Exclusion Principle for Fermions and can move through the material without significant interaction; hence, the electrical resistance vanishes. In the transition region between normal and superconductivity, the films have a stable but very steep dependence of resistance on temperature. A TES is held within this transition region to provide an extremely sensitive thermometer.
The resistance of a TES is so low that it cannot deliver significant power to JFETs and MOSFETs. Instead, the signals are fed into superconducting quantum interference devices (SQUIDs). A SQUID (Figure 8.3) consists of an input coil that is inductively coupled to a superconducting current loop. Two Josephson junctions - junctions of superconductors with an intervening insulator - interrupt the loop. Because of quantum mechanical interference effects, the current in the loop is very strongly affected by the magnetic field produced by the coil. Thus, changes in the bolometer current produce a large modulation of the SQUID current - i.e., when its output is made linear by using feedback, the device works as an amplifier. SQUIDs are the basis for a growing family of electronic devices that operate by superconductivity.

TES bolometer arrays use SQUIDs for the same readout functions that we have discussed for photo-detector arrays. The operation of two units of a simple SQUID time-domain multiplexer is illustrated in Figure 8.3 (Benford et al. 2000). The biases across the SQUIDs are controlled by the address lines. Each SQUID can be switched from a normal operational state to a superconducting one if it is biased to carry about 100μA. The address lines
are set so all the SQUIDs in series are superconducting except one, and then only that one contributes to the output voltage. By a suitable sequence of bias settings, each SQUID amplifier can be read out in turn.

When the TES temperature rises due to power from absorbed photons, its resistance rises, the bias current drops, and the electrical power dissipation decreases. These changes partially cancel the effects of the absorbed power and limit the net thermal excursion. This behavior is called electrothermal feedback. It can make the bolometers operate tens or even hundreds of times faster than implied by Equation (8.4), because it reduces the physical temperature change in the detector. In fact, if the TES is too fast, the bolometer/SQUID circuit can be unstable and measures must be taken to slow the response.

An important feature of these devices is that the superconducting readouts operate with very low power dissipation and at the ultra-low temperature required for the bolometers. Therefore, integration of detectors and readouts is simplified and the architecture can potentially be scaled to very large arrays. The SCUBA-2 bolometer arrays are an example. Each of these arrays is made of four sub-arrays, each with 1280 transition-edge sensors. The design is illustrated in Figure 8.4. The detector elements are separated from their heat sinks by a deep etched trench that is bridged by only a thin silicon nitride

![image](image.png)

Figure 8.4. Design features of the SCUBA-2 bolometer array, pixel size about 1.1mm.
membrane. The absorbing surface is blackened by implanting it with phosphorus. The dimensions of the array pixels are adjusted to form a resonant cavity at the wavelength of operation, to enhance the absorption efficiency. The superconducting electronics that read out the bolometers are fabricated on separate wafers. The two components are assembled into an array using indium bump bonding.

There are two basic approaches to multiplexing TES signals. We have described the time-domain approach, but multiplexing in the frequency domain is also possible. In this case, each TES is biased with a sinusoidally varying voltage and the signals from a number of TESs are encoded in amplitude-modulated carrier signals by summing them. They are read out by a single SQUID and then brought to room-temperature electronics that recovers each of the signals by synchronous detection.

8.3. Heterodyne Receivers for the Submm and MM-Wave
8.3.1 Heterodyne Principles of Operation
Heterodyne receivers mix the electric fields of the target source photons with those of a local source (the local oscillator (LO)) operating at a specific frequency. For simplification, we assume the target source also emits at a specific but slightly different frequency. If two such signals are multiplied together within an electronic mixer, they beat against each other due to alternating constructive and destructive interference. The resulting signal contains frequencies not only from the original two signals, but also at the difference or intermediate frequency (IF). The IF signal is isolated and amplified to provide power for convenient processing of the signal. Because the IF signal retains phase information about the target source photons, heterodyne receivers are termed coherent detectors in contrast to the incoherent detectors discussed above and in Chapter 3.

8.3.2 The Antenna Theorem
Achieving interference of the source and LO fields imposes requirements on the detection process that are summarized in the antenna theorem. Entendue must be preserved through the telescope, i.e., the $A\Omega$ product cannot decrease. Therefore, 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 (see Figure 1.3). Because of their range of tilts, the incoming wavefronts have a range of phases over the extent of the mixer. The requirement that interference occurs between the LO (at a single phase and frequency) and the signal photons sets a requirement on
the useful range of acceptance angle for the heterodyne receiver: \(2 \theta \leq \lambda/d\), where \(d\) is the diameter of the mixer. Since the optical system must conserve \(A\Omega\), this condition can be expressed as

\[
2 \Phi \approx \frac{\lambda}{D} \quad (8.5)
\]

where \(D\) is the diameter of the telescope aperture and \(\Phi\) is the angular radius of the field of view on the sky (i.e., \(2\Phi\) is the angular diameter of the FOV); this result is identical to our derivation of the diffraction limit (equation 1.12). Thus, a coherent receiver must operate at the diffraction limit of the telescope. That is, for any lossless antenna operating with a heterodyne receiver at wavelength \(\lambda\),

\[
A\Omega = \frac{\lambda^2}{\Phi} \quad (8.6)
\]

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 LO power; i.e., only a single polarization of the source emission produces any signal. These two requirements together are termed the antenna theorem.

