New Concepts for Far-Infrared and Submillimeter Space Astronomy

Edited by:
Dominic J. Benford and David T. Leisawitz
Goddard Space Flight Center, Greenbelt, MD

Proceedings of the Second Workshop on New Concepts for
Far-Infrared and Submillimeter Space Astronomy, held
March 7-8, 2002 at the University of Maryland Inn and
Conference Center, College Park, MD

The NASA STI Program Office … in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA Scientific and Technical Information (STI) Program Office plays a key part in helping NASA maintain this important role.

The NASA STI Program Office is operated by Langley Research Center, the lead center for NASA's scientific and technical information. The NASA STI Program Office provides access to the NASA STI Database, the largest collection of aeronautical and space science STI in the world. The Program Office is also NASA's institutional mechanism for disseminating the results of its research and development activities. These results are published by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA's counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or cosponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and mission, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services that complement the STI Program Office's diverse offerings include creating custom thesauri, building customized databases, organizing and publishing research results . . . even providing videos.

For more information about the NASA STI Program Office, see the following:

- Access the NASA STI Program Home Page at http://www.sti.nasa.gov/STI-homepage.html

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA Access Help Desk at (301) 621-0134

- Telephone the NASA Access Help Desk at (301) 621-0390

- Write to:
  NASA Access Help Desk
  NASA Center for AeroSpace Information
  7121 Standard Drive
  Hanover, MD 21076-1320

Available from:

New Concepts for Far-Infrared and Submillimeter Space Astronomy

Local Organizing Committee:

Dominic Benford
David Leisawitz
Lee Mundy
Robert Silverberg

Science Organizing Committee:

Rachel Akeson
Thijs de Graauw
Martin Harwit
Thomas Henning
Antoine Labeyrie
David Leisawitz
John Mather
Takao Nakagawa
George Rieke
Wes Traub
Harold Yorke

Preface

The "Second Workshop on New Concepts for Far-Infrared and Submillimeter Space Astronomy" aimed to highlight the groundbreaking opportunities available for astronomical investigations in the far-infrared to submillimeter using advanced, space-based telescopes. Held at the University of Maryland on March 7-8, 2002, the Workshop was attended by 130 participants from 50 institutions, and represented scientists and engineers from many countries and with a wide variety of experience. The technical content featured 17 invited talks and 44 contributed posters, complemented by two six-person panels to address questions of astronomy and technology. A summary of the Workshop can be found in "Proceedings of Far-IR, Sub-MM & MM Detector Technology Workshop", (J. Wolf, J. Farhoomand & C.R. McCreight, eds.), NASA/CP-211408, 2003, under the title "New Concepts for Far-Infrared and Submillimeter Space Astronomy" (paper #i-02, by D.J. Benford).

The National Research Council's Decade Report, "Astronomy and Astrophysics in the New Millennium," assigned a high priority to a Single Aperture Far-Infrared (SAFIR) observatory and encouraged the subsequent development of space-based far-infrared interferometry. With community guidance from the Origins and Structure and Evolution of the Universe Subcommittees of the Space Science Advisory Committee, NASA recently incorporated SAFIR and a kilometer maximum baseline far-IR interferometer into the Space Science roadmap. The interferometer is widely known as SPECS, the Submillimeter Probe of the Evolution of Cosmic Structure.

An important outcome of this workshop was the development of a "Community Plan for Far-IR/Submillimeter Space Astronomy," which appears on pages xv - xxvi of these proceedings. The name "Community Plan" was adopted because this paper gives the consensus view of the workshop participants. The Community Plan addresses practical considerations, such as the tradeoffs associated with alternative mission designs and the flowdown from scientific objectives to measurement requirements, engineering requirements, and technology needs. It recommends an implementation strategy for technology development and validation, and recommends specific science and technology pathfinder missions that would pave the way for the "roadmap missions" SAFIR and SPECS. The community plan concludes by saying:

> *The time is right to place SAFIR on the NASA plan as one of the successors of SIRTF and JWST, to set our sights on a longbaseline far-infrared/submillimeter interferometric imaging telescope, to further develop far-infrared/submillimeter single-aperture and interferometric mission concepts, and to invest strategically in the technology that will enable future far-infrared/submillimeter missions. Supporting studies and smaller mission opportunities should be actively pursued.*

The practical support of the University of Maryland Inn and Conference Center and of Westover Consultants – particularly Kortney Stevens – was indispensable. Funding to support the Workshop was provided by Juan Roman of NASA/GSFC and Eric Smith of NASA/HQ; the funds to produce the published proceedings were supplied by Juan Roman. The cover art and design theme were developed by Trudi Benford, and the cover for the White Paper was produced by Jay Friedlander. Julie Noonan and Justus Brevik proofread these proceedings, catching many typographical errors. Publication support for these proceedings was led by Molly McDonough.

Writing this preface, now a year after the Workshop was held, it is gratifying to see the impact that this community has had on far-infrared science at NASA and elsewhere. Mission concepts for SAFIR and interferometry have been developed (see, e.g. the paper on SAFIR by Amato et al. in the proceedings of SPIE #4850). A SAFIR Science Working Group has been established, and has productively advanced the SAFIR cause. Technologies suitable for large, cryogenic, far-infrared telescopes are being supported by NASA's Office of Aerospace Technology and are under consideration for flight validation as part of the New Millennium Program mission ST-9. In Europe, a similar workshop is planned for September 2003. It is evident that there is global momentum towards the ultimate goal to provide unprecedented advances in scientific capability at far-infrared and submillimeter wavelengths in the post-SIRTF, SOFIA, and Herschel era. We feel that this momentum is reflected in the papers that appear in these proceedings, and close by expressing our deep appreciation to the authors.

Dominic Benford & Dave Leisawitz
March 5, 2003

Table of Contents

Author List

Participants

LAST NAME	FIRST NAME	AFFILIATION	EMAIL ADDRESS
Ade	Peter	Cardiff University	Peter.ade@astro.cf.ac.uk
Akeson	Rachel	ISC	rla@ipac.caltech.edu
Ali	Shafinaz	U. Wisconsin-Madison	ali@wisp.physics.wisc.edu
Amato	Michael	NASA / GSFC	mamato@mscmail.gsfc.nasa.gov
Arendt	Richard	SSAI / GSFC	arendt@stars.gsfc.nasa.gov
Baker	Charles	NASA / GSFC	cbaker@mscmail.gsfc.nasa.gov
Benford	Dominic	NASA / GSFC	Dominic.J.Benford@nasa.gov
Blain	Andrew	Caltech	awb@astro.caltech.edu
Bock	James	JPL	James.Bock@jpl.nasa.gov
Borne	Kirk	Raytheon ITSS	Kirk.Borne@gsfc.nasa.gov
Boucarut	Ray	NASA / GSFC	ray.boucarut@gsfc.nasa.gov
Boulanger	François	IAS, Université Paris	boulanger@ias.u-psud.fr
Bradford	Matt	Caltech	bradford@submm.caltech.edu
Breon	Susan	NASA / GSFC	susan.breon@gsfc.nasa.gov
Butler	Charles	NASA / GSFC	cbutler@mscmail.gsfc.nasa.gov
Calzetti	Daniela	STScI	calzetti@stsci.edu
Canavan	Edgar	NASA / GSFC	ecanavan@proxima.gsfc.nasa.gov
Caroff	Larry	SAIC	lcaroff@attbi.com
Cheng	Edward	NASA / GSFC	ec@cobi.gsfc.nasa.gov
Chervenak	Jay	NASA / GSFC	james.a.chervenak@gsfc.nasa.gov
Cox	Lucien	NASA / GSFC	cwalters@pop700.gsfc.nasa.gov
Crane	Philippe	NASA Headquarters	pcrane@hq.nasa.gov
Danchi	William	NASA / GSFC	danchi@snoopy.gsfc.nasa.gov
Darrin	Ann	Johns Hopkins	ann.darrin@jhuapl.edu
de Korte	Piet	SRON	P.A.J.de.Korte@SRON.nl
Dean	Bruce	NASA / GSFC	bruce.dean@gsfc.nasa.gov
Devlin	Mark	U. Pennsylvania	devlin@physics.upenn.edu
DiPirro	Michael	NASA / GSFC	michael.dipirro@gsfc.nasa.gov
Doiron	Terence	NASA / GSFC	terence.a.doiron.1@gsfc.nasa.gov
Dooley	Jennifer	JPL	jennifer@squid.jpl.nasa.gov
Dragovan	Mark	JPL	mrk@jpl.nasa.gov
Dudley	Chris	NRL	dudley@vivaldi.nrl.navy.mil
Duncan	William	UKATC	william.duncan@roe.ac.uk
Dwek	Eliahu	NASA / GSFC	edwek@stars.gsfc.nasa.gov
Edgar	Mick	Caltech	mick@submm.caltech.edu
Egan	Michael	AFRL	Michael.Egan@bmdo.osd.mil
Feinberg	Lee	NASA / GSFC	LFeinberg@hst.nasa.gov
Fischer	Jacqueline	NRL	Jackie.Fischer@nrl.navy.mil
Folta	David	NASA / GSFC	cwalters@pop700.gsfc.nasa.gov
Freund	Minoru	NASA / GSFC	mino@stars.gsfc.nasa.gov
Friedman	Edward	Boeing	svsinco@qwest.net
Gao	Jianrong	SRON / DIMES	j.r.gao@tnw.tudelft.nl
Glenn	Jason	CASA	jglenn@casa.colorado.edu
Gmurczyk	Greg	NASA / SAIC	ggmurczy@hq.nasa.gov
Griffin	Matthew	Cardiff University	matt.griffin@astro.cf.ac.uk
Harwit	Martin	Cornell	harwit@verizon.net
Hasan	Hashima	NASA Headquarters	hhasan@hq.nasa.gov
Henning	Thomas	U. Jena	henning@astro.uni-jena.de
Henry	Ross	NASA / GSFC	rhenry@pop500.gsfc.nasa.gov
Heyminck	Stefan	KOSMA	heyminck@ph1.uni-koeln.de

x

LAST NAME	FIRST NAME	AFFILIATION	EMAIL ADDRESS
Hollander	Sam	Space Systems Assoc.	samhollander@msn.com
Howard	Joseph	NASA / GSFC	joseph.m.howard@gsfc.nasa.gov
Hsieh	Wen-Ting	NASA / GSFC	whsieh@pop500.gsfc.nasa.gov
Kaplan	Mike	Boeing SVS	michael.s.kaplan@boeing.com
Kasdin	Jeremy	Princeton University	jkasdin@princeton.edu
Kashlinsky	Sasha	GSFC / SSAI	kashlinsky@stars.gsfc.nasa.gov
Kogut	Alan	NASA / GSFC	kogut@stars.gsfc.nasa.gov
Kutyrev	Alexander	NASA / GSFC	kutyrev@stars.gsfc.nasa.gov
Labeyrie	Antoine	Obs. Haute Provence	labeyrie@obs-hp.fr
Lagache	Guilaine	IAS	guilaine.lagache@ias.u-psud.fr
Lakew	Brook	NASA / GSFC	brook.lakew@gsfc.nasa.gov
Lamarre	Jean-Michel	Observatoire de Paris	jean-michel.lamarre@obspm.fr
Lawrence	Charles	JPL	charles.lawrence@jpl.nasa.gov
Lawson	Peter	JPL	lawson.peter.r@jpl.nasa.gov@jpl.nasa.gov
Leisawitz	David	NASA / GSFC	leisawitz@stars.gsfc.nasa.gov
Lester	Dan	U. Texas	dfl@astro.as.utexas.edu
Lillie	Charles	TRW Space & Elect.	chuck.lillie@trw.com
Markley	Landis	NASA / GSFC	landis.markley@gsfc.nasa.gov
Mather	John	NASA / GSFC	john.mather@gsfc.nasa.gov
Matsumura	Mark	NASA / GSFC	mark.m.matsumura.1@gsfc.nasa.gov
Melnick	Gary	SAO	gmelnick@cfa.harvard.edu
Menzel	Michael	Lockheed Martin	domenick.tenerelli@lmco.com
Miller	David	MIT	millerd@mit.edu
Molinari	Sergio	IFSI-CNR	molinari@ifsi.rm.cnr.it
Moseley	Harvey	NASA / GSFC	samuel.h.moseley@gsfc.nasa.gov
Muller	Ronald	QSS Group	rmuller@qssmeds.com
Mundy	Lee	U. Maryland	lgm@astro.umd.edu
Nakagawa	Takao	ISAS	nakagawa@ir.isas.ac.jp
Nast	Ted	Lockheed Martin	ted.nast@lmco.com
Naylor	Bret	Caltech	naylor@submm.caltech.edu
Nguyen	Hien	JPL	hien.t.nguyen@jpl.nasa.gov
Omont	Alain	IAP	omont@iafr.fr
Pajot	François	CNRS / IAS	francois.pajot@ias.fr
Perrygo	Charles	Swales Aerospace	charles.perrygo@gsfc.nasa.gov
Poglitsch	Albrecht	MPIfEP	alpog@mpe.mpg.de
Prober	Daniel	Yale University	daniel.prober@yale.edu
Puget	Jean-Loup	Université Paris Sud	puget@ias.u-psud.fr
Quadrelli	Marco	JPL - Caltech	marco@grover.jpl.nasa.gov
Quijada	Manuel	Swales	mquijada@swales.com
Radford	Simon	NRAO	sradford@nrao.edu
Rajagopal	Jayadev	STScI	jayadev@stsci.edu
Redding	David	JPL	dcr@huey.jpl.nasa.gov
Rice	Frank	Caltech	rice@submm.caltech.edu
Richards	Paul	U.C. Berkeley	richards@physics.berkeley.edu
Rieke	George	U. Arizona	grieke@as.arizona.edu
Rinehart	Stephen	NASA / GSFC	rinehart@rosette.gsfc.nasa.gov
Rioux	Cyrille	CNRS / IAS	cyrille.rioux@ias.fr
Roman	Juan	NASA / GSFC	juan.roman@gsfc.nasa.gov
Sandell	Göran	USRA	gsandell@mail.arc.nasa.gov
Saraceno	Paolo	IFSI - CNR	saraceno@ifsi.rm.cnr.it

LAST NAME	FIRST NAME	AFFILIATION	EMAIL ADDRESS
Shafer	Rick	NASA / GSFC	richard.a.shafer.1@gsfc.nasa.gov
Shirron	Peter	NASA / GSFC	peter.shirron@gsfc.nasa.gov
Shu	Peter	NASA / GSFC	peter.k.shu@gsfc.nasa.gov
Silverberg	Robert	NASA / GSFC	robert.silverberg@gsfc.nasa.gov
Smith	Howard	Center for Astrophysics	hsmith@cfa.harvard.edu
Smith	Eric	NASA Headquarters	esmith1@mail.hq.nasa.gov
Staguhn	Johannes	NASA / GSFC & SSAI	staguhn@gsfc.nasa.gov
Stahl	Phil	NASA / MSFC	phil.stahl@msfc.nasa.gov
Stahle	Carl	NASA / GSFC	carl.stahle@gsfc.nasa.gov
Stevenson	Thomas	NASA / GSFC	thomas.r.stevenson@gsfc.nasa.gov
Storey	John	U. New South Wales	j.storey@unsw.edu.au
Stringfellow	Guy	U. Colorado	guy.stringfellow@colorado.edu
Strong	Keith	Lockheed Martin	keith.strong@lmco.com
Swain	Mark	JPL	swain@huey.jpl.nasa.gov
Thronson	Harley	NASA Headquarters	hthronso@mail.hq.nasa.gov
Timbie	Peter	U. Wisconsin-Madison	timbie@physics.wisc.edu
Traub	Wes	CfA	traub@cfa.harvard.edu
Vastel	Charlotte	Caltech	vastel@submm.caltech.edu
Velusamy	Thangasamy	JPL	velu@rams.jpl.nasa.gov
Walker	Christopher	U. Arizona	cwalker@as.arizona.edu
Weinberger	Alycia	CIW	weinberger@dtm.ciw.edu
Werner	Michael	JPL	mwerner@sirtfweb.jpl.nasa.gov
Wolf	Jürgen	DLR - USRA / SOFIA	jwolf@mail.arc.nasa.gov
Wollack	Edward	NASA / GSFC	edward.j.wollack.1@gsfc.nasa.gov
Wootten	Alwyn	NRAO	awootten@nrao.edu
Yorke	Harold	JPL	Harold.Yorke@jpl.nasa.gov
Young	Erick	U. of Arizona	eyoung@as.arizona.edu
Zmuidzinas	Jonas	Caltech	jonas@submm.caltech.edu
Zubko	Viktor	NASA / GSFC	zubko@stars.gsfc.nasa.gov

The "Community Plan for Far-Infrared/Submillimeter Space Astronomy" derives from the presentations in these proceedings and subsequent discussions. It represents the consensus view of the participants in the "Second Workshop on New Concepts for Far-Infrared and Submillimeter Space Astronomy," except for the participants from NASA Headquarters who recused themselves.

The "Community Plan" should be cited as follows:
Community Plan for Far-Infrared/Submillimeter Space Astronomy, in "New Concepts for Far-Infrared and Submillimeter Space Astronomy," D.J. Benford & D.T. Leisawitz, eds. (Washington, DC: NASA), NASA/CP—2003-212233, pp. xv - xxvi (2003).

Community Plan for Far-Infrared/Submillimeter Space Astronomy

February 21, 2003

This paper represents the consensus view of the 124 participants in the "Second Workshop on New Concepts for Far-Infrared/Submillimeter Space Astronomy," which was held on 7 – 8 March 2002 in College Park, Maryland. The participants are listed below.

Peter Ade	Jason Glenn	Marco Quadrelli
Rachel Akeson	Matthew Griffin	Manuel Quijada
Shafinaz Ali	Martin Harwit	Simon Radford
Michael Amato	Thomas Henning	Jayadev Rajagopal
Richard Arendt	Ross Henry	David Redding
Charles Baker	Stefan Heyminck	Frank Rice
Dominic Benford	Sam Hollander	Paul Richards
Andrew Blain	Joseph Howard	George Rieke
James Bock	Wen-Ting Hsieh	Stephen Rinehart
Kirk Borne	Mike Kaplan	Cyrille Rioux
Ray Boucarut	Jeremy Kasdin	Juan Roman
François Boulanger	Sasha Kashlinsky	Goran Sandell
Matt Bradford	Alan Kogut	Paolo Saraceno
Susan Breon	Alexander Kutyrev	Rick Shafer
Charles Butler	Guilaine Lagache	Peter Shirron
Daniela Calzetti	Brook Lakew	Peter Shu
Edgar Canavan	Jean-Michel Lamarre	Robert Silverberg
Larry Caroff	Charles Lawrence	Howard Smith
Edward Cheng	Peter Lawson	Johannes Staguhn
Jay Chervenak	David Leisawitz	Phil Stahl
Lucien Cox	Dan Lester	Carl Stahle
William Danchi	Charles F. Lillie	Thomas Stevenson
Ann Darrin	Landis Markley	John Storey
Piet de Korte	John Mather	Guy Stringfellow
Thijs de Graauw	Mark Matsumura	Keith Strong
Bruce Dean	Gary Melnick	Mark Swain
Mark Devlin	Michael Menzel	Peter Timbie
Michael DiPirro	David Miller	Wes Traub
Terence Doiron	Sergio Molinari	Charlotte Vastel
Jennifer Dooley	Harvey Moseley	Thangasamy Velusamy
Mark Dragovan	Ronald Muller	Christopher Walker
Chris Dudley	Lee Mundy	Alycia Weinberger
William Duncan	Takao Nakagawa	Michael Werner
Eliahu Dwek	Ted Nast	Juergen Wolf
Mick Edgar	Bret Naylor	Edward Wollack
Michael Egan	Hien Nguyen	Alwyn Wootten
Lee Feinberg	Alain Omont	Harold Yorke
Jacqueline Fischer	François Pajot	Erick Young
David Folta	Charles Perrygo	Jonas Zmuidzinas
Minoru Freund	Albrecht Poglitsch	
Edward Friedman	Daniel Prober	
Jianrong Gao	Jean-Loup Puget	

Community Plan for Far-Infrared/Submillimeter Space Astronomy

We recommend that NASA pursue the vision for far-IR astronomy outlined in the NAS Decadal Survey, which said: "A rational coordinated program for space optical and infrared astronomy would build on the experience gained with NGST[1] to construct [a JWST-scale filled-aperture far-IR telescope] SAFIR, and then ultimately, in the decade 2010 to 2020, build on the SAFIR, TPF, and SIM experience to assemble a space-based, far-infrared interferometer." SAFIR will study star formation in the young universe, the buildup of elements heavier than hydrogen over cosmic history, the process of galaxy formation, and the early phases of star formation, which occur behind a veil of dust that precludes detection at mid IR and shorter wavelengths. The far-infrared interferometer will resolve distant galaxies to study protogalaxy interactions and mergers and the processes that led to enhanced star formation activity and the formation of Active Galactic Nuclei, and will resolve protostars and debris disks in our Galaxy to study how stars and planetary systems form.

The following unified plan addresses practical issues and makes recommendations that would lead to the fulfillment of the Decadal Report's vision. This plan gives the consensus view of the participants in the "Second Workshop on New Concepts for Far-Infrared/Submillimeter Space Astronomy," which was co-sponsored by NASA Headquarters and the Goddard Space Flight Center and held at the University of Maryland on 7 – 8 March, 2002. The workshop participants were representatives of the community of scientists and technologists who would implement the plan.

We make these recommendations to NASA because information vital to the attainment of major SEU and Origins scientific objectives is uniquely available in the far-IR and submillimeter (FIR/SMM), a spectral range that spans the gap between the longest wavelength accessible to the JWST (formerly NGST), ~25 μm, and the shortest wavelength continuously accessible to ALMA through the atmosphere, ~800 μm. For example, to "understand the structure of the universe, from its earliest beginnings to its ultimate fate," we will need measurements of the emissions from protogalactic objects and galaxies most intimately related to the star formation process, namely emissions that reveal the physical conditions (elemental abundances, temperatures, densities) in the interstellar medium; we will need extinction-free views of the universe complementary to those provided by telescopes that operate at shorter wavelengths; and we will need telescopes that can measure the emissions from the pristine hydrogen clouds that collapsed to form the very first generation of stars. To "explore the ultimate limits of gravity and energy in the universe" we will need to peer into the dust-enshrouded nuclei of galaxies to see how matter behaves in the presence of a supermassive black hole. To "learn how stars and planets form" we will need to observe these objects where they emit most of their light – in the infrared – with telescopes that provide high enough spectral resolution to constrain theoretical models, and sufficient acuity to resolve extrasolar planetary systems. Some of the objects we wish to study – the youngest stars and galaxies – are not even visible at optical and near-IR wavelengths.

---

[1] See Appendix A for acronym definitions.

The value of infrared spectroscopy is evident in the information-rich data from NASA's Submillimeter Wave Astronomy Satellite (SWAS) and ESA's Infrared Space Observatory (ISO). Spectroscopic data give us information on chemical species, velocities, and the interaction of these in the chemodynamics that is virtually certain to play a dominant role in the next stage of astrophysical thought. High spectral resolution will therefore be an essential capability for future FIR/SMM missions.

The FIR/SMM is the least-explored spectral region in astronomy, even though COBE told us that half the luminosity of the universe, and 98% of the photons (aside from the cosmic microwave background), appear in the far-IR. Far-IR astronomy has been hampered by angular resolution worse than that of Galileo's first telescope, and by sensitivity limited by small apertures and early generations of detectors. Even SIRTF is still modest in size, only a little larger than the IRAS flown in 1982. The Herschel telescope at 70 K and the SOFIA telescope at 250 K will be about 4x and 3x larger, respectively, but are both relatively warm for this wavelength regime, and thus have limited sensitivity. The scientific yield of the upcoming missions SIRTF, SOFIA, and Herschel will whet our appetites for more sensitive FIR/SMM telescopes that can be used to detect the faint emissions from young, and therefore distant, galaxies. A further gain in sensitivity by several orders of magnitude is necessary to see these sources, and is attainable with a large aperture space observatory whose sensitivity is limited only by the insurmountable photon noise from astrophysical background radiation. New technology and mission concepts now enable a major breakthrough in this area, with the confident expectation of new discoveries.

First Step: SAFIR

The first step is to develop the technology and start the planning for a cooled JWST-class far-IR observatory called SAFIR (Single Aperture Far-IR telescope), to be operated like HST for a wide user community with a launch by the middle of the JWST lifetime in 2015. The scientific motivation and concepts for SAFIR are presented in the white paper *Charting the Winds that Change the Universe* (Appendix B). SAFIR should be background limited over a wavelength range from about 15 to 600 μm to overlap slightly with JWST and ground-based capabilities, and could be diffraction limited at around 40 μm. With a 10 m aperture (a little larger than JWST's) it would have 150 times the collecting area and an order of magnitude greater angular resolution at a given wavelength than SIRTF. Figure 1 shows the relevance of this improvement in angular resolution to the measurements needed to achieve Origins and SEU science objectives, and the striking gap in resolving power in the FIR/SMM that will be left in the wake of the next generation of telescopes. SAFIR will provide our first deep view of the sky at far-IR wavelengths that does not suffer the ill effects of extragalactic source confusion (multiple galaxies per resolution element), enabling detailed studies of the individual sources that give rise to the cosmic IR background.

SAFIR instruments would provide imaging and spectroscopic capabilities with maximum spectral resolution $\lambda/\Delta\lambda \sim 10^6$. To achieve the goal of natural background-limited performance, the SAFIR mirror will have to be cooled to about 4 K, and new generations of detectors, operating at about 0.05 K, will have to have NEP $<10^{-20}$ W Hz$^{-1/2}$. With such

extraordinary sensitivity SAFIR could readily detect spectral line emission and spectral features from gas and dust in galaxies at redshift $z \sim 4 - 5$, as illustrated in Figure 2, and probe the gas dynamics and chemistry in forming planetary systems. Figure 2 further shows the importance of improving sensitivity by 4 - 5 orders of magnitude beyond the capabilities of SIRTF, SOFIA, and Herschel. Japan's SPICA mission, which is planned to have a 4 K, 3.5 m diameter primary mirror, will take a huge step in this direction. SAFIR, with its larger mirror, will bring the most distant galaxies into range.
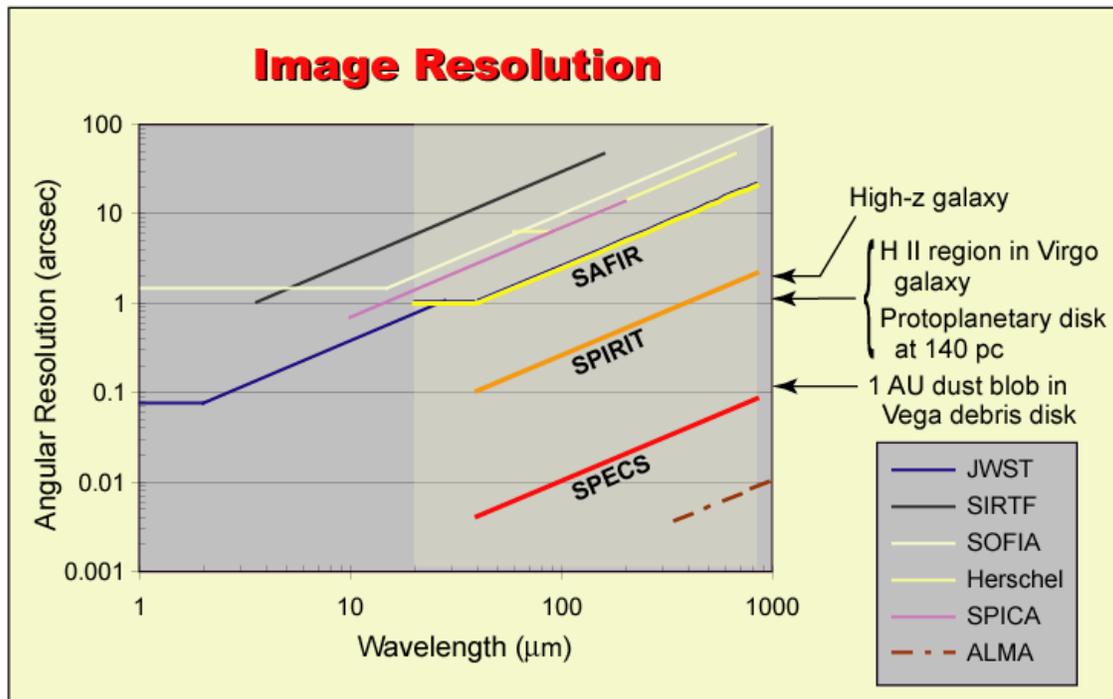
Figure 1. Vast improvements in angular resolution beyond those provided by the next-generation FIR/SMM missions will be needed to beat extragalactic source confusion, resolve the individual sources of interest, and achieve the science goals of NASA's Origins and SEU themes. Such improvements are also needed to align the FIR/SMM (shaded region) measurement capability with that available in the surrounding spectral regions, where JWST and ALMA will make complementary observations. ALMA provides complete spectral coverage at wavelengths longer than ~800 μm and observing capability into the submillimeter through atmospheric windows. With a 10 m aperture diameter SAFIR will take the first big step; interferometers like SPIRIT (assumed maximum baseline $b_{max} = 40$ m) and SPECS ($b_{max} = 1$ km) will be needed to provide the full resolution gain desired. SAFIR, SPIRIT, and SPECS are recommended in this plan.
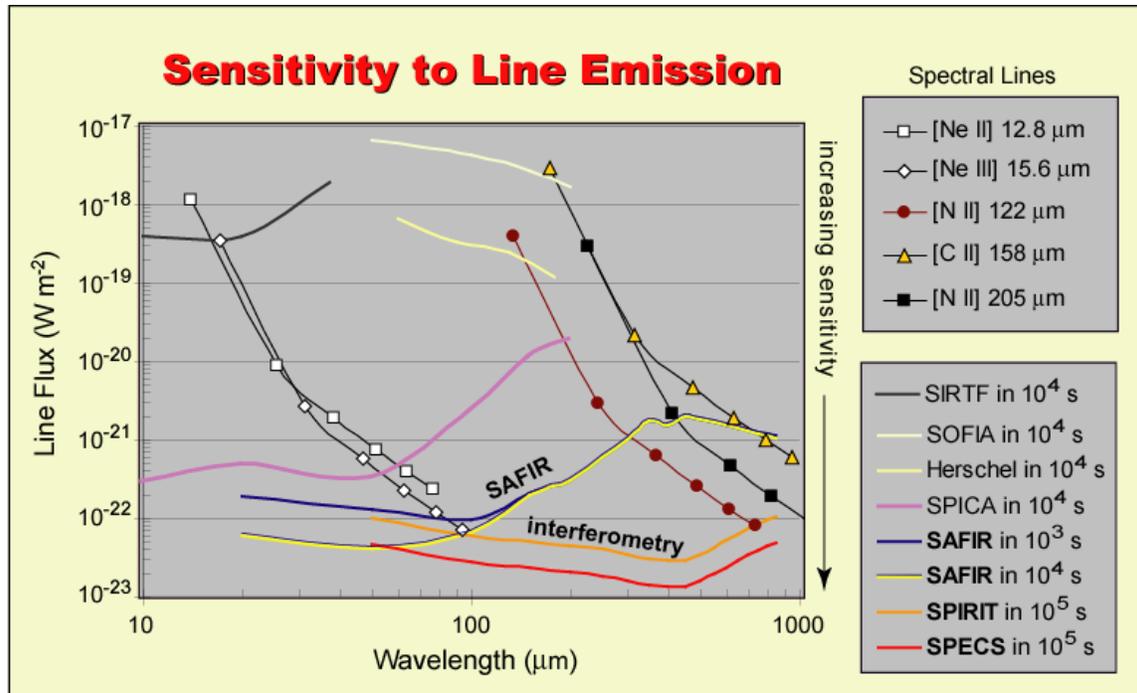
Figure 2. A spectrometer on SAFIR (with assumed $\lambda/\Delta\lambda = 10^3$) would be 4 to 5 orders of magnitude more sensitive than the corresponding instruments on SIRTF, SOFIA, and Herschel, and 1 to 2 orders of magnitude more sensitive than SPICA, enabling unprecedented studies of the star formation process and astrophysical conditions in distant, young galaxies. Estimated strengths of five important diagnostic and interstellar gas cooling lines are shown for a hypothetical "Milky Way" galaxy at redshifts of 0.1, 1, 2, 3, 4, and 5 (symbols along each curve, with redshift increasing from the upper left to the lower right). The rest wavelengths of the spectral lines are given in the inset. SAFIR could, for example, measure the [Ne II] and [Ne III] lines in "normal" galaxies out to z = 5 in modest exposure times. The relative intensity of these lines can be used to discriminate between AGN-dominated and star formation-dominated emission. Many galaxies are much more luminous than the Milky Way, making them even easier to see. At $\lambda > 200$ μm, SAFIR would reach the confusion noise "floor" in about 100 seconds[2]; longer exposure times would not help. However, because of their greater resolving power and still substantial total aperture areas, the interferometers SPIRIT and SPECS will break the confusion barrier and probe the universe to comparable depth (redshift z ~ 5) in the spectral lines that dominate the cooling of interstellar gas and allow the gas clouds to collapse and produce stars.

SAFIR will require investment in several technology areas. High-sensitivity far-IR detector technology, which is very promising but far from flight readiness, is one of the top priority items for SAFIR preparations. A total investment of the order of $50 M spread out over 10 years could produce large superconducting detector arrays with sensitivity one or two orders of magnitude beyond those now available, satisfying the performance goal. Even this budget is small compared with the investments made in

---

[2] We assume that the emission found in a single diffraction-limited beam cannot be reliably apportioned between multiple component sources. Therefore, we show the continuum confusion limit in a single $\lambda/1000$ spectral channel. The sensitivity would be less severely compromised by confusion if the redshift of each component were known from independent observations.

detectors at other wavelengths, but NASA is the only effective funding source in this area. Depending on the results and scientific opportunities developed from the Herschel mission, SAFIR might require coherent receivers that approach quantum-limited performance, as well as direct detectors.

Sensitivity in the FIR/SMM depends strongly on the temperature of the telescope, and it is imperative that improvements in detector technology be matched by efforts to cool large mirrors to very low temperatures. Because the optical tolerances on the SAFIR mirror are greatly relaxed relative to the JWST requirements, the possible use of precision machined, replicated mirror segments should be explored. Replication has the potential to speed fabrication and reduce cost. Based on JWST experience, we conclude that such a mirror study would cost $25M over three years, and the mirror technology could reach maturity (TRL 6) in a total of six years. The Advanced Cryocooler Technology Demonstration will provide a small TRL 5 cooler in 2005, but a much more powerful cooler will be needed to cool the SAFIR mirror. A development program for new cryocooler systems would cost about $30 M. Other projects, like JWST and TPF, and other government agencies (DoD and NRO) are already providing technology development funds for advanced coolers and deployable mirrors.

Whereas a JWST-like configuration flying at 3 – 4 AU from the Sun was shown in the Decadal Survey report, better thermal and sunshield engineering will permit SAFIR to operate at the JWST L_2 orbit and provide a larger aperture with the same launch vehicle. Other configurations should also be explored (Fig. 3).
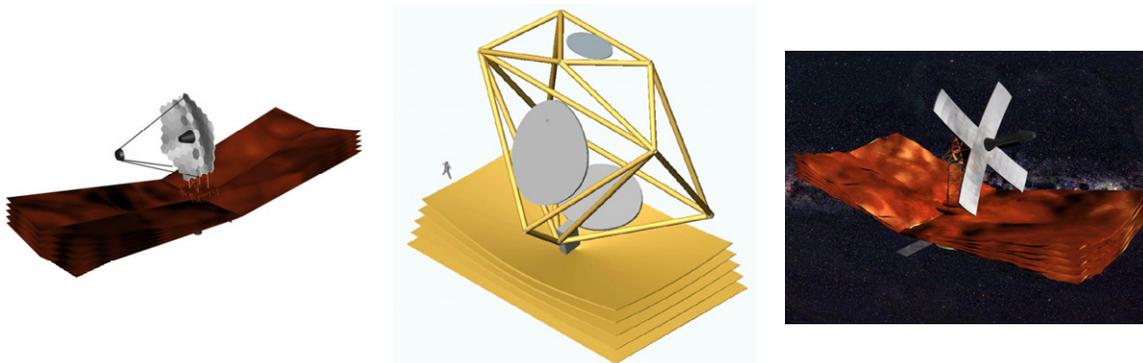
Figure 3. Further study of SAFIR mission designs will be needed before a single approach that accomplishes the highest priority science goals with ready technology, subject to programmatic considerations, can be selected. Three possible concepts are shown here: (left) based on JWST for maximum heritage and fidelity; (middle) based on stretched membrane mirrors to reduce aereal mass; and (right) based on a sparse aperture telescope to improve angular resolution.

Depending on the progress made with other planned instruments, a FIR/SMM all-sky survey mission with a 2 m class cryogenic telescope might be a scientifically compelling precursor to SAFIR. At other wavelengths, survey missions have greatly enhanced the science return of successive observatories. In the far-IR, the last sky survey, IRAS, was so insensitive that SAFIR will be blinded by every IRAS source. Given the immense discovery potential in this relatively unexplored spectral region, it is reasonable

to think that a deep FIR/SMM survey will be needed, and to plan accordingly. It is presently believed that there are many submillimeter-luminous galaxies at very high redshifts that do not have noteworthy optical counterparts, but which make up a substantial fraction of the total luminosity of the post-recombination universe. If these sources are not identified by SIRTF, Astro-F, NGSS, SOFIA, Herschel, or SPICA - because of their limited sensitivity, shorter wavelength range, or smaller survey area - then a FIR/SMM survey should be undertaken, and we believe that such a mission would be a strong competitor in the Discovery class. A sensitive FIR/SMM sky survey would answer important questions about the evolution of galaxies and would find the rare but important objects that act as signposts to the early universe, providing a rich database to build on with SAFIR and other missions.

Second Step: SPECS

The imaging sensitivity of SAFIR will be limited by the overlap of distant galaxy images at wavelengths greater than about 100 μm because its angular resolution will be ~ 3 arcsec, comparable to that of binoculars and to the spacing of galaxies seen in the Hubble Deep Field. A very large increase in angular resolution would be possible with an interferometer (Fig. 1). The scientific motivation and concepts for two FIR/SMM interferometry missions are presented in the white paper *Probing the Invisible Universe* (Appendix C). **A commendable long-range goal is to achieve HST-class resolution in the FIR/SMM, which would require an imaging interferometer with a 1 km maximum baseline**. Although this is not currently technically feasible, it is less challenging than X-ray or optical interferometry in space because the FIR/SMM wavelengths are much longer, and is comparable in overall difficulty to the other interferometry missions deemed meritorious by the astronomical community and included in the NASA Roadmap. We outline below the technical steps that will be required to build the km-baseline interferometer we call the Submillimeter Probe of the Evolution of Cosmic Structure (SPECS).

A FIR/SMM interferometer with superconducting detectors, cold mirrors, and a total light collecting aperture in the tens of m^2 would provide ample sensitivity. To derive the sensitivity curve shown for SPECS in Figure 2, the interferometer was assumed to combine the light collected with three 4 m diameter mirrors. SPECS is a natural successor to SAFIR, as it would employ the same detector, mirror, and cooler technologies, although the interferometer mirrors would be smaller than the SAFIR mirror and could be monolithic.

Three additional technologies or techniques will be needed to enable long-baseline imaging interferometry: a long-stroke cryogenic delay line; highly-reconfigurable formation flying; and the technique of wide-field imaging interferometry. Initial funding for these technologies was awarded through competitive programs during the past several years, and continued support for these mission-enabling technologies is needed. High-risk technologies, such as tethered formation flying, should be demonstrated in space with inexpensive nanosats.

