COHERENT DETECTION AND SIS MIXERS

J. Zmuidzinas
Division of Physics, Mathematics, and Astronomy
California Institute of Technology, 320–47, Pasadena, CA 91125

ABSTRACT

Submillimeter spectroscopy is a unique and very rich source of astrophysical information. While many important results have been obtained using ground based telescopes, the Kuiper Airborne Observatory, ISO, and SWAS, this field will soon experience rapid growth and development due to projects such as SOFIA, ALMA, and the Herschel Space Observatory (HSO). High spectral resolution is and will continue to be important for many observations. This paper discusses the relative merits of coherent and direct detection for high resolution spectroscopy. In addition, the paper gives a brief review of the status of coherent detection, and superconducting tunnel junction (SIS) mixers in particular. Coherent detection can be expected to play a key role in submillimeter and far–infrared astrophysics well into the future.

INTRODUCTION

Many interesting astronomical objects, such as galaxies, molecular clouds, PDRs, star–forming regions, protostars, evolved stars, planets, and comets, have rich submillimeter spectra. These spectra contain a great deal of detailed information about these objects, including their structure, energetics, dynamics, elemental and isotopic abundances, chemistry, and physical conditions such as density, temperature, and ionization. Most of this information cannot be obtained using other techniques. Combining spectroscopy with other observations, such as maps of the dust continuum emission, can be especially powerful. Some aspects of the science enabled by submillimeter spectroscopy are described elsewhere in these proceedings.[1,2] Very often it is necessary to use high spectral resolution in order to avoid line blending, and to resolve line shapes. High resolution spectroscopy is carried out primarily using coherent receivers, even though direct detection in principle has a sensitivity advantage. This is because practical factors, such as the sensitivities of existing detectors, and constraints on the instrument volume, make it difficult for direct–detection spectrometers to compete at high resolution.

COHERENT DETECTION

Coherent detection is used primarily at long wavelengths, from the radio into the far-infrared. In comparison to direct detection, coherent detection offers several important advantages, including the ability to obtain very high spectral resolution. Figure 1 shows a block diagram of a typical coherent receiver. The submillimeter signal is first downconverted into the microwave band, where it is amplified and spectrally analyzed. In the centimeter and millimeter bands, low–noise preamplifiers are used prior to downconversion.[3] In the submillimeter band, sensitive superconducting mixers (SIS and HEB) are used.

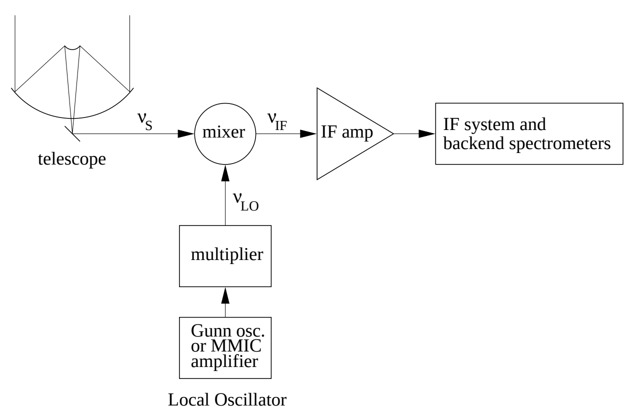

Figure 1: Block diagram of a heterodyne spectrometer. The telescope signal ν_S is combined with a local oscillator ν_{LO} in a mixer, to yield the intermediate frequency $\nu_{IF} = |\nu_S - \nu_{LO}|$.

As compared to direct detection, coherent detection has a fundamental disadvantage in sensitivity. In essence, coherent instruments amplify the electromagnetic field into the classical domain prior to detection, and preserve information about both the amplitude and phase. According to quantum mechanics, these are non–commuting quantities, and the corresponding uncertainty principle is the fundamental reason for the “quantum limit” for the sensitivity of coherent detection. The value of this quantum noise can be expressed as a noise temperature, $T_n = h\nu/k_B$, which is 0.05 K/GHz, or 50 K/THz. Equivalently, it corresponds to the photon shot noise from a background of 1 photon per second per Hertz of bandwidth. While this “quantum limit” does not play a significant role for (warm) ground–based or airborne telescopes, it would become an important issue for cold telescopes in space.