8.3.3 Performance Description

The performance of a heterodyne receiver is quoted in terms of the noise temperature, \(T_N\), defined as the temperature of a black body placed over the receiver input that would be detected at signal-to-noise of 1.\(^1\)

FOOTNOTE 1:The figures of merit used previously do not work: 1.) counting statistics are not applicable because individual photons are not detected; and 2.) the definition of NEP breaks down for a detector that imposes a spectral bandwidth.

Coherent receivers are subject to the noise associated with the background emission, just as with all other types of detectors, termed the thermal limit. However, they have an additional source of noise, because retaining phase information is equivalent to measuring accurately the time of arrival of a photon. By the Heisenberg Uncertainty Principle, there is an unavoidable minimum noise in the measurement of both the energy and time of arrival of the photon:

\[
\Delta E \Delta t \geq \frac{\hbar}{4\pi} \quad (8.7)
\]

From this expression, one can derive the quantum limit of a receiver (e.g., Wilson, Rohlfs, and Hüttemeister 2009):
Submm and Radio

\[ T_N = \frac{h\nu}{k} \quad (8.8) \]

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

\[ T_S = \frac{A_e S_\nu}{2k} \quad (8.9) \]

where \( A_e \) is the effective area of the antenna or telescope, \( S_\nu \) is the flux density from the source, and the factor of \( \frac{1}{2} \) is a consequence of the sensitivity to a single polarization. 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 (8.9) 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:

\[ \left( \frac{S}{N} \right)_c = K \frac{T_S}{T_N} (\Delta f_{IF} \Delta f)^{1/2} \quad (8.10) \]

where \( \Delta t \) is the integration time of the observation, \( \Delta f_{IF} \) is the IF bandwidth, and \( K \) is a constant of order one. Of course, life is not quite as simple as this equation implies; the signal to noise can be degraded relative to the prediction by instability in the receiver or the atmosphere, or by confusion noise (Section 1.5.4).

Further discussion of receiver performance characterization can be found in Section 8.5.1.

**8.3.4 Comparison of Incoherent and Coherent Detection**

Equations (8.8), (8.9), and (8.10) give us the means to compare the performance of coherent (heterodyne) and incoherent (e.g., bolometer) detection, as long as we also keep in mind the antenna theorem and related restrictions. From equation (8.8) and the definition of NEP, the signal-to-noise ratio with an incoherent detector system operating at the diffraction limit is
Therefore, using equation (8.10), we obtain the ratios of signal to noise achievable with the two types of system under the same measurement conditions:

\[
\frac{(S/N)_c}{(S/N)_i} = \frac{2kT_B \Delta v (\Delta t)^{1/2}}{NEP} \quad (8.11)
\]

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, \( A = \lambda^2 \) 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 given in equation (8.2). The photon incidence rate, \( \varphi \), can be shown to be

\[
\varphi = \frac{2\pi k T_B \Delta \nu}{h \nu} \quad (8.13)
\]

where \( T_B \) is the equivalent blackbody temperature of the background. If we assume the bolometer is operated at 25% spectral bandwidth, \( \Delta \nu = 0.25 \nu \), and that the IF bandwidth for the heterodyne receiver is \( 3 \times 10^9 \) Hz (a typical value) then

\[
\frac{(S/N)_c}{(S/N)_i} \approx \frac{2.4 \times 10^6}{\nu} \left( 4 \Delta f_{IF} \right)^{1/2} \approx \frac{2.6 \times 10^{11}}{\nu} \quad (8.14)
\]

Thus, the bolometer becomes more sensitive near \( 2.6 \times 10^{11} \) Hz and at higher frequencies, or at wavelengths shorter than about 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 at wavelengths out to 2 to 3mm. Hence, large-scale bolometer cameras have been developed for mm-wave and submm telescopes. Conversely, at wavelengths longer than a couple of millimeters, equation (8.14) shows why coherent detectors are the universal choice. Of course, coherent detectors are preferred for high resolution spectroscopy and for interferometry (as we will discuss later).

### 8.3.5 Basic receiver elements

Figure 8.5 shows the components of a millimeter or submillimeter heterodyne receiver. The signal photons (from the telescope) are combined at the beam splitter (sometimes called a diplexer) with the LO signal, and then conveyed to the mixer, which downconverts the combined signal to the intermediate frequency. This signal is then amplified and sent (possibly via a spectrometer stage) to a detector stage where it is converted to a slowly varying direct current.
Submm and Radio

We now describe the individual components of this receiver.

Local Oscillators start with a lower frequency tunable oscillator and put its output through a highly non-linear circuit element (e.g., a diode). The resulting waveform has substantial power in frequency overtones, which can be isolated and made the input to an amplifier, then taken to another non-linear device, from which the frequency overtones can again be isolated and amplified. The result is that the original oscillator frequency is multiplied up to the operating frequency of the receiver. As we will see, LO power is a critical asset for a receiver, and the power that can be delivered through such a multiplier chain is limited.

Mixers: In the submillimeter-, millimeter-wave, and radio regions the mixer is a diode or other nonlinear electrical circuit component. Figure 8.6 shows the current conducted by some hypothetical mixers as a function of the voltage placed across them. If the mixer has a linear $I-V$ curve, then the conversion efficiency is zero (Figure 8.6a) because the negative and positive excursions of the voltage produce symmetric negative and positive current excursions that cancel. 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, although conversion can occur for operation away from zero (e.g., A in Figure 8.6b). If $I \propto V^2$ (Figure 8.6c), the output current is proportional to the

![I-V curves of possible mixers.](image)

Figure 8.6. I-V curves of possible mixers.
square of the voltage signal amplitude, producing a down-converted output. In addition, the signal power is proportional to its electric field strength squared. \( I \propto V^2 \propto \xi^2 \propto P \), where \( \xi \) is the strength of the electric field; hence, it is attractive to use square law devices as fundamental mixers because their output is linear with input power.