Following a recommendation made by the attendees of the first community workshop in this series, "Submillimeter Space Astronomy in the Next Millennium," which took place in February 1999, concepts were developed for a FIR/SMM interferometer on a boom. We call this the Space Infrared Interferometric Telescope (SPIRIT). Figures 1 and 2 show that such a telescope would have very powerful measurement capabilities. A suitable goal for SPIRIT would be a system with a 40 m span, which provides the same angular resolution as a filled aperture telescope nearly twice this diameter and matches the JWST's resolution, but at a wavelength 10 times longer (see Fig. 1). Such a structure is within reach of the expected technology base 5 to 10 years from now, as much larger booms are already in use in space. Good metrology and active control are required, but diffraction limited performance at 40 μm is quite easy to achieve relative to the performance required for an optical interferometer, and SIM is paving this path. The apertures should be as large and as cold as can be afforded, but for this step it is clear that angular resolution is more important than sensitivity. To estimate the sensitivity shown in Figure 2 we assume that SPIRIT has two 3 m diameter light collecting mirrors, and the mirror temperature is 4 K. **If the cost of SPIRIT is much less than that of a "roadmap mission" like SPECS, as preliminary studies indicate, then SPIRIT should precede SPECS. We recommend that FIR/SMM interferometry mission concept studies be continued.**

As illustrated in Figure 4, the new technology requirements for FIR/SMM interferometry largely overlap the requirements for SAFIR.
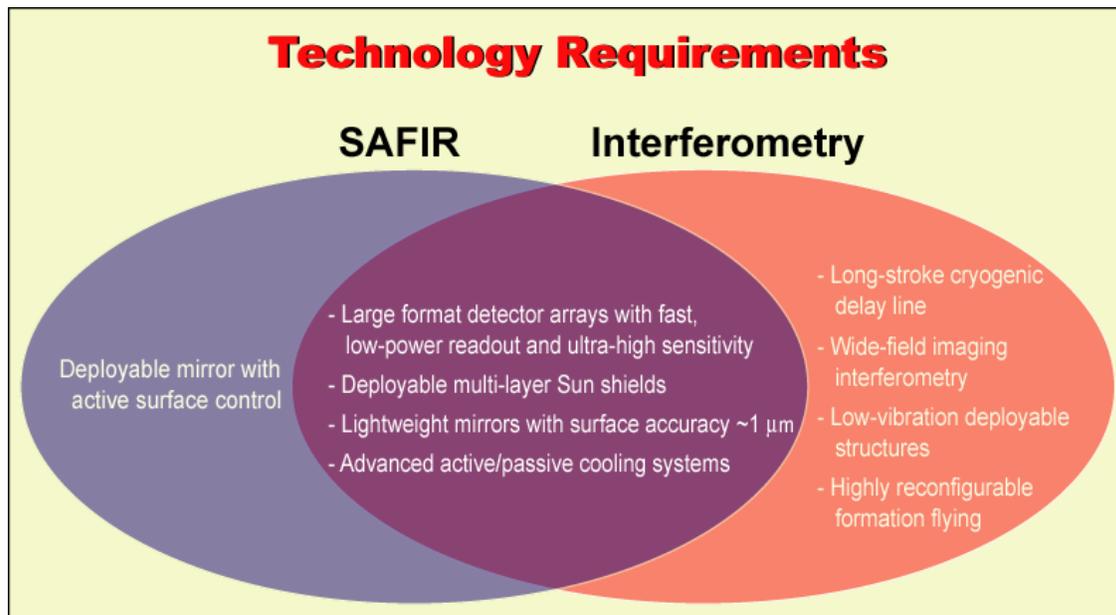
Figure 4. Investments in SAFIR technology will go a long way toward enabling FIR/SMM interferometry, and some of the interferometry technologies will be needed for LISA, SIM, and possibly TPF. New detectors and large mirror cooling technologies are particularly critical for FIR/SMM space astronomy, but a coordinated technology program would invest early in all the long lead time technologies shown here.

Supporting Research, Technology, and Missions

Several years will pass before it becomes necessary to choose design details for SAFIR and SPECS, and some decisions, such as the option to fly a far-IR sky survey mission, and the relative timing of the SPIRIT and SAFIR missions, will depend on future scientific progress, on experience with JWST, ALMA, and ground-based optical interferometers, on the available technology, and on budgetary constraints. However, a relatively modest investment is needed to prepare. **We recommend a coordinated technology program** that provides support for the development of:

1. 10^4 pixel arrays of direct detectors with NEP ~ 10^{-20} W Hz$^{-1/2}$,
2. advanced, high-efficiency cryocoolers capable of providing ~100 mW of cooling power at 4 K for mirror cooling, and capable of cooling the detectors to ~50 mK,
3. low-cost, low areal density (<15 kg/m^2) mirror technology for the FIR/SMM in a mirror development program that includes
   a) demonstration of fabrication techniques,
   b) demonstration of cooling strategies,
   c) demonstration of wavefront sensing and control, and
   d) development of test procedures,
4. interferometry testbeds that can be used to develop procedures, algorithms, and control systems,
5. broadband tunable coherent THz array receivers that approach quantum-limited performance,
6. long-stroke cryogenic delay lines and compact spectrometer technology for broadband FIR/SMM spectroscopy and wide-field imaging using direct detectors,
7. low-vibration deployable structures, and
8. highly-reconfigurable formation flying to enable interferometric (u, v) plane filling, and therefore high contrast imaging.

Successful development of the technologies listed above would serve many NASA purposes. All of the FIR/SMM missions demand new detector development, and there is a strong overlap of technology with some types of X-ray and UV detectors. For example, Con-X needs superconducting detectors and cryo-coolers. LISA, like SPECS, requires formation flying. Some of the interferometry technologies are shared with other planned NASA missions, such as SIM and the interferometer version of TPF, although SIM and TPF require much better precision to work at shorter wavelengths and make astrometric measurements or null out starlight.

We recommend that NASA support initial concept studies for SAFIR, SPIRIT, SPECS, and a sensitive FIR/SMM all-sky survey mission. SPIRIT and the sky survey mission would cost less than SAFIR or SPECS, but more than the current MIDEX cap, and would therefore require either $400M-class Space Science mission opportunities or collaboration with partners to bring down the NASA cost. NASA has been asked to collaborate on the planned Japanese SPICA mission, which would provide a 3.5 m class cold far-IR telescope at the Lagrange point L$_2$. NASA would benefit from the opportunity to deploy new generations of far-IR detectors and instruments in space. New detector technology can be tested and used first on SOFIA, but SPICA may be the nearest term opportunity to use next-generation far-IR detectors in space.

Finally, **we recommend that NASA be receptive to proposals for laboratory and theoretical astrophysics related to far-IR studies**. Relatively little is known about the chemical reactions that form large interstellar molecules or dust, or their role in the physical processes that govern star and planet formation. As new species and phenomena are recognized from observations, it will be very important to interpret them accurately. It is also clear that simulations are critically important in establishing the right observing strategies and instrument requirements.

European and Japanese astronomers are eager to collaborate with their US colleagues on SAFIR and FIR/SMM interferometry. The European community will desire opportunities to follow up the Herschel mission, and the Japanese community to follow up the SPICA mission.

Conclusion

Information needed to answer some of the most compelling astrophysical questions is uniquely available in the FIR/SMM spectral region. The time is right to place SAFIR on the NASA plan as one of the successors of SIRTF and JWST, to set our sights on a long-baseline FIR/SMM interferometric imaging telescope, to further develop FIR/SMM single-aperture and interferometric mission concepts, and to invest strategically in the technology that will enable future FIR/SMM missions. Supporting studies and smaller mission opportunities should be actively pursued.

Appendix A. Acronyms

AGN – active galactic nucleus
ALMA – Atacama Large Millimeter Array
Astro-F – Japanese Infrared Imaging Surveyor
Con-X – Constellation X
ESA – European Space Agency
FIR/SMM – the far-infrared and submillimeter wavelength range from ~25 – 800 μm
IRAS – Infrared Astronomical Satellite
ISO – Infrared Space Observatory
JWST – James Webb Space Telescope, formerly NGST
LISA – Laser Interferometer Space Antenna
MIDEX – NASA's Medium-class Explorer program
NEP – Noise Equivalent Power
NGSS – Next Generation Sky Survey, a MIDEX mission
NGST – Next Generation Space Telescope, now called JWST
SAFIR – Single Aperture Far-Infrared Telescope
SEU – NASA's Structure and Evolution of the Universe theme
SIM – Space Interferometry Mission
SIRTF – Space Infrared Telescope Facility
SOFIA – Stratospheric Observatory for Infrared Astronomy
SPECS – Submillimeter Probe of the Evolution of Cosmic Structure
SPICA – Space Infrared Telescope for Cosmology and Astrophysics, formerly HII/L2
SPIRIT – Space Infrared Interferometric Telescope
SWAS – Submillimeter Wave Astronomy Satellite
TPF – Terrestrial Planet Finder
TRL – technology readiness level

Appendix B. *Charting the Winds that Change the Universe, II. The Single Aperture Far Infrared Observatory (SAFIR)*

This document can be found on pages 157-166 of these proceedings.

Appendix C. *Probing the Invisible Universe: The Case for Far-IR/Submillimeter Interferometry*

This document can be found on pages 167-177 of these proceedings.

Session I

Strategic Planning

Astronomy & Physics Division

NASA Space Sciences Strategic Planning

Philippe Crane
ORIGINS THEME SCIENTIST

NASA / HQ

NASA Strategic Planning

- NASA Strategic Planning, Roadmaps, GPRA, and all that.
- Schedule, People, Origins/SEU
- Issues:
  - Decadal report priorities
  - Breadth of the Research
  - Content

NASA ADVISORY STRUCTURE

NASA Advisory Council

Advisory Committees:

International Space Station	Aero-Space Technology	Earth System Science and Applications	**Space Science (SScAC)**	Minority Business Resource	Life & Microgravity Sciences & Applications	Technology and Commercialization

Subcommittees:

Sun-Earth Connections

Astronomical Search for Origins

Structure and Evolution of the Universe

Solar System Exploration

The Strategic Planning Process

National Academy
Fundamental Science Questions

Roadmap Teams

NASA Strategic Plan

Space Science Roadmaps
Long Range Program Alternatives

Community Strategic Planning

SSE Strategic Plan
Science Questions/Science Goals

President's Budget
Congressional Actions

Budget & Guidance

SSAC & Subcmtes

SSE Performance Plan & Metrics

External Community Assessment

SSE Performance Report

4

Purpose

- Fulfill the strategic planning requirements.
- Provide a guide to the science community in presenting research requests to NASA.
- Inform and inspire.
- Focus investments in technology and research for future missions.
- Provide the scientific and technical justification for augmentation requests.

Relevant Time Scales

- GPRA is an annual event, so the Roadmaps feed this exercise through the Enterprise Strategic Plan.
- Strategic Planning occurs on a 3 year cycle where the near term outlook is for 5 years.
- Roadmapping also is on a 3 year cycle, but looks out 10 years and 20 years.
  - For Origins, the next 5-8 years is well defined.
  - For SEU, funding beyond GLAST for LISA, and Con-X is uncertain.

Activity Name	Start Date	Finish Date	2001	2002
			Jul Aug Sept Oct Nov Dec	Jan Feb Mar Apr May Jun Jul Aug Sept Oct Nov Dec
Kickoff Telecon to set workplan	8/15/01			
Identify/recruit committee members	8/15/01	12/1/01		
Prepare Guidelines & Distribute	8/15/01	11/10/01		
Prepare Plan for OS Meeting	8/28/01	11/19/01		
Prepare synthesize/issues	11/20/01	12/1/01		
Roadmap Committee Workshop	2/5/02	2/6/02		
Roadmap Committee				
Write/Submit 1st Draft	12/10/01	3/2/02		
Prepare synthesis for OS Review	3/3/02	3/16/02		
Review 1st draft material	3/10/02	3/11/02		
Hear selections from NRA	2/11/02	2/11/02		
Write 2nd Draft	3/25/02	6/7/02		
Prep for OS/SEUS Meeting	6/10/02	6/14/02		
Briefing on NRA Concept Studies	6/10/02	6/11/02		
Write Final Draft	6/24/02	8/16/02		
Prep for OS/SEUS Meeting	8/19/02	8/30/02		
Final Publication	9/15/02	11/1/02		
Roadmap Update Content Delivered	9/15/02			
Code S Strategic Planning Workshop	11/15/02			
OS/SEUS Meetings				
OS/SEUS Meeting	12/3/01			
OS/SEUS Meeting	3/19/02	3/24/02		
OS/SEUS Meeting	6/18/02			
OS/SEUS Meeting	9/2/02	9/7/02		
Mission/Technology concept studies	1/7/02	7/15/02		
TPF Architecture Downselect to two	4/1/02			
			Jul Aug Sept Oct Nov Dec	Jan Feb Mar Apr May Jun Jul Aug Sept Oct Nov Dec

Some Related Activities

- March: **2nd Workshop on New Concepts in IR Submm Astronomy--Univ of Maryland**
- April: Hubble Legacy Workshop--Chicago
- April: Astro-biology Conference--Ames
- May: Laboratory Astrophysics Workshop--Ames
- May: Astrophysics of Life--STScI
- May: Brown Dwarfs--Kona
- May: Origins 2002--Grand Tetons
- June: International Virtual Observatory--Munich
- June: Research in Extra-solar planets--Washington

ASTRONOMY & PHYSICS DIVISION ROADMAPPING ACTIVITIES

- **ORIGINS** Roadmapping basically an update and redirection of the 2000 Roadmap

  - Led by the Origins Subcommittee.

  - Community input through individuals selected for expertise and from meetings like this.

- **SEU** is preparing both a Roadmap, and a new initiative

  - Group of 12 people led by Sterl Phinney.

  - Aimed at generating a new initiative for LISA, Con-X, and several selected smaller missions.

  - Community Input from White Papers and meetings.

ORIGINS Roadmap Organization

1 Co-Chairs

Phil Crane (HQ)

Alan Dressler (Carnegie)

2 Science Leads

Alan Dressler (Carnegie) – Lead

Lou Allamandola (ARC)

Adam Burrows (U of A)

3 Technology/Instruments

Rich Capps (JPL) – Lead

4 Research & Analysis

Hashima Hasan (HQ) – Lead

5 Astrobiology

Carl Pilcher (HQ) – Lead

6 Missions

Mike Devirian (JPL) – Lead

7 Outreach

Carl Pilcher (HQ) – Lead

SEU Roadmap Team

Roadmap team:

Sterl Phinney* (Chair)

Sean Carroll	Sarah Church	Roy Gould
Craig Hogan	Steve Kahn	Dan Lester*
Robert March	Mike Shull	Simon Swordy*

Nick White*

Rocky Kolb* (SEUS Chair)

Paul Hertz   (SEU Theme Scientist)

Paul DeMinco   (SEU Program Integration Manager)

ROADMAPPING Cont'd

- ORIGINS Roadmap will be very similar to the 2000 version
  - Origins has a funded program.
    - HST, SIRTF, SIM, NGST, SOFIA, KEPLER, TPF.
    - New mission content in the next several years from Explorer, Discovery, or possibly from a new competed mission line.
  - Future strategic missions must define their science goals and technology needs.
    - Large filled aperture IR mission
    - UV/Optical mission in the 2020 time frame

Structure and Evolution of the Universe 2003 Roadmap

The SEU Roadmap Team solicited community input in the form of white papers describing mission concepts

Category 1 - Missions

- **Advanced Compton Telescope (ACT)**
- **Constellation-X**
- **Cosmic Microwave Background Polarization Experiment (CMBPOL)**
- **3a.   Cosmic Microwave Background Polarization Experiment (CMBPOL) Addendum**
- **Energetic X-ray Imaging Survey Telescope (EXIST)**
- **Fresnel Microarcsecond Gamma Ray Imager**
- **Generation-X**
- **Gravitational Echoes Across Time Mission (GREAT)**
- **High-resolution Spectroscopic Imaging Mission (HSI)**
- **International Advanced Radio Interferometry between Space and Earth (iARISE)**
- **Laser Interferometer Space Antenna (LISA)**
- **MicroArcsecond X-ray Imaging Mission (MAXIM)**
- **Next Generation High-Energy Gamma-Ray Astrophysics Mission**
- **Orbiting Wide-angle Light-collectors (OWL)**
- **Probing the Invisible Universe: The Case for Far-IR/Submillimeter Interferometry**
- **Single Aperture Far InfraRed Observatory (SAFIR)**
- **Space UltraViolet/Optical Observatory (SUVO)**
- **The Stellar Imager (SI)**
- **A Facility Far-Infrared Spectrometer for SOFIA**
- **Supernova/Acceleration Probe (SNAP)**
- **Tests of Relativistic Gravity via Solar System Laser Ranging**
- **Ultra-High-Throughput X-ray Telescope Observatory (UXT)**

Category 2 - Non-Mission Activities

- **"Amicus Brief"**
- **Balloon Program**
- **Bridging the Gap From New Instruments and Data to New Science and Understanding**
- **The DART System for Far-IR/Submillimeter Space Missions**
- **Theory of Rotation in Big Bang Universe**

Roadmap Issues

- Decadal Report Priorities and Requests
  - Specific Missions & Priorities
  - Competed Missions-- Call for medium size mission like Discovery
  - Other Issues-- Theory, NVO, Lab-Astro
- Origins/SEU Priorities
  - Astronomy & Physics Div. needs a funded SEU line for LISA, Con-X and for medium size competed missions
  - How to accommodate new content in the next several years

Roadmap Guidelines

Reviews and Recommendations that guide the Roadmap priorities:
- NAS Committee on Gravitational Physics (1999)
- OSS/SEU 2000 Roadmap
- NAS Physics Survey Overview Committee (2001)
- NAS Astronomy and Astrophysics Survey Committee (2001)
- NAS Committee on Physics of the Universe (2002)

Community Perspective on NASA Strategic Planning

Dan Lester
Department of Astronomy and
McDonald Observatory C1400
University of Texas, Austin TX 78712

The process that leads from a mission concept to flight hardware in space has many dimensions, involving the astronomical community, the Administration, the Congress, and even the private sector. The development of coherent space science policy and the strategic planning needed to follow through with it is a crucial factor in the success of such endeavors. In this contribution, I will try to outline some of these efforts as they apply to space infrared and submillimeter missions, report on status as it relates to our discipline, and share some thoughts about where we need to go.

1.0 Building NASA Space Science Policy – Some Background

To many in our community, especially the younger ones among us with fewer bullet holes in them, the development of space science missions is a matter of community zeitgeist. When the community really feels that the time is ripe for a mission, strings are pulled, and money just flows. In fact, the process is a little more involved than this, and history shows that it requires a level of contribution from the research community that extends well beyond straightforward science justification.

The infrared and submillimeter spectral regimes have been blessed, in the last decade, with a hearty zeitgeist indeed. While I have taken the liberty of being wavelength specific here, our work has clearly blossomed to involve astronomers of all photonic persuasions. There is probably no other spectral regime that can claim the discovery potential of ours, in which foreseeable technology improvements will provide us with many orders of magnitude improvement over the tools that we now use. Some of the most compelling major problems (galaxy seeding and formation, and the astronomy of life) in astrophysics appear to have answers in this spectral regime. We have a Great Observatory that is about to be launched, and a major airborne platform in the works that will drive technological creativity for decades. The natural advantages of space platforms (potential for low temperatures and concomitant low background) are superbly suited to extrapolation of present efforts.

The most recent Decade Report by the Astronomy Survey Committee points the way most persuasively. I want to emphasize that these reports are profoundly important tools for space science marketing efforts, in that they are not only documents that stand for community consensus, but make hard choices through real priority lists. The importance of hard choices and conspicuous concessions in curiosity-driven science cannot be overstated, and these documents, developed with inputs from a substantial fraction of the community over nearly half a century have engendered a favorable reputation for our

community on the Hill and in the Executive Branch. Shopping-list driven science might work well for NIH, but it will not work well for us. It is important for us to stand behind these documents as we move ahead.

While the Decade Report lays out a clear path to scientific excellence over the next ten years, the community must look further ahead, in order to lay the scientific and technological groundwork for missions whose readiness and success potential cannot be easily evaluated. For space science, NASA tackles this job as part of the triennial, and agency-wide strategic planning effort. The resulting plan for the Office of Space Science (OSS) itself culminates more than a year of effort by the Space Science Advisory Committee (SScAC), and the individual roadmapping efforts of its associated subcommittees. At this writing, subcommittees of scientists representing the two scientific "themes" in the Astronomy and Physics division of OSS have begun deliberations on a new twenty five year roadmap for these themes – "Astronomical Search for Origins", and "Structure and Evolution of the Universe" (SEU). These roadmaps will be digested by the SScAC later this year, and contribute to an integrated resulting 2003 OSS Strategic Plan. While such long range planning in a document with a few year lifetime might appear misguided, especially in view of the annual cycle of budget proposals for the agency and outyear projections in these proposals that apply only to the next five years, this activity forces a continual reevaluation of major scientific goals and technical thrusts. At any given time, the agency can refer to a single clear (though arguably unique) path leading to future success.

While there is no formal linkage between the Decade Report and the OSS Strategic Plan, it has always been clear that these documents have to talk to one another, and underscore a consistent picture. The OSS Strategic Plan takes off from the Decade Report in that the former is a specific touchstone in agency budget proposal preparations, and is therefore perhaps more aware of NASA-specific management and technology readiness issues than the latter. Because of its pedigree and breadth of community involvement, the latter has been termed a blueprint for decision makers in the Congress and the Executive Branch. It is the match between these that drives policy. The relationship of these documents has some similarity to that between congressional authorization and appropriation bills.

The OSS "theme" structure replaced the former Astrophysics Division almost a decade ago as an organizational construct. It helped wean space science planning efforts from the wavelength-specific cliques under which they were born. These themes encouraged the development of roadmaps that were science-driven, rather than technology- or wavelength-specific. Again, these themes garnered significant respect in the Administration and on the Hill, and were powerful marketing tools for broad mission lines. The theme structure was not, however, a particularly powerful management tool, as the lines of budget authority were functionally distinct from those guiding the construction of the strategic plans.

2.0 The Challenge of Marketing Space Science and the Articulation of Value

While at one time popularly derided by the science community, the marketing aspects of space science are increasingly important. These include not just popular outreach, but focused lobbying or advocacy to the leaders of our nation for whom what we do is very much on the far horizon of their perspective.

While it is widely agreed that astrophysics is one of the more intellectually compelling pursuits of modern science, our community has not been particularly successful in articulating the value of this curiosity-driven work to the nation in general, and to the Congress in particular. This sentiment has been explicitly stated by not only the new NASA Administrator Mr. O'Keefe, but also by Dr. Marburger, the new Presidential Science Advisor and OSTP Director. The policy milieu in which we now work matured under the federal role for basic science crafted by Vannevar Bush at the end of World War II. As federal funding for basic science requires larger and larger dollar amounts, this paradigm of curiosity for the sake of curiosity is seen as being less defensible, especially surrounded by the ashes of the cold war.

As Dr. Marburger put rather delicately to the American Astronomical Society, *"the technology that makes [our] quest so exciting today has created unprecedented opportunities for nearly every other field of science"*. About a month later, in his address to the AAAS, Dr. Marburger was more pointed – *"Today the frontiers of the large and the small – of astronomy and particle physics remain unconquered. But they have receded so far from the world of human action that the details of their phenomena are no longer very relevant to practical affairs."* He went on to say that *"for the most part, the great accelerators and observatories have been well built and well managed … but the greatest opportunities of science today are not to be found at these remote frontiers."*

While Dr. Marburger is a leading scientist, and as passionate a devotee of basic science as we are, he is serving notice from the Administration that our message is not getting out. The Value (and I like to capitalize this word to emphasize the dollar-value trade space that we're working in) of what we do is not being effectively articulated to the folks who pay for it. A recent Brookings Institution survey of economists and political scientists found that "promoting space exploration" was one of the least important priorities for the federal government in the future. While one might argue that such professionals are simply not smart enough to get the picture, I would just point out that these are the folks who build federal policy, and often go on to public office to carry it out. Their problem is very much our problem.

These sentiments, in which the rationale for basic science is being reshaped as a performance-driven, results-oriented task, first appeared in the GPRA efforts of the last decade, from the last Administration, and are now becoming formally adopted by this Administration. Note that Mr. O'Keefe is one of the authors of the President's Management Agenda, a document which came out of his tenure at OMB, and every indication is that he will run NASA by these rules. In fact, as principal witness in the

annual NASA posture hearing before the House Science Committee this month, the importance of this document to his outlook for NASA were among the first words that came out of his mouth. In this document, you will find the following directives – *"The Administration expects that these investment criteria will better focus the government research programs on performance … Basic research programs will better target improving the quality and relevance of their research."* You'll also find the following words that hit close to home – *"The federal government will spend approximately $90 billion in 2001 on R&D, an investment representing 14 percent of all discretionary spending. The ultimate goals of this research need to be clear. For instance, the objective of NASA's space science program is to 'chart our destiny in the solar system', and the goal of the US Geological Survey is to 'provide science for a changing world'. Vague goals lead to perpetual programs achieving poor results."*

In the aftermath of September 11, one might be pressed to try to validate space science missions in the context of homeland security. Such a direct connection is, of course, ludicrous. But to those who would point at the premiere marketing successes of astronomy and space science – the scientific engagement of the public, and sparking wonder and curiosity in a nation – I would say that the case must be made that such accomplishments do have national Value, and the economic details just need to be laid out. Is it important that we are a nation that tries to do hard things? Is it a national priority for us to be explorers and learn to think like ones? Is it important to be curious about the world in which we live? I'm comfortable saying that homeland security is dependent on citizens having these skills and this kind of perspective. So would Vannevar Bush, whose first recommendation in his 1945 report was that *"Science, by itself, provides no panacea for individual, social, and economic ills. It can be effective in the national welfare only as a member of a team, whether the conditions be peace or war. But without scientific progress no amount of achievement in other directions can insure our health, prosperity, and security as a nation in the modern world."*

A scientifically literate body of taxpayers is, of course, one of the most powerful ingredients in a successful marketing thrust, and our nation has yet to come to grips with the educational deficiencies that leave us without one. NSF and NASA have, using astronomy in public and educational outreach, made great strides in contributing to K-12 science education. But of course the issue is much deeper than education about astronomy per se. It's about building a citizenry that is curious about the world around them, and for whom basic research has Value in satisfying that curiosity. It is a humbling thought that as our roadmaps reach out at least a generation into the future, the kids of this generation are going to be the taxpayers who we will ask to pay for the missions that are actually in our roadmaps! The role of educational outreach in a NASA strategic planning context is intimately connected with the likelihood of our success for specific vision missions like SPECS!

Over the last decade the role of EPO in roadmapping efforts has seen some remarkable, and gratifying changes. It used to be that EPO was grudgingly acknowledged by policy-developing members of our community. The EPO section of the roadmaps and strategic plan was, somewhat to the relief of the scientists who wanted to keep it at arms length,

simply tacked inelegantly onto an otherwise well integrated document. This has changed dramatically. EPO is now not only integrated into the effort, but with strong support by the scientists involved who have come to recognize the importance of this to our future, and understand that such words make for a more approachable document to the leaders who need to appreciate it.

3.0 Far IR and Submillimeter in Current NASA Roadmapping Efforts

Within the Astronomy and Physics (A&P) Division in OSS, the SEU and Origins themes have been working hard to develop the roadmaps that will feed the 2003 version of the strategic plan. In the case of Origins, the result will be to build on a successful mission line (NGST, SIM, TPF) that has been formally adopted as an agency commitment, and which has been approved as such by the Congress for out year funding. The science-focused mission-line approach to funding, rather than a scientifically disjoint mission-by-mission approach to funding (e.g. the Great Observatory series of missions, which included HST, CGRO, Chandra, and SIRTF) has major advantages to OSS and the community in level-budget management programmatics and opportunities for long range technology development as well as a higher profile in the public and Congressional eye. Such mission lines have also been established in the other two OSS themes – the "Living with a Star" initiative in the Sun-Earth Connection theme, and the Mars initiative in the Solar System Exploration theme. The SEU is the only OSS theme that has yet to develop a similarly successful mission-line initiative, and this is a major goal of the current SEU roadmap team.

In the context of these roadmapping efforts, new far infrared and submillimeter mission concepts have, however, found themselves awkwardly planted. While these missions, which include the filled aperture space observatory concepts SAFIR and FAIR as well as the space interferometers SPIRIT and SPECS, have fundamental roles to play in what is considered to be Origins science (which has been defined by that theme as the origin of planets, stars and galaxies in our universe), the present Origins mission line will not be completed until, at best, the middle of the next decade. While these missions also have fundamental roles to play in what is considered to be SEU science (which includes the cycles of matter in the universe, and the structure of the universe up to the formation of galaxies out of cosmic structure), recent efforts to develop a mission line in that theme (e.g. the "Cosmic Journeys "draft initiative from the last roadmap cycle) have been more oriented toward fundamental physics and the earliest history of the universe (z>10), where these infrared missions are less relevant than CMB, high energy, and gravity wave telescopes. There is no evidence that this orientation will change soon. While the SEU theme has historically taken responsibility for several of what we would consider our missions such as COBE, SWAS, and the US part of Herschel and Planck, it has not taken on full thematic responsibility for any large infrared mission.

While the early efforts by the roadmapping teams are only just now being vetted by the SScAC subcommittees that they report to, it is clear that far infrared and submillimeter space missions of the future will have a challenge in finding a near term place in the

queue. That at least a single aperture far infrared observatory concept was considered by the Decade panel to be deserving of a new start toward the end of the decade cannot be assumed to have established for it this place in the queue. The position of these missions in the theme roadmaps will be an important input to management decisions about technology investments, whether through focused programs or more broadly aimed R&A programs. Some of the current focused programs will address many of our needs. For example, development of large apertures in space is part of the NGST effort, and the foundations of space interferometry will be built by SIM. Should our vision needs include formation flying, experience from ST3 will be invaluable. But aside from some obvious opportunities on SOFIA, and near term efforts on FIRST/Planck, far infrared– and submillimeter– specific investments will require a clear path towards future missions for which we don't yet have an agency commitment.

4.0 Looking Ahead

What does the future hold for space science policy planning and community input? Several changes loom large. First of all, in response to an FY02 summary budget document from this Administration expressing concerns about optimal management of ground- and space-based observatory investments, the NSF and NASA asked the NRC to do a prompt and broadly based assessment of the organizational effectiveness of the whole federally-funded astronomy and astrophysics enterprise. The committee that was formed to do this by the NRC, in a unanimous set of recommendations, strongly rejected any effort to consolidate astronomy and astrophysics under a single agency, a recommendation that has important ramifications for submillimeter astronomy. Of similar importance to submillimeter efforts is their recommendation that an OSTP board be constituted to annually review an integrated astronomy and astrophysics strategic plan for at least these two agencies. This Interagency Astronomy and Astrophysics Planning Board (IAAPB) is now being formed and, when activated, will provide short range planning that the Decade Reports cannot. We can be sure that space infrared and submillimeter efforts will as a result, in the eyes of the Executive Branch and Congress, more clearly complement the ground based research that parallels it.

Of some special significance to the SEU community is the imminent release of the final report from the NRC Committee on the Physics of the Universe. The CPU was formed from the physics and astronomy communities at the request of former NASA Administrator Goldin to examine the roles that NASA should play in scientific research at the intersection of physics and astronomy. This report will, much like the Decade Report, identify scientific problems that deserve NASA resources, and spell out a prioritized suite of missions that would constitute the best attack on them. This report will, by its charter, focus on basic physics in the universe such as dark energy and black holes. By the breadth of membership of the CPU and the NRC's imprint on the document, it will represent community consensus with great authority.

Another potential change on the horizon is the NASA strategic planning picture itself. While the theme structure in OSS has been highly successful in breaking down

wavelength barriers in the pursuit of major scientific goals, the separation of astronomy and astrophysics into two themes (OSS and SEU) may have outlived its usefulness, especially as we see more examples of individual mission concepts that would contribute profoundly to both our understanding of the origin of the chemicals of and sites for life and the evolution of the universe. Such missions, of which our SAFIR and SPECS are among the awkwardly positioned examples noted above, are lately considered to have "fallen into the crack" between perceived Origins and SEU theme boundaries, to their programmatic disadvantage. I believe that a good management structure will see through this, and understand the overlap to be a major scientific, programmatic, and marketing advantage. Our community must convey the understanding that these missions should get strength, not weakness, from the breadth of science that they cover.

Several changes in the Astronomy and Physics program are being discussed by Origins and SEU that could lead to entirely new opportunities. Both themes recognize the importance of "cycles of matter " science to their thematic goals, and design of an entirely separate mission-line initiative may be called for. While this would clearly require a major source of new funds, which is not an easy option, the maturation of the field and the relative health of the OSS enterprise may make this timely. A fiscally similar but structurally distinct route might be to create a competed mission line that would provide for non-flagship Astronomy and Physics missions at a $300-$500M level. Such a mission line would be modeled on the successful Explorer and Discovery programs, except that the scientific focus would be predetermined, and specified in the RFP.

The future of far infrared and submillimeter space astronomy is, though by no means clearly roadmapped, blessed with a rich and powerful set of mission concepts. It is the responsibility of the community to articulate these blessings to the public and national leaders as well as to the rest of the science community. That the SAFIR mission has been given a priority nod from the Decade Committee for a new start at the end toward the decade is an important and enthusiastic endorsement that the community must build upon. The excellent words in the panel reports on far infrared interferometry can be considered similarly powerful as a policy vector pointing towards SPECS. Our scientific and technical cases for these missions should be debated, refined, and heard … loudly. This workshop and the words that come out of it will help lead the way.

I would like to thank Paul Hertz, Alan Bunner, and current and former co-members of the SEUS for recent opportunities to learn, and come to terms with some of these ideas. Work on the Steering Committee of the AAU Space Science Working Group has offered additional incentives for these perspectives. I would like to give special thanks to Harley Thronson who, over the last dozen years or so, has certainly heard and encouraged, if not contributed to, the evolution of these thoughts.
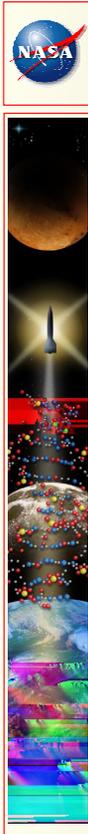
NASA Office of Space Science
Office of the Technology Director

Enabling Telescopes of the Future:
Long-Range Technology Investing

Dr. Harley Thronson
OSS Technology Director

1
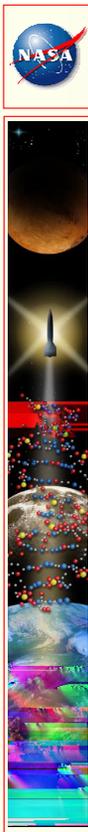
Snapshot of OSS: April, 2002

The Office of Space Science at NASA Headquarters has a current staff of about 60 professionals (aka, scientists, engineers, budget analysts) and an annual budget of $2.5 B out of NASA's $15.0 B.

About 35 missions or programs in various stages of development or operation are managed by OSS, notable among them are Hubble Space Telescope, Mars Global Surveyor, Mars 2001 Odyssey, Chandra X-ray Observatory, TRACE (solar observatory), Cassini (mission to Saturn), Galileo (mission at Jupiter), and Next Generation Space Telescope.

OSS has an annual technology budget of several hundred million dollars.

So, what is it that we are doing?

2

Office of Space Science
Selected Current/Near-Future Technology Priorities

In-Space Propulsion:
 initiated by Decade Planning Team (FY02)

Nuclear Power and Propulsion:
 initiated by Decade Planning Team (FY03)

Advanced Materials/Smart Materials:
 proposed by NASA Exploration Team as future initiative

Large Telescope Systems:
 proposed by A&P Division for future initiative

3

NASA CHARTER

• **To advance and communicate scientific knowledge and understanding of the Earth, the solar system, and the universe, and**
• **To advance human exploration, use, and development of space, and**
• **To research, develop, verify, and transfer, advanced aeronautics and space technologies.**

➡ *New agency "vision" to be presented on 14 April.*

Space Science Enterprise | Earth Science Enterprise | Biological and Physical Research Enterprise | Human Exploration and Development of Space Enterprise | Aerospace Technology Enterprise
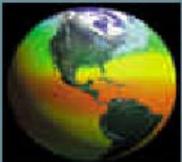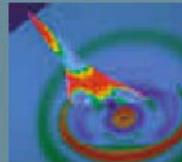
TECHNOLOGY PRINCIPLES
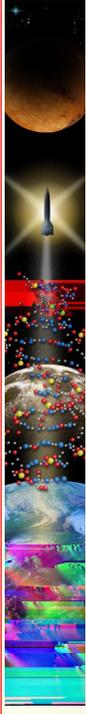
- Technology priorities are determined by science program requirements,

- Manage the technology program effectively, with milestones, deliverables . . .

- Implement space demonstrations of selected technologies:
        precursors and demos

- Use technologies in multiple missions and as "stepping stones"

- Promote partnerships with other Enterprises, agencies, industry, and academia

- Use open competition and external peer review wherever possible

5

NASA LONG-RANGE TECHOLOGY PLANNING PROCESS (ca. 2002)
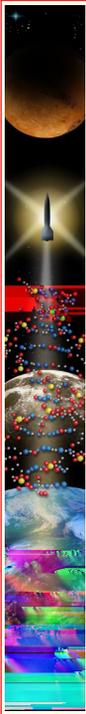
1. Science/Mission priorities determined within Divisions via Strategic Planning processes, plus Administrators new "vision"

2. [Mission concepts derived from #1]

3. Technology priorities derived from #2

4. [Evaluation of "gaps", priorities, . . . within OSS and OAT.]

5. New initiatives or re-programming, as appropriate

6

NASA FUTURE TECHOLOGY and MISSION PLANNING
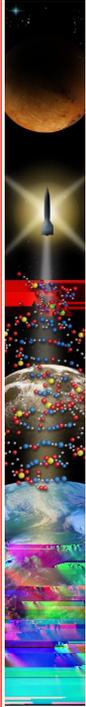
What are the primary challenges to future new missions/new technologies?

- **Priorities for the science program**
  - ➤ Determining the <u>major</u> science goals for OSS via the Strategic Plan

- **Priorities among the mission options**
  - ➤ The priority missions to achieve these goals

- **Developing mission concepts**
  - ➤ Sufficient information to identify long-term technology priorities

- **Priorities for the technology investments**
  - ➤ Given limited funding, which are the "must haves"?

- **Near-term mission needs versus long-term technology programs**
  - ➤ Technology funding as "bank account"

7

NASA FUTURE TECHOLOGY and MISSION PLANNING

What are the primary challenges to future new missions/new technologies? [continued]

- **Technology maturation**
  - ➤ Sustained investment, coordination with other Enterprises

- **Technology infusion into missions**
  - ➤ OSS missions favor established technologies

- **Flight demonstrations**
  - ➤ Insufficient opportunities for flight demonstration

- **Significant technologies outside OSS . . . and NASA**
  - ➤ OSS projects subject to vagaries external to OSS management

- **Mission software consistently a mission's "Achilles Heel"**
  - ➤ Unmanaged largest technology element

8

A FEW DO'S AND DON'TS IN GETTNG FUTURE MISSIONS
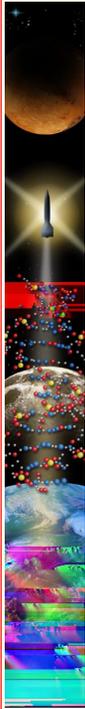
DO at least get your science goals, if not your mission concept, into the OSS Strategic Plan (and associated roadmaps)

DO play close attention to the priority goals of OSS: make sure that your mission demonstrates clear relevance to these goals.

DO get your key technologies as priorities into the Division technology roadmap and Strategic Plan,

DO follow and offer input on technology funding, management, etc, but . . .

DON'T advocate the **wrong** (or lower-priority) technologies.
For example, detectors are obvious and trivial.
What about materials, optical systems, precision structures . . .?