The background–limited sensitivity of a direct detector receiving a single mode (i.e. a diffraction–limited beam and a single polarization) is expressed by the following equation:

$$\sigma_P^d = \frac{h\nu}{\sqrt{\Delta\nu T}} \sqrt{\frac{n_0(1+\eta^d n_0)}{\eta^d}}\,\Delta\nu \tag{1}$$

where σ_P^d is the uncertainty in the incident power in a detection bandwidth $\Delta\nu$ after an integration time T using a detector with effective quantum efficiency η^d (including optical losses). Here n_0 is the mean photon occupation number of the thermal background radiation. If the background can be characterized by a single temperature T_{bg} and a total emissivity ϵ, we would have $n_0 = \epsilon \left[e^{h\nu/kT_{bg}} - 1\right]^{-1}$. In contrast, the sensitivity of a coherent receiver is

$$\sigma_P^c = \frac{h\nu}{\sqrt{\Delta\nu T}} \frac{1}{\eta^c}(1+\eta^c n_0)\Delta\nu \tag{2}$$

The two terms in the sum $(1 + \eta^c n_0)$ correspond to quantum and background noise, respectively. This expression assumes that an amplifier or a single–sideband mixer is being used; a factor of 2 must be inserted on the background term for a double–sideband mixer. *Thus, for low backgrounds,* $n_0 << 1$, *direct detection is more sensitive by a factor* $\sim \sqrt{n_0}$. For high backgrounds, $n_0 > 1$, both systems are limited by background fluctuations and have comparable sensitivity. The comparison between coherent and direct detection sensitivity is shown in Fig. 2.

Achieving a spectral resolution of $R = \nu/\Delta\nu$ *requires that the instrument have some method of delaying the signal by a time* RT, or a distance $R\lambda$, where $T = \lambda/c$ is the period of the wave. This is a simple result of the Fourier transform relationship between frequency and time. The basic reason that coherent detection is capable of very high spectral resolution, easily in excess of 10^6, is that the spectroscopy is actually done after downconversion by the “backend”, at radio or microwave frequencies, using amplified signals. A wide variety of devices, such as filterbanks, acousto–optic (AOS) spectrometers, analog correlators, and digital correlators have been developed, which use various tricks to reduce the volume associated with the time delay RT. In essence, digital correlators store the digitized signal into memory for retrieval at a later time. Filterbanks and analog correlators use guided–wave (transmission line) propagation and dielectric materials to drastically reduce the volume. AOSs and CTSs (chirp–transform spectrometers) rely on the slow velocity of sound propagation in solids.

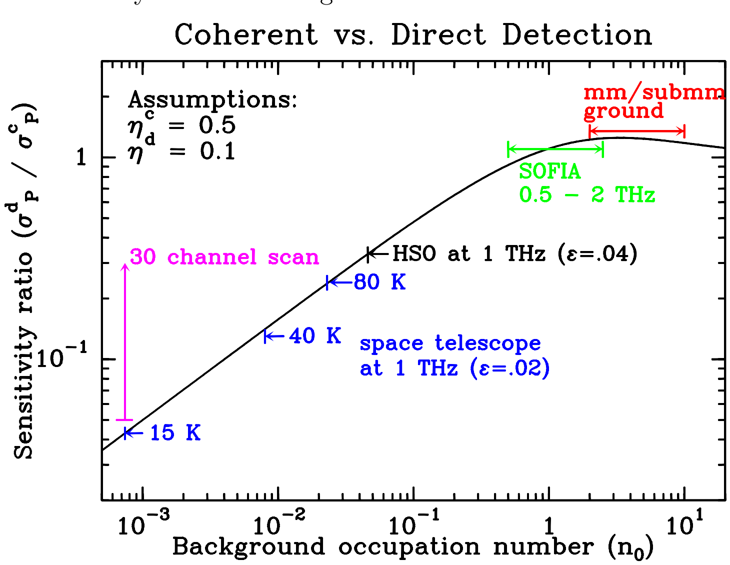

Figure 2: An idealized comparison of the relative sensitivities of coherent and direct detection. The vertical axis is the sensitivity ratio; a ratio less than unity favors direct detection. The vertical arrow shows the sensitivity penalty associated with a 30–channel sequential spectral scan.

