Progress in Coherent Detection Methods

J. Zmuidzinas[1] *

[1]California Institute of Technology, 320–47, Pasadena CA 91125, U.S.A.

Coherent detection techniques are used almost exclusively in radio astronomy, while direct detection predominates in the optical and infrared bands. The millimeter, submillimeter, and far–infrared bands represent a transition region from coherent to direct detection. In this paper, the relative merits of coherent and direct detection in the mm through far–IR are compared, the techniques of coherent detection are described, and recent technological advances in coherent detection systems are discussed.

1. Introduction

Coherent detection refers to the radio technique of amplifying, downconverting, and filtering a signal prior to detection. Technology for sensitive coherent detection in the millimeter, submillimeter, and far–infrared continues to be developed rapidly, driven by the needs of projects such as FIRST, SOFIA, and the MMA/LSA. Coherent detection is essential for the large ground–based interferometric arrays, and is also the method of choice for high resolution spectroscopy, where resolutions better than 1 km/s are often necessary for resolving the spectral lines of galactic objects. Indeed, coherent systems can easily produce high resolution spectra with thousands of simultaneous frequency channels. On the other hand, due to bandwidth limitations, coherent receivers are not very well suited for the detection of broadband continuum radiation. For this task, bolometer array cameras are the obvious choice.

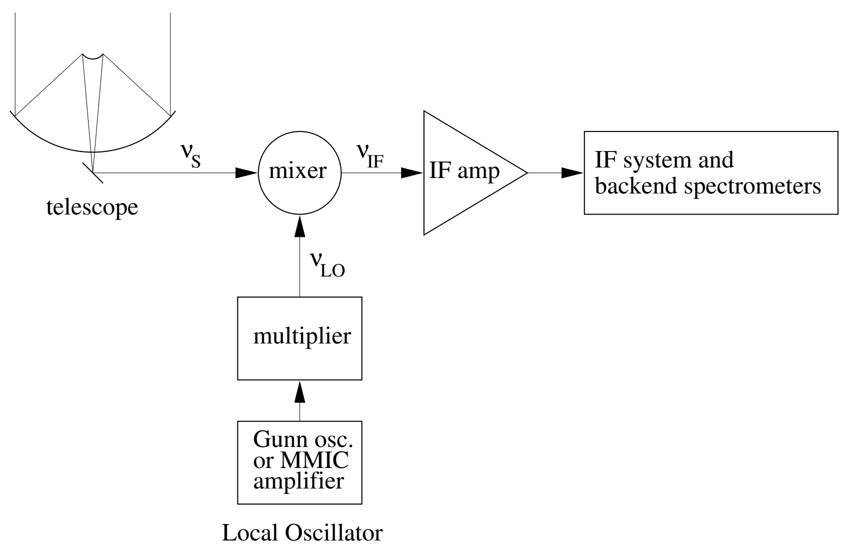

Fig. 1: A block diagram of a submillimeter heterodyne receiver system. The signal from a telesope at a frequency ν_S is combined with a local oscillator at frequency ν_{LO} in a "mixer", which is a nonlinear device, to yield the difference or "intermediate" frequency $\nu_{IF} = |\nu_S - \nu_{LO}|$, which is then amplified and and spectrally analyzed.

A block diagram of a heterodyne receiver system is shown in Fig. 1. In such a system, frequency downconversion is the first operation performed on the signal. This is the only viable possibility for submillimeter wavelengths since low–noise amplifiers are not (yet) available. The

* *e-mail:* <jonas@submm.caltech.edu>

principal components are the local oscillator, the mixer, the intermediate–frequency (IF) amplifier, and the backend spectrometer. All of these components are undergoing intensive development and are improving very rapidly. In particular, there has been a tremendous (nearly two orders of magnitude) improvement in the sensitivity of submillimeter mixers over the past decade ! Some of the more recent progress (especially mixers) will be described in the later sections. In view of this rapid development, it seems useful to assume that the technology will continue to advance until fundamental limits are reached. This point of view is adopted in the next section in order to make a comparison between the ultimate sensitivities of direct and coherent detection.

