The Institute of Space and Astronautical Science
Report SP No.14, December 2000

Science with a Large, Far-Infrared Telescope

By

George H. RIEKE*

(November 1, 2000)

Abstract: The behavior of interstellar gas and dust is responsible for many intriguing mysteries of the Universe, giving the far-infrared/submm spectral region a special importance. A new generation of FIR/Submm space instruments is needed to realize the full potential of this broad spectral range. We show how these advanced FIR/Submm missions can make basic contributions to our understanding of questions ranging from the birth of stars and planets and the process of comet and asteroid impact on the Earth to the history of star formation and active galactic nuclei over the life of the Universe.

1. THE SCIENTIFIC ROLE OF THE FAR IR/SUBMM

Winds and flows in the interstellar medium convert a potentially static scene into our mysterious and fascinating Universe. A supermassive black hole lurks unseen in the nucleus of a galaxy until interstellar gas collects into a central accretion disk and spirals in, causing an active galactic nucleus (AGN) to blaze up. Galaxy collisions spray stars in intriguing patterns, but the fundamental consequences arise from the ability of the interstellar medium (ISM) to lose angular momentum and collapse to fuel nuclear starbursts. Stellar populations everywhere are established and renewed by the formation of new stars in molecular clouds. The heavy elements that shape stellar evolution and make life possible are transported by interstellar material to the sites of star formation, awaiting incorporation into new stars and planets.

The far-infrared and submm are critical for probing the interstellar medium. Regardless of the original emission process, cosmic energy sources glow in the far infrared due to the effectiveness of interstellar dust in absorbing visible and ultraviolet photons and reemitting their energy. For example, the Milky Way and other galaxies show two broad spectral peaks, one produced directly by stars and extremely thoroughly studied in the visible and near infrared and the second comparatively unexplored in the far infrared. Warm, dense interstellar gas cools predominantly through low energy fine structure lines and also emits profusely in rotational transitions of the most abundant molecules; both systems of lines emerge predominantly in the far infrared and submm. These lines are key participants in the process of collapse that regulates formation of stars and AGNs. They also provide detailed insights to the temperature, chemical composition, density, and ionization state of the collapsing clouds.

* Steward Observatory, University of Arizona, Tucson, AZ 85721, USA; grieke@as.arizona.edu

1.1 How Stars and Galaxies Emerged from the Big Bang

The history of star formation determines the evolution of galaxies and the generation rate for heavy elements. It has been traced by a combination of deep Hubble Space Telescope (HST) imaging along with photometry and spectroscopy using the Keck Telescope. However, even at modest redshifts, these techniques only probe the rest frame ultraviolet. Interstellar dust can absorb nearly all the UV in star forming galaxies. In the best-studied starburst galaxies such as M82, a debate raged for more than a decade regarding how to correct even the near infrared emission for the effects of interstellar extinction. Such corrections are poorly determined for galaxies at high redshift. Consequently, there are significant uncertainties in the star forming rate for $z > 1$.

These uncertainties could be removed by measuring the far-infrared emission emitted by dust heated by young stars in these galaxies. The importance of this approach is underlined by the Cosmic Background Explorer (COBE) discovery of a background in the submm with an energy density comparable to the visible and near-infrared background. This background has been partially resolved by ISO in the very far-infrared and is thought to arise from starburst galaxies at $z = 1$ to 3 (Puget et al. 1999). A 10 m telescope with detection limits of 0.1mJy or less would probably resolve most of this high redshift background into individual galaxies, thus showing the dominant phases of dust embedded star formation and nuclear activity throughout the Universe.

Ultradeep optical images (e.g. Hubble Deep Field) reveal many galaxies too faint to contribute significantly to the submm diffuse background. To complete the study of star formation in the early Universe requires that we extend our understanding to these small systems and possible galaxy fragments. In this luminosity range and over $1 < z \leq 5$, ALMA and other groundbased submm telescopes are mostly sensitive to infrared cirrus emission and the output of cold dust that are not necessarily heated by recent star formation. The rate of star formation in modest galaxies for $1 < z \leq 5$ can best be determined through high sensitivity imaging from 20 to 200μm. Angular resolution of about 1" (10 m telescope at 50μm) will be adequate; sensitivities of $\sim$10μJy would allow measurements to galaxy luminosities well below 10^{11} $L_\odot$.