DIRECT DETECTION SPECTROSCOPY

For direct–detection spectroscopy, *there is only one approach that gives the best sensitivity: grating spectrometers or their equivalent*. Such spectrometers use a grating to disperse the light onto an array of detectors, and each detector pixel responds to a different wavelength channel. Because the spectroscopy is done prior to detection, the spectrometer must have high efficiency (transmission). Furthermore, the spectrometer must be cold, $kT << h\nu$, to avoid additional thermal noise. These constraints severely limit the range of design approaches. In contrast, neither of these constraints applies to backends for coherent receivers.

For best sensitivity, direct–detection spectrometers must obey a simple principle: they must *extract the necessary information from every photon*. In a grating spectrometer, the absorption of a photon by a given detector pixel corresponds to a measurement of its wavelength, to within the resolution of the instrument. This is not true for other types of spectrometers. For instance, the absorption of a photon by a detector in an FTS is not equivalent to a unique measurement of its wavelength. This corresponds to a loss of information, and therefore sensitivity. Another example is a Fabry–Perot spectrometer, which violates this principle by reflecting or "throwing away" photons outside of its resolution bandwidth. The information carried by those photons is lost. Fabry–Perots must be scanned to obtain a spectrum, which is the time penalty that is paid for throwing away photons.

While grating spectrometers are very interesting for moderate–resolution submillimeter spectroscopy using cold space telescopes,[4] they cannot provide high spectral resolution. *The difficulty with grating spectrometers is that their size grows as the spectral resolution increases.* For a resolution R, the linear size must be of order $R\lambda$, according to the time–delay principle described earlier. Achieving $R = 10^6$ at $\lambda = 200\,\mu$m, which can readily be done using a heterodyne spectrometer, would require a 200 m grating. Furthermore, this grating must be cold, to avoid a sensitivity degradation. While there are no fundamental limitations, there are obviously enormous practical problems with this approach. One can solve this size problem by folding the optical path onto itself. This is exactly what is done in a Fabry–Perot, and does indeed give a large volume reduction, and can achieve resolutions approaching $R = 10^6$. The price is reduced sensitivity since photons are thrown away.

COMPARISON OF COHERENT AND DIRECT DETECTION

Rather than attempting to take into account "real–life" factors, such as detector noise, filter efficiencies, etc., which are always improving due to technological innovation, I will compare the sensitivities of nearly ideal instruments, since it is likely that over time the fundamental limits will be approached quite closely. The sensitivity comparison, shown in Fig. 2, comes from a simple calculation using Eqs. (1,2). The one arbitrary "real–life" adjustment I have made is to give the coherent instrument a better overall efficiency, than the direct–detection spectrometer, since in reality it would have a much simpler optical system.

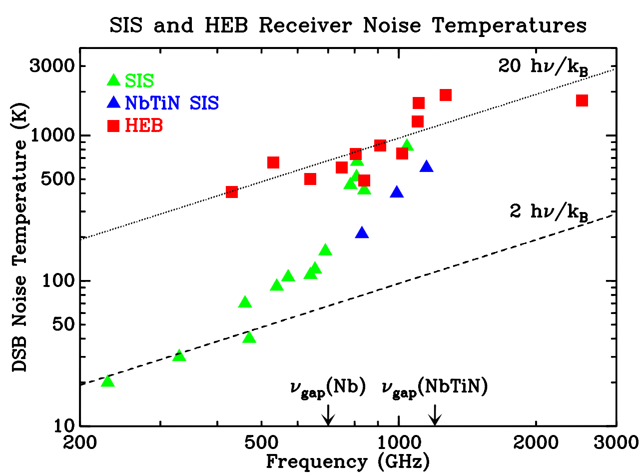

Figure 3: A selection of measured sensitivities for SIS and HEB receivers.

Fig. 2 demonstrates that direct detection is superior for low backgrounds. However, for high spectral resolution, say $R \sim 10^6$, direct detection instruments would use Fabry–Perots and would need to be scanned. For a scan consisting of M spectral channels, the sensitivity penalty would be $\sqrt{M}$, as shown by the vertical arrow on Fig. 2 for the case $M = 30$, which would give a modest amount of information about the line shape.