The literature in this field is quite extensive and it is impossible to provide a complete review here. The interested reader is referred to the article by Carlstrom and Zmuidzinas 1996 for a more detailed review. More recent developments are well represented by the papers in "Advanced Technology MMW, Radio, and Terhertz Telescopes", ed. T. G. Phillips, SPIE proceedings volume 3357, of the conference held in Kona, Hawaii on 26–28 March 1998.

2. Sensitivity of coherent and direct detection

2.1. The quantum limit for coherent detection

Coherent detection is fundamentally different than direct detection. In a coherent system, the incoming signal photons (collected by the telescope) are first amplified before they are detected. This photon amplification process, which in some cases may be simultaneously accompanied by frequency conversion (for heterodyne detection), imposes a limitation on the sensitivity known as the "quantum limit", which is usually expressed as an equivalent noise temperature $T_{QL} = h\nu/k$. Quantum noise was discovered over forty years ago (Shimoda, Takahasi, and Townes 1957) in connection with the development of the maser. As shown in Fig. 2, the maser amplifier provides a simple intuitive picture which explains quantum noise as the result of spontaneous emission by the inverted states which produce the amplification. It is easy to see that (noise) photons would emerge from the amplifier output even when no signal is present at the input. Subsequent work (e.g. Haus and Mullen 1962; Caves 1982) has shown that quantum noise is a general limitation on all "phase insensitive" linear amplifiers and not a peculiar feature of maser amplifiers.

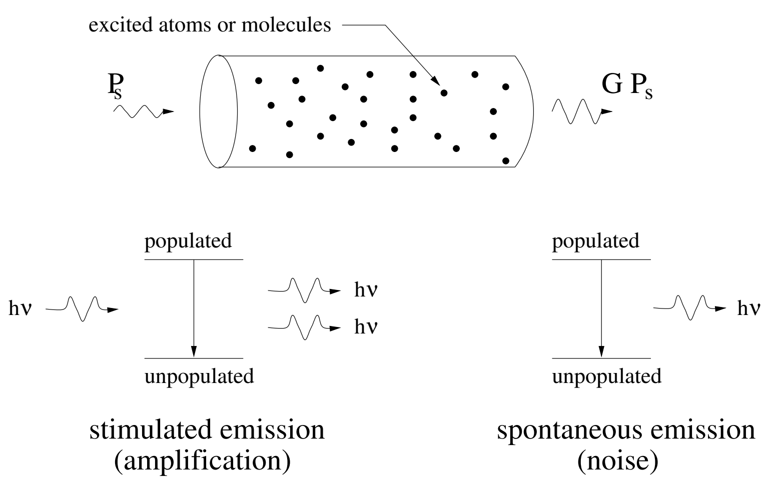

Fig. 2: An illustration of quantum noise in a maser amplifier. This (fictitious) maser amplifier consists of a tube filled with a gas of molecules or atoms, which are pumped in a way that causes some transition with frequency ν to be inverted. A signal arriving at the input with power P_s is amplified by stimulated emission and emerges with power GP_s, where G is the power gain of the amplifier. However, due to spontaneous emission, noise photons emerge from the amplifier output even when $P_s = 0$.

In contrast, there is no fundamental limit to the sensitivity of direct detection. It is possible in theory to make a direct detector which does not produce any signal at its output unless photons are being absorbed. The difference between coherent and incoherent detection can also be understood in terms of the excitation temperature of the quantum states involved in the amplification or detection processes: an ideal direct detector has a small positive excitation temperature, $T_x = +\epsilon$, while an ideal amplifier has a small but *negative* excitation temperature, $T_x = -\epsilon$, where in both cases $\epsilon << h\nu$.

2.2. Background fluctuations and direct detection

The limitations to the sensitivity of direct detection are largely practical. Detectors always produce some noise, although in many cases the detector noise is small compared to the fluctuations in the thermal background radiation from the telescope or the atmosphere. However, detector noise can be an important issue for spectroscopy, where background fluctuations are reduced due to the narrow bandwidth received by the detector. In fact, background–limited high resolution spectroscopy using a cooled space telescope in the submillimeter or far–infrared is still largely out of reach with current detector technology. This partially explains the choice of a Fourier transform spectrometer (FTS) for the spectroscopic capability of the SPIRE bolometer instrument on FIRST (Griffen, Vigroux, and Swinyard 1998).