We now assume a square law device, i.e.,

\[
I = \alpha U^2 \quad (8.15)
\]

where \( I \) is the output current for an input voltage of \( U \) and \( \alpha \) is a constant, e.g., the gain. Let the mixer be illuminated by a signal and a LO both of the form

\[
E_s \sin(\omega_s t + \phi_s) + E_{LO} \sin(\omega_{LO} t + \phi_{LO}) \quad (8.16)
\]

where \( \omega \) represents angular frequencies, \( \phi \) represents phases, and \( E \) represents voltage amplitude. These two signals are added in the mixer. The resulting output current is

\[
I = \alpha \left[ E_s \sin(\omega_s t + \phi_s) + E_{LO} \sin(\omega_{LO} t + \phi_{LO}) \right]^2
\]

With use of trigonometric identities, equation (8.13) becomes

\[
I = \frac{1}{2} \alpha \left[ \frac{E_s^2 + E_{LO}^2}{2} \right] - \frac{1}{2} \alpha E_s^2 \sin(2\omega_s t + \phi_s + \frac{\pi}{2})
- \frac{1}{2} \alpha E_{LO}^2 \sin(2\omega_{LO} t + \phi_{LO} + \frac{\pi}{2})
+ \alpha E_s E_{LO} \sin((\omega_s - \omega_{LO}) t + (\phi_s - \phi_{LO} + \frac{\pi}{2})
- \alpha E_s E_{LO} \sin((\omega_s + \omega_{LO}) t + (\phi_s + \phi_{LO} + \frac{\pi}{2})
\]

The first term is a constant that can be removed with an electronic filter. The second and third terms are the second harmonics of the signal and LO, respectively. The fourth term, oscillating at the IF frequency \( \omega_s - \omega_{LO} \), is the IF current, with behavior illustrated in Figure 8.7. The fifth term is at the sum frequency and is also easily removed.

The IF current has a mean-square-amplitude proportional to the product of the signal and LO power

\[
\langle I_{IF}^2 \rangle = 2I_i I_s \quad (8.19)
\]

where \( I_i \) is the current in the detector from the LO signal and \( I_s \) is that from the source. Because the signal strength depends on the LO power, many forms of noise can be overcome by increasing the output of the local oscillator, putting a premium on delivering high LO power to the mixer. 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. The ability to provide an increase in power while downconverting the input signal frequency (conversion gain > 1) is characteristic of quantum mixers. Classical mixers do the downconversion with conversion gain < 1 - 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. In fact, to achieve better stability, quantum mixers are usually operated with gains less than one also.

The signal encodes the spectrum of the incoming signal over a range of input frequencies equivalent to the bandwidth of the IF. 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. However, in the simplest case there is no way of telling in the mixed signal whether $\omega_S > \omega_{LO}$ or $\omega_{LO} > \omega_S$. Because the initial information regarding the relative values of $\omega_S$ and $\omega_{LO}$ is lost, many of the derivations of receiver performance assume that the input signal contains two

Figure 8.7. Multiplicative mixing of two signals (S1 light solid and S2 light dashed) of slightly different frequencies produces a signal (heavy line) oscillating at the difference frequency.
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 can result in ambiguities in interpreting the data. Therefore, more complex approaches have been developed to separate the sidebands (see Section 8.3.2).

**SIS Mixers:** In the millimeter-wave regime, the highest-performance mixers are based on SIS devices. SIS stands for "superconductor-insulator-superconductor. A sandwich of these materials produces the "diode or other nonlinear device" described previously as the heart of a mixer. Its operation is illustrated in Figure 8.8.

The two superconductors (S) in Figure 8.8 are shown in a pseudo-bandgap diagram where the bandgap is the binding energy of the Cooper pairs. Unlike semiconductors, the number of available states is huge at the top of the "valence" and bottom of the "conduction" band because the pairs are not subject to the Pauli Exclusion Principle. The superconductor layers are arranged on either side of a thin insulating layer (I). Panel (a) shows the device without a bias voltage; even in this state, Cooper pairs can flow from one side to the other (the Josephson effect) because the two superconductors share the same energy states. However, only a small current flows until a bias is applied that is large enough to align the "valence" band on the left with the "conduction" one on the right (panel (b)). At that point, the tunneling of Cooper pairs through the insulator suddenly increases dramatically because of the large number of available states at the bottom of the conduction band and the large number of filled states at the top of the valence one on the opposite side of the device (panel (c)). The sudden onset of this current results in a sharp inflection in the $I$-$V$ curve; as the bias is increased further the

![Figure 8.8. Operation of a SIS junction.](image)
device behaves resistively \( (I \propto V) \). When the SIS junction is biased to put the operating point just at this inflection, the operation can be visualized as that of a switch; whenever the mixed signal (Figure 8.7) exceeds a threshold value, a significant current is conducted. Thus, the current is dominated by a signal just at the IF. Because the inflection in the bias curve is so sharp at \( 2\Delta/q \), relatively little local oscillator power is needed to get a strong IF signal.