The overall result is that direct detection is still more sensitive, but not by an overwhelming factor. There are several other important factors to consider:

- backends can easily provide thousands of simultaneous channels
- backends can provide a wide range of spectral resolutions
- submillimeter heterodyne receivers are within a factor ~ 10 of the quantum limit (see Fig. 3)
- mixer noise temperatures degrade at higher frequencies
- tunable local oscillators are not yet available above ~ 1.5 THz
- background–limited $R \sim 10^6$ spectroscopy will require detectors that are $\sim 10^3$ times more sensitive than are now available ($10^{-21}\,\mathrm{W\,Hz^{-1/2}}$ vs. $10^{-18}\,\mathrm{W\,Hz^{-1/2}}$)
- submillimeter mixers operate at Helium temperatures; direct detectors for high–resolution spectroscopy, with NEP $\sim 10^{-21}\,\mathrm{W\,Hz^{-1/2}}$, would operate at ~ 100 mK, or lower.

At present, these factors combine to strongly favor coherent detection for high–resolution spectroscopy up to $\sim 1.5 - 2$ THz. The current lack of tunable local oscillators limits the role of coherent detection at higher frequencies. In the future, we can expect the issue of detector sensitivity to disappear (for both direct and coherent). Even in this idealized case, it appears that coherent detection will retain substantial advantages in at least some situations, such as wideband, high–resolution line surveys.

STATUS OF SUBMM RECEIVER TECHNOLOGY

Figure 3 shows the impressive improvements in receiver sensitivities that have been achieved over the last decade using superconducting mixers, with both tunnel junctions (SIS) and hot–electron bolometers (HEB). Nonetheless, there remains substantial room for improvement to reach the quantum limit, particularly at frequencies above 1 THz.

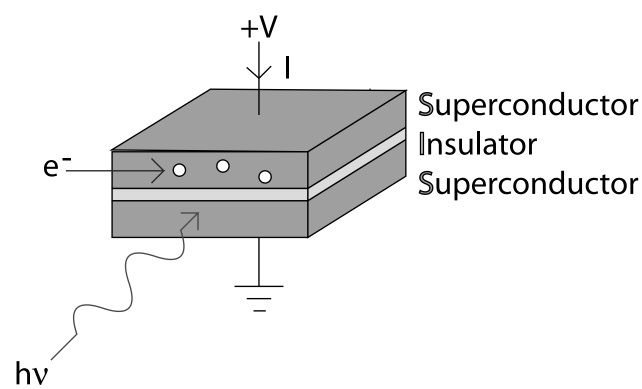

Figure 4: A superconducting (SIS) tunnel junction. Junctions are made from a Superconductor/Insulator/Superconductor trilayer, or SIS. A typical junction area is $\sim 1\,\mu\mathrm{m}^2$. Submillimeter photons are coupled to the junction via leads attached to the electrodes.

At millimeter wavelengths, SIS mixers offer the best sensitivities, closely approaching the quantum limit. The structure of an SIS junction is shown in Fig. 4, and is basically a sandwich of two superconductors with a very thin (10–20Å) insulating barrier in between. Electrons can tunnel across the barrier, provided that the energy provided by bias voltage V exceeds the superconducting gap energy: $eV > 2\Delta$. This explains the rapid turn–on of tunneling current at the gap voltage, $V_g = 2\Delta/e$, as shown in Fig. 5. Alternatively, at lower voltages, a photon can supply the missing energy, if $h\nu + eV > 2\Delta$. In this mode, the junction behaves like a photodiode, providing one electron of tunneling current per photon absorbed.

In theory, SIS mixers can approach the quantum limit, if the photons can be coupled efficiently to the tunnel junction. This requires efficient waveguide probes or planar antennas, as well as inductive tuning circuits for compensating the parallel–plate capacitance of the junction. SIS mixer development is largely focused on these areas. The second issue, providing a tuning circuit, becomes increasingly difficult at high frequencies, for two reasons: (1) the RF impedance of the junction capacitance scales inversely with frequency; and (2) the superconductors or metals used for the tuning inductor become increasingly lossy. Thus, SIS mixers become increasingly difficult to produce at higher frequencies, and their performance deteriorates with frequency. SIS mixers will not operate above 1.5–1.6 THz with current materials. Nonetheless, SIS mixers have been pushed to 1.2 THz,[5] by using highly transparent AlN tunneling barriers, and very high quality (epitaxially grown) metal films in the tuning circuit.