DON'T forget the systems studies: launch systems, operations, orbits, costs

9

Example Technology Recommendations
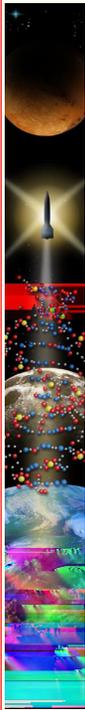
- **Current capabilities appear to be inadequate for autonomously-deploying precision structures in space significantly larger than NGST**

- **No capabilities are planned to recover, service, upgrade science facilities beyond LEO**
  10 - 30% of satellites suffer life-limiting failures

- **Enabling technology investments include precision lightweight structures and large lightweight optics, plus innovative (and poorly-studied) orbits**

- **Enhancing technology investments include detector and communication systems**

10

Science Questions for the Post-SIRTF and Herschel Era

Presented at the 2nd Workshop on New Concepts for
Far-IR/Submillimeter Space Astronomy

Michael Werner, SIRTF Project Scientist
Jet Propulsion Laboratory

March 7, 2002

[http://sirtf.caltech.edu/]

An Outline of the Talk

- **SIRTF**

- **Polarimetry**

- **Confusion**

- **Extragalactic Science**

- **Galactic Science**

SIRTF: Status Summary

- **Optical/thermal/cryogenic performance of now-completed CTA appear excellent**

- **Performance of instruments within CTA is excellent: consistent with that needed on-orbit**

- **Spacecraft hardware completed; software supporting S/C testing**

- **We are proceeding with revised plan: Launch on January 9, 2003**

- **Next major milestone is integration of CTA and Spacecraft, set for this month**

- **First call for GO Proposals – November, 2002.**
  - *Consult sirtf.caltech.edu for updates to plan and details of submission process*

CTA Arrives at Lockheed

LONG WAVELENGTH SURVEYS
PLANNED FOR SIRTF*

TYPE	AREA (DEG^2)	5-σ LIMITING FLUX, mJy		
		24um	70um	160um
WIDE*(GTO)	9	0.6	3.6	33
DEEP‡ (GTO)	2	0.15	1	30
REAL DEEP† (GTO)	0.02	0.06	0.7	
FIRST LOOK§ (SSC)	5	1.1	3.8	33
SWIRES (LEGACY)	≈70	≈0.45	≈2.75	≈(17.5)

* NOAO Deep Field
‡ HDF-N, Groth strip, CXO-S, SSA –13, Lockman Hole, XMM Deep
† Groth strip
§ North Ecliptic Pole

See Werner, Reach, Rieke paper in Manchester IAU Background Symposium. Note that due to confusion we may fall a factor of ~3 short of reaching the deep survey limits

Galaxy Discovery Rates
for Future Missions (Blain et al)
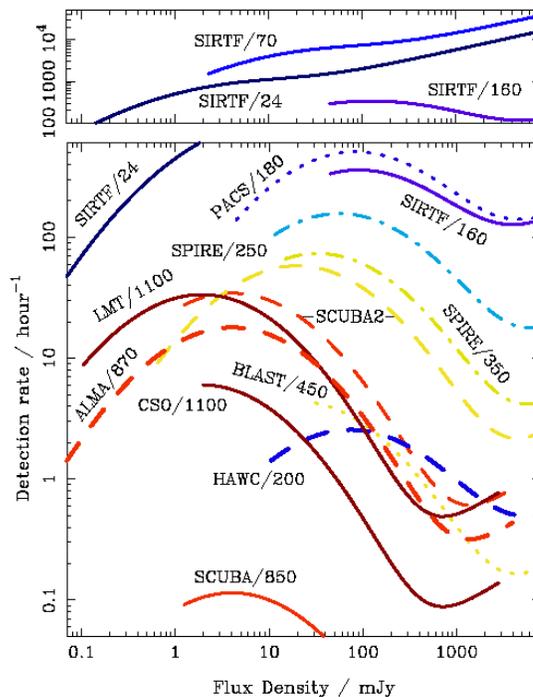
**SIRTF Fills the Gap.SIRTF survey limits compared
with data from ground-based near ir and SCUBA surveys.**
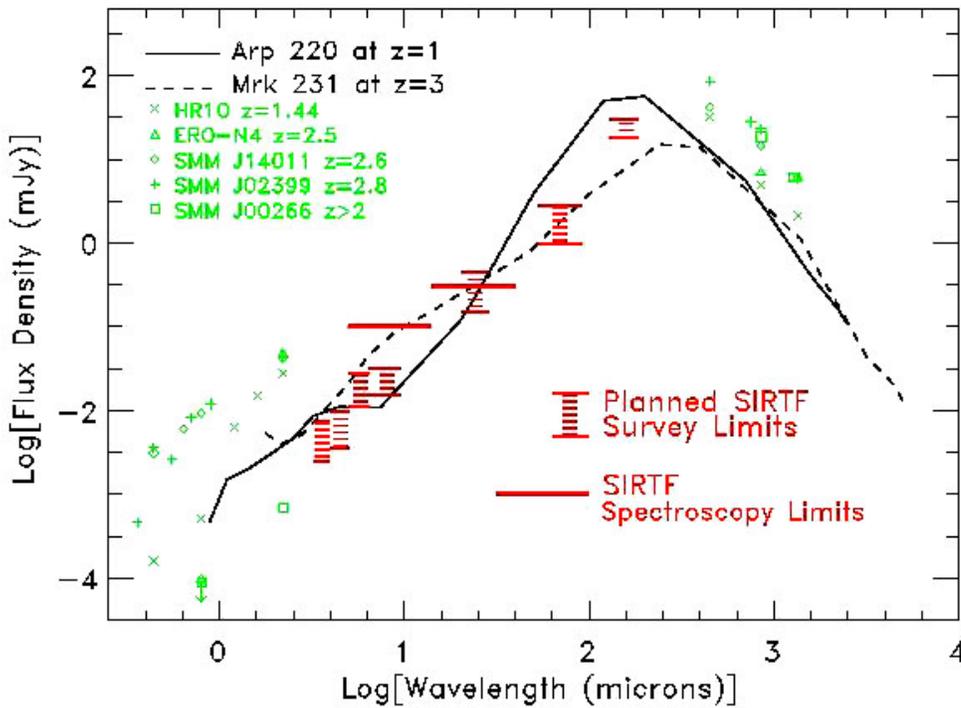
Impact of SIRTF's Improved
Resolution at 160um:
Resolving the Background *(H. Dole, MIPS)*

Simulations of a 34' x 34' sky at 160 μm
ZOOM on a 6.7 ' x 6.7 ' square

*Extragalactic Sources: ~ 600.000 sources between 1 μJy and 2 Jy (Dole H. et al astro-ph/0002283)
Foreground: Galactic Cirrus N_H ~10^{20} cm^{-2}*
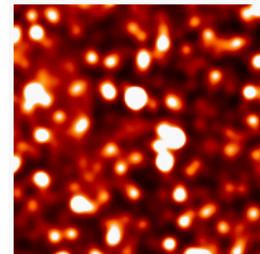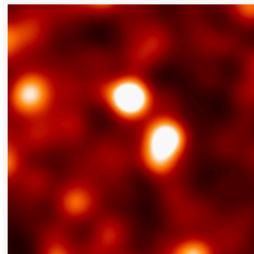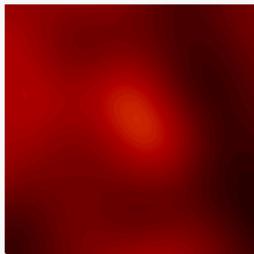
IRAS resolution   ISOPHOT resolution   MIPS resolution

These are noiseless simulations indicative of ultimate gain to be achieved in long integrations.

SIRTF Follow on

- SIRTF will provide major advances in areas such as number counts, resolving the background, the IR-Xray-AGN connection, and leave a legacy of 1000's of far infrared-selected galaxies at redshifts z. SIRTF will complement its long wavelength observations with very sensitive 3-8um surveys sensitive to redshifted starlight, so that the relationship between the distant universe as seen in the near ir and the far ir can be probed.
- Spectroscopic follow-on will be very important and will be carried out from SIRTF on the brighter sources, but higher sensitivity and spectroscopy beyond the SIRTF limit of ~40um will be needed.
- Extrapolating from nearby examples, we estimate the following line fluxes (units are 1e-18 w/m+2) for the faintest SIRTF detections

	Z=1	Z=3
F(OIV, 26um)	1	0.2
F(CII, 158um)	5	0.9

- These fluxes are not bright compared to Herschel limits, particularly if redshift is not known, but within range of larger telescopes, WaFIRS, etc.

- ***Spectroscopic follow-on of SIRTF discoveries should be an important objective of upcoming submillimeter missions***

Submillimeter Polarimetry – The State of Play

- Magnetic fields are like Ross Perot's crazy aunt in the basement: everyone knows she's there, but nobody talks about her.

- Polarimetry is the best means of probing them, and the phenomena to be studied in the far-infrared and sub-millimeter may be particularly affected by magnetic effects

- Hildebrand et al (and others) have shown that the far-ir/submm emission from dust in bright galactic sources shows strong and coherent polarization (up to ~10%)! – and can be used to study:
  - *Field strength and orientation*
  - *Effects of field on gas dynamics/turbulence/cloud evolution/etc.*
  - *Grain properties*
  - *Alignment mechanisms/cloud properties*

- Extending this work by including Zeeman and linewidth measurements from molecular lines may allow field strength and three-dimensional field orientation to be teased out (Houde et al)
  - *With sufficient spatial resolution, this could be a very powerful probe of cloud collapse, star formation, jet formation, etc.*

Magnetic Vectors Across the Orion
Molecular Cloud Core (Dowell et al)
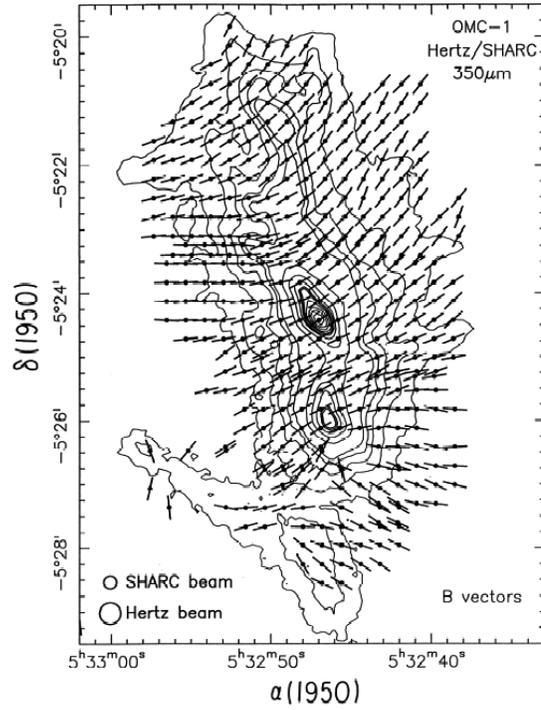
Neutral & Ionized Molecular
Spectral Lines (Houde et al)

HCN

HCO$^+$

N$_2$H$^+$
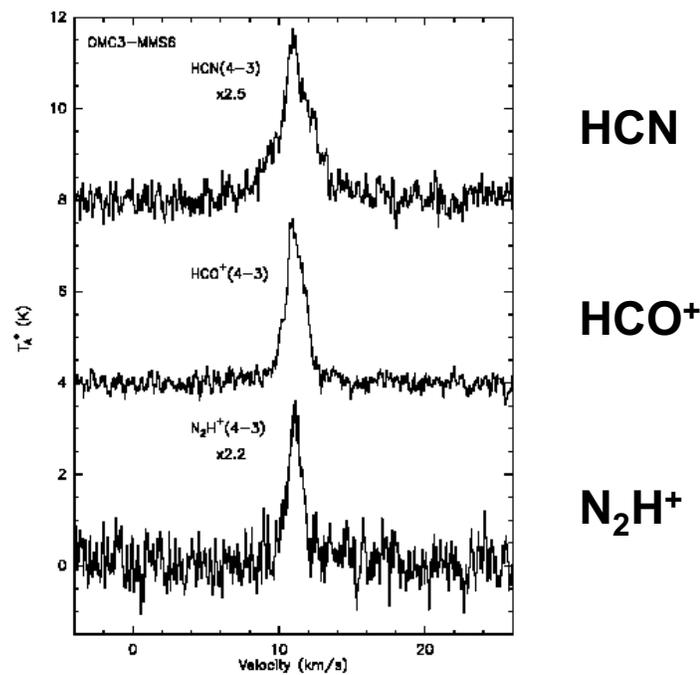
Variation of Polarization
With Wavelength (Hildebrand et al)
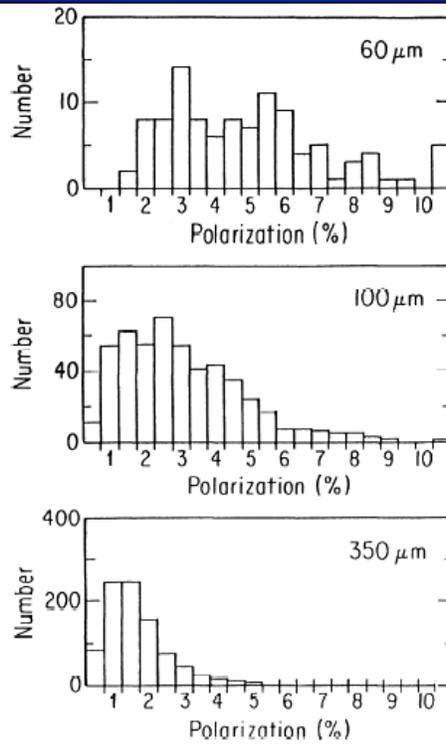
The Polarization Spectrum
(Hildebrand et al)
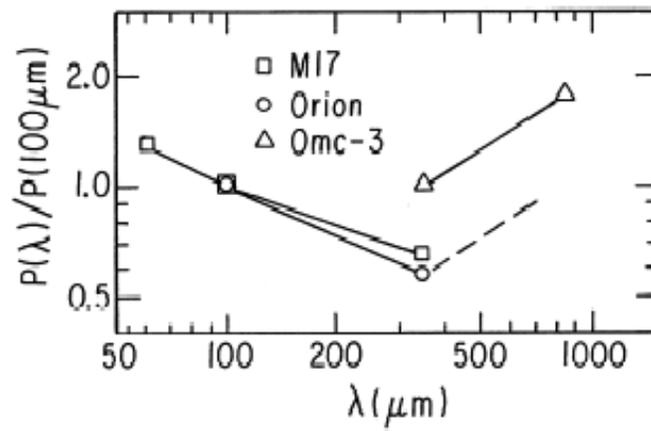
Submillimeter Polarimetry –
Looking Ahead

– *Work to date has been compromised by warm telescopes, atmospheric effects, and – in the far-infrared – limited spatial resolution*
  - ***A polarimeter is a must for SOFIA, but it could probe only the local Universe***

– *In addition to major contributions to studies of star formation, solar system evolution and jet formation, these capabilities could provide surprises*
  - For example, if the polarization characteristics of edge-on spirals vary with redshift, we might be witnessing evolution of cosmic magnetic fields and/or grain properties

– *Understanding the polarization of the galactic foreground seems a necessary precursor to polarimetry of the cosmic microwave background*
  - Fractional effects – relative to cosmological signal - could be larger than for imaging studies carried out to date
  - Jackson, Werner, and Gautier have produced catalog of high latitude cirrus filaments which might be starting point for foreground polarization studies

– ***A robust program of polarimetry should be a component of future far-ir/sub-mm programs***
  - Options beyond SOFIA include both a dedicated polarimetric explorer and serious polarimetric instrumentation on any large free-flyers, or perhaps a combined photometric/polarimetric survey

Confusion

- A small cryogenic telescope like SIRTF reaches its confusion limit for photometric observations rather quickly:
  - *~10s of s at 160um, ~100s of s at 70um for 1-sigma photon noise to equal 1-sigma confusion noise*

- Time to reach confusion limit and confusion limit as a function of beam size should be borne in mind in designing next generation of instruments and in defining cooling strategies.

- Modelling by Blain et al suggests that flux at confusion limit drops more quickly than D^{-2} for telescopes larger than Herschel and SIRTF in far-ir and sub-millimeter.

- This suggests that there are gains in photometric speed to be achieved with cooling larger apertures
  - ***High priority should be given with SIRTF and Herschel to whatever can be done to determine confusion and extrapolate it into the >~10m aperture ranges***
  - *Important to understand use of spectroscopy to break distance degeneracy and further combat confusion*
  - *Observing strategies which go deep into confusion should be evaluated*

Confusion at 500 μm
(1σ, Blain et al)
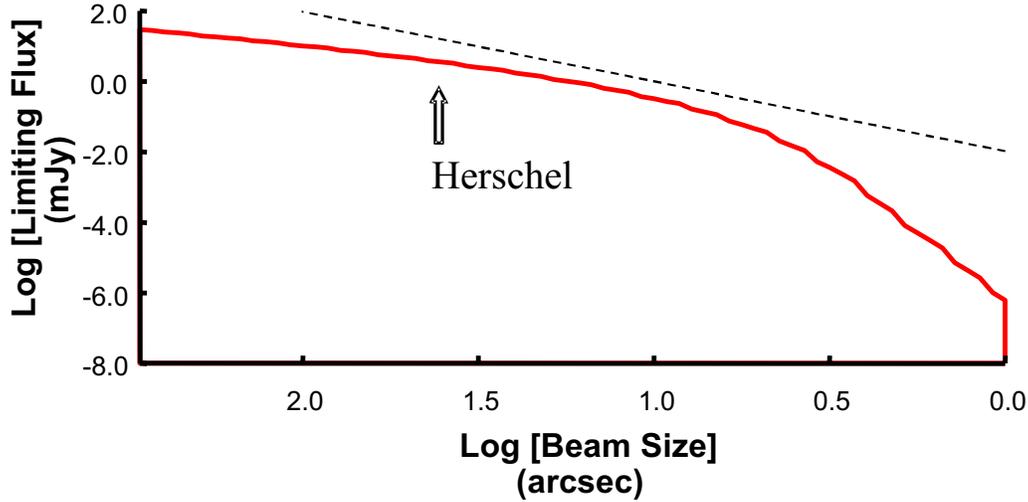
The diagonal line has a slope of –2. The fact that slope of the predicted confusion
 limit vs. beamsize is steeper suggests that for beam sizes smaller than ~10" confusion
 it will take increasingly longer [with the same instrumental+sky background] to reach the
 confusion limit than it will with Herschel. A colder telescope would be well-justified if these
 predictions are true. They should be studied by SIRTF and Herschel to the greatest
 extent possible.

Confusion at 160 μm
(1σ, Blain et al)
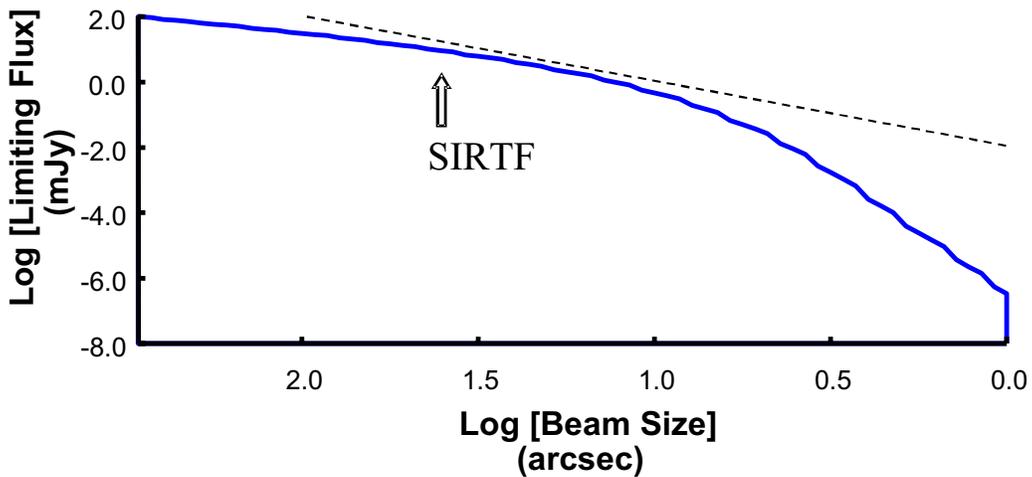
The diagonal line has a slope of –2. The fact that slope of the predicted confusion
 limit vs. beamsize is steeper suggests that for beam sizes smaller than ~10" confusion
 it will take increasingly longer [with the same instrumental+sky background] to reach the
 confusion limit than it will with SIRTF. Thus a ~10-m class cold telescope will be considerably
 less plagued by confusion than SIRTF, if these predictions are true. They should be studied
 by SIRTF and Herschel to the greatest extent possible.

Do Massive Black Holes and Galaxy Bulges form Together?

- Correlation of central black hole mass with stellar bulge magnitude and velocity dispersion suggests they may
  - *Assuming they do, and that the bulge is formed in a burst of star formation, this links nuclear and gravitational energy release – the two main forms of energy generation in the Universe*
  - *Results from Kormendy et al suggest E(starburst)/E(AGN) ~ 5*

- Comparing far-ir/submillimeter and x-ray emission is one way of probing this
  - *Page et al claim to have seen the process in action via SCUBA detections of 4 of 8 ROSAT sources in range 1<z<3*
  - *By contrast, Severgnini et al claim that submillimeter and x-ray background come from different populations*

- Exploring this connection is a prime near-term goal for far-ir/sub-mm astronomy
  - *Exploration out to Z~3 – region of maximum AGN activity – may do*
  - *Spectral diagnostics of starbursts and AGN contained within the band*
  - *Comparisons of SIRTF/Herschel/SCUBA/Chandra/XMM will be a good start*

Blackhole Mass – Bulge Mass/Velocity Correlation   *(Kormendy et al)*
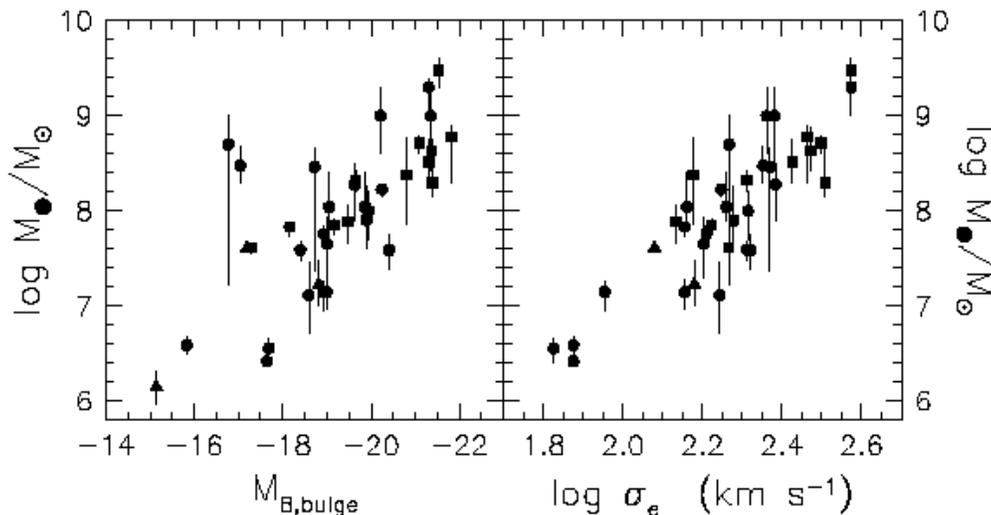
Infrared Spectral
Diagnostics
(IRS Team)
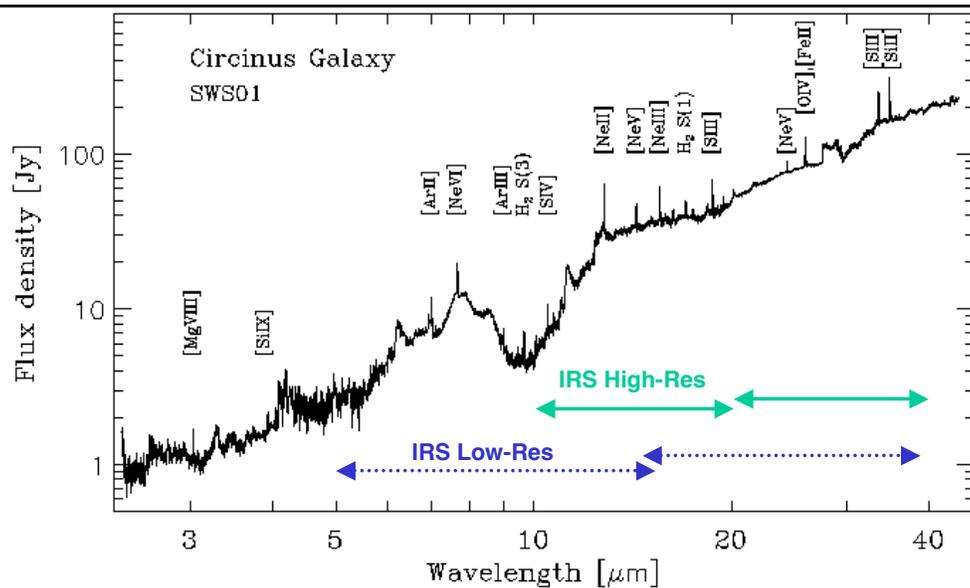
The 2.5 to 45 micron spectrum of the Circinus galaxy

The richness of the mid-infrared spectrum is such that infrared spectral observations alone may be able to assess the relative importance of starburst and AGN activity in distant objects.

Can We See the First Generations
of Stars and Metal Formation?

- The far-infrared and submillimeter is a place to look for the earliest action as objects condense in the Universe….but who knows where or when?
- Options include:
  - *Redshifted H2 lines from collapsing objects*
  - *Starlight reprocessed by first generation of dust*
  - *………*
- Serious theoretical work has begun on this intriguing question:
  - *Spergel*
  - *Ciardi and Ferrara*
  - *Abel et al*
  - *…………..*
- The emission may be within reach of next generation instrumentation

WaFIRS Sensitivity to
Redshifted H_2 Lines
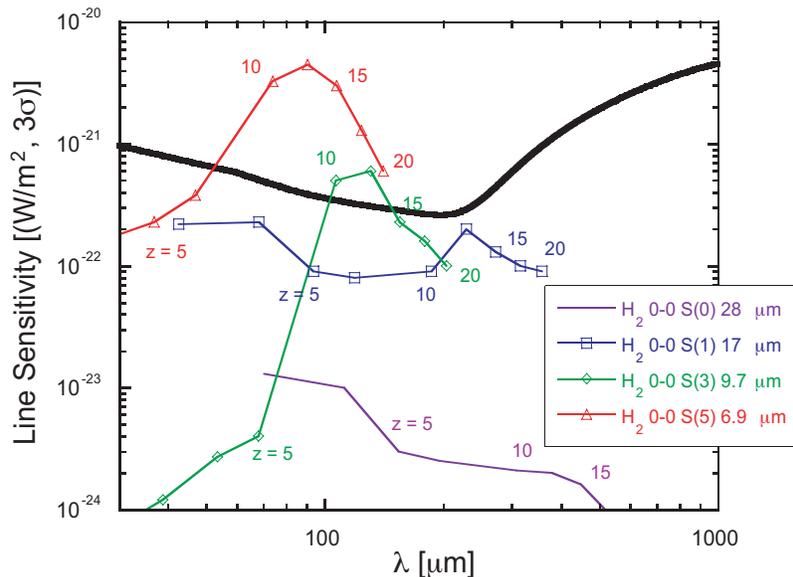
Molecular hydrogen line emission driven by the first generation of star formation, calculated according to the model of Ciardi and Ferrara (2000). Line emission is detectable out to large redshift (z = 20) in moderate integration time. WaFIRS is a concept for a highly efficient cooled spectrograph presented at this meeting by Jamie Bock et al

The Birth of Planets and the
Origins of Life

- The increasingly improved spatial resolution of successive generations of instruments will allow us to zoom in on the star and planet formation process (will also be invaluable for extragalactic studies)

- There will always be more objects detectable than resolvable, so detailed spatial studies and modelling of those which can be resolved are critical along the way – cf. SIRTF program to study Fabulous Four Debris Disks

- Far-ir imaging, polarimetry, and spectroscopy of planetary systems at all stages from the initial collapse phase through to the last stage evolution of solar systems like our own, we can attack fundamental questions of our origins and our fate
  - *Condensation of planets within the protosolar nebula*
  - *Composition of condensing material*
  - *Timescales and other influences – jets/magnetic fields*
  - *Prevalence of planetary systems in the solar neighborhood*

- Spectroscopy at high spatial resolution may even indicate what the organic/biogenic carryover into the forming planets might be, as key markers of organic molecules like in the far-ir/sub-mm

Spatial Resolution at 100um

SIRTF

SOFIA

Herschel

SPIRIT

SPECS

With the 20marcsec angular resolution of a
~1km baseline SPECS interferometer, it will be
possible to image disks like that around HH-30
~2 astronomical unit resolution.
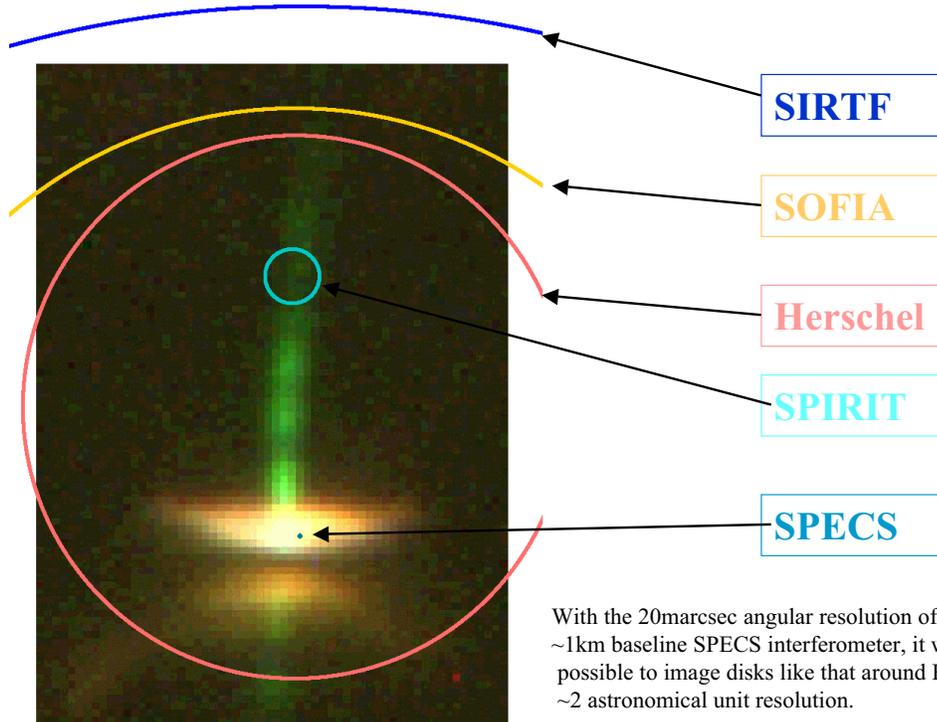
Far-ir/Sub-mm Transitions of
Linear Carbon Clusters

Lowest Bending Modes of Linear Carbon Clusters

and HIFI Receiver Bands

$C_8 - C_{13}$ center frequencies from ab initio calculations.

Rotational constants from IR-data.

C_3

C_{13} C_{11} C_{10} C_9 C_8

1   2   3   4   5   6

HIFI Bands

0.5   1.0   1.5   2.0

Frequency [THz]
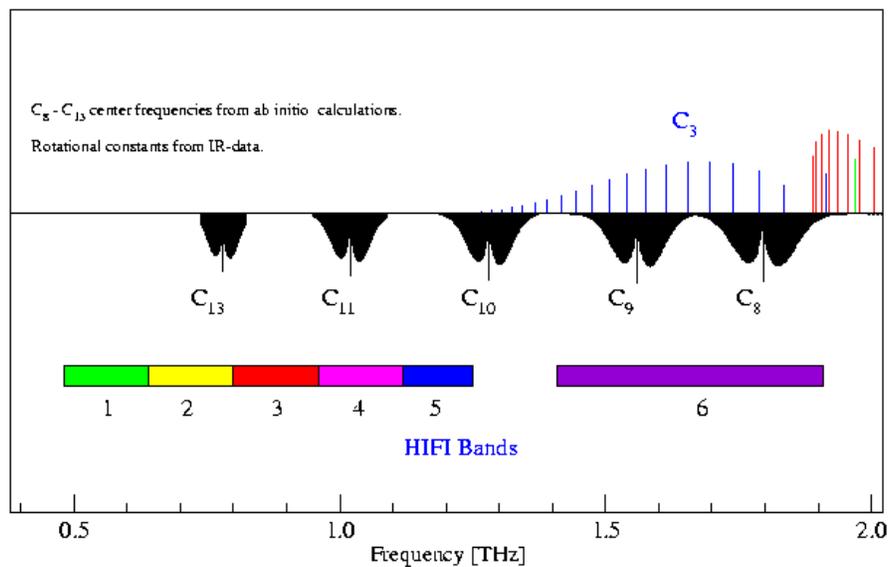
Predicted Spectra of Glycine

THZ Spectra of Glycine Conformers
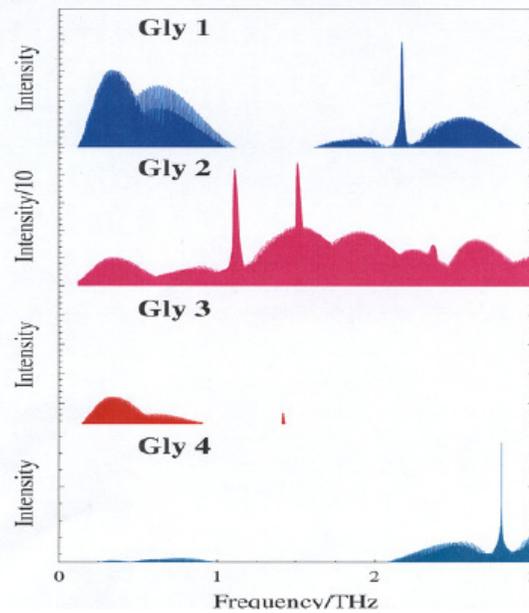
SIRTF Science Plans
Keep abreast of your opportunity to propose for observing time with SIRTF at sirtf.caltech.edu

Substantial Progress Made in Planning Year One
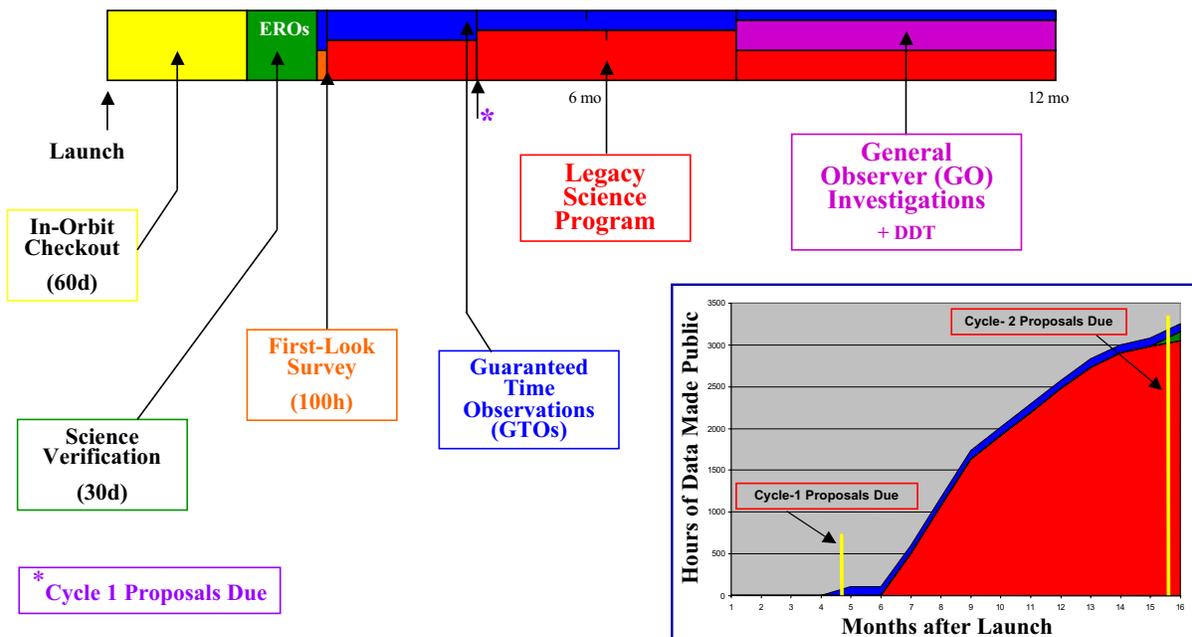
Technology Needs for Far IR, SMM, and MM Missions

S. Harvey Moseley

Mar. 7, 2002
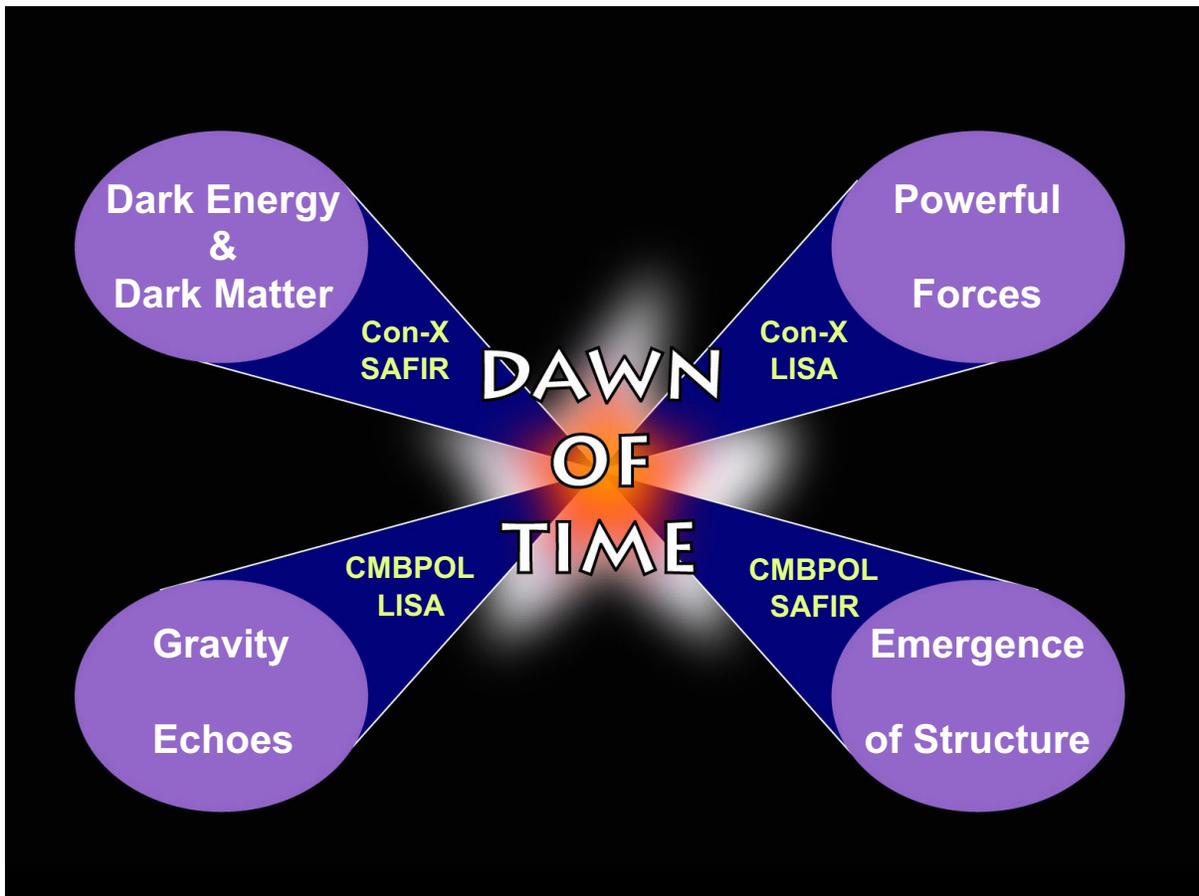
SAFIR: SINGLE APERTURE FAR-INFRARED OBSERVATORY

"SAFIR...will:

- Study the important and relatively unexplored region of the spectrum between 30 and 300μm.
- Enable the study of galaxy formation and the earliest stage of star formation by revealing regions too enshrouded by dust to be studied by NGST
- Be more than 100 times as sensitive as SIRTF or the European [Herschel] mission.

"The committee recommends SAFIR...

- The combination of its size, low temperature, and detector capability makes its astronomical capability about 100,000 times that of other missions
- It [has] tremendous potential to uncover new phenomena in the universe."

– pages 12, 39 & 110 of *Astronomy and Astrophysics in the New Millenium*, National Research Council, National Academy Press, 2001.

SAFIR is projected to cost around $600M total. The decadal review committee recommends that $100M be allocated in this decade to start the SAFIR project, and that additional technology developments be funded separately:

- Far-Infrared Array Development ($10M[†])
- Refrigerators ($50M[†])
- Large, Lightweight Optics ($80M[†])

[†] Funding levels recommended by decadal review, page 47.

SAFIR features:
- ✓8m diameter telescope cooled to 10K
- ✓Located at L2
- ✓Lifetime 5 years
- ✓30μm-800μm wavelength range
- ✓Instrument complement including cameras and imaging spectrometers

POSSIBLE **SAFIR** CONCEPT BASED ON NGST

SAFIR IS WITHIN REACH AND IN NEED OF NEAR-TERM MISSION CONCEPT STUDIES.