1.2 Formation of AGNs

It appears that central supermassive black holes are a universal component of galactic bulges. At the current epoch, galaxy mergers produce huge far-infrared fluxes through a combination, evidently, of violent starbursts and of AGNs associated with these black holes. Both types of energy source are hidden within cocoons of interstellar dust that are impenetrable in the optical and near-infrared. What happens during the much more common mergers that build galaxies in the early Universe? Is the strong cosmic evolution of quasars an indication that their formation is favored at early epochs? Is much of the far-infrared luminosity in the early Universe derived from dust embedded AGNs? Do AGNs at high redshift differ in basic properties from nearby ones?

The fine structure lines of NeII (12.8μm), NeIII (15.6μm) and NeV (14.3μm) are the best tool to distinguish unambiguously whether the ISM of a dusty galaxy is ionized by a starburst or by an AGN (e.g. Spignolio & Malkan 1992). Not only are the line ratios very well separated, but their extinction is reduced by more than a factor of thirty compared with the visible. At the epoch of peak quasar activity, these lines will be redshifted to the 45 to 55μm range. A 10 m far-infrared telescope would have both the necessary resolution and sensitivity to use this tool to determine the relative roles of star formation and nuclear activity in the early Universe.

The full suite of infrared fine structure lines probes a very wide range of excitation energy and, with a large far-infrared telescope, could establish the UV spectra of AGNs over large lookback times, extending work with the Infrared Space Observatory (ISO) on a few nearby Seyfert galaxies. In addition, many of these lines have relatively high critical densities (up to 10^{10} cm^{-3}), so they have a unique ability to probe the density of the gas around AGNs.

1.3 Dynamical and Chemical Evolution of Galaxies and Stars

How do the first gas clouds form? What chemical processes occur within them and how do their characteristics change as the first traces of metals are injected into them by stellar processing?

The low lying H_2 lines at 17 and 28.2μm are one of the few conceivable ways to study molecular gas prior to the formation of metals (and to determine if molecules can even form during this epoch). ISO has demonstrated that these lines can be detected in the interstellar medium (Thi et al. 1999). In this case, fits to their absolute and relative strengths require a molecular gas temperature of about 100K and a density of about 3000 cm^{-3}. However, these lines are undetectable from the ground until z $>$ 20 (particularly since both need to be measured for physical interpretation of temperatures and densities).

Once even traces of metals have formed, the C^+ line at 157μm becomes very bright. Its luminosity in nearby spiral galaxies is typically a few tenths of a percent of the entire bolometric luminosity of the galaxy. Although this line is partially accessible in the poor atmospheric windows between 300 and 700μm, it will be routinely observed from the ground only at z $\geq$ 4, when beyond 800μm. Study of the molecular hydrogen and C^+ emission of gas clouds in the early Universe and as a function of redshift promises to reveal many of the processes occurring in the gas clouds that collapse into the first galaxies. Space-borne observations in the FIR/Submm must be a major component of this study.

1.4 Birth and Evolution of Stars and Planetary Systems

Stars are born in cold interstellar cloud cores that are so optically thick they are undetectable even in the mid-infrared. In about 100,000 years, a young star emerges, ejecting material along powerful jets and still surrounded by a circumstellar disk. The subsequent evolution is increasingly well studied, but the star formation event has occurred hidden from view. How does the cloud core collapse? How does subfragmentation occur to produce binary stars? What are the conditions within protoplanetary disks? When, where, and how frequently do these disks form planets?

The birth of stars and planets can be probed thoroughly at FIR/Submm wavelengths. A far-infrared 10 m telescope provides a resolution of 1 arcsec at 50μm ($\leq$ 100 AU for the nearest star forming regions), so imaging could probe the density and temperature structure of these 1000 AU collapsing cores on critical physical scales. The gas in the core is warmed until its primary transitions lie in the FIR/Submm. Spectroscopy in molecular lines such as H_2O and the J>6 high series lines of CO, as well as in FIR atomic lines of OI, C^+, and NII, can probe the physical conditions in the collapse. In addition, 100 AU resolution would reveal the steps toward binary formation. Far-infrared polarimetry is a powerful probe of magnetic field geometries, both for studying core collapse and mapping the fields that must play an important role in accelerating and collimating jets.