Future developments can be expected in several areas. First, and most importantly, the push toward quantum–limited sensitivities must be continued. For frequencies above $\sim$ 1 THz, HEB mixers are used; these are reviewed by McGrath *et al.*[6] Whether or not HEB mixers can ultimately reach the quantum limit is still an open issue; new device concepts may be required. Another area of development is to continue to expand the mixer instantaneous bandwidths. For ALMA, the goal is 8 GHz; work at Caltech is pushing toward 12 GHz.[7] A third area, important for future space missions, is to look at integrating the later stages of the local oscillator with the mixer, in order to simplify the local oscillator injection problem.

Local oscillators are another area in which dramatic improvements have been made. Electronically tunable, all solid–state local oscillators are being developed for Herschel. High–power transistor amplifiers at frequencies near 100 GHz are being used to drive diode multiplier chains to produce usable output power at frequencies as high as 1.5 THz. These developments are described in detail by Mehdi *et al.*[9] Nonetheless, there is still substantial room for improvement of local oscillators.

Also contained in these proceedings is a review of backend spectrometer technology.[10] Here, a wide variety of technologies are available, which can provide high spectral resolution, wide bandwidths, and low power operation. Any two of these characteristics can be obtained simultaneously; combining all three remains a challenge.

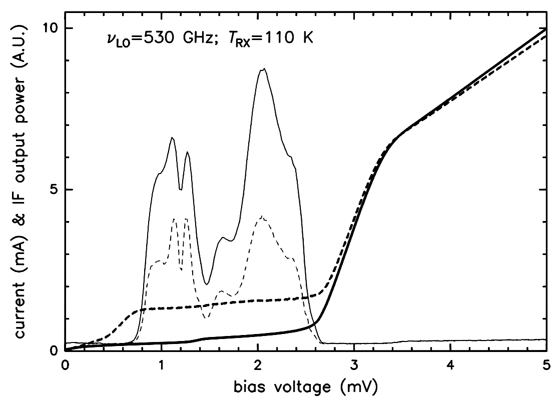

Figure 5: Measurement of a 530 GHz SIS mixer using a Nb/AlN/Nb SIS junction.[8] The unpumped current–voltage (IV) is the lowest solid line; the dashed curve just above is the pumped I–V curve. The double–peaked curves show the IF response to hot (295 K) and cold (80 K) blackbody loads. For this example, the sensitivity is a factor of $\sim$ 10 above the quantum limit.

THE FUTURE OF SUBMILLIMETER SPECTROSCOPY

Despite more than two decades of effort, submillimeter astronomical spectroscopy remains a small and underdeveloped field, mainly due to the lack of sensitive instruments and telescopes. However, the situation is changing rapidly. Increasingly sensitive and higher–frequency receivers are being deployed on ground–based single–dish telescopes. Large–scale projects such as ALMA will play a very important role in developing this field. However, large parts of the submillimeter spectrum are blocked by the atmosphere, especially above 1 THz, and can only be observed from the stratosphere (SOFIA) or from space (HIFI/HSO). SOFIA and HSO, in combination with ALMA, will completely transform this field, and one can expect that there will be rapid growth in this area along with a number of very interesting scientific results covering a broad range of astrophysical topics. In this context, it is intriguing to consider the use of high spatial resolution maps provided by ALMA in combination with high spectral resolution observations using SOFIA or HSO. For instance, what might this combination provide for the study of protostellar/protoplanetary disks ?

What will be the role of submillimeter spectroscopy beyond SOFIA and HSO ? It is clear that there are very exciting possibilities for moderate–resolution spectroscopy of distant galaxies using cold space telescopes and direct detection instruments; this should be a major new area for future development. However, high resolution spectroscopy will continue to be of substantial scientific interest, particularly for the study of star formation, and for nearby galaxies, which serve as templates for the high–z objects. The scientific results from SOFIA and HSO will undoubtedly produce a strong desire for observations with higher sensitivity and with better angular resolution. These can be provided with a larger aperture telescope, such as SAFIR, along with the continued development of more sensitive submillimeter receivers. Overall, one can expect a factor of $\sim 5 - 10$ further improvement in receiver sensitivity, which when coupled to a telescope with ~ 5 times larger collecting area, yields an overall sensitivity improvement of ~ 30, a speed advantage of $\sim 10^3$, and a factor of $2 - 3$ better angular resolution. The speed advantage for line surveys will be larger still, due to wider receiver bandwidths.