Half the luminosity and 99% of the photons in the post-Big Bang Universe are in the far-infrared and submillimeter

Spectrum of Milky Way Galaxy
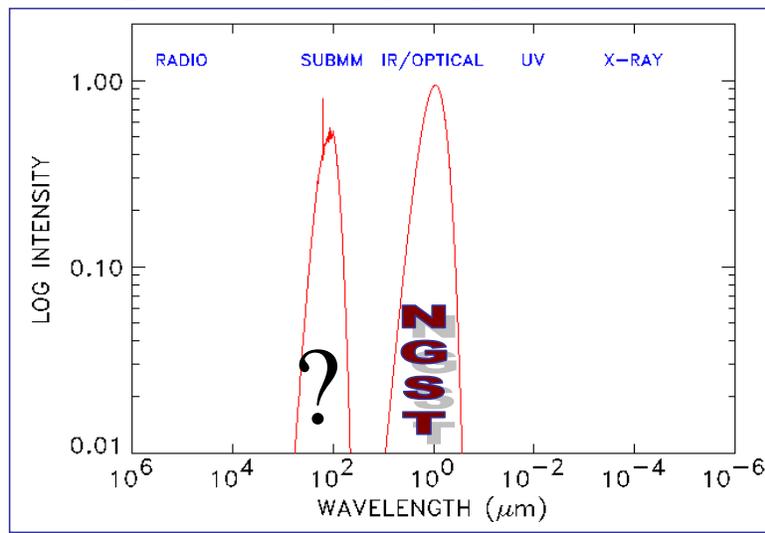
Ultimate Performance Requires Cold Telescope

NEP for Diffraction Limited Beam, 20% Bandwidth
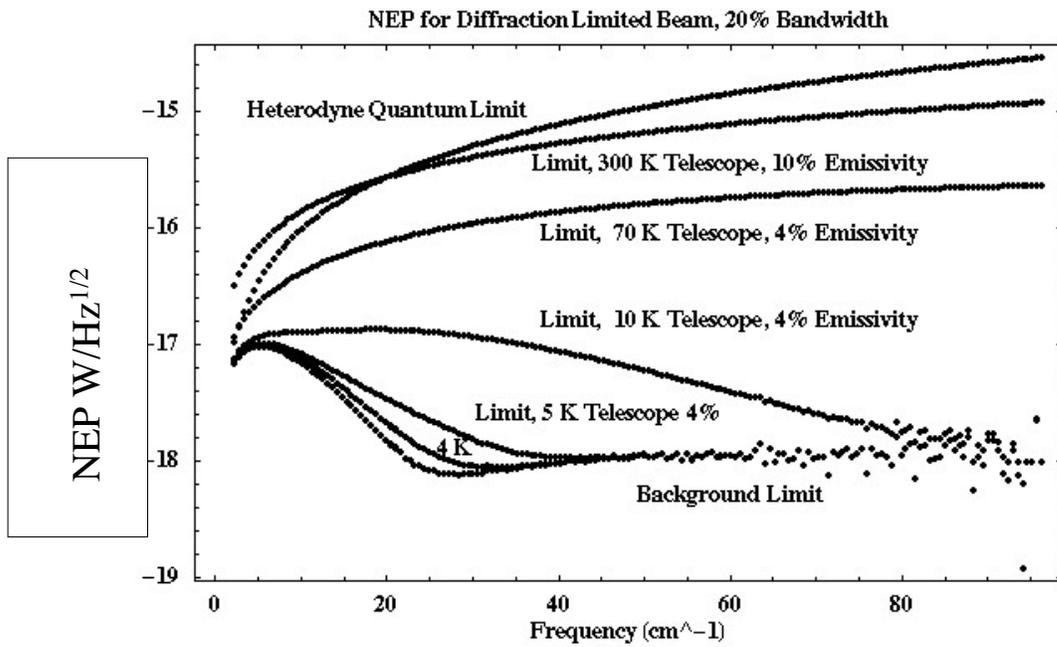
Roadmap

Illuminating the Cosmic Dark Age

(1)

(2) SPECS
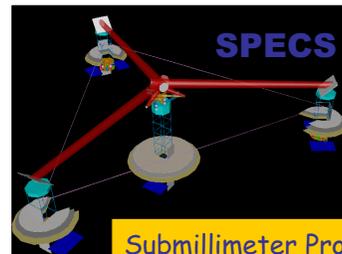
Submillimeter Probe of the Evolution of Cosmic Structure

2020

SAFIR

Single Aperture Far-IR Observatory

2010
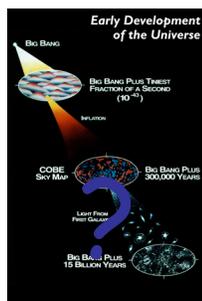
COBE IRAS

"Astronomy and Astrophysics in the New Millennium" Astronomy and Astrophysics Survey Committee

Gravitational waves excited during the first instants after the Big Bang should have produced effects that polarized the background radiation. More precise measurements of the properties of this polarization—to be made by the generation of CMB missions beyond Planck—will **enable a direct test of the current paradigm of inflationary cosmology**, and at the same time they will shed light on the physics of processes that occurred in the early universe at energies far above those accessible to Earth-bound accelerators.

Future microwave background experiments, such as measuring the polarization, are of great importance, but the committee recommends that prioritization of such experiments await the results from MAP…

"Astronomy and Astrophysics in the New Millennium" Astronomy and Astrophysics Survey Committee

…NASA's MAP mission, to be launched in spring 2001, will revolutionize knowledge of the microwave background, and the committee believes that no decision on the next major or moderate microwave background project should be made until the results from that mission are available. ESA's Planck mission later in the decade will also provide important information, but **it will be possible to decide on the next step before its results are available**. Together, MAP and Planck will test the most promising ideas about the very early universe as well as determine cosmological parameters to high precision. **The next frontier is to measure the polarization of the cosmic microwave background, which has the potential of probing even earlier times, close to the Big Bang itself**.

Technologies for Far IR/mm Missions

- Cryogenic Cooling
  - Efficient use of radiative cooling
    - Choose a good orbit
  - Refrigerator for cooling to ~10 K
  - ADR or Dilution Fridge for lower temps required by detectors
- Lightweight large optics
  - Less demanding figure compared to (SIRTF and NGST)
  - Colder
  - Lighter

Mission Technologies (cont.)

- Large Format background limited arrays
  - 10^3 to 10^4 or more elements
  - Integrated low power, low temperature readout
- Coherent receivers will have a role for high resolution spectroscopy

Current Developments

- SIRTF
  - Long-life cryostat with efficient radiative cooling
  - Cryo-friendly orbit
  - Large format Far IR detector arrays.
- MAP
  - Radiative cooling in good environment

Current Developments

- NGST
  - Low mass deployable mirrors with figure control
  - Deployable sunshield
- Herschel
  - SiC mirror
  - Bolometer Arrays

43

Current Developments

- SOFIA and Suborbital
  - Instruments and detectors

- Constellation-X
  - The Con X detector requirements are very similar to, and perhaps more difficult than, those of SAFIR or CMBPOL.
    - We must watch and participate in this development so we are prepared to transfer it to our missions.

Cooler Technologies

Cooler Performance, Circa 2002
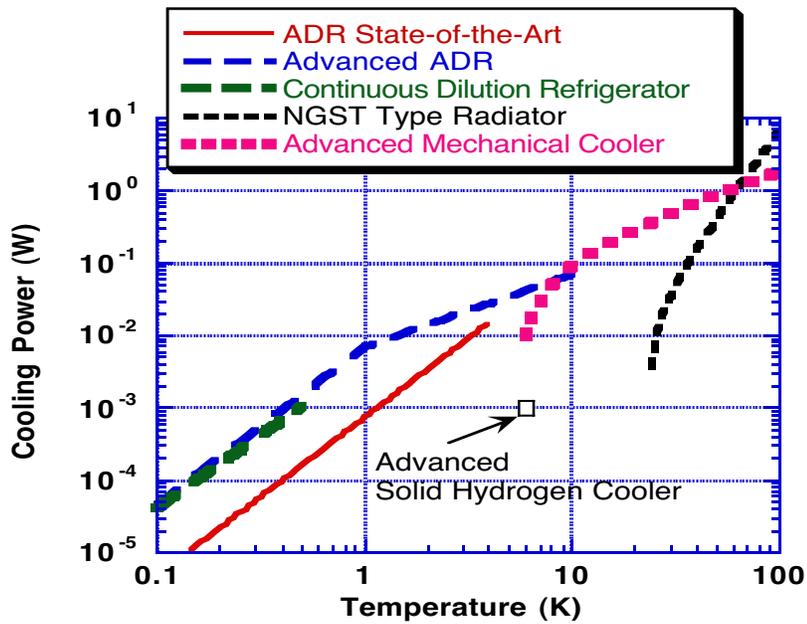
Cryostats Can Provide Long Life, Modest Cooling Capacity
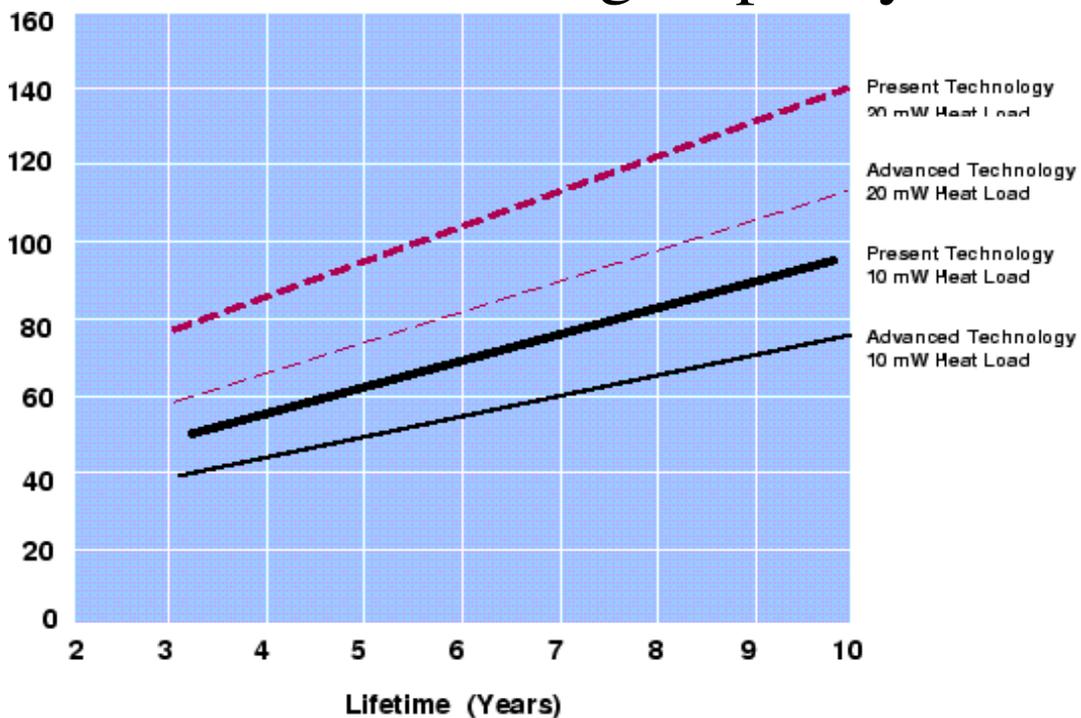
Heat Switch Options
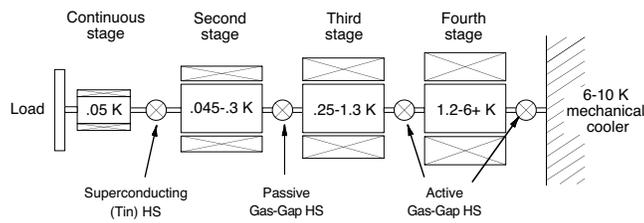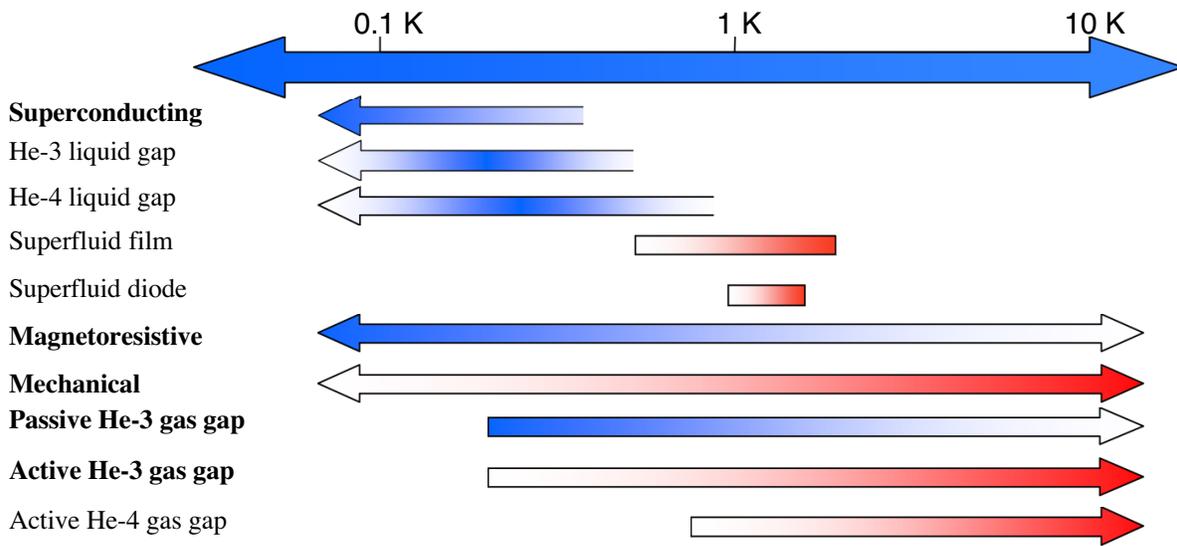
Actual Performance
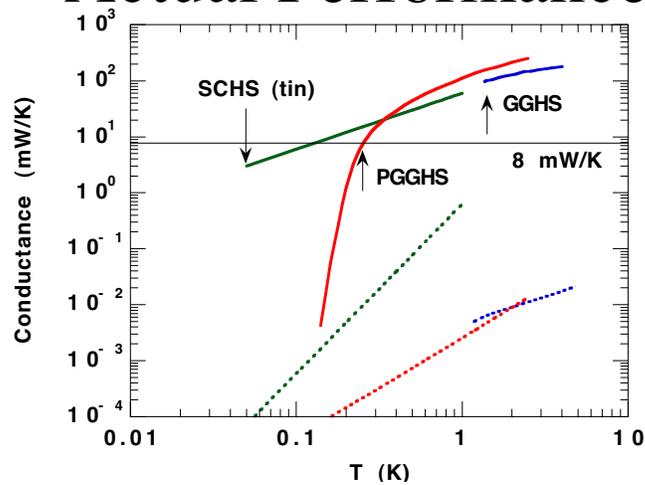
- Superconducting switch is 2x less conductive than desired
  – Limits present system to proportionally lower cooling power

3-Stage ADR

- **Continuous Stage**
  - 42 g CPA salt pill
  - .1 T magnet
  - .05 cm thick magnetic shield
  - Superconducting heat switch
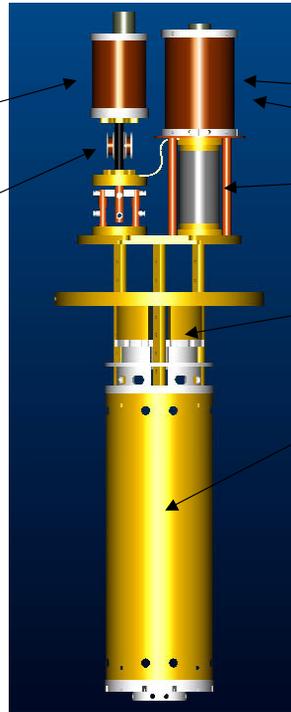
- **Second stage**
  - 100 g CPA salt pill
  - .5 T magnet
  - Passive gas-gap HS

- **Third stage**
  - Connected to 1.2 K helium bath by an active gas-gap HS
  - 730 g FAA salt pill
  - 0.8 T magnet (not shown)

SIRTF Demonstrates Key Technologies for Future Missions

Warm launch

Efficient radiative design (dewar shell 40 K)

Earth-trailing Solar orbit

Long life cryostat

Large format far IR array
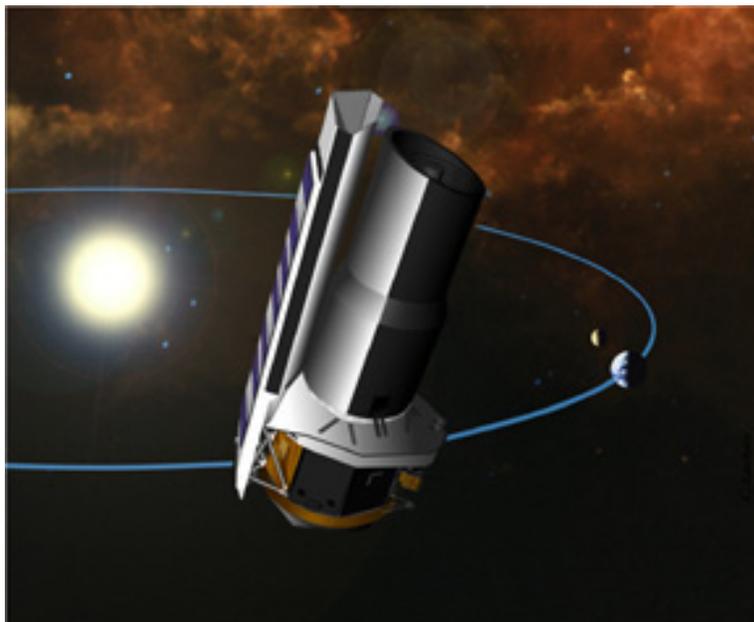
NGST Sunshield Development Status
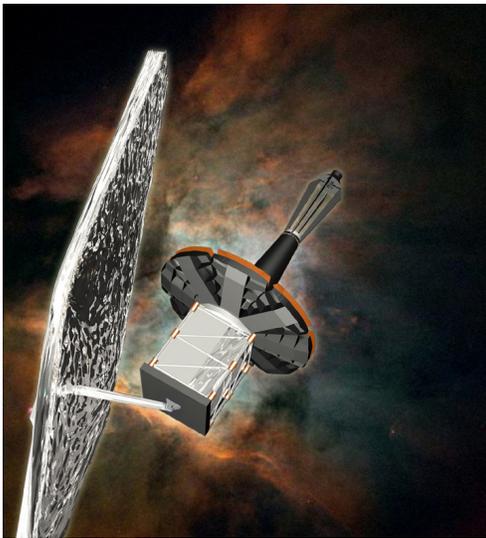
Lockheed Martin NGST Concept
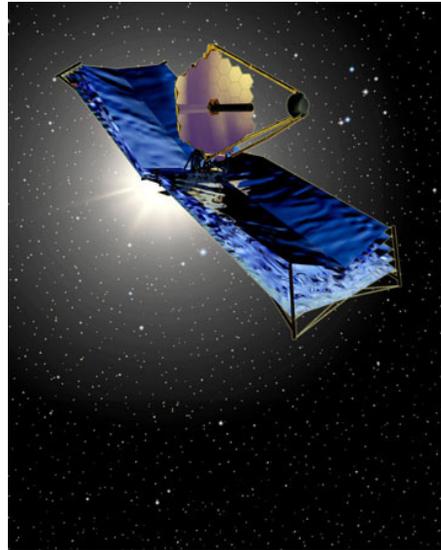
TRW/Ball NGST Concept

NGST Passive Cooling Development Status

Key requirements

• Telescope optics less then 50 K passively
• Instrument near infrared detectors < 30 K passively
• Heat loads on mid infrared instrument compatible with volume and mass allocation for 10 year cryogen lifetime
• Passive cooling heat load margins > 50 %

Key passive cooling technologies

1. Sunshield utilizing deployed multiple separated membranes for sunshield

2. Large 30 K radiator (>8 m^2) and thermal strap system for cooling of detector arrays

Sunshield Technology Development

NGST 2-m Sunshield Model

Sunshield stowage, deployment & film management approaches have been explored and demonstrated using numerous models of 1/12 to >1/2 scale

NGST Yardstick 1/2-Scale Sunshield

- •4 film layers
- •Inflatable deployment booms
- •Film launch restraint & tensioning mechanisms
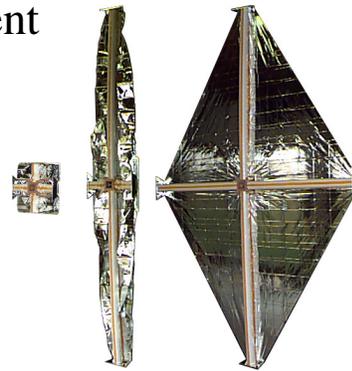
Optics Technologies

Herschel Telescope

3.5 m diameter
Collecting area > 9m^2
FOV +/-15arcmin
f/# 8.68
Mass 280 kg

Fabrication is in progress
SiC green bodies (blocks of SiC)
    are being cold pressed
Machining of green bodies will
    start later this month

Demonstration Petal

Primary will be
made from 12 petals
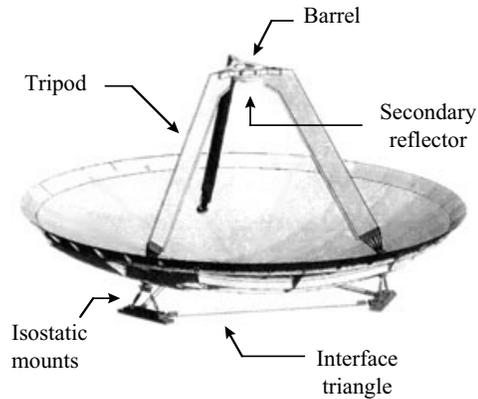
2-Meter Lightweight Mirror Demonstrator and FIRST Telescope Mock-Up
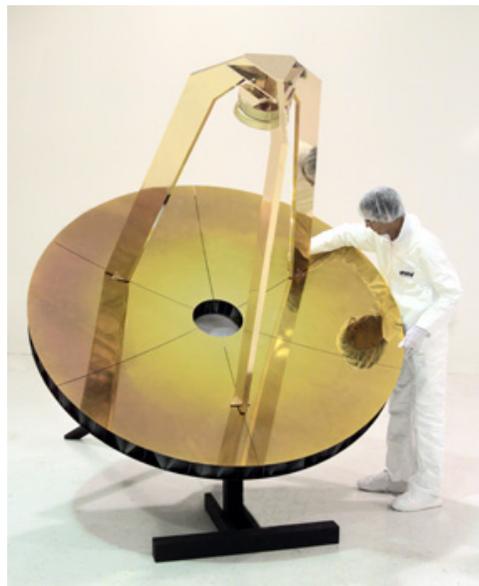
Evolution of Figure vs. Temp
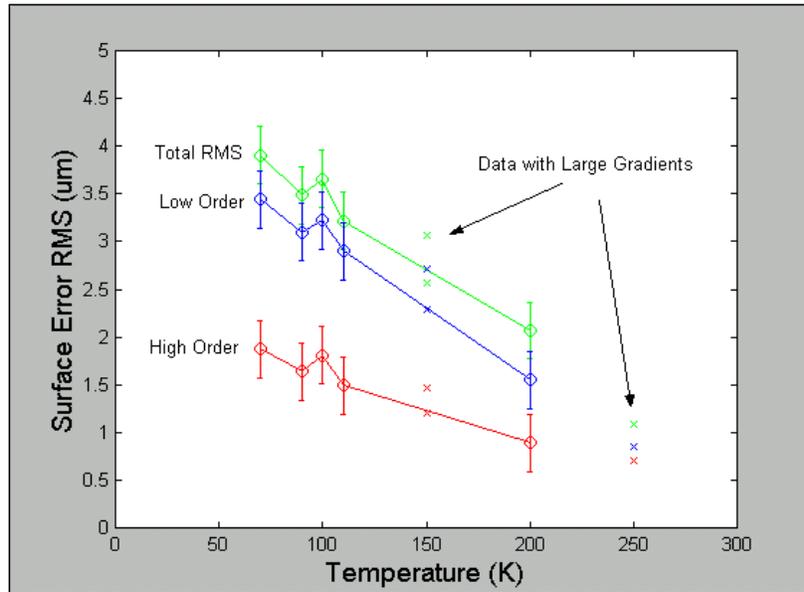
Modest Variation Over Operating Temperature

Delta Figure: 293K to 70K
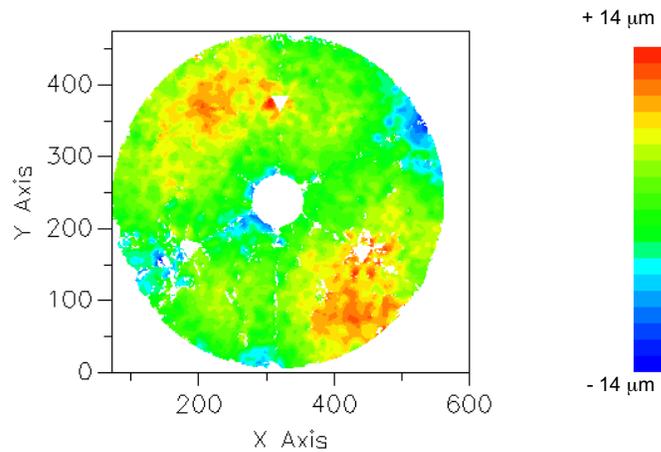
- ◆ $S_5 - S_2$
- ◆ 3.9 μm RMS

New Ideas Can Change Optimization

DART telescope, Dragovan
et al., provides imaging over a
100 x 100 beam area using crossed
cylindrical mirrors.

Detector Technologies

SIRTF
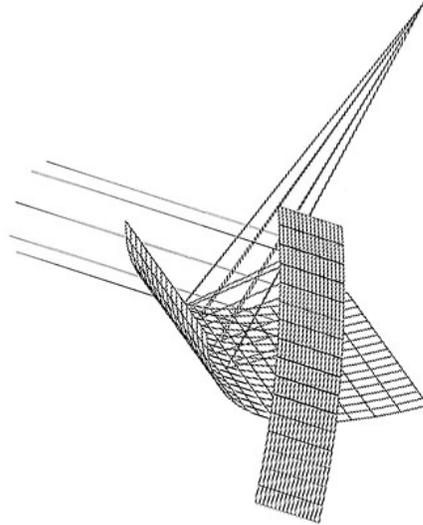
MIPS 70 μm Flight Array

- Wavelength Coverage:
  50 - 115 μm

- Detector Material: Ge:Ga

- Format: 32 x 32

- Pixel Size: 750 x 750 μm

- Readout Technology:
  Cryogenic CTIA

- Read Noise: 92 electrons

- Dark Current: <190 e/s

- NEP: 1.2×10^{-18} W Hz$^{-1/2}$
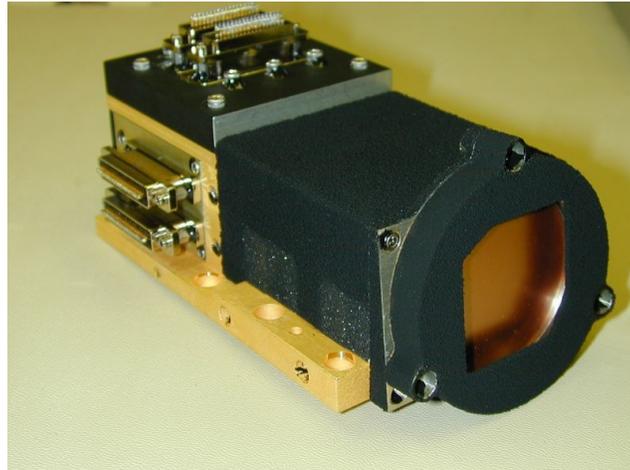
SPIRE Uses Arrays of Spider Web Bolometers for Submillimeter Imaging and Spectroscopy

Feedhorn coupled

NTD Ge thermometer

Silicon Nitride Structure

Si backing wafer

AC biased detectors read out by
JFET preamps for low 1/f noise
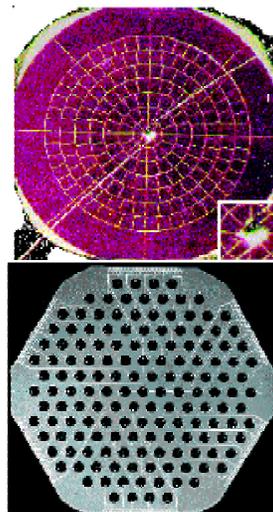
Large Format Far IR Arrays for SOFIA
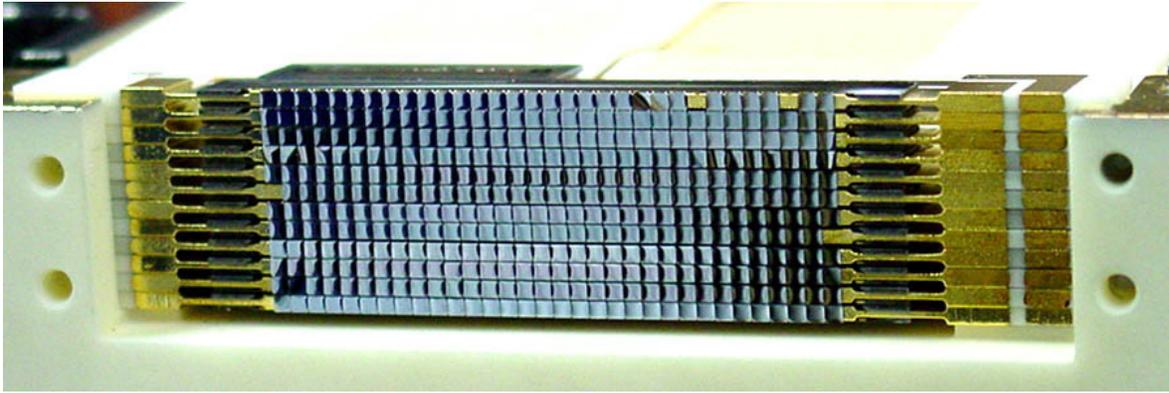
12 x 32 pop up array, prototype for HAWC, to be demonstrated in the SHARC II instrument at 350 μm on the CSO.

Superconducting TES Bolometers

- TES detectors and SQUID amplifiers offer significant benefits for large format arrays.
    - High performance thermometer
    - Couples naturally to SQUID amps
    - Possibility of multiplexing
- Several groups have demonstrated near-ideal noise in "IR" type TES detectors
    - Irwin et al., Lee et al., Bock et al.
- Multiplexed TES array operation demonstrated in lab and FIBRE instrument
    - First light on CSO in the May 2001 (Benford et al.)
- Pop-up versions planned for SAFIRE on SOFIA, SPIFI, and other applications
- Array technology development
    - First arrays built in pop-up architecture
    - Planar arrays mid-term goal (SCUBA II, Con-X, and GSFC/Penn (GBT) developing such devices with TDM mux)

NIST measurement of AlAg Bilayer
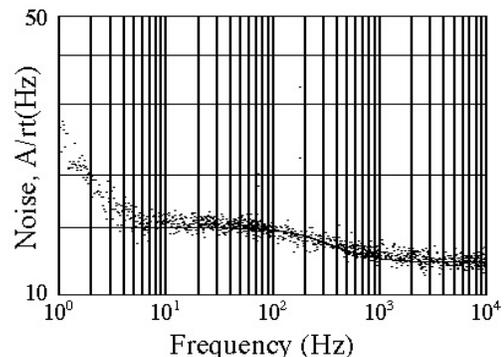
Superconducting Transition in Bilayer

- Transitions remain sharp, but transition temperature changes with relative thickness of normal metals and superconductors.

$R_n = 0.2209\ \Omega$
$T_c = 0.4442\ K$
Sensitivity = 178
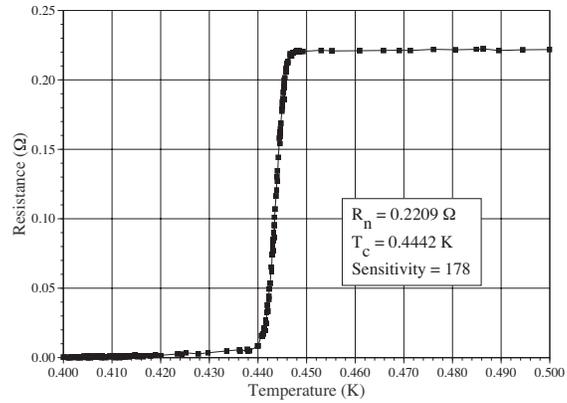
Unfolded TES Pop-up
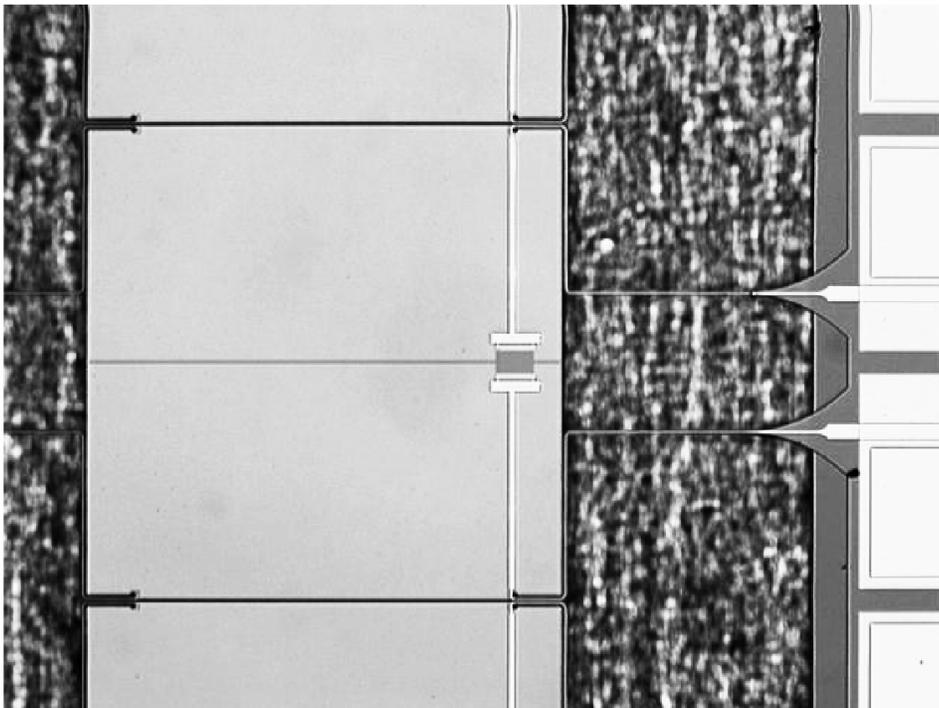
Alternative Configurations Are Under Study for Large Format Arrays

- Superconducting version of Saclay/CEA detector shows promise for scaling to large arrays
- Front end SQUIDs can operate at detector temperature
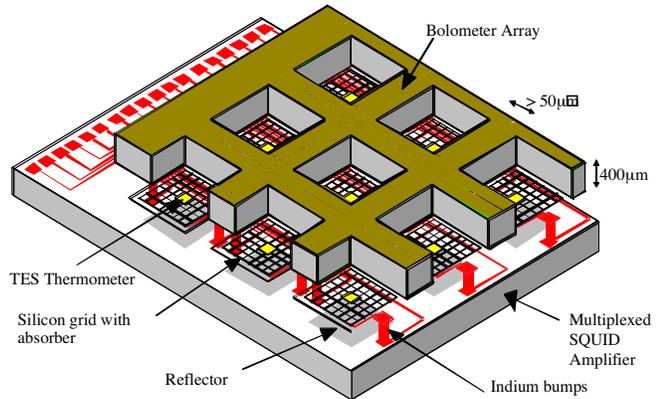
First Steps Towards Mechanical Structure
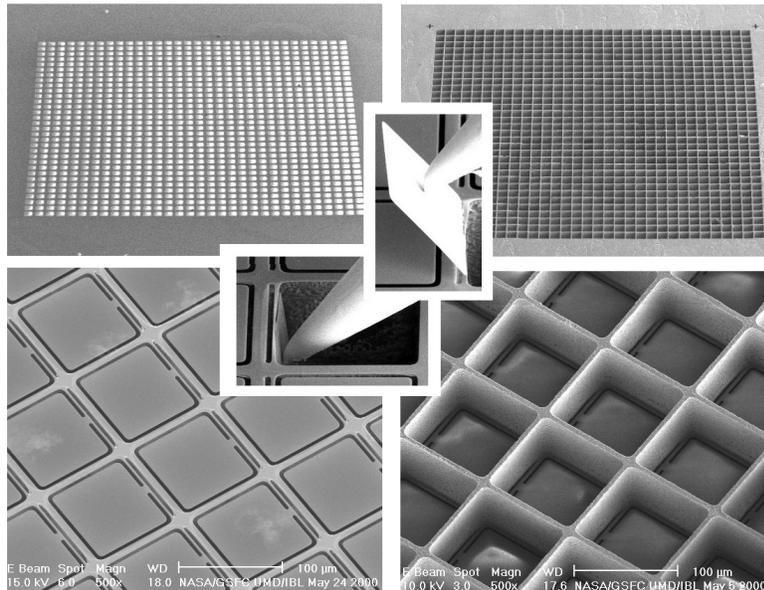
Multiplexing Concept

- Multiplexing is effected by sequential sampling of individual input SQUIDS

- We see full noise of SQUID, band limited noise from detectors
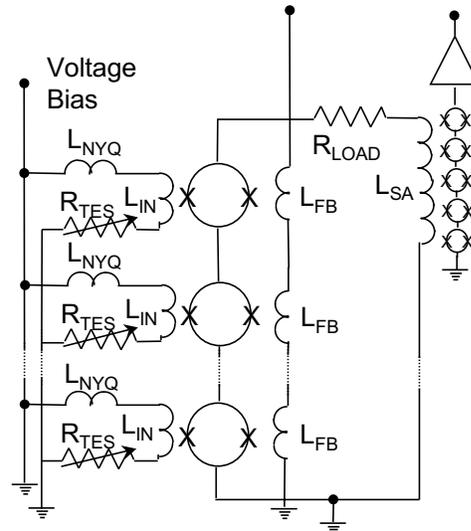
Steps to Required Technologies

- Long term industrial development on refrigerators, optics, and cryogenic systems must continue with community involvement
- We must identify unique needs for Far IR/ SMM missions that we must do ourselves
  - HQ-chartered detector panel is doing this in our most critical area, detectors
- Establish close connections with other relevant development programs (NGST, Con X, Herschel)
  - Avoid duplication, save time

Technology Concerns

- Basic concepts for SAFIR, e.g., are established. Significant engineering effort is required to make progress
    - Detailed engineering is typically only done after performance requirements are established.
    - Much important (and schedule-risky) work must be done after initiation of project.
    - We need to make progress on critical elements, such as detectors, before "meter is running" to lower cost risk on missions.

Summary

- We need to refine mission concepts, agree on science goals and requirements
- We need to identify performance requirements and pursue them to demonstrate readiness
- Far IR, SMM, and CMB science is so compelling that given a clear community voice and concerted technical effort in critical areas, it will compete successfully with other fields.

SESSION II

SCIENCE INVESTIGATIONS IN THE FAR-INFRARED AND SUBMILLIMETER

Star and Galaxy Formation at High Redshifts:
What is to be Learned from Future Far Infrared and Submillimeter Observations

J.L. Puget, G. Lagache
Institut d'Astrophysique Spatiale, Université Paris Sud

1- Status of observations and empirical models

The observations relevant to the problem of star and galaxy formation at high redshift available at present are of three types:
- the cosmic infrared background of radiation from all generations of stars and galaxies (see. Hauser & Dwek 2001 for a review)
- deep number counts of extragalactic sources from mid infrared to millimeter wavelengths (e.g. Elbaz et al. 2002, Dole et al. 2001, Serjeant et al. 2001, Scott et al. 2002 and references therein)
- power spectra of the unresolved background in the far infrared (Lagache & Puget 2000, Miville-Deschênes et al. 2002)

Identifications and detailed studies of individual sources at long wavelengths are coming very slowly. Partial redshift distribution starts to be available.

To use these data to find the discrepancies with present theories of structure formation in the universe and plan future observations it is convenient to gather all these data in a phenomenological model which has for basic inputs the luminosity functions of a small number of populations (typically two to three) of infrared galaxies as a function of redshift and a set of templates of spectral energy distribution (SEDs) which fit the whole set of observations mentioned above.

Such models have been built recently (e.g. Devriendt & Guiderdoni 2000, Dole et al. 2000, Wang & Biermann 2000, Chary & Elbaz 2001, Franceschini et al. 2001, Malkan & Stecker 2001, Pearson 2001, Rowan-Robinson 2001, Takeuchi et al. 2001, Xu et al. 2001, Wang 2002) and gives extrapolations of the number counts to low fluxes. We use in this paper the empirical model of Lagache et al. (2002). The number counts and redshift distribution at wavelengths where present observations give strong constraints are shown in figures 1 and 2. These empirical models are very useful to investigate the basic capabilities of the future missions which will be able to answer some of the open questions in the present galaxy formation models.