The spectrum for a collapsing cloud core has been predicted by Ceccarelli, Hollenbach, & Tielens (1996). The OI lines have narrow components from the infalling envelope and broad ones from outflow shocks. They are the main coolant of the gas in the intermediate regions of the cloud. Bright H_2O lines between 25 and 180μm are the dominant coolant in the inner cloud, where a broad component is expected from the accretion shock and a narrow one from the disk. The CO lines from 170 to 520μm are the main coolant for the outer cloud; warmer CO from within the cloud can also be studied because of velocity shifts due to the collapse. This suite of lines therefore would allow us to probe the process of star formation thoroughly.

1.5 Nature and Formation of the Solar System

What were the conditions in the early solar nebula, as the protoplanetary disk formed and planets and small bodies accreted out of it? All the bodies in the inner solar system have been so heavily processed that they no longer reflect clearly the conditions at their formation. The discovery of many small bodies outside the orbit of Neptune, or crossing that orbit, gives access to objects where accretion proceeded slowly and its products should be primitive and still reflect conditions in the early solar nebula. For brevity, we refer to all these objects as Kuiper Belt Objects (KBOs).

KBOs are being discovered rapidly, from deep CCD images that catch their faint reflected light. It has become clear that there is a large population, including objects of large size, rivaling the largest asteroids. Because of selection effects, the system probably extends much farther than indicated by currently known objects.

Surprisingly, the KBOs and related objects appear to have a broad variety of surface characteristics revealed in a variety of colors in photometry between 0.5 and 2.5μm (e.g. Green et al. 1997). To interpret the clues they provide for formation of the solar system requires that we understand how this variety of surface chemistry has come about. Two very important parameters are: 1.) the albedoes of the surfaces (important to help identify the substances that cover them); and 2.) surface temperatures (both to help understand what chemical reactions can occur and to determine the escape rates for different molecules). Both of these parameters can be determined in the far-infrared, through measurements of the thermal emission. It is for this reason that the 1998 United States National Academy of Sciences study on "Exploring the Trans-Neptunian Solar System" placed a very high priority both on large, far-infrared telescopes and on development of high performance far-infrared detector arrays.

1.6 Comet and Asteroid Impacts and the Origin of Life

The Kuiper Belt is thought to be the source of short period comets and hence has a central role in the comet impacts that brought water to the earth and made life possible here. However, most traces of this process have been erased by time. How can we understand the conditions that regulated the early formation and evolution of the KB and its release of comets toward the inner solar system?

The Infrared Astronomy Satellite (IRAS) discovered debris disks around Vega, β Pic, and other stars, with evidence for inner voids that might have resulted from planet formation. The Kuiper Belt is therefore similar in many ways to these systems and should be interpreted as the debris disk of the solar system. Taking an example, β Pic is thought to be only about 20 million years old. Transient and variable absorptions by the CaII H&K lines in its spectrum have been interpreted as the infall of small bodies from the debris system. This system contains fine grains that heat sufficiently to be detected in the mid-infrared and scatter enough light to be seen at shorter wavelengths. Because it should be drawn into the star quickly, this fine dust may be produced in recent collisions between planetesimals. Thus, this system and others like it demonstrate the potential of examining the early, violent evolution of debris disks and the infall of comets.

Debris disks are bright in the far-infrared, where they can be imaged to identify bright zones due to recent planetesimal collisions, as well as voids. The radial zones sampled will vary with wavelength, from a few AU near 20μm to hundreds of AU in the submm. Spatially resolved spectroscopy with such a telescope could probe the mineralogy of the debris disks in the 20–35μm region where the Infrared Space observatory (ISO) has found a number of features diagnostic of crystalline and amorphous silicates, and can locate ice through its 63μm emission feature. Giant planets similar to Jupiter and Saturn could be detected to compare their placement with the debris disk structure.

1.7 Discovery Potential

To compare the discovery potential of different missions operating over the same spectral region, the Bahcall Committee developed a figure of merit they called "Astronomical Capability". This parameter is defined as the product of the lifetime, the efficiency, and the number of pixels divided by the square of the sensitivity. It scales the amount of data can be obtained in a given mission, if all other things were equal such as optical parameters defining the projection of the pixels onto the sky. Compared with our current knowledge about the far-infrared sky (from ISOPHOT between 20 and 100μm) and what could be achieved with the best currently available detector arrays on a 10 m cold telescope, the gain in Astronomical Capability is about a factor of 10^9! With improvements in detectors, a factor of well over 10^{10} should be possible.