Up to about \( 4 \times 10^{11} \) Hz, SIS mixers can operate close to the quantum limit. However, the mixer performance degrades dramatically if used with photons capable of breaking the Cooper pairs. Operation up to \( 1.2 \times 10^{12} \) Hz is possible with the relatively high gap energy (and hence large pair binding energy) of NbTiN, and SIS mixers remain within a factor of 2 to 3 of the quantum limit up to \( 10^{12} \) Hz. At higher frequencies alternative types of mixer can be used, with the best performance achieved with very small hot electron bolometers coupled to the signal and local oscillator energy through small antennae. Current performance levels are a factor of four or more above the quantum limit at these very high frequencies.

**Sideband Separation:** We emphasized that a simple mixer produces indistinguishable outputs from two ranges of input frequencies (two sidebands). More complex arrangements can separate the sidebands. An example is to use two mixers driven either by LO signals that have a 90° phase difference (see Figure 8.9), or to use LO signals in phase but to delay the input to one of the mixers by 90°. If the upper and lower sideband signals are exactly in phase at the input for one of the mixers, they will be exactly at opposite phase for the other. Circuitry can combine these two signals so either the upper or lower sideband is canceled; generally, there are two outputs, one

![Figure 8.9](image-url)
with the lower and the other with the upper sideband. Although in principle this added complexity might increase the noise temperature of the receiver, in many cases state-of-the-art receivers are limited by atmospheric noise, so the overall performance is not degraded by separating the sidebands. High performance radio telescopes may also use a pair of receivers to capture both polarizations.

**Amplifiers and Back Ends:** The properties of the remaining components in Figure 8.5 have a lot in common with similar components in radio receivers, so we will discuss them in Section 8.4.2.

### 8.4. Radio Astronomy

#### 8.4.1 The Antenna

The basic principles of operation for super heterodyne radio receivers are similar to those for submm receivers, but the longer wavelengths permit a more direct handling of the electric fields. One example is that it is generally no longer necessary to use optical methods (lenses, complex imaging optical trains) to handle the photons after they have been collected. Instead, their electric fields excite currents in antennae and the signals can be amplified and conveyed large distances through waveguides and related devices.

A half wave dipole (Figure 8.10) is the simplest example of an antenna. It is made of two conducting strips each ¼ wavelength long, with a small gap between them. The electric field of a photon excites a current in the antenna wires. The outputs of these wires are brought through a shielded cable (indicated schematically in Figure 8.10) to a resistor: in principle, the voltage due to the photon field can be sensed as a voltage across the resistor.

We can visualize the performance of such an antenna in terms of the pattern of the radiation it would *emit* if excited at an appropriate frequency. This reversal of the point of view is based on the Reciprocity Theorem. If we excite the antenna with a varying signal, the electric charges in it are

![Figure 8.10. A half-wave dipole antenna](image)
accelerated and according to Maxwell’s equations they will emit electromagnetic radiation. Emission at wavelength $\lambda$ from one end of a half-wave dipole will cancel emission from the other end, yielding a pattern peaked in the direction perpendicular to the antenna.

We demonstrate this behavior quantitatively. The motions of electrons in a wire are sub-relativistic, so the emission pattern can be calculated using the approach introduced by Larmor (in 1897). The idea is that a particle moving at constant velocity will always have a purely radial electric field, but when the velocity changes, by continuity arguments the field must develop a tangential component (Purcell 1984). The situation is illustrated in Figure 8.11. For simplicity, we assume that the charge was decelerated uniformly from velocity $v_0$ over time $t_0$ (acceleration $a = v_0 / t_0$), a time of $T$ ago. The field from the time of this event has expanded to $r=ct$, where to maintain continuity with the radial behavior before and after the deceleration there must be a kink in the field line spanning the distance $ct_0$. This kink represents the transverse component of the field; from the geometry illustrated in Figure 8.11,

$$\frac{E_{\text{tran}}}{E_r} = \frac{a T \sin \theta}{c} \quad (8.20)$$

where $a$ is the acceleration of the particle, $T >> t_0$ is the time of the observation (after the acceleration has occurred), $E_r$ is the usual radial field at distance $r$, and $\theta$ is the angle between the acceleration and the direction toward the observer. Substituting for $E_r$ and setting $T=r/c$ yields

$$E_{\text{tran}} = \frac{qa \sin \theta}{4\pi \varepsilon_0 c^2 r} \quad (8.21)$$

The energy flux density of the radiation is given by the Poynting vector

![Figure 8.11 Generation of Larmor radiation.](image)
\[ \vec{S} = \frac{E_{\text{trans}}}{\mu_0} \hat{r} = \frac{q^2a^2 \sin^2 \theta}{16\pi^2 \varepsilon_0 r^2} \hat{r} \]

(8.22)

where \( \mu_0 \) is the magnetic constant (or permeability of free space), \( \varepsilon_0 \) is the permittivity of free space, and \( \hat{r} \) is a unit vector. The Poynting vector also gives the power radiated by the particle per unit solid angle. The field of the dipole antenna is therefore the integral of equation (8.22) over the length of the antenna, which gives a power output similar in angular behavior to the simple Larmor radiation, since the half-wave dipole confines the charge motions to much less than the wavelength:

\[ P \propto \sin^2 \theta \]

(8.23)

The resulting polar pattern of the radiation is shown in Figure 8.12, both viewed perpendicular to the antenna axis (left) and parallel to it (right).