Whether or not the quantum limit for coherent detection is actually a significant factor depends largely on the level of the background radiation. For instance, a lower limit to the thermal background is provided by the cosmic microwave background (CMB) radiation. This means that the quantum limit is actually irrelevant at radio frequencies ($h\nu/k < 2.7$ K or $\nu < 56$ GHz) regardless of the temperature and location of the radio telescope. In the submillimeter and far–infrared ($\lambda > 100$ μm), the quantum limit is important for cooled space telescopes but is not a limiting factor for ground–based or airborne observations, as we shall see below.

2.3. A comparison of detection methods

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:

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

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 associated with the thermal background radiation entering the system. For example, in the case that the background arises from objects at a single temperature T_{bg} with a total emissivity ϵ, n_0 would be given by $n_0 = \epsilon \left[e^{h\nu/kT_{bg}} - 1\right]^{-1}$. For a coherent receiver, the corresponding expression is given by

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

The two terms in the sum $(1+\eta^c n_0)$ correspond to quantum and background noise, respectively. Thermal emission from the optics inside the instrument has been neglected for both types of detection systems. The expression for a coherent receiver 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.

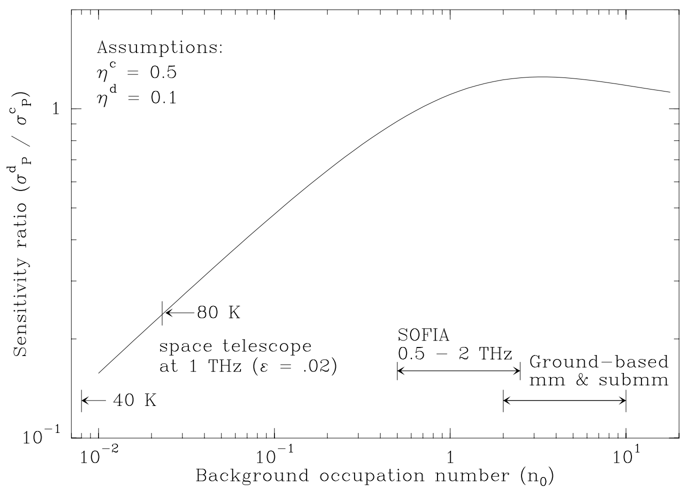

Fig. 3: A comparison of the relative sensitivity of heterodyne and direct detection. Both methods have the same background–limited sensitivity in the high occupation number limit ($n_0 > 1$), while direct detection is more sensitive in the low occupation limit ($n_0 << 1$). The labels "40 K" and "80 K" are appropriate for a passively cooled space telescope with a 2% total emissivity.

We must make assumptions about the quantum efficiencies in order to make a comparison. In general, coherent systems have extremely simple optical paths. In contrast, direct detection systems, particularly high resolution spectrometers, have rather complicated filtering systems to restrict the bandwidth of the radiation landing on the detector, which inevitably reduces the optical transmission and the effective quantum efficiency. For this reason, we assume $\eta^c = 0.5$, which is readily achievable, while for the direct detection spectrometer we assume $\eta^d = 0.1$, which is fairly optimistic for resolutions of 10^4 or higher. The resulting sensitivity comparison under these assumptions is shown in Fig. 3.

2.4. Advantages of coherent detection

Coherent detection has several important practical advantages:

- high spectral resolution ($\Delta\nu/\nu \leq 10^{-6}$) is readily obtained
- the output (IF) signal can be amplified and copied (split) without adding a significant amount of noise
- time–domain "lag" correlators can be used for spectral analysis without incurring a noise penalty

These features are closely related. High spectral resolution is a natural result of frequency down-conversion: the spectral analysis is being carried out at a frequency that is $\sim 10^{-3}$ of the original signal frequency, and so a filter with a relatively modest resolution of 10^{-3} yields a resolution of 1 part in 10^6 at the signal frequency. Furthermore, the transmission loss of the filter is irrelevant since the IF signal can be amplified prior to filtering. The fact that the IF signal can be copied and amplified is extremely useful in interferometry, since the signals from N telescopes are replicated to feed the correlators for the corresponding $N(N-1)/2$ baselines. The fundamental underlying reason that the IF signal can be amplified and copied is that the IF signal is classical, i.e. it is in the high photon occupation limit ($n >> 1$).