These basic capabilities are
- sensitivity
- resolving power to beat confusion limit
- capabilities to cover large enough areas to find the rare distant sources
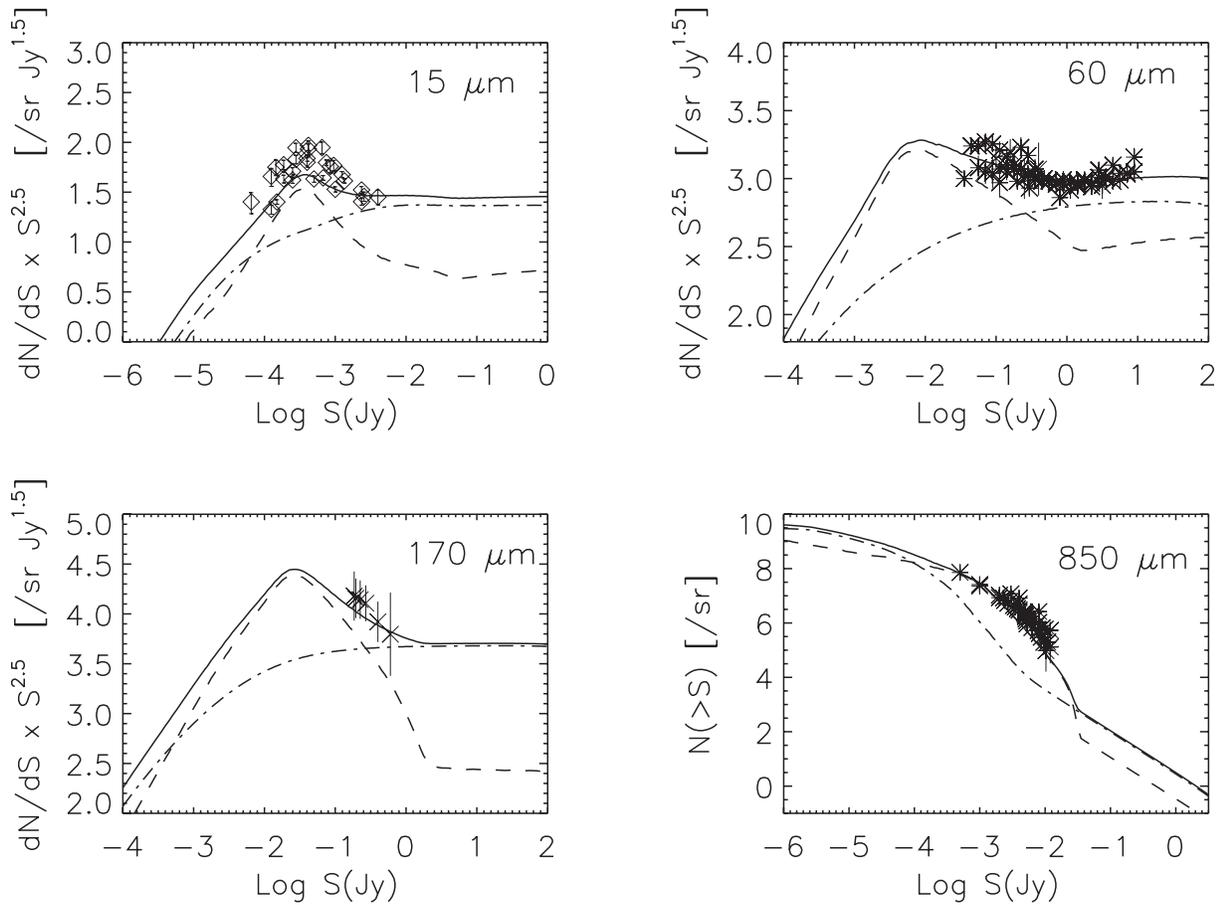
Fig. 1: Number counts at 15, 60, 170 and 850µm together with the model predictions (starburst galaxies: dash line, normal galaxies: dot-dash line, both normal and starburst galaxies: continuous line). Data at 15µm are from Elbaz et al. (1999), at 170µm from Dole et al. (2001), at 60µm from Hacking & Houck (1987), Gregorich et al. (1995), Bertin et al. (1997), Lonsdale et al. (1990), Saunders et al. (1991) and Rowan-Robinson et al. (1990) and at 850µm from Smail et al. 1997, Hughes et al. 1998, Barger et al. 1999, Blain et al. 1999, Borys et al. 2001, Scott et al. 2002 and Webb et al. 2002
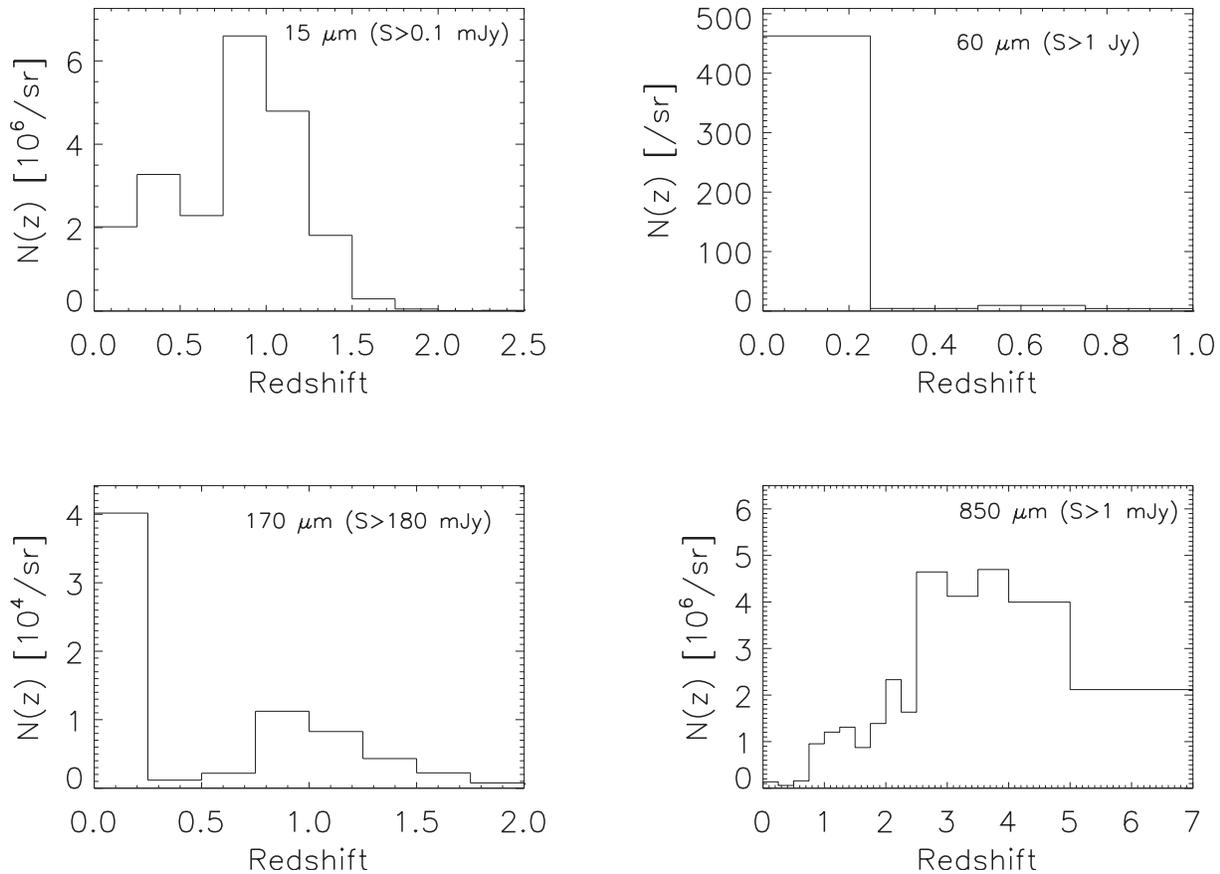
Fig. 2: Predicted redshift distribution for resolved sources at 15, 60, 170 and 850 μm.

2- Main unresolved problems in galaxy formation

There is a standard model of cosmological structure formation which is very successful in the description of the formation of large scale structures in the universe as very small adiabatic perturbations which originate from the early universe and are amplified by self-gravity. Random adiabatic fluctuations with a Harrison- Zeldovich spectrum and in the presence of a large amount of cold dark matter lead to a power spectrum on large scale which explains naturally the observations on the large scale structure of the distribution of galaxies. The predictions of such a model for the anisotropies of the cosmic microwave background (CMB) are in excellent agreement with the simplest of this class of models with a Euclidean geometry. The linear development of the density perturbations can be modeled by well known physics as long as the amplitude of the perturbations remains small and the different components are either well decoupled or rather close to thermal equilibrium.

The description of the nonlinear phase of the development of the density fluctuations and particularly of its baryonic component is more complicated. It requires modeling the thermal balance which in turn depends on chemistry and the hydrodynamics of the baryonic gas flowing in the dark matter potential wells.
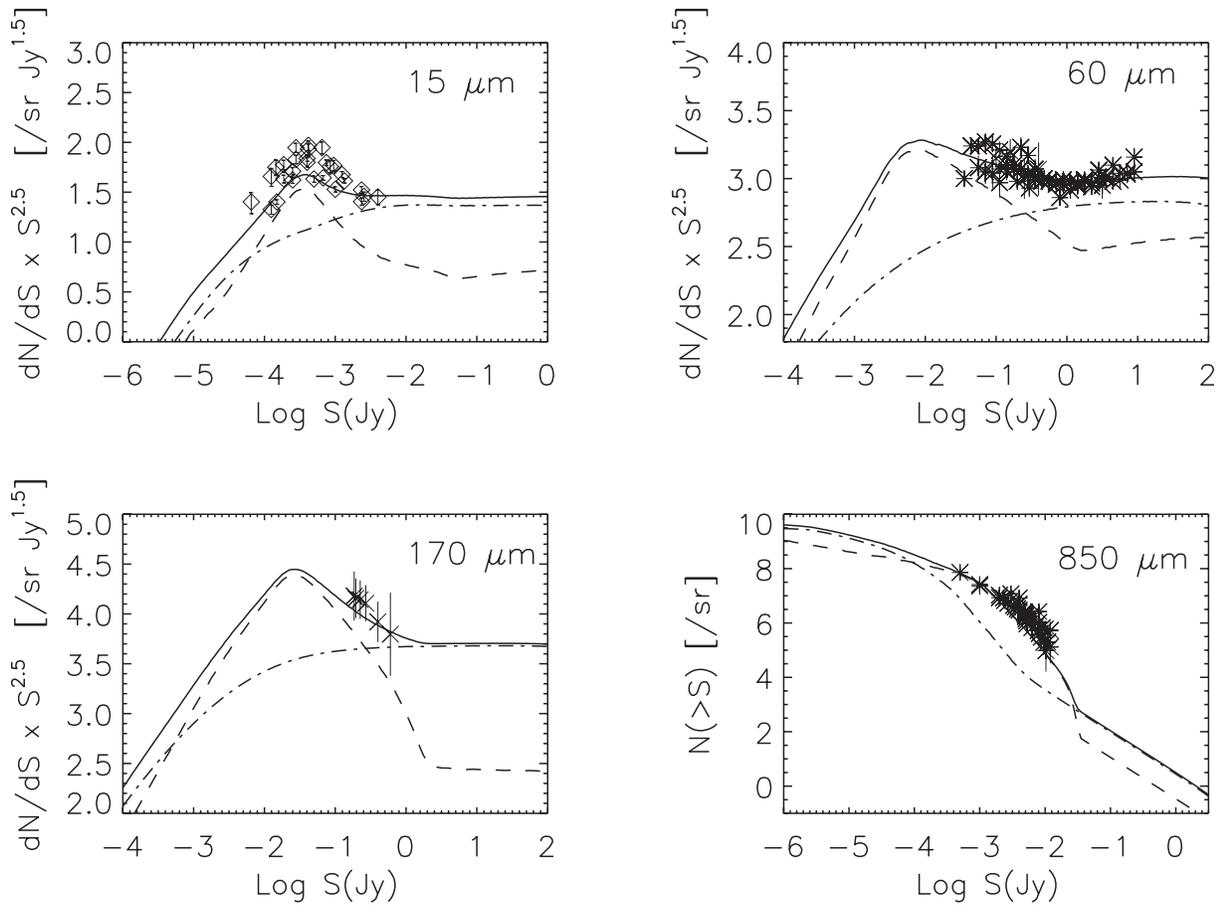
Fig. 1: Number counts at 15, 60, 170 and 850μm together with the model predictions (starburst galaxies: dash line, normal galaxies: dot-dash line, both normal and starburst galaxies: continuous line). Data at 15μm are from Elbaz et al. (1999), at 170μm from Dole et al. (2001), at 60μm from Hacking & Houck (1987), Gregorich et al. (1995), Bertin et al. (1997), Lonsdale et al. (1990), Saunders et al. (1991) and Rowan-Robinson et al. (1990) and at 850μm from Smail et al. 1997, Hughes et al. 1998, Barger et al. 1999, Blain et al. 1999, Borys et al. 2001, Scott et al. 2002 and Webb et al. 2002

We thus would like to observe the substructure and some of these very highly luminous sources at high redshifts at the time they assemble to constrain the physics of the subcomponents and of the merging process.

3- Observations needed at long wavelengths

At low redshift it has been discovered through the study of the luminous and ultra luminous infrared galaxies first found by IRAS and then by ISO (ISOCAM and ISOPHOT) and in millimeter (MAMBO-IRAM) and submillimeter (SCUBA-JCMT) deep extragalactic surveys that these sources are, most of the time, associated with mergers or interacting galaxies. Furthermore, as already mentioned, their energy output is dominated at high redshift by high luminosity sources. These are thus compelling reasons to study in detail such sources in the long wavelengths range as a tool to understand the merging process and the physics of the first non linear structures. The requirements on future missions will thus derive from the following:
- the wavelength range should be the one of the redshifted dust emission
- the dust emission dominates the emission of spiral and irregular galaxies at low z between 5 microns and 1 millimeter
- the corresponding wavelength range at the redshift of formation of the first structures is thus typically 40 microns to 1 cm with a peak emission in the 300 microns to 1 millimeter range.

The study of the star formation process in our own galaxy to understand the physics of the gravitational instability in a highly inhomogeneous, turbulent, magnetized medium relies also on searches for cold condensations radiating in the continuum at wavelength between 60 and 300 microns. Similarly in the galaxy formation process we'll need to be able to study the substructure of the early galaxies we study.

4- Requirements on future space missions at far-infrared and submillimeter wavelength

We leave aside the requirements on future missions coming from the study of the star formation which is a closely related astrophysics problem but outside the scope of this paper.

For galaxy formation at high z, the requirements are naturally split in 2 groups.

(I) The first one is related to the observations of high redshift mergers including the detailed observations of the subcomponents. The size of such subcomponents is typically that of a virialised structure of mass in the range 10^7 to 10^9 M_o i.e. 2 to 10 kpc or 0.25 to 1 arc seconds, in the redshift range 1 to 6.

The sensitivity requirement is such that luminosities of 10^{10} L_o (1 to 3 % of the total luminosity of the infrared galaxies dominating the luminosity function at large redshifts) should be detectable. Furthermore enough spectroscopic capabilities are required to study the physics of these subcomponents. Therefore, in the far-infrared, to observe the high redshift mergers and their subcomponents, interferometers are clearly needed. Part of the useful wavelength range,

thanks to the redshift effect, is accessible from the ground. The ALMA millimeter-submillimeter interferometer is clearly a very powerful tool for such studies. An interferometer working in space in the far infrared will probably be the next generation tool.

Single dish observations are possible only in the near or mid infrared. In this wavelength domain, NGST is clearly the tool for such observations. However, there are mainly two limitations: (i) the redshift up to which the dust component can be observed with NGST is limited to about 4 or 5 and (ii) the stellar component can be observed at higher z but the experience of the deep HST observations has shown how difficult it is to identify the galaxies which have most of their energy output in the far infrared.

(II) The second type of requirement comes from the necessity of finding the interesting high redshift sources which are very early mergers made of building blocks not affected much yet by star formation and evolution. When these are found, they must be studied in detail with telescopes in the first group.

Basing ourselves on the phenomenological model described in Sect. 1, we compute the density of sources as a function of redshift and luminosity. The results are given in Table 1. It gives the minimum size of the surveys to be conducted to find enough objects to do statistical studies of them. We see that, at high z, we need surveys of hundreds of square degrees!

The set of Table 2 gives, at different wavelengths, the flux of typical starburst galaxies for given redshifts. We see that the next generation telescopes will have enough sensitivity to detect such objects, especially at long wavelengths where the K-correction becomes negative. However, there observations will be limited by the confusion[1]. The angular resolution (and the associated telescope diameter) required to have observations not limited by the confusion, for different fractions of resolved IR background, are given in the set of Table 3. For example, to resolve 80% of the background, at 850 μm, we need a telescope diameter of about 113 m and at 350 μm a telescope diameter of about 23 m.

Table 1: Sky area in square degrees to be covered to detect more than N sources in a redshift range $\Delta z/z = 0.3$

redshift	$3\ 10^{11}$ Lo Sky area for N>100	$3\ 10^{12}$ Lo Sky area for N>10
0.9	0.5	0.6
3	5.4	1.8
5.1	71.6	37.7
7.2	469.6	314.8

[1]: A detailed study of the confusion at long wavelength is given in Lagache et al 2002.

Table 2: Flux of typical starburst galaxies for given redshift

170 μm

z	S_υ for L=3 10^{11} L_o mJy	S_υ for L=3 10^{12} L_o mJy
1	5.90	54.2
2	1.03	12.2.
6	0.028	0.39
8	0.011	0.14

550 μm

Z	S_υ for L=3 10^{11} L_o mJy	S_υ for L=3 10^{12} L_o mJy
1	1.0	7.0
2	0.72	5.4
6	0.25	2.5
8	0.12	1.5

350 μm

z	S_υ for L=3 10^{11} L_o mJy	S_υ for L=3 10^{12} L_o mJy
1	2.95	22.1
2	1.49	12.7
6	0.11	1.67
8	0.04	0.66

850 μm

Z	S_υ for L=3 10^{11} L_o mJy	S_υ for L=3 10^{12} L_o mJy
1	0.30	2.1
2	0.24	1.7
6	0.25	2.1
8	0.18	1.7

Table 3: Angular resolution and telescope diameter needed to have the confusion limit at a flux level S_{min} such that the sources above S_{min} contains 10, 30, 60 and 80 % of the cosmic background

170 μm	S_υ (mJy)	log N (/sr)	θ (arcsec)	D (meters)
80% of CIB	2	7.65	5.0	7.1
60% of CIB	6	7.27	7.9	4.4
30% of CIB	22	6.57	17.8	2.0
10% of CIB	65	5.57	50.7	0.7

350 μm	S_υ (mJy)	log N (/sr)	θ (arcsec)	D (meters)
80% of CIB	0.9	7.99	3.2	22.6
60% of CIB	2.5	7.69	4.9	14.8
30% of CIB	8	7.07	10.2	7.07
10% of CIB	18	6.29	24.4	2.96

550 μm	S_υ (mJy)	log N (/sr)	θ (arcsec)	D (meters)
80% of CIB	0.2	8.26	2.1	54.7
60% of CIB	0.8	7.92	3.7	30.8
30% of CIB	3	7.26	8.5	13.3
10% of CIB	6	6.56	19.1	6.0

850 μm	S_υ (mJy)	log N (/sr)	θ (arcsec)	D (meters)
80% of CIB	0.05	8.35	1.6	113
60% of CIB	0.2	8.03	3.0	59
30% of CIB	1	7.45	7.0	25
10% of CIB	2	6.77	15.3	12

5- Conclusions:

With the discovery of the extragalactic infrared background, and the first observations of its contributing galaxies, the past decade has brought new and surprising highlights in the formation and evolution of galaxies. The next generation of telescopes has to allow us to study the detailed physics of the IR galaxies in order to constrain the formation process. Since, up to now, there is no way to find the starburst galaxies and measure the star formation rate from other wavelength domains than in the Mid-IR, far-IR and submm, the observations from the next generation of IR/submm telescopes have to cover enough area with enough spatial resolution to be able to find these sources at high z. This is a very challenging task due to the very large telescopes needed to beat the confusion. This leaves open an alternative to be studied: how to make interferometers efficient enough to carry out large surveys?

Once the starburst galaxies are found, the detailed physics have to be done using the interferometers, both from the ground at large wavelengths (with ALMA) and from space, near the peak of the dust emission.

References:

Chary R., Elbaz D., 2001, ApJ 556, 562

Devriendt J., Guiderdoni B., 2000, A&A 363, 851

Dole H., Gispert R., Lagache G. et al., 2000, in «ISO view of a dusty Universe », Ringberg, Springer Lecture Notes

Dole H., Gispert R., Lagache G. et al., 2001, A&A 372, 364

Elbaz D., Cesarsky C.J., Chanial P. et al., 2002, A&A 384, 848

Franceschini A., Aussel H., Cesarsky C.J. et al. 2001, A&A 378, 1

Guiderdoni B., 2001, in « SF2A : Scientific Highlights 2001», Eds : Société Française d'Astronomie et d'Astrophysique, EdP-Sciences, Conference Series

Hatton S. et al., MNRAS, to be submitted

Hauser M.G., Dwek E., 2001, ARAA 37 , 249

Lagache G., Puget J.-P., 2000, A&A 355, 17

Lagache G., Dole H., Puget J.-L., 2002, to be submitted

Malkan M.A., Stecker F.W., 2001, ApJ 555, 641

Miville-Deschênes M.-A., Lagache G., Puget J.-P , 2002, A&A, in press

Pearson C.P., 2001, MNRAS 325, 1511

Rowan-Robinson M., 2001, New astronomy Reviews 45, 631

Serjeant S., Efstathiou A., Oliver S. et al.,, 2001, MNRAS 322, 262

Scott S.F., Fox M.J., Dunlop J.S. et al., 2002, MNRAS 3331, 839

Takeuchi T.T., Ishi T.T., Hirashita H., et al. 2001, PASJ 53? 37

Xu C., Lonsdale C.J., Shupe D.L., et al. 2001, ApJ 562, 179

Wang Y.P., Biermann P.L., 2000, A&A 356, 808

Wang Y.P., 2002, A&A 383, 755

Star Formation in Local Galaxies from the UV to the Far-IR

D. Calzetti

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, U.S.A.; calzetti@stsci.edu

Abstract

The star formation rates and histories of galaxies are gauged via a variety of indicators, from the direct stellar light (e.g., UV luminosity) to the dust-reprocessed light (far–infrared luminosity). Yet, despite all the advances brought by a number of missions, including infrared ones like IRAS and ISO, the cross–calibration between indicators at different wavelengths is still incomplete. The largely unknown physical connection between stellar populations (the heaters) and the surrounding ISM (the emitters) in the vast zoo of extragalactic environments is the main culprit. Addressing this issue is a crucial step for understanding the nature of high redshift galaxies and galaxy evolution in general, but it is currently hampered by the angular resolution mismatch between UV/optical and far–infrared/submillimiter telescopes. Far-IR/sub-mm missions with arcsec angular resolution or better are the obligatory course of action to solve this conundrum.

1 Introduction

One major missing piece in our quest to understand the evolution of galaxies is a theory of star formation (Silk 2000). What are the regulating (driving and inhibiting) mechanisms of star formation in galaxies? What determines the duration, and the spatial and temporal evolution of a star forming event? What are the effects of star formation feedback on the host galaxy? Are there multiple modes of star formation? If so, which are the physical mechanisms that drive each of them?

Answering these questions requires unraveling the temporal and spatial evolution of star forming or post-star-forming regions as a function of galactic location (e.g., Calzetti et al. 1997, Harris et al. 2001, Tremonti et al. 2001). Measurements of

star formation rates (SFRs) as a function of time and position are among the most powerful diagnostics to accomplish the task.

In recent years measurements of SFRs have also enjoyed popularity as a tool for tracking the evolution of star formation in galaxies as a function of cosmic time (Madau et al. 1996). Models (e.g., hierarchical CDM models, Baugh et al. 1998) can predict effectively the evolution of the dark haloes of structures, but only loosely the evolution of the stellar populations in galaxies. Again, this is because of the absence of a theory of star formation. Observational data attempt at filling this crucial gap, in order to link galaxy evolution to, e.g., the origin of the Hubble sequence.

Measurements at intermediate and high redshift have so far concentrated on tracers of current or 'instantaneous' SFR, as obtained from massive stars' emission. Observational constraints have mainly provided SFR diagnostics from the rest–frame UV (Madau et al. 1996, Lilly et al. 1996, Steidel et al. 1999, Cowie et al. 1999), nebular line emission (Glazebrook et al. 1999, Yan et al. 1999, Tepliz et al. 2000, Pettini et al. 2001, Moorwood et al. 2000), and mid/farIR and sub-mm emission (Hughes et al. 1998, Blain et al. 1999, Flores et al. 1999, Barger et al. 2000). General-purpose conversion formulae from the measured emission to SFRs are given in the review by Kennicutt (1998).

The UV emission from galaxies probes the stellar continuum of young, massive stars, and is, in principle, the most direct indicator of current SFR. It suffers, however, from large random and systematic uncertainties: dust obscuration is difficult to gauge and it decreases the UV emission of typical galaxy samples by an average factor 5–20 (Calzetti 1997, 2001, Buat et al. 2002, Bell 2002); the unknown ages of the probed stellar populations add to the systematic uncertainty, because old stellar populations can mimic the UV colors of a young, dusty stellar population. Variations in the metallicity and stellar IMF from galaxy to galaxy may provide an additional factor 2–3 random uncertainty.

The nebular line emission probes the gas ionized by the most massive stars; it is a more indirect probe of current SFR than the UV emission, because a fraction of the ionizing photons may escape the galaxy rather than excite the gas in it. Current observational evidence suggests that in local galaxies such fraction is small (Leitherer et al. 1995, Deharveng et al. 1997, but see Hurwitz et al. 1997), but the verdict is still out for galaxies at increasignly higher redshift (e.g., Steidel et al. 2001, see, however, Giallongo et al. 2002). Variations in the stellar IMF can in principle give random uncertainties as large as one order of magnitude; this uncertainty is narrowed down by recent studies indicating that regions of recent or current star formation are consistent with a 'universal' IMF, represented by a Salpeter IMF with upper mass cut-off around 50–100 $M_\odot$ (see Leitherer 1999 for a review). Finally, dust obscuration produces systematic underestimates of the SFR(Hα) by factors 2–3.

Measurements of galaxy emission between the mid-IR and the sub-mm are probing the stellar radiation reprocessed by dust. This indicator can be considered complementary to the two discussed above, in the sense that it will measure whatever UV and

optical stellar radiation does not directly come out of the galaxy. There are two main limitations to the SFR(IR) indicator. The first is observational in nature: while the UV and Hα probes are monochromatic, reliable determinations of SFR(IR) require sampling the *entire* spectral energy distribution (SED) of the dust emission, typically from a few μm all the way to 500–1000 μm. The second limitation is physical; not all the far-infrared emission from galaxies is directly associated with recent star formation, as older stars can also heat the dust, producing a generally cooler dust emission component ('cirrus'). To date, attempts at discriminating the different dust components and correlating each to a specific heating population has not yielded unique results (Lonsdale–Persson & Helou 1987, Trewhella et al. 2000; see, however, Devereux & Young 1991; also, Dunne & Eales 2001). The main consequence is a poorly constrained, systematic overestimate of the SFR from the far-infrared emission by factors between 0.3 and 5. For instance, only about 30% of the far-infrared emission in our own Galaxy is directly associated with current star formation (Sodroski et al. 1997).

2 Dust Corrections at UV–Optical Wavelengths

When estimating SFRs of galaxies from their UV–optical emission, the largest source of systematic uncertainty is dust obscuration; dust-induced underestimates of SFRs range from a factor of a few to a few tens depending on the wavelength and the galaxy observed. In the local Universe, there is a correlation between the SFR of a galaxy and the amount of dust obscuration it suffers, in the sense that the more active the galaxy, the dustier it is (Heckman et al. 1998, Hopkins et al. 2001); moderate (i.e, non-ULIRG) starbursts obey the correlation:

$$10^{A(UV)} \sim 250 SFR^{2.2}, \tag{1}$$

(Calzetti 2001). In other words, there are no local examples of a powerful starburst galaxy like Arp220, which has SFR$\sim$100-200 M$_\odot$/yr, that is also dust-free. If similar properties hold true at high redshift (e.g., Adelberger & Steidel 2000), presence of dust in galaxies coupled with observational biases may lead to a distorted view of the evolution of SFRs as a function of time.

Estimates of the dust reddening and obscuration affecting the SEDs of galaxies are especially challenging to obtain, because of the degeneracies produced by complex dust–star geometries (Witt, Thronson & Capuano 1992, Calzetti et al. 1994) and by multiple-age stellar populations. Selecting specific galaxy classes helps reducing the number of free parameters and, thus, of degeneracies; this is for instance the case of starburst–dominated galaxies, where the dust-age degeneracy is eliminated by selecting galaxies whose energy output is dominated by the recent star-forming event (Calzetti et al. 1994, Calzetti 1997). In this case, the free parameter of the stellar

populations age is removed, as the starburst–formed young population dominates the light output from the galaxy.

For those galaxies, systematic investigation of the impact of dust on the SEDs has led to a galaxy-wise, geometry–independent obscuration curve (Calzetti et al. 1994, 2000, Calzetti 1997, 2001, Meurer et al. 1999, Leitherer et al. 2002). Figure 1 shows a representation of that curve, compared to standard stellar extinction curves. The curve is applicable to the continuum continuum only, while the ionized gas emission suffers approximately twice the total obscuration of the stellar emission (Calzetti et al. 1994). This recipe is effective, for instance, at explaining the absence of differential reddening between the UV stellar flux and the Hβ nebular emission of z$\sim$3 galaxies, even in the presence of increasingly redder UV SEDs (Pettini et al. 2001, Figure 2).
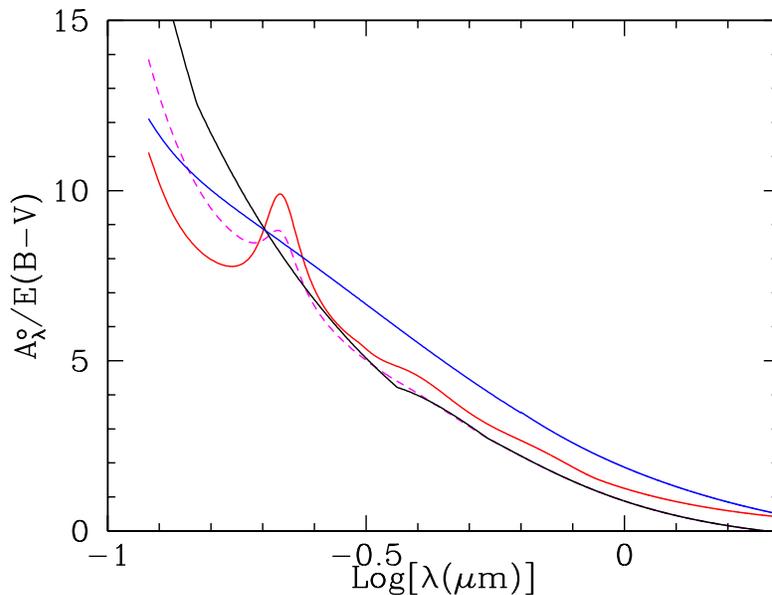
Figure 1: The starburst obscuration curve (blue) is plotted together with representative extinction curves of the Milky Way (red), the Large Magellanic Cloud (magenta) and the Small Magellanic Cloud (black). The Milky Way extinction curve is shown for R$_V$=3.1 and the Magellanic Clouds' extinction curves have R$_V$=2.7 (Gordon & Clayton 1998, Misselt et al. 1999). The comparison between the starburst obscuration curve and the extinction curves should not be taken at face value. The extinction curve measures strictly the optical depth of the dust between the observer and the star, while the obscuration curve expresses a general attenuation, in that it is derived from the spatially integrated colors and fluxes of galaxies and folds in one expression effects of extinction, scattering, and the geometrical distribution of the dust relative to the heaters.

However, the fact that the obscuration curve was obtained from a selected class of galaxies will place limitations on its applicability to other samples. For instance, samples of 'quiescent' star-forming galaxies show lower average obscuration values
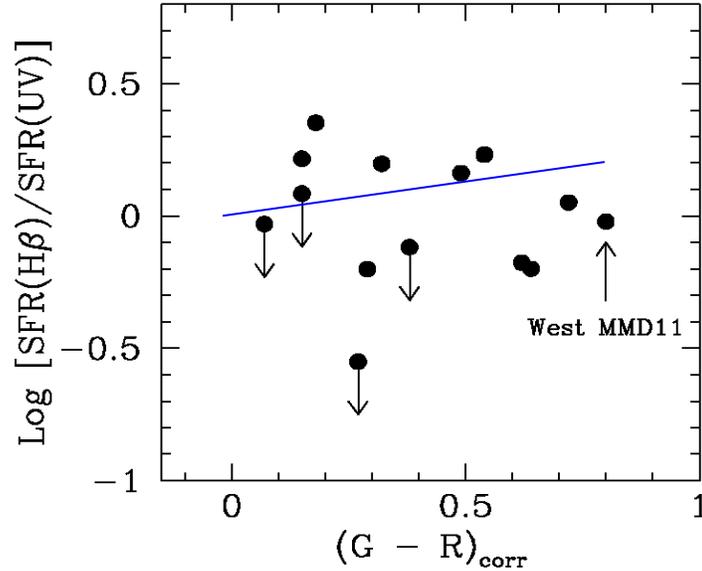
Figure 2: The ratio of the SFRs derived from the Hβ nebulal emission and the UV stellar continuum, respectively, for a small sample of z$\sim$3 galaxies (figure from Pettini et al. 2001). The ratios are uncorrected for effects of dust reddening, and are plotted as a function of the observed *restframe* UV colors of the galaxies, (G−R)$_{corr}$: increasing values of (G−R) correspond to redder UV SEDs. The continuous line is the expected trend for dust distributions that follow the starburst obscuration curve (Calzetti et al. 1994). If the absence of a trend in the SFR(Hβ)/SFR(UV) ratio were interpreted as evidence for absence of dust in the high redshift galaxies, the only possible explanation for the red UV SEDs would be aging of the stellar population; however, this would imply ages >100 Myr for the UV stellar populations of the reddest galaxies, which are incompatible with the presence of Hβ emission.

than starbursts, by about a factor of 2 in the UV (Buat et al. 2002), although their UV SEDs can be as red as those of the most dust–affected starbursts (Bell 2002). Mixtures of stellar populations of varying ages in the more quiescent samples may play a large role in the observed discrepancy with the dust reddening/obscuration measurements of the starburst samples.

Radiative transfer models have evolved in recent years into increasingly sophisticated tools to treat dust reddening and obscuration in a general galaxy (e.g., Efstathiou et al. 2000, Gordon et al. 2001). In principle, these models do not require all sample galaxies to be dominated by a common, known stellar population. They overcome some of the limitations of the approach used in the case of starburst galaxies, but they have shortcomings of their own. Among others, a large number of free parameters needs to be constrained either by a large body of observational data or by simplifying assumptions on the dust/star geometry and/or on the stellar population mix. In addition, it is not always clear whether the models or assumptions that work

for a particular galaxy or class of galaxies are applicable to samples with different characteristics.

In the end, modelling the dust absorption of UV–optical stellar radiation in galaxies has intrinsic difficulties that defeat generalization, and progress on this front is bound to be slow. Where progress could be leaping forward in giant steps is at the other end of the electromagnetic spectrum: in the often overlooked far–infrared/submillimeter region, where the bulk of the dust emission from galaxies emerges.

3   Dust Emission from Galaxies: Terra Incognita

The limited sensitivity, angular resolution ($\sim$arcmin), and array size (one to a few tens of pixels in side) of past mid/far–infrared missions, like IRAS and ISO, have provided us with only a cursory picture of the Universe in the 'dust emission' regime, and have left behind more questions than answers. For instance, we don't know what the fundamental parameters are that govern the shape of the dust emission SED; what the relative importance of different stellar populations and of the dust content and distribution is in determining that shape; whether global galaxy properties, such as metallicity, have a role. Ultimately, we would like to unveil the *physical and spatial* connection between heaters (stellar populations) and emitters (the multiple dust and gas components), and to quantify the physical relationship between SFR and ISM properties.

Addressing the questions above is one of the main science goals of the SIRTF Legacy Program SINGS, the SIRTF Nearby Galaxy Survey (P.I. R. Kennicutt). The unprecedented angular resolution and expected sensitivity of SIRTF, indeed, promise major steps forward in the direction of tackling the heaters–emitters connection. SINGS will obtain images and spectra in the range 4–160 μm of a physically–selected sample of 75 galaxies closer than about 30 Mpc. The angular resolution offered by the telescope ranges from about $2''$ in the mid–IR to $\sim40''$ at 160 μm, and corresponds to linear scales in the range 50 pc to 1 kpc for a galaxy at 5 Mpc distance. In terms of sensitivity, a 10 sec exposure will provide a 10 σ detection of a 10^4 to 10^5 M$_\odot$ cloud, depending on the wavelength. In the mid-IR, thus, we will be resolving basic galaxy constituents, like HII regions.

The SINGS sample has been selected to ensure the full parameter space coverage of type (morphology, SFR, gas fraction), mass (luminosity, metallicity), and infrared-to-optical luminosity ratio. Table 1 lists the range of the main galaxy parameters and Figure 3 provides a graphic representation of some of the primary selection parameters. To aid relating stellar populations and gas distribution to dust emission components, we are complementing the SIRTF observations with a full set of ancillary data: UV (GALEX) imaging, optical/near–IR broad-band imaging, Hα imaging, optical spectroscopy, and CO and HI mapping.

Although SIRTF will represent a leap forward in our understanding of dust emis-

Table 1: Parameter Space of the SINGS Sample

Parameter	Range
Hubble Type	E – Irr
L_V (L^*)	$0.0003 - 10$
L_{IR}/L_V	$<0.02 - 42$
F(60 μm)/F(100 μm)	$0.16 - 1.2$
O/H (O/H)$_\odot$	$0.05 - 3$
M_{gas}/M_{star}	$<0.001 - 5$
SFR ($M_\odot$/yr)	$0 - 15$

sion from galaxies, still the telescope's capabilities are limited by the mirror's small size and by the wavelength coverage. Beyond 20 μm, the angular resolution degrades to $>5''$, and resolving a nearby galaxy's population constituents becomes increasingly difficult at longer and longer wavelengths. In addition, the longest wavelength probed by SIRTF is 160 μm, which is barely beyond the peak emission of dust from galaxies.

Ideally, one would need to achieve at least the same level of performance, in terms of angular resolution, as ground–based optical images, and couple it with large–format arrays. This would offer a one–to–one match with the wealth of visible/near–IR data already available for galaxies. With a $1''$ resolution at 100 μm, the peak region of dust emission could be probed in 50 pc extended regions as far as 10 Mpc away, thus resolving galaxy constituents in a representative volume of the local Universe.

In the long wavelength regime, coverage to about 350 μm would sample the Rayleigh–Jeans tail of cool ($T_{dust} \sim 20$ K, e.g. cirrus) dust emission all the way to about 10% of its peak intensity. Past and current far–infrared missions have concentrated on warmer dust constituents; however, the bulk of a galaxy's dust mass is carried by its cool components. Cirrus are a crucial ingredient of the make–up of 'quiescent' galaxies' dust, but even active galaxies like ULIRGs have been shown to have substantial cool dust emission (Dunne et al. 2001). An accurate calibration of SFR(IR) will be achieved only by studying the spatial and physical relation of this component to other dust constituents and to the relevant heating populations.

4   Summary

Despite that about half or more of the energy from galaxies is re-processed by dust into the far–infrared and submillimiter (Hauser & Dwek 2001), investigations of the dust emission from galaxies have so far suffered from three main shortcomings: (1) limited angular resolution, (2) small–format arrays, and (3) low sensitivity. All three together have been contributing to widen the gap in our understanding between galaxies stellar
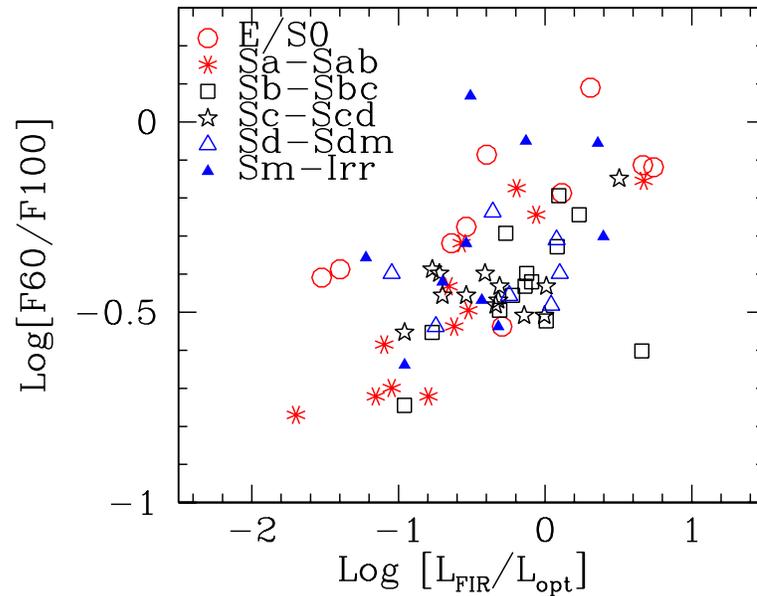
Figure 3: The dust temperature proxy, F(60)/F(100), is plotted as a function of the IR–to–optical luminosity ratio for the 75 SINGS galaxies. F(60)/F(100) is the ratio of the IRAS fluxes at 60 μm and 100 μm. Different plotting symbols identify different morphological types, as indicated in the figure's legend. Coverage in the parameter space of galaxy morphology, dust temperature, and infrared–to–optical ratio are three of the main selection criteria of the SINGS sample.

populations and their dust emission.

Studies that exploit the direct stellar light from galaxies are limited by their capability to model and remove the effects of dust from the observables. Progress will be slow on this front, because of the difficulty of models to properly account for varying dust–star geometries and multiple–age stellar populations in general cases.

On the contrary, progress in the far–infrared and submillimiter detector technology coupled with large–mirror space missions will allow us to readily answer basic questions on the physical connection between stellar populations and the dust they heat, and will provide major leaps forward in our understanding of the physics of a major constituent of galaxies' energy balance.

Acknowledgements.

I would like to thank the Organizing Committe of the Second Workshop on New Concepts for Far–IR and Submillimiter Space Astronomy, and in particular David Leisawitz, for the invitation to this very stimulating meeting.