The directivity (or maximum gain) of the antenna is measured in dBi, the dB above the signal that would be received from a perfectly isotropic radiator. Db (short for decibel) is a logarithmic unit of signal above some reference level:

\[ G_{\text{dB}} = 10 \log \left( \frac{P_{\text{out}}}{P_{\text{in}}} \right) \]

(8.24)

where \( G_{\text{db}} \) is the gain in dB, \( P_{\text{in}} \) is the input power level, and \( P_{\text{out}} \) is the output. A parallel definition applies to losses of power; for example a loss of half of the power in a signal corresponds to 3 dB. Losses are indicated with \( L = 1/G \).
Since dB are logarithmic, the gains and losses in a series of devices can be added to get their net effect.

The gain for the simple dipole is $G \approx 1.76 \text{ dBi (dB above isotropic)}$; it is only barely better than an isotropic radiator. If we use more than one dipole, the polar diagram becomes more peaked as shown for two dipoles arranged parallel to each other and $\frac{1}{2} \lambda$ apart in Figure 8.13. Viewed perpendicular to the dipoles, the pattern is identical to the single dipole. However, viewed parallel to them, and assuming that they are powered in phase, the pattern is as shown in Figure 8.13, left. Any wave launched from the right antenna toward the left one arrives there $180^\circ$ out of phase compared with the wave from the left antenna, so there is no propagation in the plane defined by the dipoles. Perpendicular to this plane, the waves from both dipoles are in phase and the emission is peaked. The radiation pattern is defined by the zone where the phases match sufficiently well to combine constructively. Adding dipoles in either direction (parallel or perpendicular to the original one) will further narrow the zone of constructive interference, that is make the system more directional in its response and increase its gain above isotropic. It is characteristic of the symmetry of dipoles that their response has two lobes one above and one below the plane of the multiple dipoles. By driving one of the dipoles in Figure 8.13 at a constant phase difference from the other, the locus

![Figure 8.13. Radiation pattern of a pair of half-wave dipoles. The dipoles are shown by the line to the left (they are projected onto each other) and by the dots at the right.](image)
of these lobes will be shifted in direction – that is the pointing of a dipole array can be controlled by controlling the phases from its elements.

By the antenna theorem, the effective area of an isolated dipole is

\[ A_e = \frac{\lambda^2}{\Omega_\lambda} \]

(8.25)

\( \Omega_\lambda \sim 4 \) for a dipole in a dense array spaced at half-wavelengths. Therefore, at long wavelengths very large collecting areas can be realized with arrays of dipole antennas. At wavelengths \( \sim 10\text{m} \), such arrays are used effectively as large-area radio telescopes (e.g., the Gauribidanur Telescope in India, LOFAR at Effelsberg, and the proposal in the US for the Long Wavelength Array (LWA)). Appropriate phase shifts on the output of each dipole can be used to point the beam and to track objects. Thus, dipole arrays have large fields on the sky within which the beam can be selected. In fact, by adding additional sets of electronics operating at different phase shifts, a number of beams can be generated simultaneously. At the low frequencies where this approach is most effective, the sky foreground is dominated by emission by the Galaxy and extremely sensitive receivers are not needed to reach the background limit; hence, low-cost commercial electronics can be used for phased dipole arrays. However, for a large collecting area at wavelengths less than a meter, equation (8.26) shows that the number of dipoles and thus the amount of electronics to process their outputs becomes very large.

Therefore, in the cm-wave (and often at longer wavelengths), a paraboloidal reflector is placed behind the antenna (now termed a feed), which causes the beam pattern to have one main lobe along the optical axis of the paraboloid. Suppressing the backward lobe intrinsic to dipoles and eliminating the multiple antenna interactions in dipole arrays substantially increases the

Figure 8.14. A feedhorn with a ground-plane vertical as the antenna. The ground plane vertical creates a full dipole antenna by reflection of one half of such an antenna.
resistance to sources of unwanted interference. Radio telescope designs of this type are described in Section 2.4.3. They may operate at prime focus, or may have a secondary mirror (called a subreflector), allowing a classic Cassegrain, an offset Cassegrain to avoid beam blockage by the subreflector, or be designed to couple efficiently to the receiver directly through a waveguide. At high frequencies (cm-wavelengths), a feed horn (Figure 8.14) is a better approach than an antenna feed because it can give better control of the angular dependence of the signal and thus helps suppress unwanted off-axis response (sidelobes). Corrugations along the wall of a feed horn increase the surface impedance and help convey the wave without losses to a dipole antenna at the exit aperture of the horn. The figure illustrates a horn that feeds a waveguide, within which the antenna is placed to optimize its absorption of energy.

Continuing with the thought exercise of using the antenna as a radiator, the response pattern of the feed over the telescope primary mirror aperture is referred to as the illumination pattern. In the design of a feed antenna or of a feed horn, a variety of illumination patterns can be created. They determine the weighting of the beam over the telescope primary mirror that is used in creating an image. To understand this behavior, we consider the field pattern of an antenna, $E(\phi)$, as discussed in the preceding chapter (equation (7.26) and surrounding discussion).