Finally, it is interesting to note that direct detection systems generally do not use the concept of the time–domain "lag" correlator which is quite popular in radio astronomy for performing the spectral analysis function of a multichannel filterbank. It is well known (at least to radio astronomers) that a correlator produces spectra with the same sensitivity as a filterbank, apart from the usual quantization noise issue (to reduce the complexity of the hardware, digital correlators typically use digitizers with only 3 or 4 quantization levels). However, in a direct detection correlator, the photons in the optical input beam would need to be shared among the multiple lags, and this would result in a sensitivity degradation. In fact, there is one and only one way to construct an ideal direct–detection spectrometer: first, filter the signal into multiple frequency channels (e.g. using a grating), and then detect each channel with a separate detector. Note that although an FTS is in fact a direct–detection correlator, it has only a single lag, and suffers in sensitivity compared to a grating spectrometer in the background–limited case. This sensitivity degradation associated with direct–detection correlation spectrometers does not occur for coherent systems since the IF signal has a large photon occupation number ($n >> 1$) and so the splitting of the signal to feed multiple lags does not incur a further noise penalty. In a sense, a noise penalty has already been paid since coherent systems are subject to the quantum limit.

3. SIS mixers

Two types of mixers are being developed for astronomical applications, both of which use superconductors: SIS (superconductor–insulator–superconductor) tunnel junction mixers and hot electron bolometer (HEB) mixers. SIS devices are well understood, and in theory can reach the quantum limit of sensitivity. An SIS junction behaves much like a photodiode: an electron in one superconducting film absorbs a photon, which gives the electron enough energy to tunnel through the insulating barrier into the other superconducting film. This process, known as photon–assisted tunneling, yields one electron of current for every photon absorbed by the junction. However, getting the junction to absorb photons is a nontrivial task: the junctions are small, which means that some kind of antenna is required; the junctions are also highly capacitive, and therefore a resonant tuning circuit is needed. These details are discussed more thoroughly by Carlstrom and Zmuidzinas 1996 and the references therein.

In practice, excellent performance is achieved up to 700 GHz using niobium devices. Recent results (e.g. Karpov 1998; Baryshev et al. 1999) are astoundingly good, and the uncorrected receiver noise temperatures are approaching the quantum limit (see Fig. 4). In fact, much of the measured receiver noise is usually contributed by the thermal emission of the receiver optics; the mixer itself can often be very close to the quantum limit. This is an important point: space instruments such as HIFI for FIRST (de Graauw et al. 1998) will not have warm optics.

Above 700 GHz, the performance of SIS mixers deteriorates fairly rapidly. This is due to the onset of loss in the niobium tuning circuits: above 700 GHz, photons have sufficient energy to break the superconducting electron ("Cooper") pairs, and niobium starts to behave like a rather poor normal metal. The 700 GHz "gap frequency" is set by the 9.2 K critical temperature of niobium. The use of good normal metals in place of niobium improves the situation somewhat, and allows reasonably low noise SIS mixers to be operated at 1 THz (e.g. Bin et al. 1996). However, what is really needed is a low–loss superconducting film with a larger critical temperature and gap frequency. Recent work indicates that niobium titanium nitride (NbTiN) films have quite low loss in the submillimeter (Kooi et al. 1998b; Zmuidzinas et al. 1998) and may provide near–quantum limit performance up to the ~ 1.2 THz gap frequency.

Now that the junction fabrication processes are reliable and the material properties are better determined, more complex SIS mixer circuits are becoming possible. A good example is the sideband–separating mixer being developed at NRAO (Kerr, Pan, and LeDuc 1998) for the MMA project. Other possibilities include balanced mixers, dual–polarization mixers, junction array mixers, multiple–beam mixers, or even combinations of the above. New software design tools are now being developed (Ward, Rice, and Zmuidzinas 1999) which will allow these more complex SIS designs to be accurately analyzed and optimized.