References

[1] Adelberger, K.L., & Steidel, C.C., 2000, ApJ, 544, 218

[2] Barger, A.J., Cowie, L.L., & Richards, E.A. 2000, AJ, 119, 2092

[3] Baugh, C.M., Cole, S., Frenk, C.S., & Lacey, C.G. 1998, ApJ, 498, 504

[4] Bell, E.F. 2002, ApJ, submitted

[5] Blain, A.W., Jameson, A., Smail, I., Longair, M.S., Kneib, J.-P., & Ivison, R.J. 1999, MNRAS, 309, 715

[6] Buat, V. Boselli, A., Gavazzi, G., & Bonfanti, C. 2002, A&A, 2002, in press

[7] Calzetti, D., Kinney, A.L., & Storchi-Bergmann, T. 1994, ApJ, 429, 582

[8] Calzetti, D. 1997, AJ, 113, 162

[9] Calzetti, D., Meurer, G.R., Bohlin, R.C., Garnett, D.R., Kinney, A.L., Leitherer, C., & Storchi-Bergmann, T. 1997, AJ, 114, 1834

[10] Calzetti, D., Armus, L., Bohlin, R.C., Kinney, A.L., Koornneef, J., Storchi-Bergmann, T., & Wyse, R.F.G. 2000, ApJ, 533, 682

[11] Calzetti, D., 2001, PASP, 113, 1449

[12] Cowie, L.L, Songaila, A., & Barger, A.J. 1999, AJ, 118, 603

[13] Deharveng, J.-M., Faiesse, S., Milliard, B., & Le Brun, V. 1997, A&A, 325, 1259

[14] Devereux, N.A., & Young, J.S. 1991, ApJ, 371, 515

[15] Dunne, L., & Eales, S.A. 2001, MNRAS, 327, 697

[16] Efstathiou, A., Rowan-Robinson, M., & Siebenmorgen, R. 2000, MNRAS, 313, 734

[17] Flores, H., Hammer, F., Thuan, T.X., Césarski, C., Desert, F.X., Omont, A., Lilly, S.J., Eales, S., Crampton, D., & Le Fèvre, O., 1999, ApJ, 517, 148

[18] Giallongo, E., Cristiani, S., D'Odorico, S., & Fontana, A., 2002, ApJ Letters, in press (astroph/0202293)

[19] Glazebrook, K., Blake, C., Economou, F., Lilly, S., & Colless, M. 1999, MNRAS, 306, 843

[20] Gordon, K.A., & Clayton, G.C. 1998, ApJ, 500, 816

[Gordon et al.(2001)] Gordon, K.D., Misselt, K.A., Witt, A.N., & Clayton, G.C. 2001, ApJ, 551, 269

[21] Harris, J., Calzetti, D., Gallagher, J.S., Conselice, C.J., & Smith, D.A. 2001, AJ, 122, 3046

[22] Hauser, M.G., & Dwek, E. 2001, ARAA, 39, 249

[23] Heckman, T.M., Robert, C., Leitherer, C., Garnett, D.R., & van der Rydt, F. 1998, ApJ, 503, 646

[24] Hopkins, A.M., Connolly, A.J., Haarsma, D.B., & Cram, L.E. 2001, AJ, 122, 288

[25] Hughes, D., Serjeant, S., Dunlop, J., Rowan-Robinson, M., Blain, A., Mann, R.G., Ivison, R., Peacock, J., Efstathiou, A., Gear, W., Oliver, S., Lawrence, A., Longair, M., Goldschmidt, P., & Jenness, T. 1998, Nature, 394, 24

[26] Hurwitz, M., Jelinsky, P., & Van Dyke Dixon, W. 1997, ApJ, 481, L31

[27] Kennicutt Jr., R.C. 1998, ARAA, 36, 189

[28] Leitherer, C., Ferguson, H.C., Heckman, T.M., & Lowenthal, J.D. 1995, ApJ, 454, L19

[29] Leitherer, C. 1999, in Galaxy Interactions at Low and High Redshift, 186th IAU Symp., J.E. Barnes & D.B. Sanders eds. (Dordrecht: KAP), 243

[30] Leitherer, C., Li, I-H., Calzetti, D., & Heckman, T.M. 2002, ApJ Supp., in press.

[31] Lilly, S.J., Le Févre, O., Hammer, F. & Crampton, D. 1996, ApJ, 460, L1

[32] Londale-Persson, C.J., & Helou, G. 1987, ApJ, 314, 513

[33] Madau, P., Ferguson, H.C., Dickinson, M.E., Giavalisco, M., Steidel, C.C., & Fruchter, A. 1996, MNRAS, 283, 1388

[34] Meurer, G. R., Heckman, T.M., & Calzetti, D. 1999, ApJ, 521, 64

[35] Misselt, K.A., Clayton, G.C., & Gordon, K.D., 1999, ApJ, 515, 128

[36] Moorwood, A.F.M., van der Werf, P.P., Cuby, J.C., & Oliva, E. 2000, A&A, 362, 9

[37] Pettini, M., Shapley, A.E., Steidel, C.C., Cuby, J.-G., Dickinson, M., Moorwood, A.F.M., Adelberger, K., & Giavalisco, M. 2001, ApJ, 554, 981

[38] Silk, J. 2000, PASP, 112, 1003

[39] Sodroski, T.J., Odegard, N., Arendt, R.G., Dwek, E., Weiland, J.L., Hauser, M.G., & Kelsall, T. 1997, ApJ, 480, 173

[40] Steidel, C.C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, ApJ, 519, 1

[41] Steidel, C.C., Pettini, M., & Adelberger, K. 2001, ApJ, 546, 665

[42] Teplitz, H.I., Malkan, M.A., Steidel, C.C., McLean, I.S., Becklin, E.E., et al. 2000a, ApJ, 542, 18

[43] Tremonti, C.A., Calzetti, D., Leitherer, C., Heckman, T.M. 2001, ApJ, 555, 322

[44] Trewhella, M., Davies, J.I., Alton, P.B., Bianchi, S., & Madore, B.F. 2000, ApJ, 543, 153

[45] Witt, A.N., Thronson, H.A., & Capuano, J.M. 1992, ApJ, 393, 611

[46] Yan, L., McCarthy, P.J., Freudling, W., Teplitz, H.J., Malumuth, E.M., Weymann, R.J., & Malkan, M.A. 1999, ApJ, 519, L47

Observing Star and Planet Formation in the Submillimeter and Far Infrared

Harold W. Yorke

Jet Propulsion Laboratory, California Institute of Technology,
MS 169-506, 4800 Oak Grove Drive, Pasadena, CA 91109
        Harold.Yorke jpl.nasa.gov

Abstract. Stars form in the densest parts of cold interstellar clouds which — due to presence of obscuring dust — cannot be observed with optical telescopes. Recent rapid progress in understanding how stars and planets are formed has gone hand in hand with our ability to observe extremely young systems in the infrared and (sub-)millimeter spectral regimes. The detections and silhouetted imaging of disks around young objects in the visible and NIR have demonstrated the common occurrence of circumstellar disks and their associated jets and outflows in star forming regions. However, in order to obtain quantitative information pertaining to even earlier evolutionary phases, studies at longer wavelengths are necessary. From spectro-photometric imaging at all wavelengths we learn about the temperature and density structure of the young stellar environment. From narrow band imaging in the far infrared and submillimeter spectral regimes we can learn much about the velocity structure and the chemical makeup (pre-biotic material) of the planet-forming regions.

1.   Introduction

Stars are the basic building blocks of larger scale structures in the visible universe. Their birth, aging and deaths affect the evolution of their host galaxies in a multitude of ways which make them key players in the universe's evolution. Understanding the formation of stars and planets and their early evolution is a necessary prerequisite for understanding the cycles of matter in the universe and the sequence of events which has led to the origin of life on Earth.

Some of the key issues that need to be addressed are:

- Formation of the first stars in the early universe and their evolution
    - How and when did the first stars form?
    - What is the subsequent star formation history?
    - Is there an unknown population of high-z IR galaxies?

- Formation of stars and the physics of the interstellar medium
    - How do stars form out of the interstellar medium?
    - Circulation / enrichment / chemical processes of the interstellar medium
    - Detailed studies of nearby (resolvable) protostars, star forming regions, "mini-starbursts", starbursts — each as templates

- Formation of new solar systems in protostellar disks
    - How are planetisimals built up out of interstellar dust?
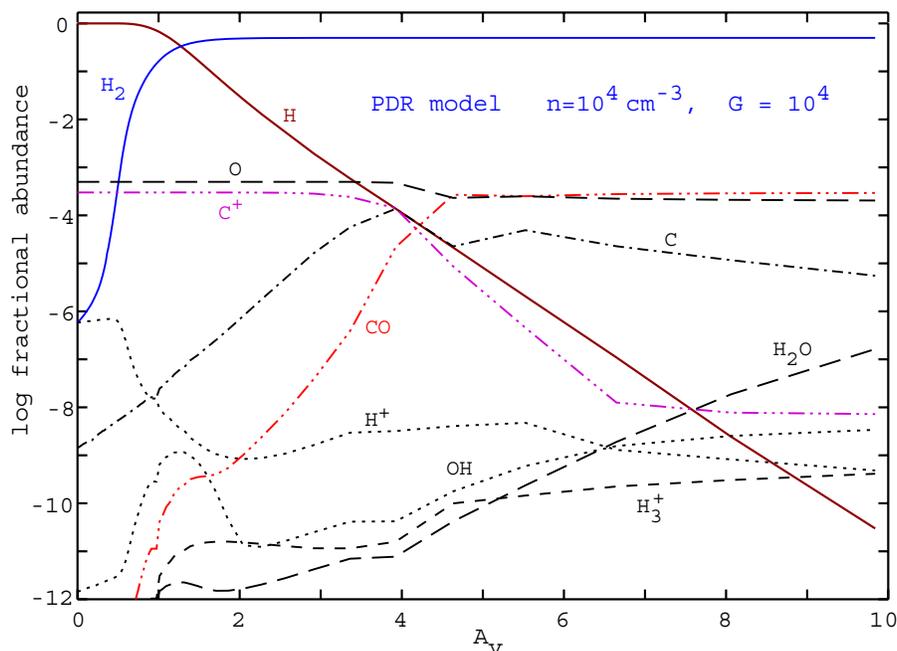
Figure 1.   Example of chemical structure within a constant density "Photodissociated Region".

    – What is the chemical state (pre-biotic?) of material that enters into planetary atmospheres?

- Cometary, planetary, and satellite bodies and atmospheres
  - History of the solar system, pristine material in comets
  - Minor bodies in the solar system

Theoretical studies of the formation of circumstellar disks via infall and their early evolution stress their importance not only as a reservoir of material to be partially accreted onto the (proto-)stellar cores and partially ejected in a collimated outflow, but as the basic environment for the growth and evolution of dust grains and ultimately for the formation of planetesimals and planets. Presumably many of the building blocks of pre-biotic chemistry, large organic molecules, are either created in this gaseous environment and subsequently adsorbed onto solid particles or are formed directly on the surfaces of the solid material.

Newly formed massive stars illuminate and begin to modify their immediate environment in a manner that is observable as the "PDR" phenomenon (see Fig. 1). Analysis of low excitation fine structure lines of these PDRs supplies vital quantitative information on global star formation in extragalactic sources (c.f. Table 1 and Figs. 2 and 3). The careful interpretation of observations of PDRs on a much smaller scale — protostellar disks being photoionized and UV-photoheated by a nearby hot star — give us quantitative information on the structure and evolution of these disks (see Fig. 4).

Further progress in our understanding of star formation is inhibited by our *inability* to make high spectrally resolved and high spatially resolved images of young objects during early evolutionary phases. The warm material in disks, infall regions, and jets cools predominantly through low excitation fine structure lines and molecular rotational transitions, many

Table 1.   Bridging the gap from local high mass star formation to Starburst Galaxies

Object	Distance	# of OB stars	Spatial scale
Orion	0.5 kpc	4	0.001pc
NGC3603	7 kpc	10's	0.015pc
30 Dor	55 kpc	100s	0.1pc
NGC248	2 Mpc	1000's	4pc
Z=1	4 Gpc	10,000's	8 kpc
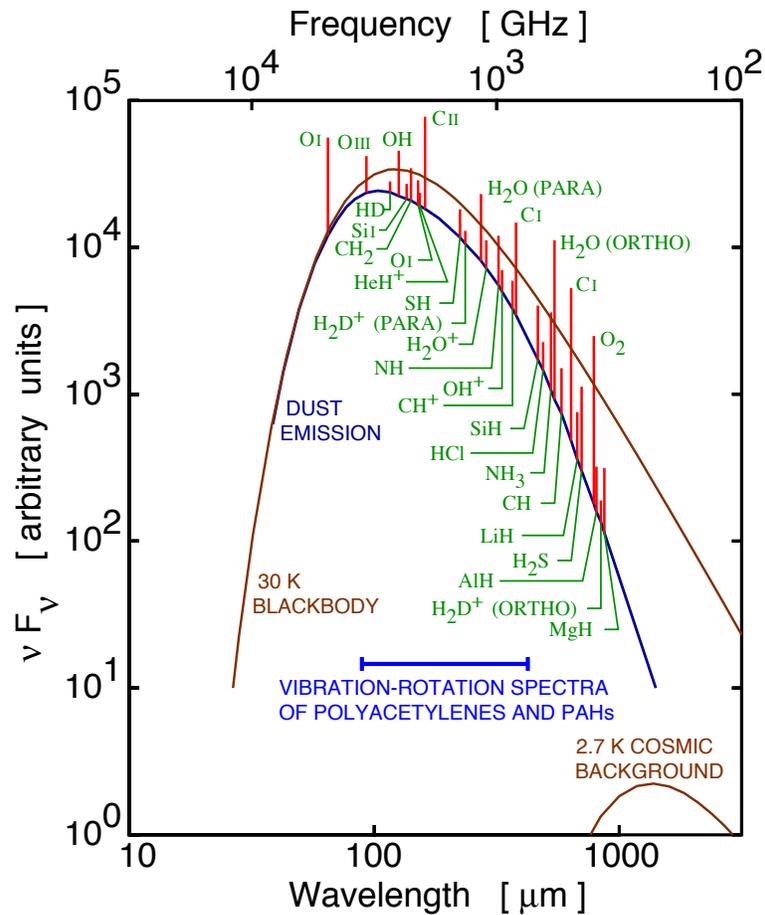
Figure 2.   Schematic spectrum of a star forming region which is generally dominated by the far infrared continuum emission of dust and line emission from fine structure transitions, molecules, and dust spectral feature.
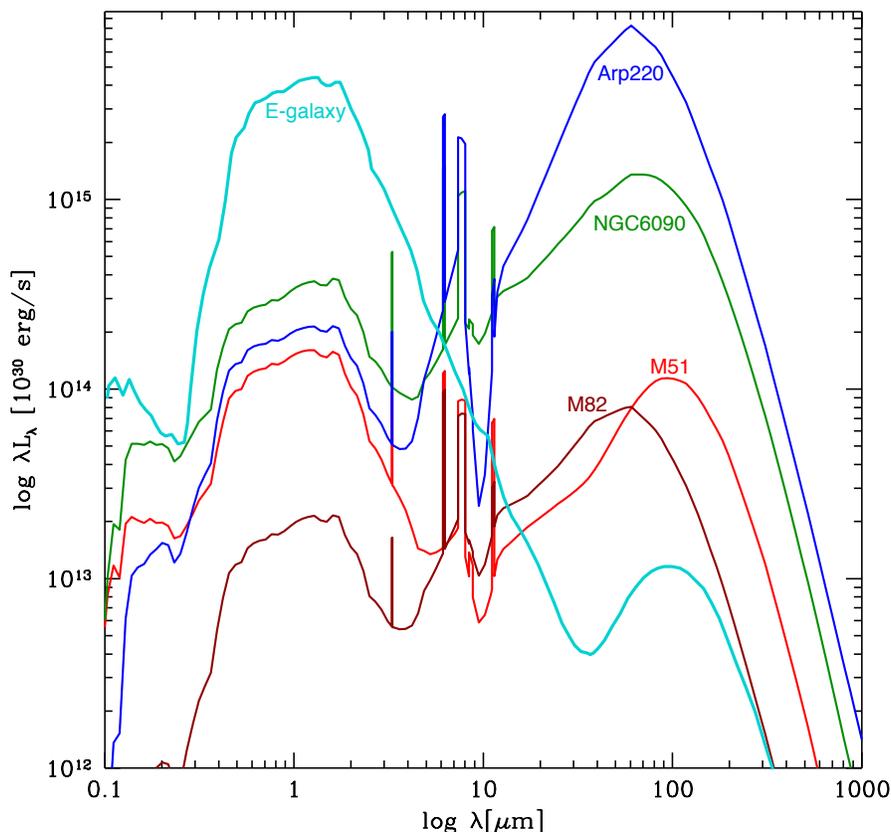
Figure 3.   Spectra of galaxies experiencing differing rates of star formation activity (adapted from Silva et al., 1998). Note that the ratio of UV+optical+near-infrared to far-infrared emission can vary by many orders of magnitude. Estimates of the total luminosity of a galaxy which do not include the far infrared are highly unreliable.

atmospherically blocked, as well as dust continuum emission in the FIR and submillimeter spectral regimes. High quality observations in this wavelength range in conjunction with detailed modeling can provide unique insights into the distributions of temperature, density, velocity, chemical composition, dust composition and ionization state during various phases of star and planet formation. The complexity of the lines of H_2O as observed with ISO in star forming regions (c.f. Fig. 5) and how they change strongly from position to position provide a compelling argument for improved spatial resolution.

The interactions of the (proto-)stars, possible close companions, mass infall, disks, jets and interstellar material in their close vicinity guarantee the occurrence of structure over a wide spectrum of spatial scales. Using the values (sensitivity, spectral and spatial resolution) believed attainable for space-based FIR/submillimeter interferometers within 10 years, the
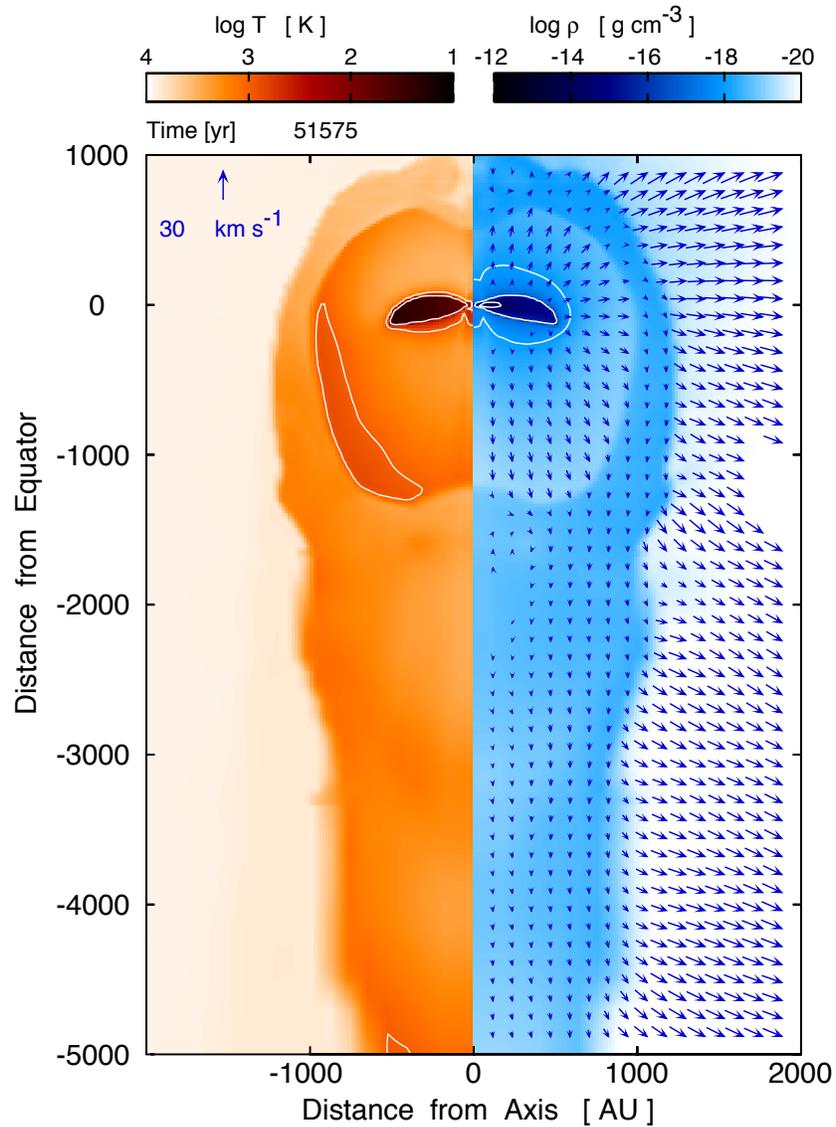
Figure 4. Temperature (left), density, and velocity structure (right) of a protostellar disk undergoing photoevaporation by an O5 star located 0.5 pc "above" the disks central star at (0,0). The vertical "z" axis is assumed to be an axis of rotational symmetry (adapted from Richling & Yorke 2000

outer regions of post-planet forming debris disks will be observable down to 1 "zodi"[1] — complimenting what can be done with TPF at shorter wavelengths in regards to the inner disk regions. Jupiter and possibly Saturn would be observable around stars out to 15 pc, perhaps their associated gaps as well. During earlier phases, jets, infall motions and disk accretion shocks are clearly detectable. With interferometric studies of proper motions (30 km s^{-1}= 1 mas per week at 130 pc) these flows and structures can be traced as they evolve over the course of weeks.

When discussing the mission capabilities which would be highly advantageous for these types of observations, it is important to note 1) there is a continuous improvement in the quality of science which can be done by improving sensitivity as well as angular/spectral resolution, 2) there are important "quantum leaps" in the types of problems which can be addressed and 3) any improvement by orders of magnitude in sensitivity and angular/spectral resolution has the potential of yielding unexpected significant discoveries. We can address points 1) and 2), but by definition can say little about point 3).

With even better spectral resolution the identification of individual complex organic species such as the Polycyclic Aromatic Hydrocarbons (PAHs) is possible. The quality of science attainable is not limited by this particular choice spatial resolution either. With increasing spatial resolution it eventually will be possible to image directly the jet formation and innermost collimation zones — indeed, protostellar jets provide us with the nearest examples of this generic phenomenon generally occurring in accretion disks. Thus by extension we will learn about the jet mechanisms in more distant active galactic nuclei and black holes.

References

Richling, S., Yorke, H.W., 2000, ApJ, 539, 258

Silva, L., Granato, G.L., Bressan, A., Danese, L., 1998, ApJ, 509, 103

---

[1]One "zodi" corresponds to the emission of interplanetary dust in our own Solar System
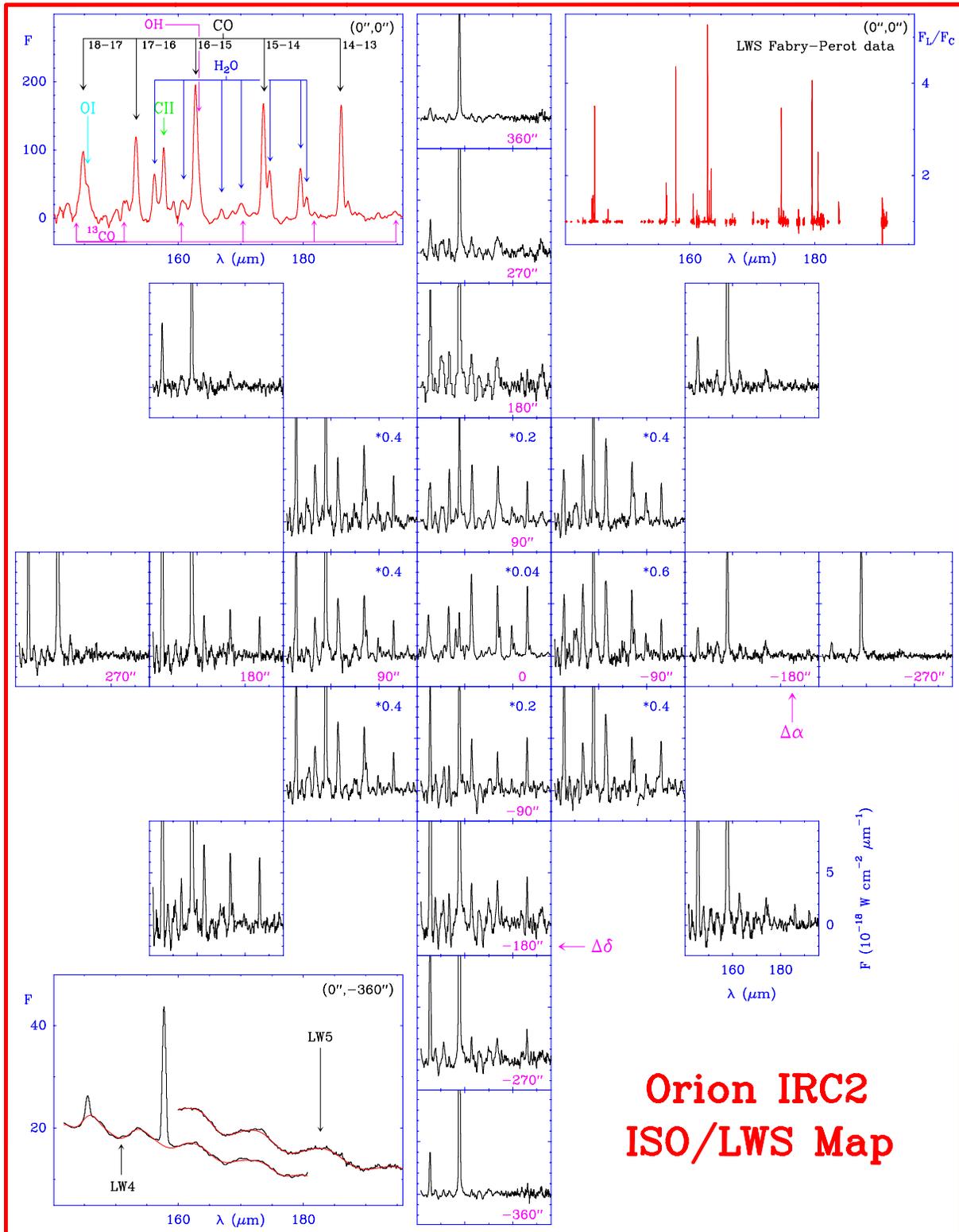
Figure 5.   H$_2$O emission in the Orion region

Future developments in far-IR and submm space observations

Andrew W. Blain

awb@astro.caltech.edu.

Astronomy, Caltech 105-24, Pasadena, CA91125, USA.

Abstract Recent observations have highlighted the importance of mid- and far-infrared surveys for studying galaxy evolution. A crucial step forward will be provided by moving to resolve galaxies at these wavelengths, and providing resolution better than 10 arcsec in order to reduce the effects of source confusion.

Introduction

Important new information about the way that galaxies form and evolve has been provided by the measurement of the thermal emission from the dust in moderate and high-redshift galaxies that is heated by star-formation activity and active galactic nuclei (AGNs), The relevant data was obtained at wavelengths of 15 and $175\,\mu$m using the cryogenic *Infrared Space Observatory (ISO)*, and using first-generation bolometer array cameras on ground based telescopes: at $450/850$-μm using the SCUBA camera at the JCMT and at 1.2-mm using the MAMBO camera at the IRAM 30-m telescope. These instruments detected a previously unknown population of luminous dust-enshrouded galaxies at moderate to high redshifts, with spectral properties that appear to be similar to those of the most luminous low-redshift *IRAS* galaxies.

With samples of only several hundred such galaxies known, it is difficult to be sure of their detailed properties, and especially of their properties in other wavebands. This is both because their optical emission is suppressed by dust absorption, and their positions are not known more accurately than a few arcsec. Counterparts at other wavelengths are still not identified for the majority of these galaxies.

Planned missions

The charted route into the future for infrared space astronomy flows from *SIRTF* in early 2003, through *ASTRO-F* to the *Herschel Space Observatory* and *Planck Surveyor*. These forthcoming missions will provide much more information about the population and individual properties of strongly evolving infrared galaxies.

SIRTF has improved resolution as compared to *ISO*, a fully-sampled focal plane, and a larger field of view due to larger format detectors. It also has an advanced spectrograph for wavelengths shorter than $40\,\mu$m. Imaging and spectroscopy of both nearby and distant galaxies in existing catalogs using *SIRTF* will provide much more information about their astrophysics, including their power sources from mid-infrared Ne emission lines (as demonstrated by Genzel et al. using data from *ISO*). Wide-field, deep survey projects will provide new samples of hundreds of thousands of infrared-selected galaxies for future study, and thus generate the first representative view of the dust-enshrouded Universe at moderate redshift ($z \simeq 1$–2). By the end of its mission, *SIRTF* should have provided the basis for determining a detailed form of evolution of dust-enshrouded galaxies through the peak of

both star-formation and AGN activity that is inferred to lie at $z \simeq 2$. To fully realize this goal, redshifts will be required for many of the detected galaxies. *ASTRO-F* will provide both a shallow all-sky survey to finally supercede *IRAS* after 20 very successful years, and to fulfil an observatory role both to continue the work of *SIRTF*, and to capitalize on the developments and discoveries made in follow-up observations after *SIRTF*'s lifetime ends.

In the atmosphere, the SOFIA airborne observatory and balloon-borne facilities, will provide additional information on individual galaxies, with better spatial and spectral resolution at a range of infrared wavelengths. *NGST* may observe in the mid-IR, and will certainly provide a complementary view of the traces of emerging UV, visible and near-IR light from high-redshift dusty galaxies. On a longer timescale, *Planck Surveyor* will provide a submm-wave map of the sky at a resolution of 5 arcmin, an equivalent to the *IRAS* survey, and *Herschel* will carry out both far-infrared and submm surveys at unprecedented depths, as well as providing follow-up imaging and spectroscopy of many galaxies cataloged by *SIRTF*.

There are key opportunities for space facilities beyond these horizons. There are three key features that they should be able to provide, and to observe.

Resolution

Telescopes with large cold apertures or interferometers are required to resolve and study the astrophysics of known distant galaxies, and to make very deep surveys for new samples of faint high-redshift galaxies at wavelengths from 30–300 mum. This a unique opportunity for space-borne instruments: telescopes that observe through and exchange heat with even a trace of the atmosphere would be too warm to achieve the sensitivity required for these observations.

Fine angular resolution suppresses confusion noise to better match the low thermal/detector noise levels that can be achieved in long integrations. Countering confusion is crucial for making deep surveys, especially at wavelengths longer than 100 μm, out to about 300 μm where the background emission from galaxies is strongest. The confusion noise conditions are expected to improve rapidly if the observing beam can be made smaller than 10 arcsec. At 100 μm 5-arcsec resolution corresponds to an aperture of about 4 m well-matched to the proposed *SPICA* and *SAFIR*. At longer wavelengths, a larger aperture is necessary to achieve this condition. Ultimately, it will be essential to image sub-structure in galaxies in order to probe the earliest objects. The effective aperture will then need to increase by about an order of magnitude. In order to improve significantly on *Herschel* images at longer wavelengths requires almost an order of magnitude increase in aperture/baseline. Arcsec resolution will be crucial if the first metal-rich galaxies are to be detected at $z \sim 5$ or more.

The ability to resolve distant galaxies is also vital for understanding their power sources, lifetimes and dynamical state. Resolution on a scale of a few arcsec will allow high-redshift galaxies detected at other wavelengths to be studied in the far-infrared waveband, moving us closer to a complete multiwaveband picture of galaxy formation. At 70 μm an L^* galaxy at $z \simeq 3$ is expected to produce a flux density of about 10 μJy, close to the confusion limit of a 5-m telescope.

Certain aspects of this resolved imaging can be carried out using ALMA from the ground, but not at rest-frame mid-infrared wavelengths where fine-structure transitions can be detected from the redshift range where the luminosity of galaxies appears to be peak.

Spectroscopy

Adding spectral information can reduce the effects of confusion in even a small aperture. In space, there are neither restrictions on the choice of observing frequencies, nor interruptions to a continuous spectral coverage. Determining redshifts for, and measuring diagnostic lines to indicate the energy source that heats the dust in existing samples of dust-enshrouded high-redshift galaxies is going to be very challenging for large samples of galaxies detected using *SIRTF*. With low thermal backgrounds in space, targeted wide-bandwidth far-infrared spectroscopy from a telescope with an cold aperture in the 1–5 m class could provide these astrophysical diagnostics and redshifts. These observations would complement spectroscopy carried out at shorter wavelengths from *NGST* and future 30-m optical/near-infrared telescopes, and at longer wavelengths using ALMA.

Studying absorption lines at mid- and far-infrared wavelengths is more difficult than at optical or radio wavelengths, owing to the lack of extremely bright beacons; however, GRB afterglows and the brightest high-redshift dusty galaxies like APM 08279+5255 could provide the necessary backlighting photons. Absorption studies could thus also provide a way to overcome limited spatial resolution and to find out about the properties of low-luminosity galaxies at far-infrared wavelengths using the same type of instrument.

Probing the first structures

The radiation from molecular hydrogen that is expected to dominate the cooling of gas settling into the first generation of cosmic structures at $z \sim 10 - -20$ appears at near- and mid-infrared wavelengths, and is be redshifted to $\lambda \simeq 200\text{-}300 \, \mu$m. The first starlight should also be detectable directly in extremely deep images at wavelengths around $20 \, \mu$m. High-resolution spectral imaging may thus be able to detect the fluctuations in the submillimeter background radiation due to the signal from the energy released during the collapse of the first generation of proto-galactic gas clouds. Spectral observations of galaxies at more moderate redshifts in the far-infrared could probe PAH features in the dust emission spectra of the first more massive galaxies, revealing the astrophysics of the star formation processes before the peak redshift of star-formation activity.

The competition from submillimeter-wave ground-based facilities to conduct simple imaging surveys is likely to be very strong. Although they must operate at longer wavelengths, ground-based instruments can easily have large apertures and benefit from the advantageous submm K correction. In very deep pencil-beam fields, ALMA should be able to observe almost all of the galaxies that exist brighter than $0.1 \, L^*$ containing dust or molecular gas, without any source confusion. Ground-based single-antenna submm telescopes with 10-arcmin fields and 10^{4-5}-element detector arrays should be in service in several years. Their sensitivities are such that a significant fraction of the sky could be surveyed to the confusion limit of existing submillimeter-wave surveys in less than a decade. However, it will be essential to go to space to probe the short-wavelength SEDs of these galaxies, and to determine their spectral line emission properties.

Summary

Astrophysical diagnostics from line detections and SED measurements at wavelengths too short to ever be observed from the ground/stratosphere, and high-resolution spectral imaging at 20-40 μm provide unique scientific opportunities for future space missions. It is important to ensure that the angular resolution is the best that can be achieved, and that a spectroscopic capability is available, both to explore the properties of known high-redshift galaxies and the signals from the first generation of stars, and the cooling clouds in which they form.

Without the blocking effect of atmospheric windows, and by operating at wavelengths sensitive to the highest excitation CO lines, it should be possible to determine a great deal about the astrophysics of distant dusty galaxies to complement ALMA.

Cold space-borne telescopes are capable of very fast deep survey imaging. They requires apertures in the 5–10-m class to provide enough sensitivity and to defeat confusion. Their wavelength range is complementary to that covered by ALMA and large single-antenna submillimeter telescopes with wide-field cameras from the ground. They are also uniquely able to carry out deep follow-up spectroscopy in crucial wavelength ranges that cannot be reached from within atmosphere. Sensitivity at 100–300 μm is required to trace fine-structure lines from cooling gas in high-redshift galaxies and from any signature of molecular hydrogen emission from the first proto-galactic clouds. Spectra at 20–40 μm can be used to probe PAH emission from moderate redshifts and any primordial starlight that might be detectable redwards of Lyman-α.

The Universe at Infrared and Submillimeter Wavelengths

E. Dwek[1] , R. G. Arendt[2], D. J. Benford[1], J. C. Mather[1], S. H. Moseley[1], R. A. Shafer[1], & J. Staguhn[2]

ABSTRACT

Far infrared and submillimeter surveys offer unique information on the early stages of galaxy formation and evolution, and the cosmic history of star formation and metal enrichment. This paper presents various model results that can be used in the interpretation of far–IR and submm surveys with different diameter telescopes.

1.   Introduction

Recent developments have shown that far–infrared (IR) and submillimeter (submm) observations can offer unique probes of the high– redshift universe. Detections and limits on the cosmic infrared background (CIB) (Hauser & Dwek 2001) show that a significant fraction of the energy generated in the universe is absorbed and reradiated by dust at far–IR and submm wavelengths. Deep surveys with the SCUBA instrument on the JCMT resolved most of the CIB at 850 μm, and revealed that most of the star formation at redshifts above $\sim$ 2 takes place behind a veil of dust. The spectral energy distribution (SED) of galaxies peaks in the $\sim$ 60 – 140 μm wavelength region. So even at large redshifts, submm observations sample the rising part of the galaxies' SED, and effect referred to as the negative K–correction. Thus, far–IR and submm surveys can be used to observe the early stages of galaxy formation and evolution, and provide the history of the star formation rate and metal enrichment in the universe.

2.   The Construction of Galaxy Number Count Models

Here we will outline the ingredients for the construction of a simple galaxy number count model in the framework of a Backward Evolution (BE) model. BE models, as opposed to Forward Evolution, Semi–Analytical, or Cosmic Chemical Evolution models, extrapolate the spectral properties of galaxies in the local universe and their comoving number density to higher redshift, using some parametric form for their evolution (see Hauser & Dwek 2001 for a review of the various models). Within this framework, the model ingredients include: (1) a complete set of galactic

---

[1]Code 685, Laboratory for Astronomy & Solar Physics, NASA/Goddard Space Flight Center, Greenbelt, MD 20771

[1]SSAI, Code 685, NASA/Goddard Space Flight Center, Greenbelt, MD 20771

SED, representing galaxies of all morphological types and luminosities; (2) the galaxies' luminosity function (LF) in the local universe; (3) a prescription for the evolution of the galaxies SED and their LF with redshift; and (4) a cosmological model. The results presented here use the set of galaxies SEDs, their 15 μm LF, and the luminosity+density evolution model described by Chary & Elbaz (2001; CE01). The cosmology adopted here is that of a flat Λ dominated universe with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, and a Hubble constant of $H_0 = 75$ km s^{-1} Mpc^{-1}.

3. Model Results

Figure 1 shows the SED of select galaxies in the local universe, bracketing the luminosities assigned to the different type of galaxies: normal galaxies with $2.6 \times 10^8 < L_{\rm IR}({\rm L}_\odot) < 1 \times 10^{11}$; luminous IR galaxies (LIRGs) with $1 \times 10^{11} < L_{\rm IR}({\rm L}_\odot) < 1 \times 10^{12}$; and ultraluminous IR galaxies (ULIRGs) with $1 \times 10^{12} < L_{\rm IR}({\rm L}_\odot) < 3.6 \times 10^{13}$. A noticeable trend in the figure is the diminishing strength of the broad unidetified IR bands (UIB), usually attributed to polycyclic aromatic hydrocarbon molecules or PAHs, as a function of the galaxies' IR luminosity.
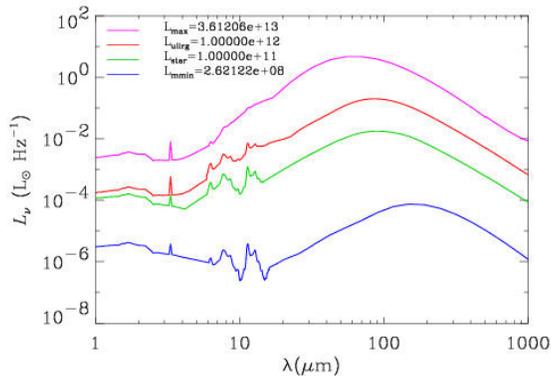
Fig. 1.— The SED of the galaxies bracketing normal galaxies, LIRGs, and ULIRGs in luminosity.

Figure 2 shows the observed flux of the galaxies depicted in the previous figure in the different IR passbands as a function of redshift. The behavior of the fluxes at redshifts above $z \approx 0.3$ is dominated by the movement of the UIB in and out of the band (at 15 μm only), and by the combined effects of the negative K-correction and luminosity evolution.

Figure 3 compares the calculated and observed differential galaxy number count, dN/dS, normalized to that expected in a Euclidean universe. The figure is essentially a reproduction of the results of the luminosity+density evolution model of Chary & Elbaz (2001). The model successfully reproduces the observed 15, 90, 170, and 850 μm galaxy number counts obtained with the *Infrared Space Observatory* (*ISO*) satellite (see Chary & Elbaz for a summary of the observations) and the SCUBA observations (Blaine et al. 1999). The figure also shows the contribution of the different

galaxy types to the number counts. The figure shows that ULIRG are the main contributors to the 850 μm number counts, whereas normal galaxies dominate the 15 and 90 μm bright galaxy counts, with LIRGs taking over the counts at lower fluxes.

Figure 4 depicts how the differential number count builds up with redshift. The vertical lines represent the 5σ confusion limts for (from right to left) 1, 2, 4, 8, and 16 m diameter telescopes. The cumulative flux from galaxies converges when the slope of $\mathrm{d}N\mathrm{d}S \times S^{5/2} \propto S^{1/2}$. Comparison of this figure with the data shown in the previous figure illustrates the redshift regions that contribute most to the observed number counts at each flux S_ν.

An important constraint on number count models is that the predicted cumulative flux from all sources be finite and converge to the cosmic IR background (CIB) at the different wavelengths. Figure 5 shows the cumulative contribution of all galaxy types to the CIB (left panel) and to the total number counts (right panel), as a function of wavelength. The shaded region in the left panel marks the limits and detections of the CIB summarized by Hauser & Dwek (2001).

In practice, the number of galaxies that can be resolved is limited by the confusion limit, defined as the flux below which an individual galaxy cannot be distinguished from fluctuations in the unresolved distribution of fainter galaxies. Figures 6 and 7 show, respectively, the contribution of resolved galaxies of different types to the CIB and the total number of galaxies using an 8 m diameter telescope (left panel), and a 16 m diameter telescope (right panel). Galaxy numbers and fluxes were integrated down to the 5σ confusion limit.