To place this gain on a more familiar scale, the improvement from the work of Shapley on the size and shape of the Milky Way to the Hubble Deep Field represents a gain of about 10^7 in Astronomical Capability. Thus, our ability to predict what will be found in future FIR/Submm missions is probably no better than Shapley would have done in predicting the contents of the HDF; recall that at the time he did not believe in extragalactic astronomy.

It would probably be foolish to try to leap a factor of 10^{10} in one step. SIRTF should advance us about half this way. That is, it will provide an increase in Astronomical Capability by about 10^5, leaving a similar factor for future missions like the ones we have discussed.

2. NEW TOOLS TO EXPLORE THE FIR/SUBMM UNIVERSE

2.1 Filled Aperture Telescopes

The science goals above suggest that a major advance over previous missions could be achieved with a telescope of about 10 m aperture and correspondingly high sensitivity and angular resolution. The investment for NGST in large scale, lightweight, deployable mirrors can enable such a FIR/Submm capability at lower cost than NGST because of the ~ 20 times reduction in surface precision for the optics. Many other aspects of NGST could be adapted to such a telescope, helping to constrain its cost. Additional thermal shields would allow the telescope to reach the required very low temperatures without a fundamental change in the NGST architecture. With the large aperture, sensitivities more than two orders of magnitude greater than with the current generation of facilities should result. Alternately, placing a FIR/Submm telescope into an orbit that reaches into the outer solar system can both provide increased radiative cooling to reduce the optics temperature and also reduce the natural far-infrared background due to thermal emission by the zodiacal dust grains, providing even better performance.

Such a telescope would be the logical next step beyond SIRTF and FIRST for very deep, moderately high angular resolution measurements in the 40 to 300μm spectral region where the atmosphere of the earth is opaque. It would have collecting area and resolution (in beam area) 100 times greater than SIRTF and 10 times greater than FIRST. Allowing for an order of magnitude gain in detector performance, its gain in sensitivity is equivalent to a gain in throughput over SIRTF by a factor of 10^5. Its operation above the atmosphere will make it far more powerful than groundbased or airborne platforms for spectroscopy across the entire 20 to 800μm spectral range. Its angular resolution and sensitivity will address a significant portion of the science problems we have identified for this spectral region.

Because the telescope architecture can benefit from engineering studies of NGST, system-level studies for this telescope should be conducted in the early-to-middle part of the decade. The heritage from both NGST and SIRTF should allow a large FIR telescope to be developed economically. By the time the mission progressed into detailed definition late in the decade, it would benefit from ten years of progress in detector and heterodyne receiver technology.

I have baselined a 10 m telescope because it provides a combination of sensitivity and resolution very well suited to the highest priority science problems for this spectral region. In addition, it has the potential to gain substantially from the technology investment in NGST. However, in the event that only a moderate far-infrared mission can be provided, then a less challenging approach would be to launch the largest feasible fixed aperture. Such a telescope would still gain from the development of lightweight mirror technology for NGST. In fact, the relaxation of surface specifications would allow even lighter mirrors, which might be important in providing economical launch capability for a large fixed aperture. Two candidates are already proposed: the HII/L2 mission and a 6 m fixed aperture concept studied by Lockheed for NGST.

A more ambitious and technically challenging approach would exploit and expand upon developments in ultra lightweight optics to allow a telescope significantly bigger than 10 m. A 30 m telescope operating at 25K would have sensitivity similar to that of a colder 10 m, but with increased angular resolution. The higher operating temperature would simplify cooling (possibly allowing an all-radiative solution to this requirement) and also reduce the problems with contamination of the optics due to thruster effluvient (particularly if a hydrogen based jet system is used).

2.2 Focal Planes and Supporting Electronics

Detection techniques have been pushed very close to theoretical limits across most of the electromagnetic spectrum. Astronomy has benefited in these spectral regions from huge investments in the technology infrastructure for arrays to be used in commercial and military applications, as well as the highly developed capabilities to process electronic circuits in silicon. Only a relatively modest additional investment has gifted us with megapixel arrays operating at high quantum efficiency and with read noises of only a few photons equivalent.