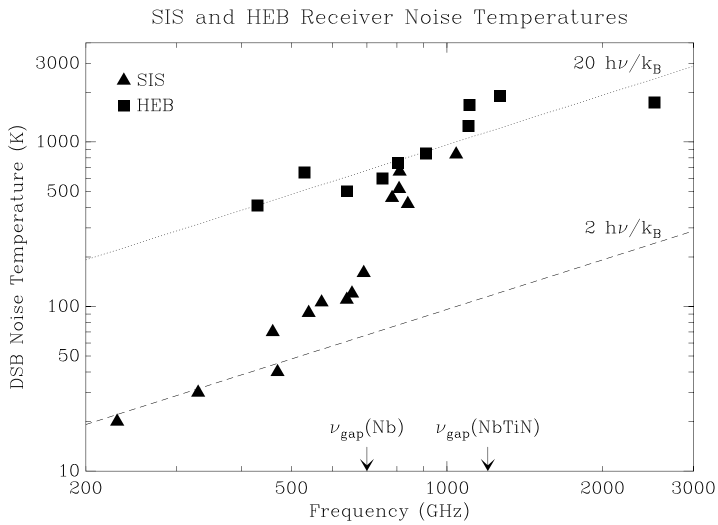

Fig. 4: A selection of SIS and HEB receiver noise temperatures measured to date. The SIS data are taken from: Baryshev et al. 1999; Bin et al. 1996; Gaidis et al. 1996; Honingh et al. 1997; Karpov 1998; Kooi et al. 1998a; Uzawa, Wang, and Kawakami 1998; and Zmuidzinas et al. 1998. The HEB data are taken from: Kawamura et al. 1999; McGrath et al. 1998; Wyss et al. 1999; and Yagoubov et al. 1998.

4. HEB mixers

It is interesting to recall that hot electron bolometer (HEB) mixers have been in use for nearly three decades – in fact, the field of submillimeter astronomy was pioneered in the 1970's by T. Phillips and others using InSb hot–electron mixers. The InSb mixers had narrow IF bandwidths (~ 1 MHz) and were replaced by Schottky and SIS mixers with much wider (~ 1 GHz) bandwidths. There has been a strong resurgence of interest in HEB mixers in the past few years due to the development of new superconducting devices with large IF bandwidths and low noise.

To a rough approximation, the new HEB mixers operate as transition–edge sensors, in which the temperature (and therefore the resistance) of a superconducting film is influenced by the absorption of radiation in the device. The IF bandwidth is directly controlled by the thermal time constant of the device. Thus, to achieve a wide bandwidth, the device must simultaneously have a very low heat capacity and a high thermal conductance. The heat capacity is made small by heating only the electrons in a very small volume ($< 10^{-2}\ \mu m^3$) of a superconducting film. The thermal conductance is made large by choosing a material with a large electron–phonon interaction, or by engineering the device so that the electron outdiffusion is very rapid (Prober 1993). Both types of devices (phonon cooled and diffusion cooled) have now demonstrated IF bandwidths of several GHz, along with noise temperatures roughly following 1 K/GHz (Fig. 4). Note that an HEB receiver has now been demonstrated on a telescope (Kawamura et al. 1999).

Compared to SIS mixers, the new superconducting HEB devices offer two major advantages: higher frequency operation (2500 GHz to date), and lower LO power requirements. However, SIS mixers currently offer up to an order of magnitude better performance at lower frequencies. It is fair to say that the new HEB mixers are not yet as well theoretically understood as the SIS mixers. Some of the recent work (e.g. Wilms Floet et al. 1998; Wilms Floet et al. 1999) has focused on obtaining an improved physical understanding of the resistive transition and the mixing process. HEB's made with lower–T_c materials may provide even lower noise and lower

LO power requirements. However, whether or not HEB mixers can ultimately reach the quantum limit remains an open question.

5. Local oscillators, IF amplifiers, and backend spectrometers

The local oscillator (LO) is a critical component for coherent receivers. For missions such as FIRST, the LO should be compact and lightweight, and ideally use solid–state devices and contain few moving parts. New advances in this field include the use 100 GHz broad–band InP HEMT transistor power amplifiers (Gaier et al. 1998) in place of Gunn oscillators, and fixed tuned, high efficiency harmonic multipliers using planar varactor diodes (Erickson 1998). Terahertz beat frequency generation (Verhese, McIntosh, and Brown 1997) using ultrafast photodectors driven by tunable, frequency stable diode lasers may offer an alternative LO technology, especially for the HEB mixers which require very little LO power. Particularly interesting is the recent demonstration of a phase–locked superconducting (flux–flow) oscillator (Koshelets et al. 1999).