SUMMARY

The main points are: (1) The submillimeter spectrum has a very high information content, much of it unique, and is best studied at high spectral resolution using coherent detection; (2) recent breakthroughs have dramatically improved the technology for coherent detection, but the technology has not yet reached fundamental limits; (3) the scientific motivation for high resolution spectroscopy will grow substantially as a result of ALMA, SOFIA, and HSO; and (4) large gains are possible beyond these missions, using a larger aperture and better receivers; this will substantially strengthen and broaden the scientific case for a large–aperture mission. Thus, it is important to continue the development of even more capable technology for coherent detection.

Acknowledgments: Submillimeter receiver development at Caltech is supported in part by the NASA/USRA SOFIA instrumentation program, the NASA/JPL HIFI/HERSCHEL program, and NASA grant NAGS–4890.

REFERENCES

[1] T. G. Phillips, "Prospects for High Resolution Spectroscopy from Future Space Platforms," In *Proceedings of the Far–IR, Sub–MM, and MM Detector Technology Workshop*, J. Wolf and J. Davidson, eds., (2002).

[2] R. Genzel, "Detectors for the Future of Extra-Galactic Astrophysics," In *Proceedings of the Far–IR, Sub–MM, and MM Detector Technology Workshop*, J. Wolf and J. Davidson, eds., (2002).

[3] T. Gaier *et al.*, "Amplifier Technology for Astrophysics," In *Proceedings of the Far–IR, Sub–MM, and MM Detector Technology Workshop*, J. Wolf and J. Davidson, eds., (2002).

[4] C. Bradford, J. Bock, M. Dragovan, L. Earle, J. Glenn, B. Naylor, H. Nguyen, and J. Zmuidzinas, "WaFIRS, a Waveguide Far-IR Spectrometer: Enabling Space-Borne Spectroscopy of High-z Galaxies in the Far-IR and Submm," In *Proceedings of the Far–IR, Sub–MM, and MM Detector Technology Workshop*, J. Wolf and J. Davidson, eds., (2002).

[5] A. Karpov, D. Miller, J. Zmuidzinas, J. Stern, B. Bumble, and H. LeDuc, "Low Noise SIS Mixer for the Frequency above 1 THz," In *Proceedings of the Far–IR, Sub–MM, and MM Detector Technology Workshop*, J. Wolf and J. Davidson, eds., (2002).

[6] W. R. McGrath, B. Karasik, A. Skalare, I. Siddiqi, and D. Prober, "Superconductive Hot Electron Bolometer Mixers," In *Proceedings of the Far–IR, Sub–MM, and MM Detector Technology Workshop*, J. Wolf and J. Davidson, eds., (2002).

[7] M. Sumner, F. Rice, R. Hu, and J. Zmuidzinas, "An Extremely Wide Bandwidth, Low Noise SIS Heterodyne Receiver Design for Millimeter and Submillimeter Observations," In *Proceedings of the Far–IR, Sub–MM, and MM Detector Technology Workshop*, J. Wolf and J. Davidson, eds., (2002).

[8] J. Kawamura, D. Miller, J. Chen, J. Zmuidzinas, B. Bumble, H. LeDuc, and J. Stern, "Very high-current-density Nb/AlN/Nb tunnel junctions for low-noise submillimeter mixers," Appl. Phys. Lett. **76,** 2119–2121 (2000).

[9] I. Mehdi, E. Schlecht, G. Chattopadhyay, and P. H. Siegel, "THz Local Oscillator Sources," In *Proceedings of the Far–IR, Sub–MM, and MM Detector Technology Workshop*, J. Wolf and J. Davidson, eds., (2002).

[10] A. Harris, "Spectrometers for Heterodyne Detection," In *Proceedings of the Far–IR, Sub–MM, and MM Detector Technology Workshop*, J. Wolf and J. Davidson, eds., (2002).