Detectors for the FIR/Submm have not gained comparably from commercial and military investments, and they tend to be based on technologies which lack the leverage from integrated circuit development. Continuum imaging and spectral mapping at modest spectral resolution require arrays of photoconductors or bolometers with many thousands of high quantum efficiency pixels operating at about ten photons read noise equivalent. High spectral resolution imaging needs the spectral and spatial multiplexing provided by coherent receiver arrays of many hundreds of pixels that approach the quantum limit at far-infrared wavelengths. Current capabilities fall short of these performance levels.

Fortunately, there are promising developments toward these goals. Particularly with increased support, future missions will be able to take advantage of rapidly improving capabilities. For example, the first true integrated high performance far-infrared photoconductor arrays have been developed for SIRTF. They will provide natural-background-limited performance to the cutoff wavelength of Ge:Ga, near 120μm. The technical approach employed for these devices should be capable of expansion beyond the current 32×32 pixel format. Stressed Ge:Ga photoconductors can provide similar levels of performance to beyond 200μm, but the complications in manufacturing large format arrays have not yet been solved. However, a larger scale array concept is under development for the Photoconductor Array Camera and Spectrometer (PACS) instrument on FIRST. Photoconductors based on GaAs rather than Ge may extend the range of this detector type to 300μm.

Bolometers operating at 50–300mK are the most sensitive continuum detectors in the submm and short mm wavelengths. The developments of metal film absorbers and lithographed support membranes such as for the GSFC 'popup' and JPL 'spider web' devices are beginning to solve the problems in producing large arrays. However, the necessity for individual high impedance JFET readout amplifiers operating near 50K creates major complexities in high performance bolometer arrays.

A new generation of bolometers is being developed which will make use of a superconducting film which is voltage-biased at the superconducting/normal transition edge to produce a detector with strong negative electrothermal feedback. This feedback increases the bolometer dynamic range and speed and suppresses Johnson noise. The current that holds the bolometer temperature constant is measured with a SQUID ammeter. The SQUID readouts operate at low temperatures, dissipate very little power, and have large noise margin. A readout multiplexing scheme involving SQUID switches and a single SQUID amplifier is under development, and a 1x8 multiplexer has been demonstrated. When large scale multiplexed readouts become available, arrays of 1000 or more bolometers will become practical.

The voltage-biased sensor and SQUID multiplexer can be used with metal film absorbers which are thermally isolated by thin legs as in the spider web or popup designs. Alternately, they can be antenna coupled and thermally isolated by the relatively weak electron-phonon interaction at low temperatures. Antenna coupled bolometers can make use of thin film transmission line filters and frequency multiplexers. They are fast enough to meet the fringe frequency requirements of interferometry with direct detectors.

A promising antenna coupled detector has been proposed in which the excited quasiparticles in a superconducting film are read out through a superconducting radio frequency single electron transistor before they can thermalize. In principle, this device could operate as a submillimeter single quasiparticle photon counter. Multiplexed arrays appear possible if the transistors are excited at individual frequencies.

Submillimeter (Terahertz) heterodyne mixers are required for high spectral resolution in this spectral region. Receivers based on SIS quasiparticle mixers are within a factor of five of the fundamental quantum limit up to the 600 GHz limit (500μm) for Nb technology. The production of high quality SIS mixer junctions from superconductors with wider bandgaps, and hence capable of operating at higher frequencies, has proceeded more slowly. However, good results are now reported with NbTiN, which has a bandgap nearly twice as large as Nb and hence can operate at twice the frequency. The antenna-coupled hot electron bolometers provide a suitable approach for high quality mixers operating at still higher frequencies. Although all these devices need relatively little local oscillator power, stable and tunable devices to generate this power also require attention.

ACKNOWLEDGMENT

This material was originally prepared with assistance from the M. Bicay, S. Casey, G. Melnick, J. Peterson, P. Richards, M. Werner, & C. Wilson.

REFERENCES

Ceccarelli, C., Hollenbach, D. J., & Tielens, A. G. G. M. 1996, ApJ, 471, 400
Green, S. F. et al. 1997, MNRAS, 290, 186
Puget, J. L. et al. 1999, A&A, 345, 29
Spinoglio, L. & Malkan, M. A. 1992, ApJ, 399, 504
Thi, F. W. et al. 1999, in: ESA SP-427, The Universe as Seen by ISO, ed. P. Cox & M. F. Kessler (Noordwijk: ESA), 767