Significant advances are also being made in the intermediate–frequency (IF) amplifiers, which amplify the mixer output before it is processed by the back–end spectrometer. For example, monolithic microwave integrated circuits (MMICs) using InP HEMT transistors should have low noise over bandwidths of 10 GHz or more (Weinreb 1998, priv. comm.). Finally, new backend spectrometers are being built to deal with the increased IF bandwidths, including array acousto–optic spectrometers (Schieder et al. 1998), very high speed digital correlators (Ravera et al. 1998), and wideband analog correlators (Harris, Isaak, and Zmuidzinas 1998).

Acknowledgements

This work was supported in part by NASA/JPL and its Center for Space Microelectronics Technology, by NASA grants NAG5–4890, NAGW–107, and NAG2–1068, by the NASA/USRA SOFIA instrument development program, and by the CSO (NSF grant AST–9615025).

References

Baryshev, A., Luinge, W., Koshelets, V., Shitov, S., and Klapwijk, T. 1999, "Zero-point fluctuations limited SIS receiver at 500 GHz", *Applied Physics Letters*, submitted.

Bin, M., Gaidis, M. C., Zmuidzinas, J., Phillips, T. G., and LeDuc, H. G. 1996, "Low–noise 1 THz superconducting tunnel junction mixer with normal metal tuning circuit", *Appl. Phys. Lett.* **68**, 1714–1716

J. Carlstrom and J. Zmuidzinas 1996, "Millimeter and submillimeter techniques", in *Review of Radio Science 1993–1996*, ed. W. Ross Stone, Oxford University Press, 839–882

Caves, C. M. 1982, "Quantum limits on noise in linear amplifiers", *Phys. Rev. D* **26**, 1817–1839

de Graauw, Th. et al. 1998, "Heterodyne instrument for FIRST (HIFI): preliminary design", *Proc. SPIE* **3357**, 336–347

Erickson, N. 1998, "Diode frequency multipliers for THz local oscillator applications", *Proc. SPIE* **3357**, 75–84

Gaidis, M. C., LeDuc, H. G., Bin, M., Miller, D. M., Stern, J. A., and Zmuidzinas, J. 1996, "Characterization of low–noise quasi–optical SIS mixers for the submillimeter band", *IEEE Trans. Micr. Theory Tech.* **44**, 1130–1139

Gaier, T., Mehdi, I., Kooi, J., Fujiwara, B., and Lai, R. 1998, "A W–band HEMT based power amplifier module for millimeter–wave LO multipliers", *Proc. SPIE* **3357**, 114–121

Griffen, M., Vigroux, L., and Swinyard, B. 1998, "SPIRE – a bolometer instrument for FIRST", *Proc. SPIE* **3357**, 404–413

Harris, A. I., Isaak, K. G., and Zmuidzinas, J. 1998, "WASP: a wideband spectrometer for heterodyne spectroscopy" *Proc. SPIE* **3357**, 384–394

Haus, H. A. and Mullen, J. A. 1962, "Quantum noise in linear amplifiers", Phys. Rev. **128**, 2407–2413

Honingh, C. E., Haas, S., Hottgenroth, D., Jacobs, K., Stutzki, J. 1997, "Low noise broadband fixed tuned SIS waveguide mixers at 660 and 800 GHz", *IEEE Trans. Appl. Supercond.* **7**, 2582–2586

Karpov, A. 1998, "Receivers for ground–based mm wave radio telescopes", *Proc. SPIE* **3357**, 593–608

Kawamura, J. H., Tong, C.–Y. E., Blundell, R., Cosmo Papa, D., Hunter, T. R., Gol'tsman, G., Cherednichenko, S., Voronov, B., and Gershenzon, E. 1999, "An 800 GHz NbN phonon–cooled hot–electron bolometer mixer receiver", *IEEE Trans. Appl. Supercond.*, in press.