In Figure 8.16, we illustrate a number of options. Uniform illumination, (a), yields a PSF following the familiar Airy function. Illuminations with reduced

![Figure 8.15. Illumination patterns (to left) compared (to right) with the resulting distribution of electric field in the telescope beam (dashed) and point spread functions (solid).]
weight at the edge of the primary (e.g., Gaussian, (b)) reduce the sidelobes (e.g., the bright rings in the Airy pattern), but they also increase the width of the central maximum, that is reduce the resolution. Patterns with increased illumination at the primary edges (e.g.,(c)) improve the resolution in the central image at the expense of increased sidelobes. For many radio telescopes, the primary illumination is reduced at the edge of the mirror; a Gaussian illumination pattern is generally a reasonable approximation. This practice both reduces sidelobes and minimizes the response of the telescope to sources behind the primary, whose signal otherwise can be picked up at the focal plane without being reflected by the primary mirror. This unwanted signal is called spillover radiation.

Because the receiver must work at the diffraction limit, any failure to deliver images at this limit by the telescope results in lost efficiency. The resulting requirements can be determined from the Maréchal formula, equation (2.3), called the Ruze formula by radio engineers:

\[ S \approx e^{-(2\pi \sigma/\lambda)^2} \]  

where in this case S is the efficiency and \( \sigma \) is the rms wavefront error. Thus, a surface accuracy of \( \lambda/28 \) rms (wavefront error of \( \lambda/14 \)) gives an efficiency of 0.8, while a surface of \( \lambda/20 \) yields an efficiency of about 0.7. A radio telescope with its feed is subject to additional losses due to: 1.) blockage of the aperture, e.g. by a Cassegrain subreflector; 2.) spillover, e.g., energy reflected by the subreflector outside the acceptance solid angle for the receiver; 3.) the feed illumination; and 4.) other miscellaneous causes. The net efficiency is typically ~ 0.4, if the antenna surface accuracy is sufficient to make it diffraction-limited.

**8.4.2. Front and Back Ends**

The overall layout of a radio receiver is shown in Figure 8.16. Most of the components are familiar from Figure 8.5. A new addition is the coupler, which is a passive device that divides and combines radio frequency signals. The directional coupler in Figure 8.16 has three ports where signals enter and leave: line in, line out, and the tap. The signal passes between the line in and line out ports. A calibration signal can be applied to the tap port, from which it is passed to the line in port attenuated by some value. It can then emerge
from the line out port just as a signal from the feed would, providing the capability to compare the signal from the telescope to an accurately controlled electronic signal.

We now follow the signal path from the feed to the mixer. This chain is called the front end; its characteristics dominate the behavior of the system. Below about 100 GHz, improved performance can be obtained by interposing high-performance amplifiers between the output of the antenna and the mixer. The effective input noise of the amplifier is lower than the mixer noise, so the receiver noise temperature is reduced with this design. A typical front end then consists of the feed, connecting cables, perhaps filters or couplers, and then one or two low-noise amplifiers, followed by the mixer.

The amplifiers used for the low-power IF output of a mm- or submm-wave mixer, or to amplify the antenna output in the cm-wave, must have very low noise. Excellent performance is obtained with high electron mobility transistors (HEMTs) built on GaAs (and with other devices of related design). The basic transistor (a metal-semiconductor field effect transistor, MESFET) is shown in Figure 8.17. The electron flow between source and drain is regulated by the reverse bias on the gate; with an adequately large reverse bias the depletion region grows to the semi-insulating layer and pinches off the

Figure 8.16. Block diagram of a radio receiver.

Figure 8.17. A HEMT.
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 junction between two different zones of the semiconductor with different bandgaps so the electrons can flow in undoped material. The very high mobility in undoped GaAs makes for very fast response, to ~ 100 GHz.

For a radio receiver, the amplifier stages are followed by a mixer, in the form of a nonlinear circuit element such as a MESFET or diode. The mixer is described as having three ports, an input, a port for the local oscillator signal, and an output port for the intermediate frequency signal.

For all heterodyne receivers (submm through m-wave), the IF signal is brought to a detector stage, shown schematically in Figure 8.18 (a variety of alternative circuit concepts can perform the same function). This stage rectifies the signal and sends it through a low-pass filter, converting it into a slowly varying indication of the signal strength suitable for interpretation by human beings. We would like the circuit to act as a square law detector because \(<I_{IF}^2>\) is proportional to \(I_S\) (see equation (8.11)), which in turn is proportional to the power in the incoming signal.

To demonstrate how the detector stage achieves this goal, we solve the diode equation (equation 3.15) for voltage and expand in \(I/I_0\).
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.

### 8.4.3. Spectrometers

Usually it is desirable to carry out a variety of operations with the IF signal itself before smoothing it in the detector stage. For example, imagine that the down-converted IF signal is sent to a bank of narrow bandpass fixed-width electronic filters that divide the IF band into small frequency intervals (Figure 8.19). Each of these intervals maps back to a unique difference from the LO frequency, i.e., to a unique input frequency to the receiver. A typical “filter

![Filter Bank Spectrometer Diagram](image)

**Figure 8.19.** A filter bank spectrometer. The input IF signal is divided among the bandpass filters and the output of each is processed by a detector/integrator stage. The outputs of these stages are switched sequentially to the output computer where the spectrum can be displayed.
bank” may have 128, 256, or 512 channels. A detector stage can then be put at the output of each filter, so the outputs are proportional to the power at a sequence of input frequencies, that is they provide a spectrum. In this manner the total IF bandwidth (perhaps ~ 4 GHz) is divided into a spectrum, even though only a single observation with a single receiver has been made; the process is called spectral multiplexing.