Figures 6 and 7 show that a 16 m diameter telescope will detect about 30% of the CIB at 500 μm, and about 3% of the total number of galaxies. The galaxies will comprise of about equal number of LIRGs and ULIRGs, at redshifts between $\sim$ 1 and 4.

REFERENCES

Blain, A. W., Kneib, J. P., Ivison, R. J., & Smail I. 1999, ApJ, 512, L87

Chary, R., & Elbaz, D. 2001, ApJ, 556, 562

Hauser, M. G., & Dwek, E. 2001, ARA&A, 39, 249
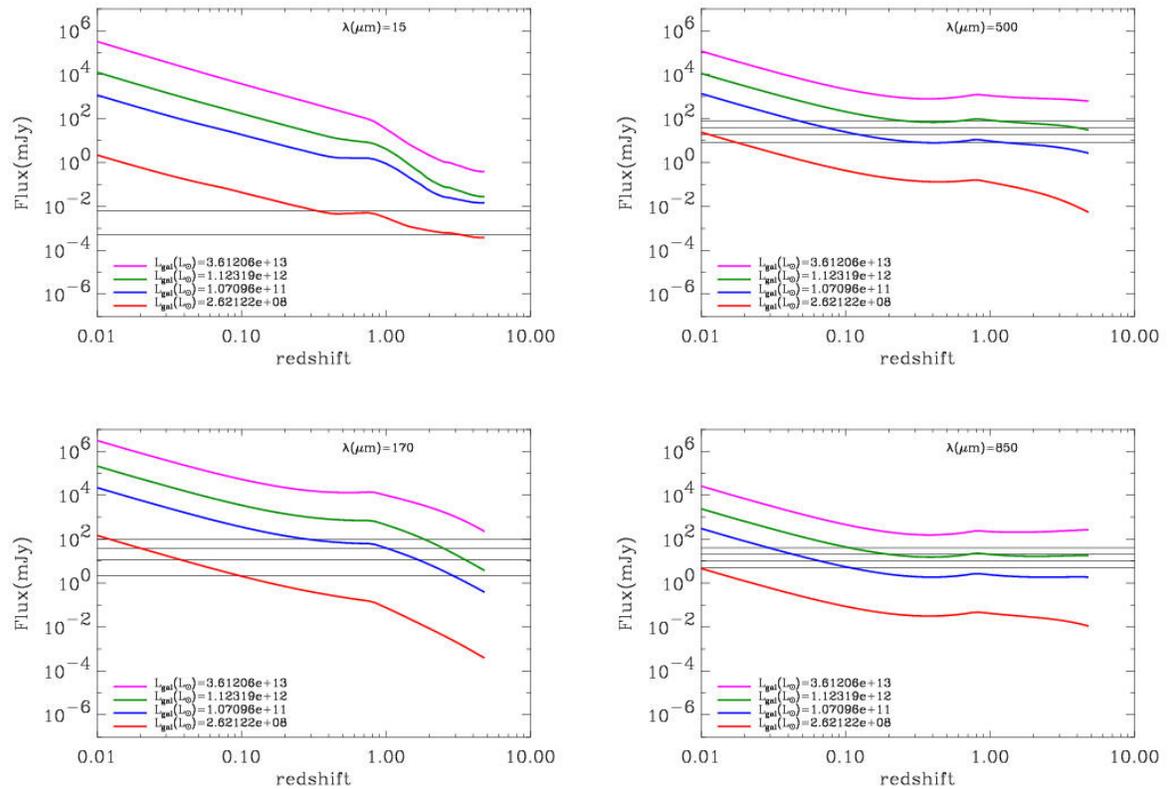
Fig. 2.— The in–band flux of the select galaxies depicted in Figure 1 as a function of redshift. Note the effect of the UIB features at 15 μm, and the negative K–correction at 850 μm on the behavior of the flux with redshift.
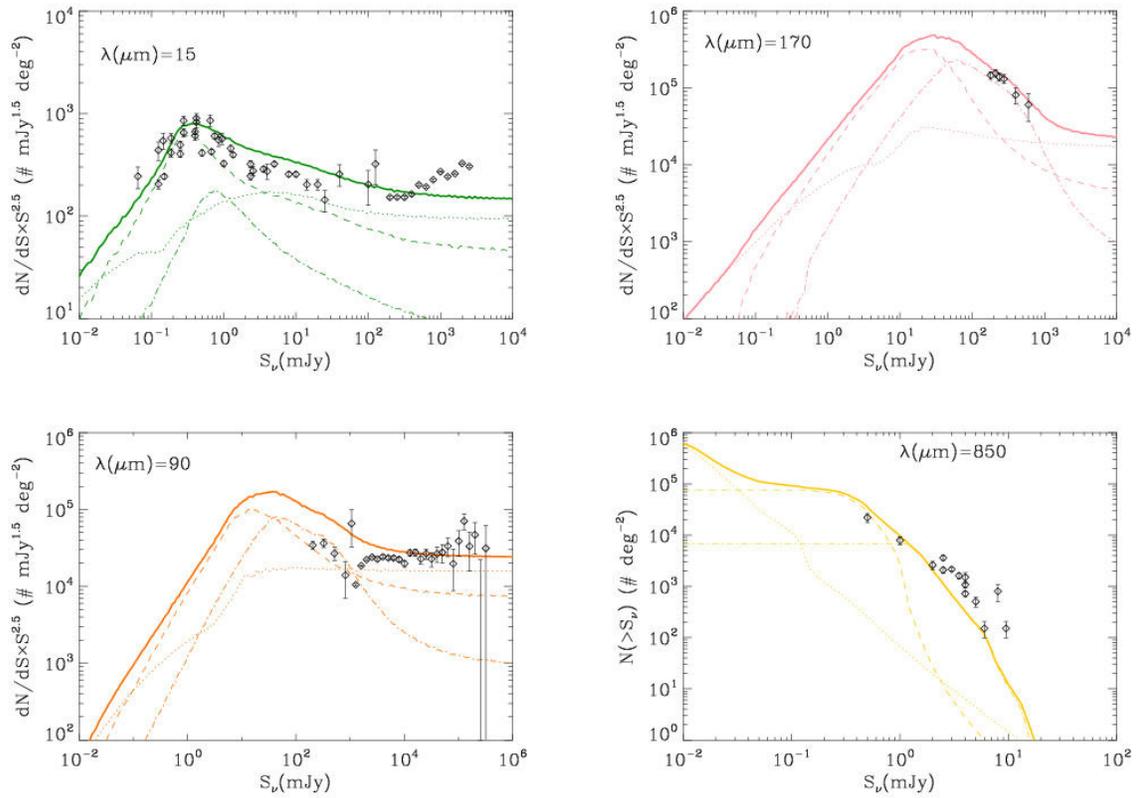
Fig. 3.— The differential number counts, calculated for the luminosity+density evolution model of Chary & Elbaz (2001), multiplied by $S^{5/2}$ is plotted as a function of galaxy flux density S_ν, and compared to observed counts. In a Euclidean universe with non–evolving galaxies, the plotted quantity should be a horizontal line. The figure also depicts the contribution of normal (......), LIRGs (——), and ULIRGs (-.-.-.-.) to the differential number counts.
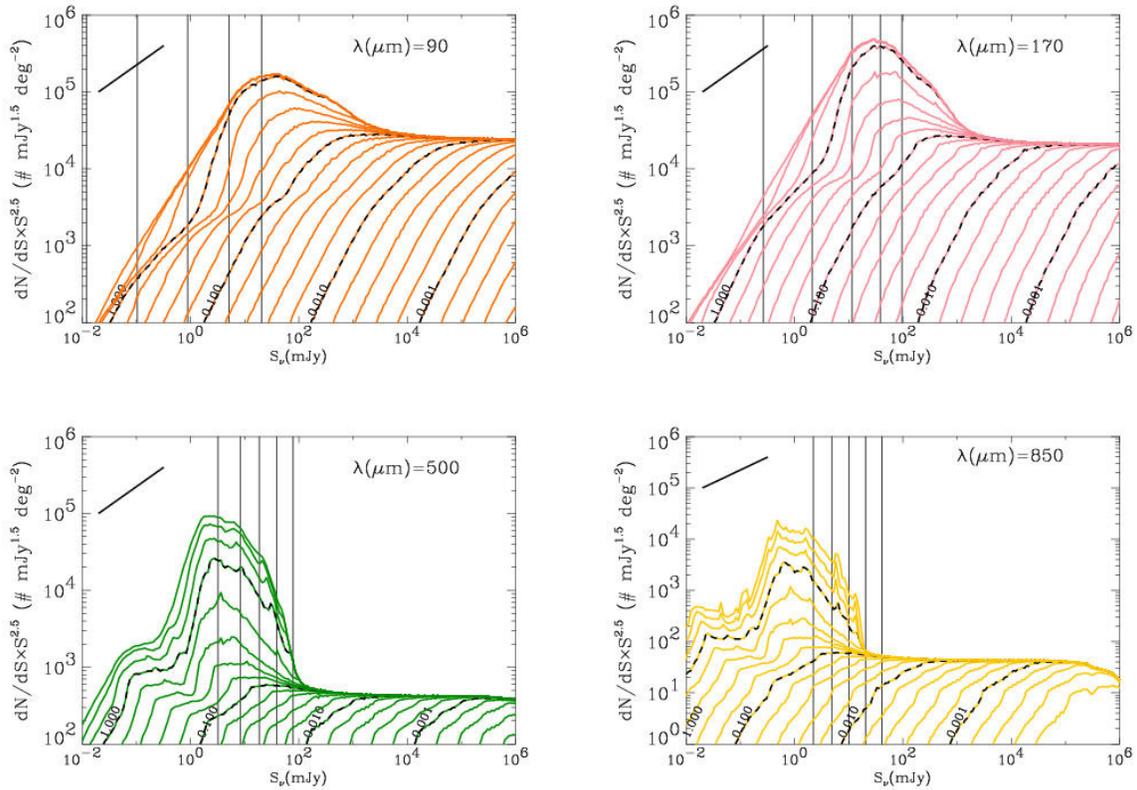
Fig. 4.— The buildup of the differential number counts with redshift. The dark contours represent redshifts of 0.001, 0.01, 0.1, and 1.0. The intervals between each pair of dark lines is divided into 5 equal logarithmic intervals. Vertical lines represent the confusion limits for various diameter telescopes. The dark line in the upper left corner has a slope of $S^{1/2}$, required for the cumuative light from the sources to converge. See text for more details.
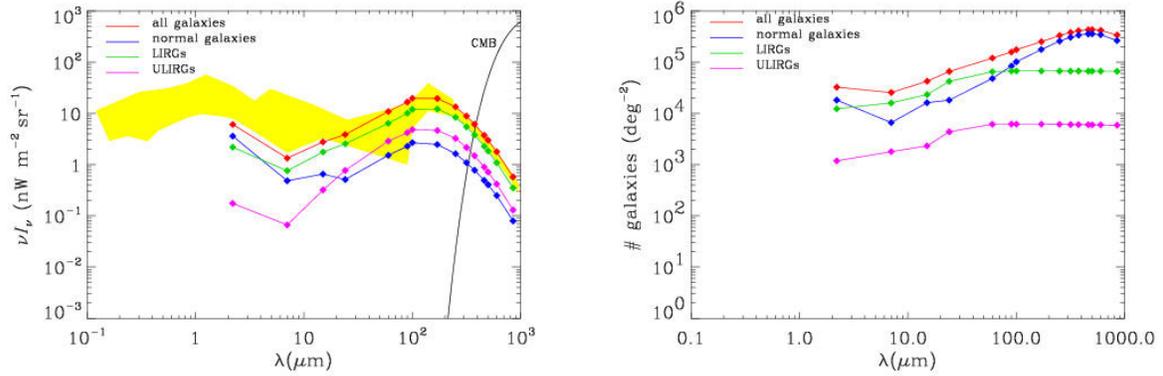
Fig. 5.— The contribution of the different type of galaxies to the CIB (left panel) and the total number of galaxies as a function of wavelength. The calculations assume that all galaxies in the universe are resolved, a zero–flux confusion limit.
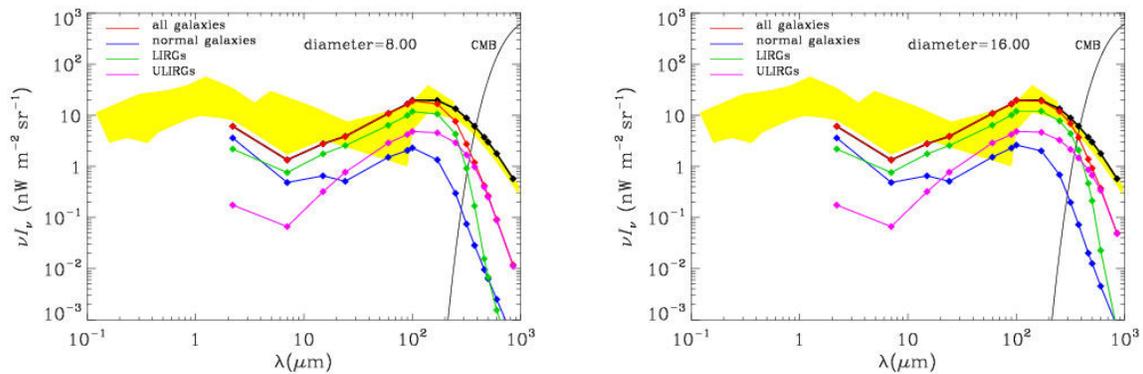
Fig. 6.— The contribution of resolved galaxies of different type to the CIB. Galaxy fluxes were integrated down to the 5σ confusion limit.
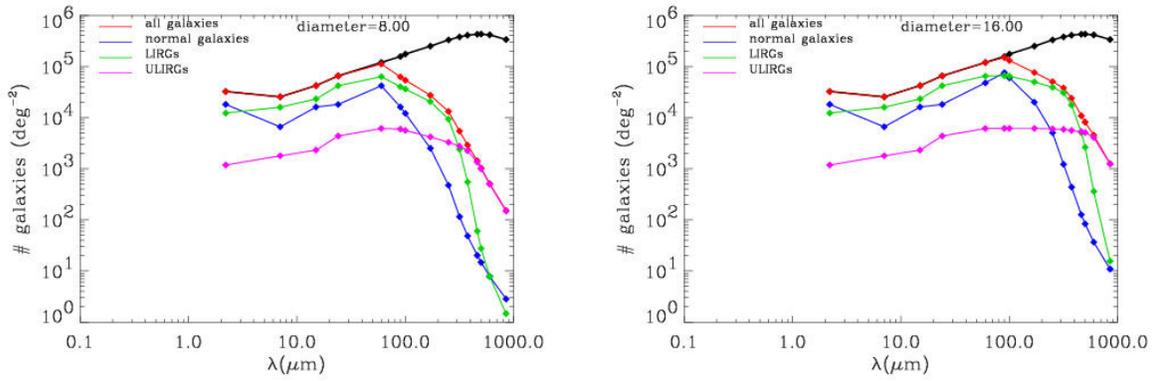
Fig. 7.— The contribution of resolved galaxies of different type to the total number of galaxies. Galaxies were counted down to the 5σ confusion limit.

The Far Infrared Lines of OH as Molecular Cloud Diagnostics

Howard A. Smith
Harvard-Smithsonian Center for Astrophysics
60 Garden Street
Cambridge, MA 02138
hsmith@cfa.harvard.edu

Abstract

Future IR missions should give some priority to high resolution spectroscopic observations of the set of far-IR transitions of OH. There are 15 far-IR lines arising between the lowest eight rotational levels of OH, and ISO detected nine of them. Furthermore, ISO found the OH lines, sometimes in emission and sometimes in absorption, in a wide variety of galactic and extragalactic objects ranging from AGB stars to molecular clouds to active galactic nuclei and ultra-luminous IR galaxies. The ISO/LWS Fabry-Perot resolved the 119μm doublet line in a few of the strong sources. This set of OH lines provides a uniquely important diagnostic for many reasons: the lines span a wide wavelength range (28.9μm to 163.2μm); the transitions have fast radiative rates; the abundance of the species is relatively high; the IR continuum plays an important role as a pump; the contribution from shocks is relatively minor; and, not least, the powerful centimeter-wave radiation from OH allows comparison with radio and VLBI datasets. The problem is that the large number of sensitive free parameters, and the large optical depths of the strongest lines, make modeling the full set a difficult job. The SWAS montecarlo radiative transfer code has been used to analyze the ISO/LWS spectra of a number of objects with good success, including in both the lines and the FIR continuum; the DUSTY radiative transfer code was used to insure a self-consistent continuum. Other far IR lines including those from H_2O, CO, and [OI] are also in the code. The OH lines all show features which future FIR spectrometers should be able to resolve, and which will enable further refinements in the details of each cloud's structure. Some examples are given, including the case of S140, for which independent SWAS data found evidence for bulk flows.

1 The Far Infrared Transitions of OH

Storey, Watson and Townes (1981) made the first far infrared detection of OH in the interstellar medium: the two 119μm lambda-doubled lines between the ground and first excited states, which they discovered in absorption from Sag B2, and in emission from the shock in Orion KL. Altogether there are 15 far infrared lines between the lowest 8 rotational levels of OH. The lines (each involving six transitions between the hyperfine-split rotational levels) are at wavelengths of (approximately) 28.9μm, 34.6μm, 43.9μm, 48.8μm,, 53.3μm, 65.1μm, 71.2μm, 79.2μm, 84.4μm, 96.3μm, 98.7μm, 115.4μm,

119.9µm, 134.8µm, and 163.1µm. The upper state excitation temperatures for these lines range from 120K to 618K. The dipole moment for OH is large, 1.668 Debye (for comparison the CO dipole moment is 0.112 Debye), and the radiative rates for OH transitions are generally fast. For example, the 119µm fundamental transition rate is about 0.1 sec^{-1}. But the FIR OH transitions also include some cross-ladder lines whose radiative rates are one hundred times weaker, providing a dataset of neighboring, far IR lines which frequently include both optically thin and very optically thick features. The OH analyses have an additional resource from which to draw: the strong hyperfine radio wavelength transitions that OH has in its ground-state, and which have been extensively observed. Maser and/or mega-maser activity is seen in many of the stars and galaxies which ISO observed.

The general properties of interstellar OH are known from thermal OH emission studies done at radio wavelengths, as well as from the far IR observations. Typically OH has the following range of properties: $N_{OH}/N_{H2} = 0.1 - 3 \times 10^{-7}$; $N_{OH}L = 2-100 \times 10^{15}$ cm^{-2}; $T_x = 100 - 275K$; and $N_{H2} = 0.1 - 3 \times 10^7$ cm^{-3} (e.g., Watson *et al.* 1985; Jones *et al.* 1994). In maser galaxies, such as the ones we observed with ISO/LWS, the OH masers regions typically have somewhat different properties: $N_{OH}/N_{H2} = 10^{-7} - 10^{-8}$; $N_{OH}L \sim 10^{15}$ cm^{-2}; $T_x = 40K - 50K$; $N_{H2} \sim 0.1-1 \times 10^4$ cm^{-3} (e.g., Henkel and Wilson, 1990). The presence of OH megamasers allow for VLBI observations in AGN, and results indicate that the sizes of the emitting regions are compact: ~ from a few to tens of parsecs. For example, in Arp220, one of the galaxies I discuss below, they seem to surround the AGN, with H_2 densities ~ 10^6 cm^{-3} (Lonsdale *et al.* 1994; Skinner *et al.* 1997).

Four of the relatively strong far IR OH transitions involve the ground state. In warm clouds the molecules absorb the strong IR dust continuum, populating effectively some of the higher lying levels. As was conclusively shown by Sylvester *et al.* (1997), the 18 centimeter radio maser emission in evolved stars is pumped by absorption of the 34µm dust continuum, while in AGN Skinner *et al.* (1997) proved the effectiveness of the IR pumping of mega-masers in Arp220. The presence of a strong far IR continuum affects all of the far IR lines, to varying degrees, and means (besides making the models more complicated) that the set of far IR OH lines also provide a sensitive measure of the local continuum conditions. One final point is worth noting about OH: unlike other some species commonly used as far IR probes of the interstellar medium, like CO, H_2O, or OI, OH emission from PDRs or shocks does not contribute a relatively dominant amount of luminosity to these other processes. In the cases I model below, just the warm gas from a dusty molecular clouds is adequate to explain the strengths we observe. OH, therefore, has a powerful combination of features that makes it a very useful species for dis-entangling cloud properties across a very wide range of conditions.

2 ISO/LWS Observations of Extragalactic OH

The ISO/LWS Extragalactic Science team has seen OH in fifteen galaxies, and has obtained potentially useful limits on approximately one hundred other galaxies observed by ISO. Our observations include a set on the 34.6μm "IR pump" transition from the OH $^2PI_{3/2}$ ground state, which was obtained using the ISO/SWS spectrometer. This line, and the less important 28.9μm line between excited upper states, are the only OH far IR lines not in the LWS wavelength coverage. ISO of course has also seen OH in numerous galactic locations, including Sag A, and Sag B2, and in particular in the evolved star IRC+10420 -- the first source to demonstrate that the 34μm continuum can effectively pump the OH maser lines in these stars.   The LWS extragalactic detections of OH are in the following sources: Arp 220, Cen A, IRAS17208-0014, IRAS20100-4156, M82, Mkn 231, Mkn 273, NGC 253, NGC 891, NGC 1068, NGC 1614, NGC 3690A, NGC 4945, NGC 7469, and 3Zw35.   Figure 1 shows all the eight OH lines detected in Arp220, along with a sample of the lines detected in other objects to illustrates some of the morphological variety.

2.1 General Characteristics of the Observed Extragalactic OH Lines

Perhaps the most striking observation about the set of extragalactic lines measured is the wide morphological range of behavior they display, even though all arise in infrared bright galaxies with either active star formation, an active nucleus, or perhaps both. In Arp 220, for example, every OH line is seen in absorption except the longest wavelength, 163μm, line which is seen weakly in emission (Fischer *et al.* 1998) . By contrast, NGC 1068, another AGN, has every detected line seen in *emission*, even the strong 119μm transition between ground and the first excited state (Spinoglio *et al.* 1999). NGC 253, a nearby starburst, has some OH lines in emission and some in absorption (Bradford *et al.* 1999), while in M82, the infrared bright, prototype starburst galaxy, the lines' equivalent widths are so small continuum that even with our high signal-to-noise ratio only the 119μm line has been conclusively seen at all, in absorption (Colbert *et al.* 1999). From the analysis of this diverse set of line strengths, several useful preliminary generalizations may be drawn for the different categories, as summarized below.

ULIGs: In the case of Arp 220, the VLBI megamaser studies together with the strength of the 34μm pump absorption provide some strong physical constraints. The OH lies in numerous small clouds which surround the AGN, and which are pumped by the local, warm, far IR continuum. Through further modeling I hope to arrive at a better sense of the shape of that continuum: is it starburst-like, or more AGN like, and is it characteristic of all ULIGs?

AGN: In the case of the bright Seyfert galaxy, NGC 1068, analyses of the strong atomic lines (Spinoglio *et al.* 1999) show the substantial presence of a PDR line emission component, along with the high ionization lines from the AGN component in the ISO beam, and a starburst component, which is seen as well in other AGN. However the [CII]/FIR ratios are strange -- often less than in PDRs. The OH lines, all seen in

53.3µm | 65.1µm | 71.2µm | 79.2µm

Arp 220 ----->

NGC 253   M 82   NGC 1068   IRC+10420

Sample of next four lines:

84.4µm | 98.7µm | 119.9µm | 163.1µm

Arp 220 ----->

NGC 253   IRC+10420   NGC 1068   M 82
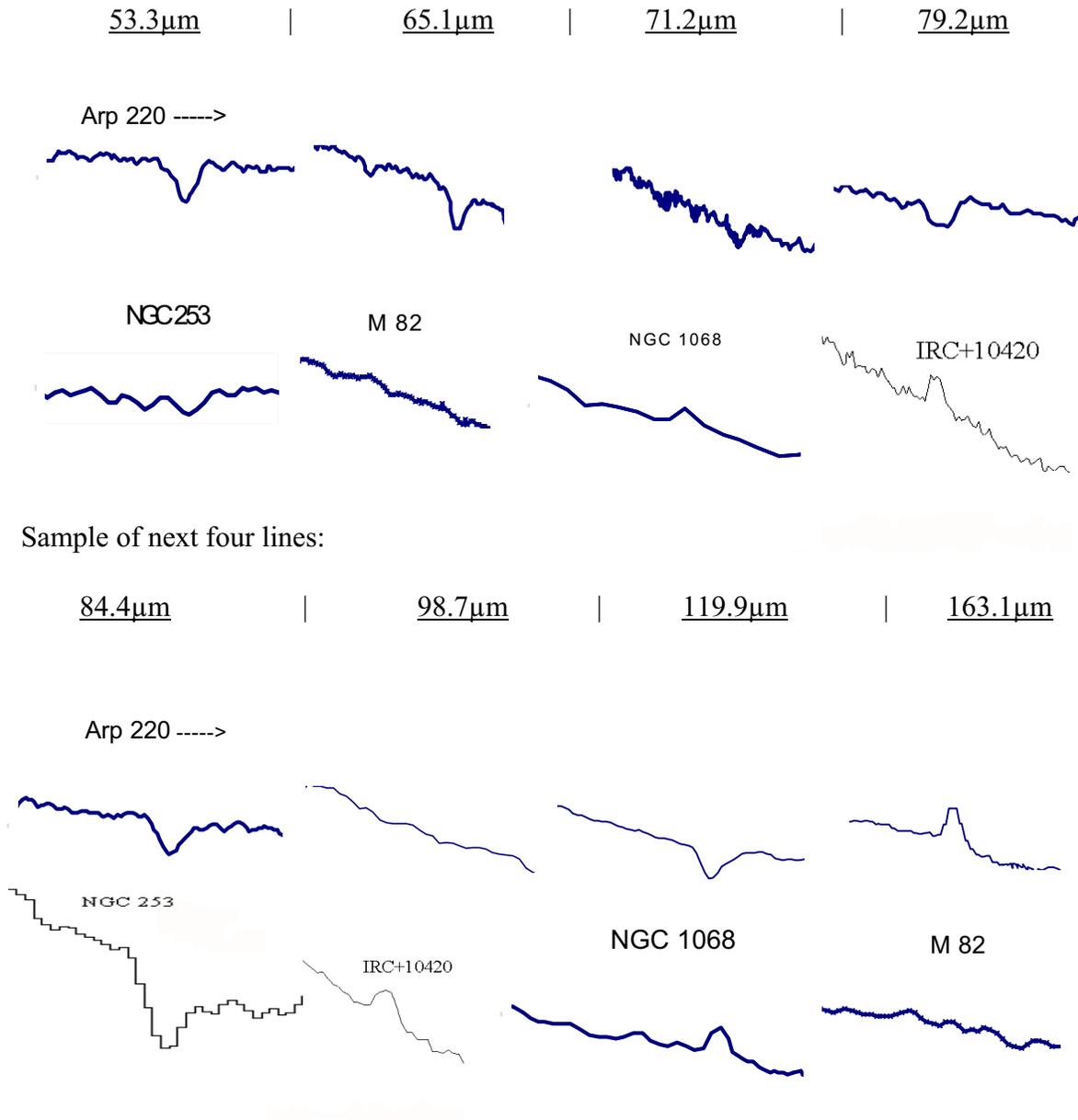
Figure 1: The 8 ISO/LWS OH lines seen in Arp 220 (the 34µm SWS-band line is not shown here). Under each OH line in Arp220, an OH line selected from another other ISO galaxy (or IRC+10420) is shown for comparison at the corresponding wavelength to illustrate some of the variety in the intensities observed.

emission, might help sort out the density and geometry of the clouds (for example, to see if the PDR regions have smaller than average filling factors) – and provide clues to their relationship to the active nucleus.

Infrared Bright Galaxies: We find that the OH 119 μm ground fundamental transition is always in absorption in these galaxies, as is the 53μm line, while the 163μm feature is always seen in emission. Otherwise there does not appear to be any consistent behavior in the lines from different objects: in some sources they are seen in emission, in others they are in absorption. Saraceno *et al.* (1996) and Benedetti *et al.* (2000), among others, have noted there seems to be a dearth of H_2O emission in some galactic clouds. Also, the [CII] fluxes are very low, perhaps due to low gas heating (but this is not conclusively demonstrated). They note that the neutral oxygen [OI] 63μm line is often abnormally weak and may be self absorbed. Finally, the SWAS satellite found that O_2 is very weak or absent (Goldsmith *et al.* 2000). OH plays a key role in the chemistry of the ISM, is sensitive to the temperatures and radiation fields, and its abundance and distribution should help in the analyses of all these issues.

3 Modeling of the Far Infrared OH Lines

3.1 The Montecarlo Code for Lines; the "DUSTY" Code for Continuum

I used the one-dimensional montecarlo radiative transfer code developed by the SWAS mission (Ashby *et al.* private comm) to model the OH line strengths. This code, a modification of the original Bernes code, adds a treatment of continuum photons from dust mixed in with the gas -- a particularly essential feature for OH, which is pumped in many cases by absorption of 34μm continuum. In addition, the code corrects for some previous errors encountered at large optical depths, also an issue of importance for OH which has a very strong matrix element. Finally, the code includes an ability to handle a wide range of molecules besides rigid rotors. The montecarlo code takes as input a series of concentric shells, each of which is specified as to size, gas and dust temperature, H_2 density, velocity and velocity width, and molecular abundance relative to H_2. The model as a whole also assumes a (specifiable) dust emissivity. In all the modeling, the dust is assumed to be 1% of the gas, and to have the same temperature as the gas everywhere in the cloud. The montecarlo code calculates the populations of the molecular levels in each shell. The output is then fed into a radiative transfer code that calculates the line profiles as seen by an external observer looking at the cloud.

The montecarlo results confirm the fact, noted above, that the OH lines are often optically thick. The peak optical depth at line center for the 119μm line exceeds 100 in some cases. As a result, many of the lines in many situations show self-absorption. As the model is tuned to fit the data by increasing the column density, these features can turn an emission line into an absorption line, one while strengthening the emission from a weaker

neighbor.   The full set of nine line intensities, which span a range of wavelengths and optical depths, enable us to derive a rather detailed self-consistent picture of the cloud conditions solely from their intensity ratios, <u>without</u> the need for velocity information. But, as seen in the example below, there is also a wealth of information in the line shapes.

ISO/LWS also observed lines of CO, H_2O, and [OI] in these galaxies, with varying degrees of success. The SWAS montecarlo code can also predict the line emission from these species. In general the model predictions for these lines are in overall agreement with the observations; the uncertainties are due to uncertainties in the ISO line fluxes themselves, in the assumed molecular abundances, and in the amounts of possible line "contamination" from shocks and PDRs in the beam. Overall the results add further confidence to the models. The [OI] 63µm line is clearly self absorbed in several instances, confirming the suspicion that the line intensity is sometimes very weak due to self-absorption (Saraceno *et al.*)

The montecarlo output is not strictly internally self-consistent; the input parameters need not conserve luminosity between shells, for example. To obtain this self consistency, and to guarantee that the final cloud model was also consistent with the observed far infrared continuum flux, I used the DUSTY code (Ivezic and Elitzur, 1997) to model the continuum and generate a set of shell parameters that provided this consistency. Then I iterated the DUSTY model with the montecarlo line outputs. While this technique does not give a unique solution for the cloud structure, it does provide a canonical model consistent with the observations.

3.2 Some Modeling Results

3.2.1 S140 - A Molecular Cloud with Bulk Inflow Motions
Ashby *et al.* (2000) used the SWAS satellite to observe S140, and the SWAS montecarlo code to model the shape of the observed submm H_2O line. They are able to obtain a detailed, if not entirely unique, model of the cloud. From their set of models they conclude, for example, that the cloud radius is 0.44pc, has an inner temperature of ~70K an inner hydrogen density of $1.4x106$ cm^{-3} , a radial profile of temperature which varies like r$^{-0.5}$, and a density profile varying like r$^{-0.8}$. They also conclude that "significant bulk flow" is required to explain the H_2O line shape, but because of the small optical depth of the 557 GHz line they could not differentiate between infall and outflow. Although the ISO observations of S140 were only able to set weak limits on the OH lines (Aannestad and Emery, 1998*)*, the detailed nature of the SWAS models made it a useful check. Taking the SWAS cloud parameters for the case of bulk infall motions, I used the montecarlo code to calculate the strengths and line shapes for the full set of OH lines; I calculated the far IR lines of H_2O and the [OI] lines as well, for comparison with the ISO values. The reasonable results obtained from the model of S140, as confirmed with the SWAS (and ISO) observations, provide some confidence in the galaxy modeling. Figure

2 shows the results for three of the OH lines, on the same scale: the 163μm line (top), the 79μm line (middle), and the 119μm line (bottom). It is clear that high spectral resolution observations of the far IR water lines can readily distinguish infall from outflow, because these lines are optically thicker than the submillimeter line.

3.2.2 Arp220 - The Molecular Cloud Component of a Peculiar Ultra Luminous Galaxy

Arp220 is a particular challenge, because so many OH lines are seen, and every one of them (except the line at 163μm) is seen in absorption -- something that happens in no other known object. Arp220 is also unusual in general in that its spectrum is characterized by numerous molecular absorptions; even the [OI] 63μm fine structure line is seen in absorption, while the [CII] 157μm line is found only in weak emission (e.g., Fischer *et al.* 1998). Skinner *et al.* (1997) showed that the 34μm OH absorption could pump the OH megamasers in Arp220, and their analysis supported the model in which many small, dense molecular clouds circulate around the nucleus. Suter *et al.* (1998) attempted to model the OH line absorptions with some less complex radiative transfer models, and were driven to consider some unusual non-LTE situations to explain the observations; .

I used the SWAS montecarlo code successfully to model (to first order) all of the observed OH lines in Arp220. Figure 3 shows a sample of these model line shapes. In order to get absorption in all the lines, including the 79μm line to the ground state $^2PI_{3/2}$ from the lowest level of the $^2PI_{1/2}$ ladder, while still getting weak emisson at 163 μm, it was necessary to have falling within the ISO/LWS beam a combination small dense clouds, and a few large, giant molecular clouds. Iterating with the DUSTY code provides a way to obtain a continuum shape that fits the observed infrared continuum, though there may also be a continuum component in our beam arising from clouds without much OH. It was not necessary in this process to resort to the non-LTE scenario postulated in Suter *et al.*

3.3.3 NGC 253 - The OH Emission from a Nearby Infrared Bright Galaxy

In NGC 253, the two strongest OH lines are seen as emission lines: the 79μm and 163μm features, each of which is about twice as strong as the 119μm fundamental absorption (Bradford *et al.* 1999). Only two other OH lines are also seen in this galaxy, and as a result the model's constraints are weaker than for Arp220. The montecarlo modeling of these OH lines implies that a few giant molecular clouds, with r ~ 100pc, can explain the observations. The clouds have H_2 densities averaging approximately $3x10^3$ cm^{-3} and an OH abundance relative to H of about $5x10^{-8}$. ISO/LWS also observed the 119μm line in NGC 253 with the high resolution Fabry-Perot, which Bradford (2001) has successfully modeled.

3.3.4 NGC 1068 - OH Emission from AGN

NGC1068 is a remarkable extragalactic source in its OH spectrum, because it is the only galaxy observed in which all the detected transitions are found in emission. The code accurately predicts the observed line flux ratios to 40%. Based on the montecarlo
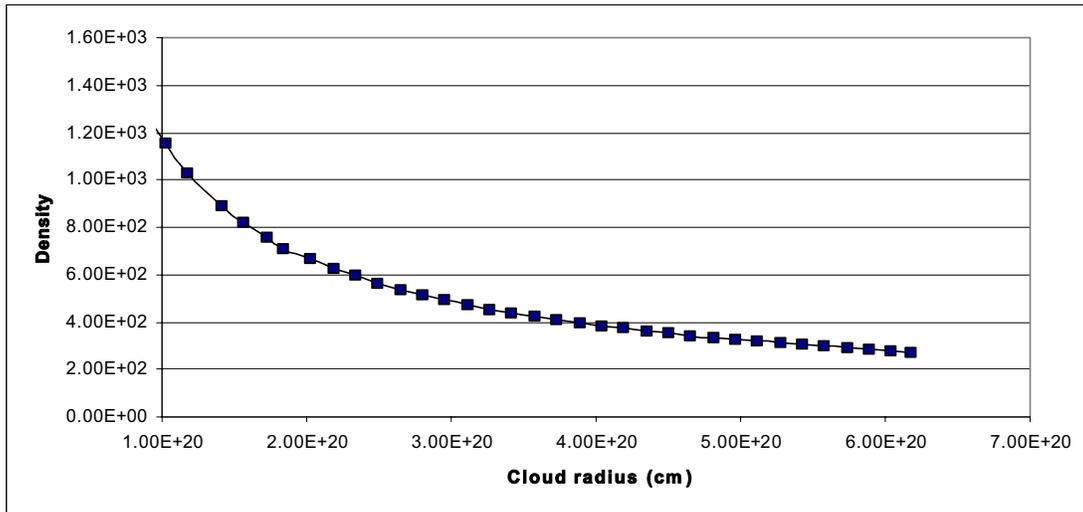
Figure 2: The hydrogen density distribution in a nominal giant molecular cloud in the galaxy NGC 253. This distribution was used in the montecarlo code modeling. The inner 50 parsecs of the GMC has warm, 250K, gas and dust, at constant density, while the bulk of the cloud has much cooler, 35K material.
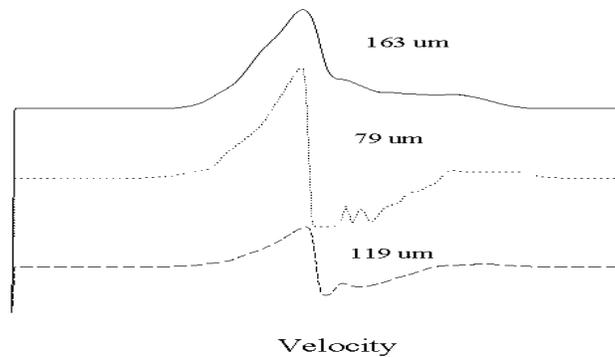
Figure 3: Predicted OH line shapes from infalling gas in S140, based on cloud profiles derived from SWAS data. The figure shows the 163 µm line (top), the 79 µm line (middle), and the 119 µm line (bottom). While all three lines show evidence for infall from their highly asymmetric shapes, only the latter two show characteristic absorption: in the redshifted (near side) material. The velocity range across the figure is ±10 km/sec, reflecting the fact that the maximum infall velocity in the model (and seen by SWAS) was 7.2 km/sec.
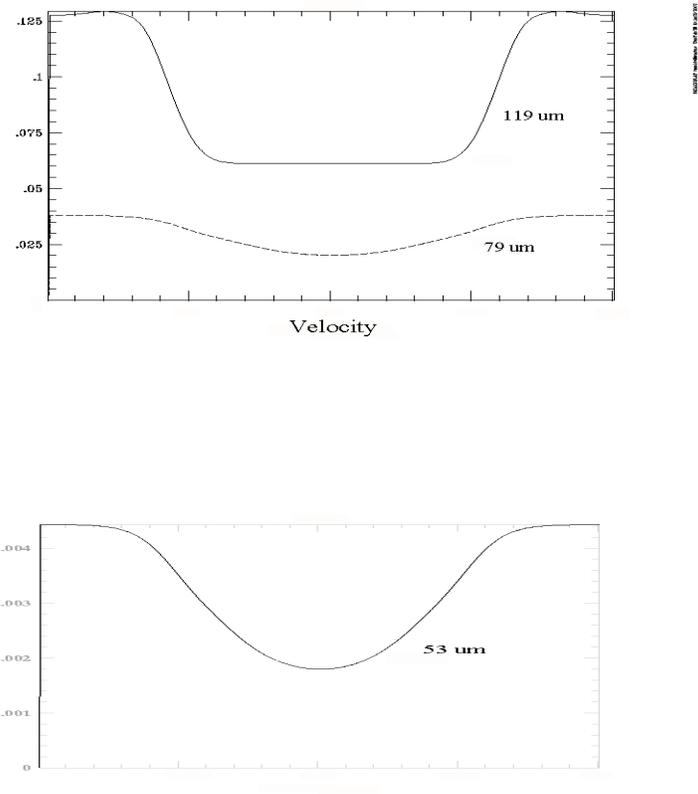
Figure 4: Montecarlo modeling results of three OH lines in Arp 220. The OH lines at 119μm, 79 μm, and 53 μm are displayed. All the OH lines are seen in absorption in Arp 220, except the weakly emitting 163 μm line – the only known object with this property.

modeling, it seems the clouds that produce these lines in NGC1068 must be relatively small and dense, and heated from the inside. They are about 0.2pc in radius, with densities of $\sim 10^4$ cm^{-3} at the outer edges, increasing towards the center with a power law behavior of $(R/Ro)^{1.25}$. The temperature in the outer shells is about 25K, increasing inwards approximately with a power law dependence $(R/Ro)^{1.47}$. The total column density of H$_2$ in each cloud is 1.5×10^{24} cm^{-2}, and the relative abundance of OH in the model is approximately 10^{-7}. OH is taken to be absent in the hotter portions of the cloud, where T>300K. With these high column densities the strong lines of OH are very optically thick, and precise radiative transfer calculations like the present ones are absolutely essential. At a distance of 16.2Mpc, the total number of equivalent number of clouds needed to produce the observed absolute flux in the lines and the continuum is $\sim 3 \times 10^7$. The limits to the observed fluxes in all the weaker lines are consistent with this modeling. It is worth noting in this context that in some galaxies the weaker OH lines in NGC 1068, like the 53μm line, are amongst the strongest, but the montecarlo modeling can successfully account for these differences. The montecarlo code also successfully reproduces the continuum emission seen from NGC 1068 to within a factor of 2-3 in absolute flux density across the entire LWS spectrum. The total mass in such an ensemble of clouds in NGC 1068 is approximately 9×10^9 M$_\odot$.