Kerr, A. R., Pan, S.–K., and LeDuc, H. G. 1998, "An integrated sideband separating SIS mixer for 200–280 GHz", *Proc. 9th Int. Symp. on Space Terahertz Technology*, Pasadena, CA, 215–221

Kooi, J. W., Pety, J., Bumble, B., Walker, C. K., LeDuc, H. G., Schaffer, P. L., and Phillips, T. G. 1998a, "A 850-GHz waveguide receiver employing a niobium SIS junction fabricated on a 1 μm Si_3N_4 membrane" *IEEE Trans. Micr. Theory Tech.* **46**, 151–161

Kooi, J. W., Stern, J. A., Chattopadhyay, G., LeDuc, H. G., Bumble, B., and Zmuidzinas, J. 1998b, "Low-loss NbTiN films for THz SIS mixer tuning circuits", *Int. J. IR and MM Waves*, **19**, 373–383

Koshelets, V. P., Shitov, S. V., Filippenko, L. V., Vaks, V. L., Mygind, J., Baryshev, A. B., Luinge, W., and Whyborn, N. 1999, "Completely phase–locked Josephson flux-flow oscillator", *Applied Physics Letters*, submitted.

McGrath, W. R., Skalare, A., Karasik, B., Gaidis, M. C., Bumble, B., LeDuc, H. G., Burke, P. J., Schoelkopf, R., and Prober, D. E. 1998, "Superconductive hot–electron mixers for terahertz heterodyne receiver applications", *Proc. SPIE* **3357**, 14–21

Ravera, L. et al. 1998, "Wide band digital autocorrelator for FIRST", *Proc. SPIE* **3357**, 368–375

Schieder, R., Horn, J., Siebertz, O., Möckel, C., Schlöder, F., Macke, C., and Schmülling, F. 1998, "Design of large bandwidth acousto–optic spectrometer", *Proc. SPIE* **3357**, 359–367

Prober, D. E. 1993, "Superconducting terahertz mixer using a transition-edge microbolometer" *Appl. Phys. Lett.* **62**, 2119–2121

Shimoda, K., Takahasi, H. and Townes, C. H. 1957, "Fluctutations in amplification of quanta with application to maser amplifiers", *J. Phys. Soc. Japan* **12**, 686–700

Uzawa, Y.,, Wang, Z.,, and Kawakami, A. 1998, "Terahertz NbN/AlN/NbN mixers with Al/SiO/NbN microstrip tuning circuits", *Appl. Phys. Lett.* **73**, 680–682

Verghese, S., McIntosh, K. A., and Brown, E. R. 1997, "Highly tunable fiber-coupled photomixers with coherent terahertz output power" *IEEE Trans. Micr. Theory Tech.* **45**, 1301–1309

Ward, J., Rice, F., and Zmuidzinas, J. 1999, "Supermix: a flexible software library for high-frequency circuit simulation, including SIS mixers and superconducting components", *Proc. Tenth Intl. Symp. Space Terahertz Technology*, Charlottesville, VA, to appear

Wilms Floet, D., Baselmans, J. J. A., Klapwijk, T. M., and Gao, J. R. 1998, "Resistive transition of niobium superconducting hot-electron bolometer mixers", *Appl. Phys. Lett.* **73**, 2826–2828

Wilms Floet, D., Miedema, E., Gao, J. R., and Klapwijk, T. M. 1999, "Hotspot mixing: a new framework for heterodyne mixing in superconducting hot–electron bolometers" *Appl .Phys. Lett.* **74**, 433–435

Wyss, R. A., Karasik, B. S., McGrath, W. R., Bumble, B., and LeDuc, H. G. 1999, "Noise and bandwidth measurements of diffusion-cooled Nb hot-electron bolometer mixers at frequencies above the superconductive energy gap", *Proc. Tenth Intl. Symp. Space Terahertz Technology*, Charlottesville, VA, to appear

Yagoubov, P., Kroug, M., Merkel, H., Kollberg, E., Gol'tsman, G., Lipatov, A., Svechnikov, S. and Gershenzon, E. 1998, "Quasioptical NbN phonon cooled hot electron bolometric mixers with low optimal local oscillator power", *Proc. 9th Int. Symp. on Space Terahertz Technology*, Pasadena, CA, 131–145

Zmuidzinas, J., Kooi, J. W., Kawamura, J., Chattopadhyay, G., Bumble, B., LeDuc, H. G., and Stern, J. A. 1998, "Development of SIS mixers for 1 THz", *Proc. SPIE* **3357**, 53–61