Although conceptually simple, a high performance filter bank can be an engineering challenge. The filters need to have closely matched properties and be robust against drift of those properties due to effects like temperature changes. A filter bank is also inflexible in use; the resolution must be set during design and construction. Finally, these devices are complex electronically and expensive to build if many channels are required.

An acousto optical spectrometer provides a many-channel spectrometer without the electronic complexity of a filter bank, since it divides the IF signal into frequency components without a dedicated unit for each component. In this device (see Figure 8.20), a piezoelectric transducer is attached to a Bragg cell, a transparent volume containing either a crystal like lithium niobate or water. When the IF signal is fed into this transducer, it vibrates to produce ultrasonic waves that propagate through the Bragg cell and produce periodic density variations. As a result, the index of refraction in the cell also varies periodically, making it act like a volume phase diffraction grating (Section 6.3.2.1). When light from a near-infrared laser diode passes through the cell,
Submm and Radio

it is deflected accordingly – the zero order path is absorbed and the first order path is focused onto a CCD by camera optics. The light intensity is proportional to the IF power injected into the Bragg cell, while the deflection angle and hence position on the CCD is determined by the ultrasound wavelength. The output signal is basically the Fourier transform of the IF signal. AOS spectrometers are capable of resolving the IF signal into more than 2000 spectral channels.

A third method to divide the IF into a spectrum is called a chirp transform spectrometer. In this case, the LO is swept in frequency, at a rate df/dt. This process is termed the expander. As a result, an input signal at a fixed frequency is modulated at linearly swept frequencies in the IF. A delay line with a delay time depending on frequency, called the compressor, is put in the IF path with a dispersion, dτ/dt, that just counteracts the frequency modulation. The result is that the fixed input frequency is converted to a sinc function output at a specific time; a different input frequency will produce an output at different time, as illustrated in Figure 8.21. Therefore, the spectrum emerges in time series. Since the full input bandwidth is being processed at all times, there is no compromise in the sensitivity of the receiver.

Figure 8.21. Operation of a chirp transform spectrometer.
The fourth method to obtain a spectrum from the IF signal is to compute its autocorrelation, the integral of the results of multiplying the signal by itself with a sequence of equally spaced delays:

\[
R(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} U(t)U(t + \tau)dt
\]

(8.29)

The Fourier transform then gives the spectrum, according to the Wiener-Khinchin Theorem (Wilson, Rholfs, Hüttemeister 2009). To measure the autocorrelation, the IF signal is first digitized. High speed performance is required; the Nyquist theorem says that the digitization rate must be at least twice the IF bandwidth. Therefore, autocorrelators often digitize to only a small number of bits. As shown in Table 8.1, the loss of information is surprisingly modest if the gains are set optimally. Many systems only digitize to two bits, and some only to a single bit. The digitized signal is taken to an electronics circuit that imposes the necessary delays in the signal using shift registers and then combines the results to provide the autocorrelation as an output. Autocorrelators are flexible in operating parameters and very stable, since they work digitally. Their biggest disadvantage is that they can only operate over a limited bandwidth.

Because it is not possible actually to carry out the limit to infinite time in equation (8.29), autocorrelators produce spectra with “ringing” due to any sharp spectral features. Similar behavior was discussed with regard to Fourier transform spectrometers (Section 6.5). As in that case, these artifacts can be reduced by filtering the signal (e.g., with a “Hanning filter”), but with a loss in spectral resolution.

### Table 8.1. Information retained as a function of digital bits

<table>
<thead>
<tr>
<th>bits</th>
<th>information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>64%</td>
</tr>
<tr>
<td>2</td>
<td>81%</td>
</tr>
<tr>
<td>3</td>
<td>88%</td>
</tr>
<tr>
<td>infinite</td>
<td>100%</td>
</tr>
</tbody>
</table>

#### 8.4.4. Receiver Arrays

Heterodyne receivers naturally provide multiplexing in wavelength – that is, by measuring multiple frequencies or wavelengths simultaneously and
without degradation in the noise, they accelerate the measurement of spectra. In the mm-wave and at longer wavelengths, receivers now operate essentially at the quantum limit. However, because of the complexity of the supporting electronics (primarily the back end) for each receiver, most telescopes have operated with just one at a time. Thus, maps in the radio are typically made “on the fly” (OTF), meaning that the single beam is scanned systematically over the field to sample the image pixels one at a time. Substantial gains in mapping speed can be achieved with an array of receivers. Making such instruments practical depends on creating cost-effective forms of multiple back end electronics. Arrays with of order 100 receivers are in operation at a number of telescopes (e.g., DIGESTIF at Westerbork operating in the GHz range and Super-cam at the UA SMT, operating in the mm-wave).

8.5. Characterizing Heterodyne Receivers
8.5.1 Noise Temperature
Although the final astronomical results from heterodyne receivers are often quoted in familiar radiometric units (e.g., Jy), a temperature-based system of characterization is also used, as introduced in Section 8.3.3 where we defined noise temperature and antenna temperature. This approach is also employed for the components of the receiver. We have also already used a common terminology in radio electronics, in which circuits are considered virtually as black boxes without detailed consideration of their components, and they are characterized in terms of the input and outputs, called ports.

For example, a 2-port device is shown in Figure 8.22. It is considered to be ideal, i.e., lossless and noise free. It has a gain for signals, $G$,

$$G = \frac{P_{out}}{P_{in}} \quad (8.30)$$
where $P_{\text{in}}$ is the input signal power and $P_{\text{out}}$ is the output. In addition to the desired signal, noise will be injected at the input port, which we characterize as the Johnson noise of a resistor (equation 8.3) at ambient temperature $T_0$ (taken to be 290K).

The power that can be delivered by our resistor to a matched load (equal resistance) at the input to the two-port device is then

$$P_{N,\text{in}} = kT_0 \Delta v$$  \hspace{1cm} (8.31)

where $\Delta v$ is the frequency bandwidth. To reproduce the sources of noise internal to the device, we assume an imaginary input that gives just this noise, $P_N$, as the output; that is, we put on the input a resistor that produces a noise of $P_N/G$. The resistance value is set to the input impedance of the device, so we have to adjust the temperature to obtain the required noise. This temperature is called the noise temperature of the device, $T_N$:

$$T_N = \frac{P_N}{kG \Delta v}$$  \hspace{1cm} (8.32)

A complex system can be represented as a linear chain of 2-port devices, see Figure 8.23. The input (i.e. from the feed) is shown as $T_0$, and each two-port device is labeled with its noise temperature and gain (this notation includes the possibility that some of the gains are less than one – e.g. a lossy cable or an attenuator). The corresponding system temperature is

$$T = T_0 + T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1G_2} + \frac{T_4}{G_1G_2G_3} + \cdots \cdots + \frac{T_0}{G_1G_2G_3 \cdots G_{n-1}}$$  \hspace{1cm} (8.33)

The external temperature (the antenna temperature) just gets added to by all of the noise temperatures of the following devices, but each stage after the first
Submm and Radio

one is divided by the total gains of the preceding stages. To achieve minimum system temperature, low noise in the first amplifier stage is critical.

8.5.2 Source Characteristics
The brightness temperature of an astronomical source, $T_b$, is the temperature an object that subtends the same solid angle as the source would need to match its flux density. For sources that are unresolved by the beam of a radio telescope, the brightness temperature must be greater than the antenna temperature, $T_s$, which is calculated on the assumption that the source fills the beam. The beam filling factor, i.e. the fraction of the beam area filled by the source, is the ratio,

$$filling \ factor = \frac{T_s}{T_b} \quad (8.34)$$

This quantity is also called the beam dilution.

8.6. Observatories
8.6.1. Submillimeter
Observations from the ground in the submillimeter are extremely sensitive to the amount of water vapor overlying the observatory site. Figure 1.6 demonstrates that the windows at 350 and 450μm close and the one at 850μm is impaired at total water vapor levels of 2mm and above. Even observatory sites considered for other purposes to be high and dry only allow submm observations under the most favorable circumstances. For example, over Paranal, the site of the VLT and at an altitude of 2635m, for half of the nights the level is above 2mm and it reaches 0.5mm only on very rare winter nights. Therefore, the selection of sites with extremely low water vapor is critical to success for a submm observatory, requiring that the observatory be either placed at high latitude (the Antarctic has been the site of a number of successful observations) or extremely high (e.g., 5000 meters elevation for the Atacama Large Millimeter Array (ALMA)).

Wind-blown fluctuations in the water vapor content of the air over a submm telescope produce a fluctuating level of thermal emission into the beam, resulting in noise. This noise rises rapidly with decreasing frequency. For modern broad-band continuum cameras (e.g., bolometric detectors), it dominates the detector noise below about 0.5 Hz (e.g., Sayers et al. 2000). The effect largely arises close to the telescope, due to the small scale height.
for water vapor. The noise can be reduced by nutating the secondary mirror at a frequency of 1 Hz or higher, so the measurements are made between the source and a blank region next to the source in rapid sequence. For small nutation angles, the two positions sample nearly the identical path through the low-lying atmosphere while modulating the source signal completely. Fitting the common-mode fluctuations and spatial behavior over the pixels of an imaging array can also be effective in reducing the impact of the sky signal.

8.6.2. Radio
Over much of the cm- and m-wave spectrum, the atmosphere is quite transparent. At very long wavelengths, radio signals are reflected by the ionosphere of the earth; this effect increases in proportion to $\lambda^2$, making the atmosphere opaque as viewed from the outside at a frequency of about $10^7$ Hz and below (wavelength of about 30m and longer). Absorption by atmospheric oxygen makes ground based radio observations impossible between 52 and 68 GHz and in a narrow range near 118 GHz. Water vapor absorption becomes increasingly strong at frequencies above $10^{11}$ Hz and is dominant above $3 \times 10^{11}$ Hz, as discussed in the preceding section.

However, the dominant obstacle to untrammeled radio astronomy is man, in the form of radio transmission devices. When there is a source of interference, spectral multiplexing can become a problem, since a strong signal at a single frequency within the range of frequencies downconverted to the IF can overwhelm the entire output of a receiver. Often, notch filters that suppress a single frequency must be employed to circumvent such issues. Radio frequency allocations are subjects of intense negotiation because substantial

---

**Figure 8.24.** Radio frequencies allocated to astronomy between 10 MHz and 10 GHz. The astronomy “preserves” are the vertical lines (seldom of a resolved width); all the open spaces between them are used for other purposes.
amounts of money can be involved in radio communications. Figure 8.24 shows the frequency allocations for radio astronomy. Clearly, operating only within these allocations would be inadequate and another strategy has to be to place major radio telescopes at sites that are relatively removed and sheltered from man-made transmissions.

Further Reading


Additional References

Craig, S. C. et al. 2010, SPIE, 7741, 43
Holland, W. S. et al. 2006, SPIE, 6275, 45