4   Conclusions to Date

What general conclusions might we hope to draw from this large set of ISO OH observations? Although the analysis is still underway, it appears that the infrared luminous galaxies can be grouped into three general categories based on the relative strengths of their infrared OH lines: (1) relatively normal galaxies like NGC 253 and M82, from which the OH lines are seen in both emission and absorption, and which are dominated by starburst activity in giant molecular clouds; (2) NGC 1068 and other AGN, whose OH is seen in emission, and whose clouds must be quite small and dense (hydrogen densities up to 10^8 cm^{-3} in the cloud cores); and (3) Arp220, and perhaps other ULIGs with active starbursts as well as an active nucleus, for which a combination of a few giant molecular clouds and a collection of moderately dense cloudlets are required. (The ISO/LWS spectrum of Mkn 231 is much noisier than Arp 220's, but hints at similar behavior in its OH.) While I have made substantial progress using only the intensities of these lines, velocity resolved spectra will enable even further refinements. Future far infrared missions with sensitive, spectroscopic capabilities should find that the set of OH lines can provide wealth of information needed to unravel the structures of complex clouds.

References:

Aannestad, P.A. and Emery, R, 1998, *BAAS*, **192**, 1009.

Ashby, M., *et al.,* 2000, *Ap.J. (Letters*), **539**, L119.

Benedettini, M., Giannini, T., Nisini, B., Tommasi, E., Lorenzetti, D., Di Giorgio, A. M., Saraceno, P., Smith, H. A., and White, G. J., 2000, *A&A*, **359**, 148.

Bradford, C. M., Stacey, G. J., Fischer, J., Smith, H. A., Cohen, R. J., Greenhouse, M. A., Lord, S. D., Lutz, D., Maiolino, R., Malkan, M. A., and Rieu, N. Q., 1999, *The Universe as Seen by ISO*. Eds. P. Cox & M. F. Kessler. ESA-SP 427.

Bradford, C.M., 2001, Ph.D.thesis.

Colbert, James W., Malkan, Matthew A., Clegg, Peter E., Cox, Pierre, Fischer, Jacqueline, Lord, Steven D., Luhman, Michael,Satyapal, Shobita, Smith, Howard A., 1999, *Ap,J.*, **511**, 721.

Fischer, J., Satyapal, S., Luhman, M., Melnick, G., Cox, P., Cernicharo, P., Stacey, G., Smith, H.A., Lord, S., and Greenhouse, M., 1998, in Proceedings of the "ISO to the Peaks Workshop: The First ISO workshop on analytical Spectroscopy", Madrid, Spain.

Goldsmith, P., *et al.,* 2000, *Ap.J. (Letters)*, **539**, L123.

Henkel, C., and Wilson, T.L., 1990, *A&A*, **229**, 431.

Ivezic, Z., and Elitzur,M., 1997, *M.N.R.A.S.*, **287**, 799.

Jones, K.N., Field, D., Gray, M.D., and Walker, R.N.F., 1994, *A&A*, **288**, 581.

Lonsdale, C.J., Diamond, P.J., Smith, H.E., Lonsdale, C.J., 1994, *Nature*, **370**, 117.

Saraceno, P., *et al.* 1996, *A&A*, **315**, L293-L296.

Skinner, C.J., Smith, H.A., Sturm, E., Barlow, M.J., Cohen, R.J., and Stacey, G.J., 1997, *Nature*, **386**, 472.

Spinoglio, Luigi, Stacey, Gordon, and Unger, Sarah J, 1999, *Ap.J.*, **511**, 721.

Spinoglio, L., Suter, J. P., Malkan, M. A., Clegg, P. E., Fischer, J., Greenhouse, M. A., Smith, H. A., Pequignot, D., Stacey, G. J., Rieu, N. Q., Unger, S., J., and Strelnitski, V., 1999, *The Universe as Seen by ISO*. Eds. P. Cox & M. F. Kessler. ESA-SP 427.

Storey, J.W.V., Watson, D.M., and Townes, C.H., 1981, *Ap.J.(Letters)*, **244**, L27.

Suter, J. P., Harvey, V., Cox, P., Fischer, J., Lord, S. D, Melnick, G. J, Satyapal, S., Smith, H. A., Stacey, G., Strelnitski, V., and Unger, S. J., 1998, *BAAS*, **193**, 9002.

Watson, D.M., Genzel, R., Townes, C.H., and Storey, J.W.V., 1985, *Ap.J.*, **298**, 316.

IR Fine-Structure Line Signatures of Central Dust-Bounded Nebulae in Luminous Infrared Galaxies

J. Fischer,[1] R. Allen,[1] C. C. Dudley,[1,2] S. Satyapal,[3] M. Luhman,[4] M. Wolfire[5] & H. A. Smith[6]

ABSTRACT

To date, the only far-infrared spectroscopic observations of ultraluminous infrared galaxies have been obtained with the European Space Agency's Infrared Space Observatory Long Wavelength Spectrometer. The spectra of these galaxies are characterized by molecular absorption lines and weak emission lines from photodissociation regions (PDRs), but no far-infrared ($\lambda > 40\ \mu$m) lines from ionized regions have been detected. ESA's Herschel Space Observatory, slated for launch in 2007, will likely be able to detect these lines in samples of local and moderate redshift ultraluminous galaxies and to enable measurement of the ionization parameters, the slope of the ionizing continuum, and densities present in the ionized regions of these galaxies. The higher spatial resolution of proposed observatories discussed in this workshop will enable isolation of the central regions of local galaxies and detection of these lines in high-redshift galaxies for study of the evolution of galaxies. Here we discuss evidence for the effects of absorption by dust within ionized regions and present the spectroscopic signatures predicted by photoionization modeling of dust-bounded regions.

[1]Naval Research Laboratory, Remote Sensing Division, 4555 Overlook Ave SW, Washington, DC 20375, USA, Jackie.Fischer@nrl.navy.mil

[2]NRL/NRC Research Associate

[3]George Mason University, Fairfax, VA

[4]Institute for Defense Analysis, Alexandria, VA, USA

[5]University of Maryland, College Park, MD, USA

[6]Harvard-Smithsonian, Center for Astrophysics, Cambridge, MA USA

1. <u>Background</u>

Prior to the launch of the European Space Agency's Infrared Space Observatory (ISO) in 1995, Voit (1992) showed how mid- and far-infrared fine-structure lines could be used to constrain the electron densities, extinction, and shape and ionization parameters of the central ionizing sources in ultraluminous infrared galaxies (ULIRGs). Moreover, the ground-based work of Roche et al. (1991) showed that the mid-IR spectra of the nuclei of galaxies could be placed into three classes: those with aromatic feature emission, featureless, and those with silicate absorption, typically associated with optically identified H II region, Seyfert 1, and Seyfert 2 nuclei respectively. Building on this early work on optically selected starburst and AGN galaxies, Genzel et al. (1998) constructed a diagnostic diagram of the ratio of high-to-low ionization fine-structure lines vs. the equivalent width of the $7.7\,\mu$m aromatic feature emission based on which they concluded that 70 - 80% of ULIRGs are powered predominantly by starbursts and 20 - 30% are powered by a central AGN. They attributed the weakness of the mid-infrared fine-structure lines relative to the infrared luminosity to the effects of extinction.

Far-infrared spectroscopy of a small sample of IR-bright and ultraluminous galaxies taken with the ISO Long Wavelength Spectrometer (LWS) has revealed a dramatic progression extending from strong fine-structure line emission from photoionized and photodissociated gas in the starburst galaxy Arp 299 (Satyapal et al. 2002) to faint [C II]$158\,\mu$m line emission and absorption in lines of OH, H_2O, CH, and [O I] in the ULIRG Arp 220 (Figure 1; Fischer et al. 1999). With the progression towards weak emission line strengths, no trend in density or far-infrared differential extinction is indicated (Figure 2), i.e. the temperature-insensitive [O III]$52\,\mu$m/[O III]$88\,\mu$m line ratio does not show a trend with the ratio [O III]$88\,\mu$m/FIR ratio as it would in either of these cases and all of the measured [O III] line ratios fall within the range 0.6 - 1.2, consistent with electron densities between 100 - 500 cm^{-3}. The sequence does show a trend toward lines with lower excitation potentials in the ratios [N III]$57\,\mu$m/[N II]$122\,\mu$m (Figure 3) and [O III]$52\,\mu$m/[N III]$57\,\mu$m (Figure 4). No FIR fine-structure line emission from species with excitation potentials greater than 13.6 eV were detected in Arp 220 or in Mkn 231 (Harvey et al. 1999), the two ULIRGs for which full ISO LWS spectra were taken. Voit (1992) discussed the possibility that even the mid- and far-infrared fine-structure lines would be weak in ULIRGs if they are formed in high ionization parameter regions. In such regions with high ratios of ionizing photon to electron densities, UV photons are preferentially absorbed by dust rather than gas, due to the high column densities of ionized gas in such regions. Bottorff et al. (1998) found that the L_{IR}/H_β ratios in such dust-bounded nebulae are
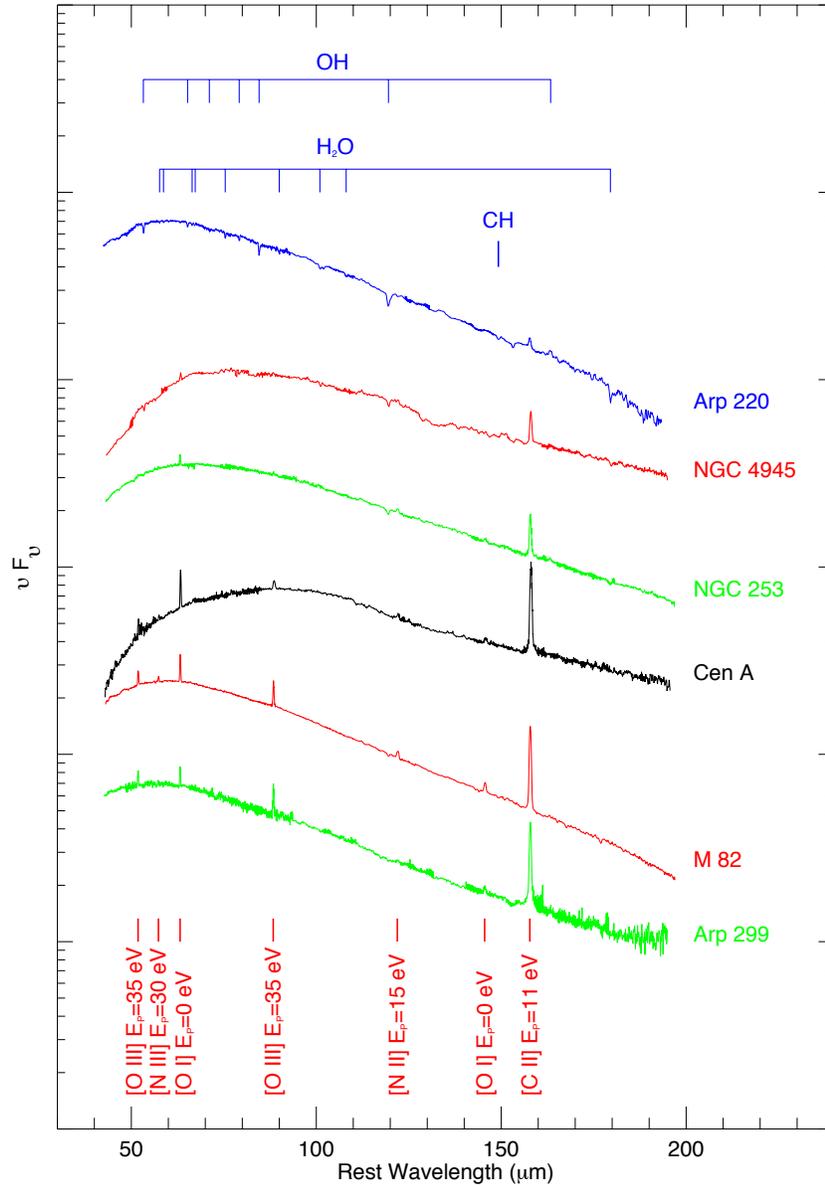
Fig. 1.— The full ISO Long Wavelength Spectrometer spectra of six IR-bright galaxies. The spectra have been shifted and ordered vertically according to the equivalent width of the [O III]88 μm line. The excitation potential, the energy required to create the species, is given in eV at the bottom of the figure. From Fischer et al. (1999).

Fig. 2.— The [O III]52 μm/[O III]88 μm line ratio versus the [O III]88 μm line to integrated far-infrared continuum flux ratio for the sample galaxies. The dashed and dotted lines show the [O III] line ratio in the low density limit (≤ 100 cm^{-3}) and for an electron density of 500 cm^{-3}, respectively (Fischer et al. 1999).
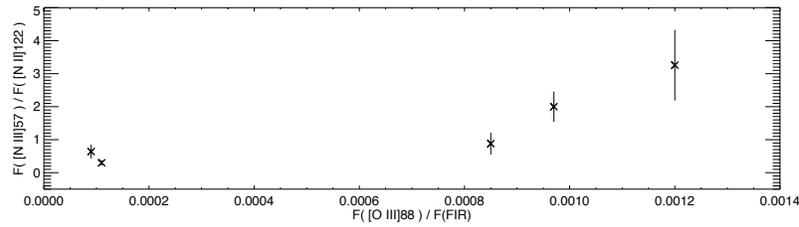
Fig. 3.— As in Figure 2 for the [N III]57 μm/[N II]122 μm line ratio.
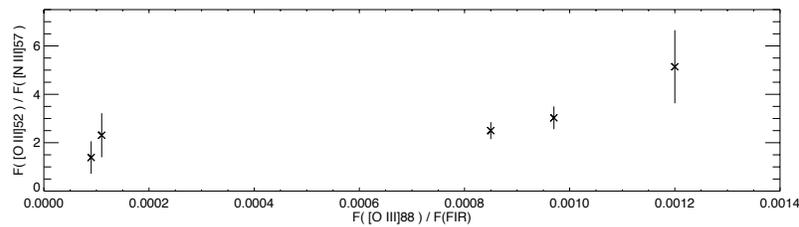
Fig. 4.— As in Figure 2 for the [O III]52 μm/[N III]57 μm line ratio.

greater than 100 for ionization parameters greater than 10^{-2}, density $= 100$ cm^{-3}, and stellar temperatures between 30,000 - 50,000 K.

Here we present photoionization models with starburst and power law ionization sources to predict the strengths of the fine-structure lines in dust-bounded nebulae and to compare them with the LWS spectra of the infrared-bright galaxies. Due to the weakness of the infrared fine-structure lines from photoionized gas in ULIRGs, ISO mid- and far-infrared spectroscopy produced mostly upper limits. Comparison of photoionization models of dust-bounded nebulae with spectra from future space missions such as SIRTF and Herschel will help to determine the conditions in the photoionized regions of these galaxies. The understanding yielded by these missions can then be used to probe the ionized media in high redshift galaxies by the missions being discussed in this workshop.

2. Photoionization modeling

The photoionization modeling was done using CLOUDY 94.01 (Ferland et al. 1998) for central power law (Figure 5) and instantaneous starburst (Figure 6) ionization sources. For power law models, the "table" power law option in CLOUDY was used. This option produces a continuum with f$_\nu \propto \nu^\alpha$ that is well behaved at both high and low energy limits (10^{-8} - 10^8 Rydbergs in CLOUDY). An index $\alpha = $ -1.5 was used for the mid-range ($10 \, \mu$m - 50 keV), while the default indices of +2.5 and -2 were used for the low and high ranges, respectively. For the starburst models, we used the instantaneous, Salpeter IMF, 3 and 5 Myr aged burst models of Leitherer et al. (1999) with solar metallicity and standard mass loss. H II region abundances were used in CLOUDY.

With a central ionizing source, the ionization parameter, defined as the ratio of ionizing photons to hydrogen atoms at the inner face of the cloud, $U = Q/4\pi r^2 nc = N_{H\,II}\,\alpha_B\,/\,c$, where Q is the central Lyman continuum rate, n is the density, r is the inner radius of the cloud, and α_B is the Menzel case B recombination coefficient. It was varied from $10^{-3} - 10^1$ by setting $Q = 4.5 \times 10^{54}$ sec^{-1} and varying the inner radius from 30 - 1600 pc.

The power law models predict a decrease in the [N III]57 μm/[N II]122 μm and [Ne V]14 μm /[Ne III]15 μm line ratios in the high U cases. For the starburst models, only the far-infrared line ratio decreased at high U, over the range of parameters explored. In both cases the line-to-luminosity ratios drop at high U, as expected. As Figures 5 & 6 show, high densities also produce both lower values of the line ratios
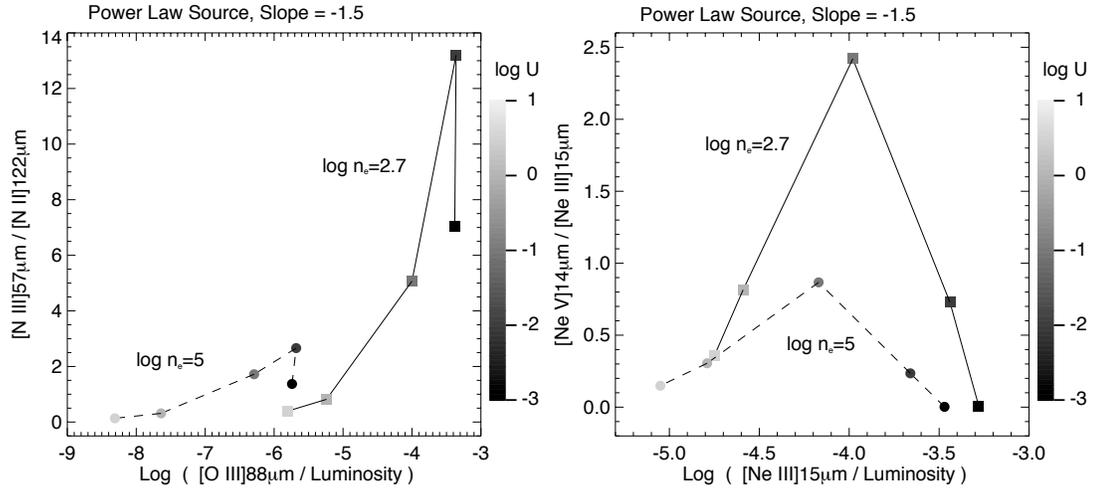
Fig. 5.— Far-infrared (left) and mid-infrared (right) line ratios vs. line-to-luminosity ratios for an AGN-like power law ionizing source for moderate (500 cm^{-3}) and high (10^5 cm^{-3}) densities are plotted for ionization parameters U, for log U = -3, -2, -1, 0, and 0.5 (as shown by the greyscale bar, right). See text for the details of the modeling.
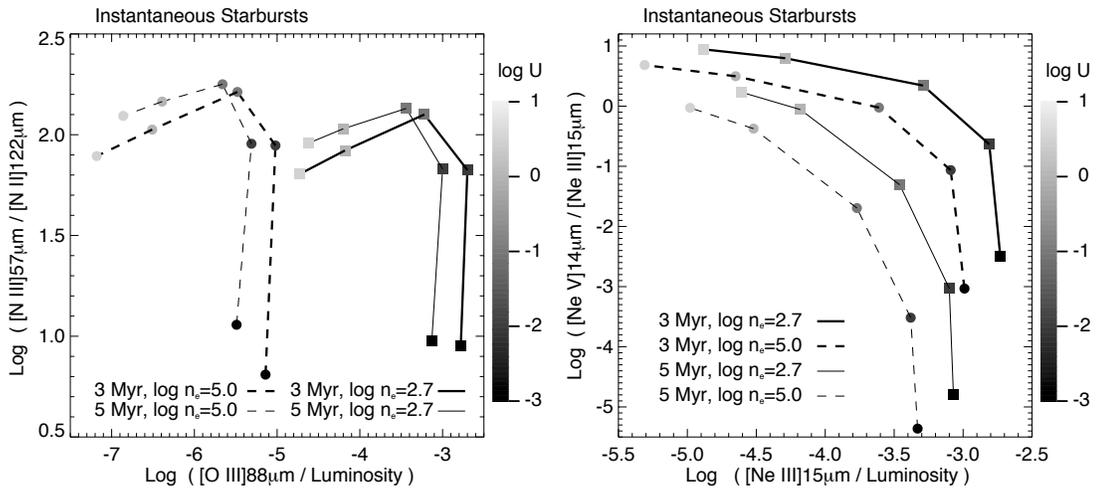
Fig. 6.— Same as Figure 5 (except for log scale in the line ratios), for ionization by instantaneous starbursts of ages 3 and 5 Myrs.

and lower line-to-luminosity ratios. Older starbursts can also produce these effects, but are unlikely to power ULIRGs (Satyapal et al. 2002).

3.   The effects of high U on photodissociation region diagnostics

The CLOUDY models presented here show that under high U conditions, line emission ratios from ionization states below the dominant stage of ionization at the illuminated face of the cloud, i.e., from constituents deeper in the cloud, are often inverted compared with moderate ionization parameters due to the effects of the strong grain absorption. Some of the other infrared characteristics of ULIRGs are warm $60/100\,\mu$m colors, low aromatic feature-to-luminosity and low [C II]$158\,\mu$m line-to-luminosity ratios, although their aromatic feature-to-[C II] ratios are normal (Dudley et al. 2002; Luhman et al. 2002). These characteristics may be the result of the effects of grains in ionized regions. For example in our own galaxy, Boulanger et al. (1988) find that $60/100\,\mu$m colors are higher and small grain emission lower, as traced by the IRAS $12\,\mu$m flux, within the California nebula H II region than outside of it.

SIRTF and Herschel studies will greatly improve our understanding of these diagnostics. SAFIR and future space-borne FIR/submillimeter interferometry will be able to exploit them in the high redshift universe to futher our understanding of the evolution of galaxies.

REFERENCES

Bottorff, M., Lamothe, J., Momjian, E., et al. 1998, PASP, 110, 1040

Boulanger, F., Beichman, C., Desert, F. X., et al. 1988, ApJ, 332, 328

Dudley, C. C., et al. 2002, in prep.

Ferland, G. J., Korista, K. T., Verner, D. A., et al. 1998, PASP, 110, 761

Fischer, J., Luhman, M. L., Satyapal et al. 1999, Ap&SS, 266, 91

Genzel, R. Lutz, D., Sturm, E., et al. 1998, ApJ, 498, 579

Harvey, V.I., Satyapal, S., Luhman, M.L, Fischer, J. et al. 1999, in *The Universe As Seen By ISO*, eds. P. Cox, M.F. Kessler, ESA SP-427, 889

Leitherer, C., Schaerer, D., Goldader, J. D. et al. 1999, ApJS, 123, 3

Luhman, M. L., Satyapal, S., Fischer, J., et al. 2002, ApJ, submitted.

Roche, P., Aitken, D., Smith, C. H., & Ward, M. 1991, MNRAS, 248, 606

Satyapal, S., Luhman, M. L., Fischer, J. et al. 2002, ApJ, submitted

Voit, G. M. 1992, ApJ, 399, 495

Star formation with Herschel

Paolo Saraceno & Sergio Molinari

saraceno@ifsi.rm.cnr.it
Istituto di fisica dello Spazio Interplanetario, CNR
Via Fosso del Cavaliere, 00133 Roma Italy

Abstract

We briefly discuss two of the major programmes that could be carried out with the Herschel satellite in the field of star formation: the photometric and spectroscopic imaging of the nearby star forming regions and the survey of the Galactic plane.

1) Introduction

HERSCHEL will give a unique contribution to the study of star formation because it will explore a spectral range crucial for these studies, with sensitivities, imaging capabilities, spatial and spectral resolution that have never been met in previous observations, which will not be achieved, below 350 µm, by the foreseen instrumentation for the next 10-15 years. The description of the Herschel satellite is given elsewhere in this conference. Here we simply summarize that HERSCHEL has broad band and line imaging capabilities in the 57-670 µm range, with a spatial resolution close to the diffraction limit and a field of view of several arcmin. High resolution spectroscopy (up to 10^7) will be carried on in the 157-212 µm and in the 240-625 µm range.

The uniqueness of HERSCHEL for star formation lies in its ability to discriminate among different evolutionary stages of young stellar objects (YSOs) and to define the temperatures and the masses of the circumstellar material and pre-stellar condensations (prior to the formation of the protostars). Fig 1 shows that the spectral energy distributions (SEDs) of both protostars and pre-stellar condensations peak in the 100–200µm range, inside HERSCHEL bands. The ALMA spectral bands, situated in the Rayleigh-Jeans portion of the spectrum, will be unable to provide dust temperature information. To further illustrate the potential of HERSCHEL, in Fig. 2 we report an ISO-LWS two-colours diagram, that reproduces the colours of the PACS instrument for a sample of YSOs of different luminosity (from ~ 1 to 10^5 Lo) and different evolutionary stage (from 10^3 years for Class 0 to more than 10^7 years for TTau and HAeBe stars). Different classes of objects populate well separated areas of the plot, making this diagram a powerful probe. If the same diagram would have been done using the ALMA colours, all the objects would be mixed in the same region because, in the Raleight Jeans regime, different temperatures have similar colours.

In conclusion, ALMA will be the ultimate instrument to study in detail the morphologies of the dust condensations, but only HERSCHEL's observations will define dust temperatures and therefore the masses. Moreover, thanks to its superior mapping capabilities, HERSCHEL will have no competitor, even in the ALMA bands, in the surveys of large areas of the sky.
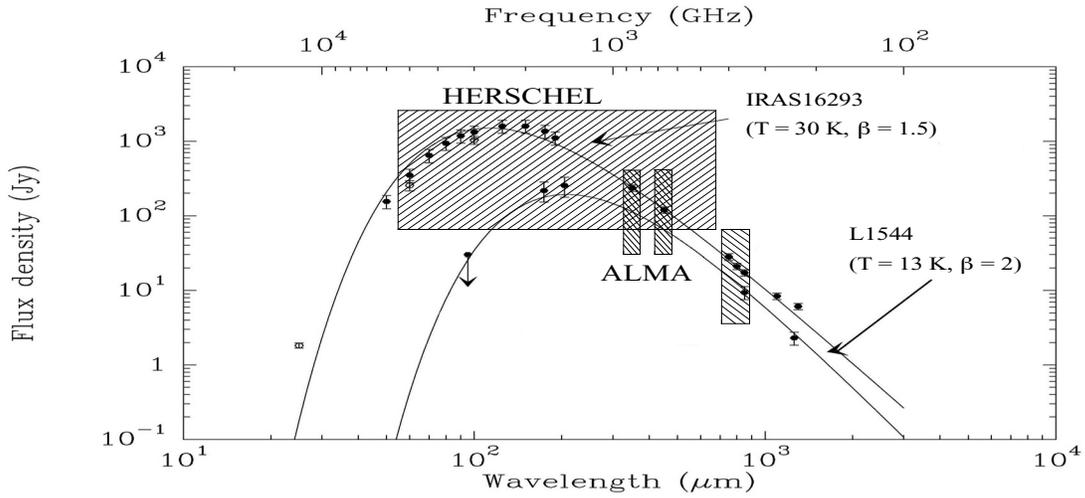
Fig.1 *(adapted from André et al. 2000): Spectral energy distribution of the Class 0 protostar IRAS 16293 and of the prestellar core L1554 obtained from ISO and JCMT observations, fitted with a simple grey body. The HERSCHEL spectral band and the three high frequencies bands of ALMA are reported on the figure.*
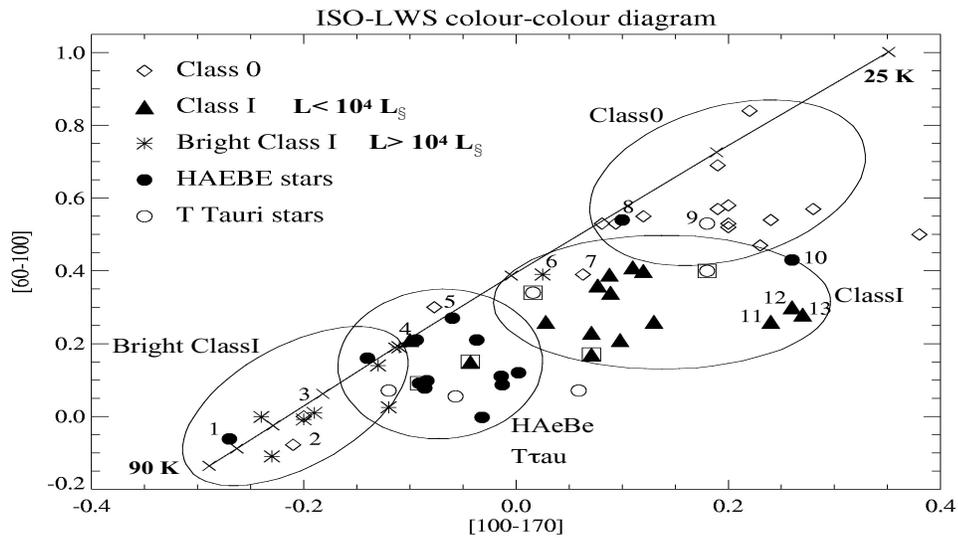
Fig.2 *(adapted from Pezzuto et al 2002): Two colours diagram of a sample of YSOs of different masses and evolutionary stages, observed with the ISO-LWS spectrometers; the spectra were averaged over 1 μm around the wavelength of the three PACS bands. The line shows the blackbody colours from 25 to 90 K.*

Starting from these considerations, in the following we will outline two key projects that we think HERSCHEL has to do: *i) the study of star formation in clusters* with PACS and SPIRE, (most of the time will be devoted to the PACS observations) and *ii) the survey of the Galactic plane at 250, 350, and 500 μm* with SPIRE, with a PACS and HIFI follow-up of selected areas.

2) Star formation in clusters

The formation of isolated stars is a process fairly well understood and several models have been proposed during the last two decades (e.g. Mouskovias & Ciolek 1999, Shu et al 1987, Palla & Stahler 1999, Bernasconi et al.1996). However, the observed large fraction of stars belonging to multiple systems and the observational evidence that most stars form in clusters (e.g. Nordh et al for Chamaleon; Lada et al. 1991 for L1630; Wilking and Lada 1985 for Taurus) has recently revolutionized the way we look at this process.

One aspect that has puzzled researchers for a long time is *the origin of the Initial Mass Function* (IMF) that is crucial to understand the evolution of the stellar populations in our and other galaxies. Millimetre surveys of protoclusters (e.g. Testi & Sargent 1998, Motte et al. 1998**.** Johnston et al. 2001), suggest that that the mechanism that fixes the final mass of the forming stars has to be found in protoclusters, which represent the first stages of cluster formation. But all these surveys are affected by large uncertainties in the determination of the masses, because the SEDs of the pre-stellar cores peak at higher frequencies (e.g. L1544 in Fig.1)

While it has been known for a long time that massive stars form in clusters, only recently it becomes more and more evident that *the formation of massive stars* is intimately related to the properties of the hosting clusters. In particular, the mass of the most massive object of a cluster seems to increase with the stellar density of the cluster (e.g. Zinnecker 1993). In the low mass star forming region of Taurus the average distance between stars is 0.3 pc (Gomez et al. 1993), while the intermediate mass stars like Herbig AeBe are found in clusters with a separation ranging from 0.2 pc, for the less massive objects, to 0.06 pc for the most massive ones (Testi et al. 1999). Finally, in the high mass stars formation region of the Trapezium cluster a stellar density in excess of 2200 pc^{-3} has been found (Herbig & Terendrup 1986) with an average distance among stars of less than 15.000 AU. Such distances are of the order of the sizes of the stellar envelopes, making highly probable that interactions among protostars occur during the star formation process.

Observations of young clusters and Protoclusters will then be a top priority of HERSCHEL whose high sensitivity (see Fig. 3) will allow to detect dust condensation of the mass of Jupiter at the distance of Orion, if they are warmer than 10 K. Observation of young clusters is also made easier because low mass stars are much more luminous during the protostellar phase than during the Main Sequence phase (e.g. Stahler 1988, Dantona & Mazzitelli 1994). All this suggest that the factor that will limit the HERSCHEL studies will be likely confusion (e.g. Franceschini et al. 1991) rather than sensitivity even with at the highest HERSCHEL spatial

resolution (~ 5"). This problem is very serious because protostars evolve fast (e.g. Saraceno et al 1996) and objects, in different evolutionary stages, with different physical properties that cannot be easily discriminated, are likely to be found in the same beam. An example of the influence of confusion is given in the following section.
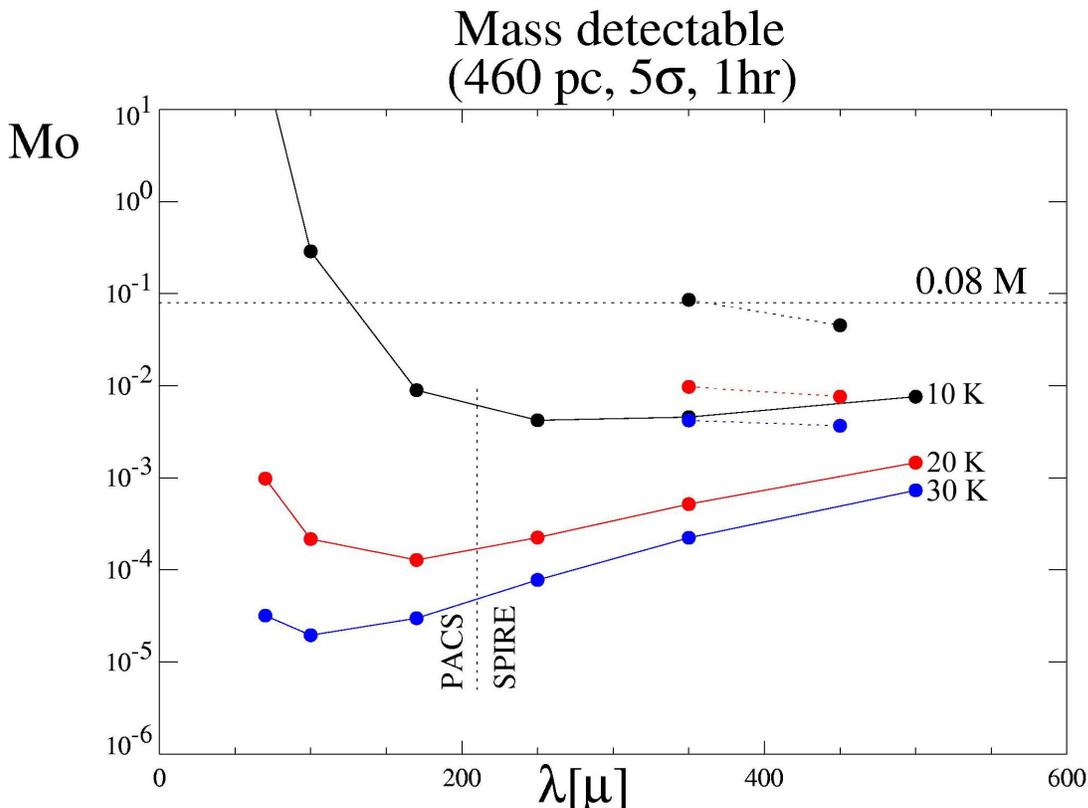
Fig.3 *Solid lines: the minimum detectable mass by HERSCHEL computed for dust temperatures of 10, 20 and 30 K, as $M_{dust} = S_v D^2 /(k B_v)$ (Hildebrand 1983) where S_v is the minimum detectable flux, D the distance of the cloud, k the dust emissivity and $B_v(T)$ the Planck function at the temperature T. The dashed lines on the right give the limits of the SCUBA camera on the JCMT (15 m) telescope. The horizontal dashed line correspond to the mass limit for Hydrogen ignition. All the spatially resolved proto-brown dwarfs should be detected by HERSCHEL.*

To conclude this point, we think that all the nearby clusters and protoclusters, with limited confusion problems, should be studied with Herschel at high priority not only in broad-band imaging, but also in spectroscopy. Indeed, traces of the interactions going on among the different members of a cluster during the various stages of formation should be found in the intra-cluster gaseous component as discussed in (Saraceno et al. 2000 and 2001) where also a possible target list is given.
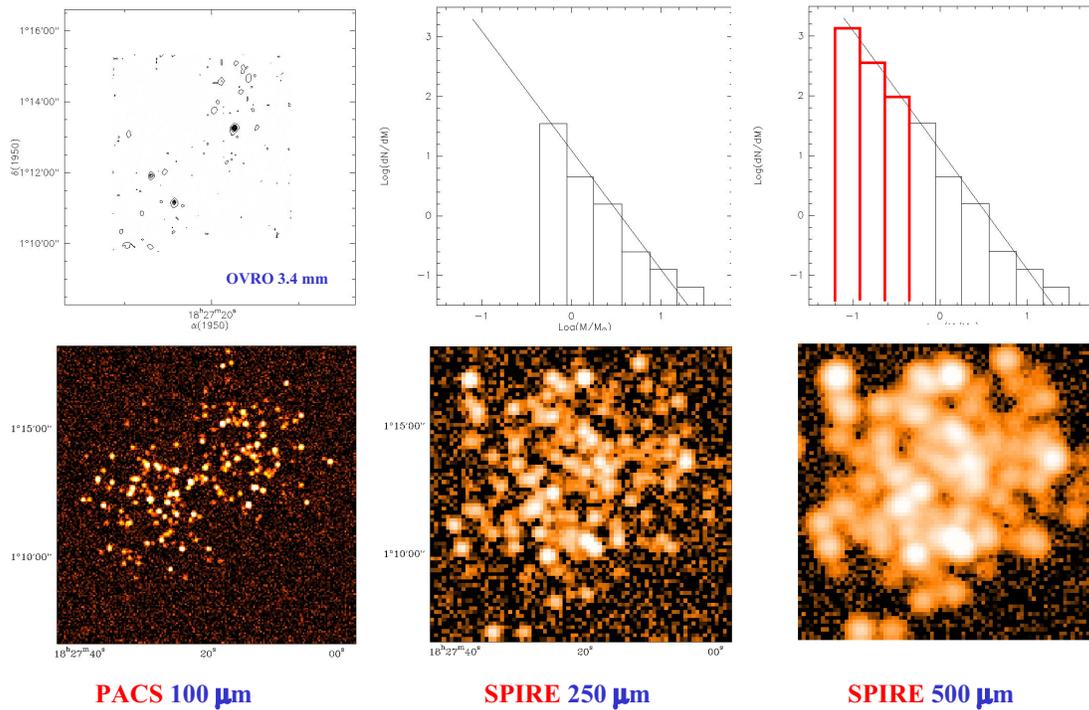
Fig.4: *HERSCHEL simulated observations of the Serpens protocluster. Top-left : the 5'x 5' map of the Serpens cloud core observed at 3mm with the OVRO interferometer (Testi & Sargent 1998); resolution is 5.5" x 4.3" and limiting sensitivity is 3 mJy. Top-centre: the observed mass function. Top-right: extrapolated mass function. In the bottom panels the three simulated field are reported.*

3) Simulated observation of a protocluster

In order to evaluate the confusion problems we simulated the HERSCHEL observations at 100 µm (with PACS), 250 and 500µm (with SPIRE) of the Serpens protocluster (d=350 pc). This region was mapped by Testi & Sargent (1998) with the OVRO interferometer at 3.4 mm with a resolution of 5.5" x 4.3" (similar to the PACS resolution) and a limiting sensitivity of 3 mJy.   Fig. 4 (upper left panel) shows the published millimetre map of the protocluster with the 26 detected condensations organised in two subclusters. The derived mass function is reported in the central upper panel. To simulate the number of possible sources that HERSCHEL could detect, we extrapolate the derived mass function down to the mass that corresponds to a 50 mJy flux at 100µm, (assuming T_{dust}=20 K, k=0.005 cm^2/g, β=1.5), resulting in 220 extra sources (right upper panel). These sources, assumed point like, were smeared with the diffraction pattern of the HERSCHEL 3.5m telescope at the three wavelengths and added to the observed field with a 2D-Gaussian distribution around on the 2 subclusters. Finally the noise expected for the needed integration times was

added to the simulated fields. The results for the 100, 250 and 500μm are reported in the three low panels of Fig 4.

We then used DAOFIND to analyze our simulated fields and construct the mass function resulting from the simulated observations, that we compared with the input one (Fig. 5). The result shows that we missed ~ 30% of sources because of the insufficient spatial resolution, considerably affecting the determination of the mass function.
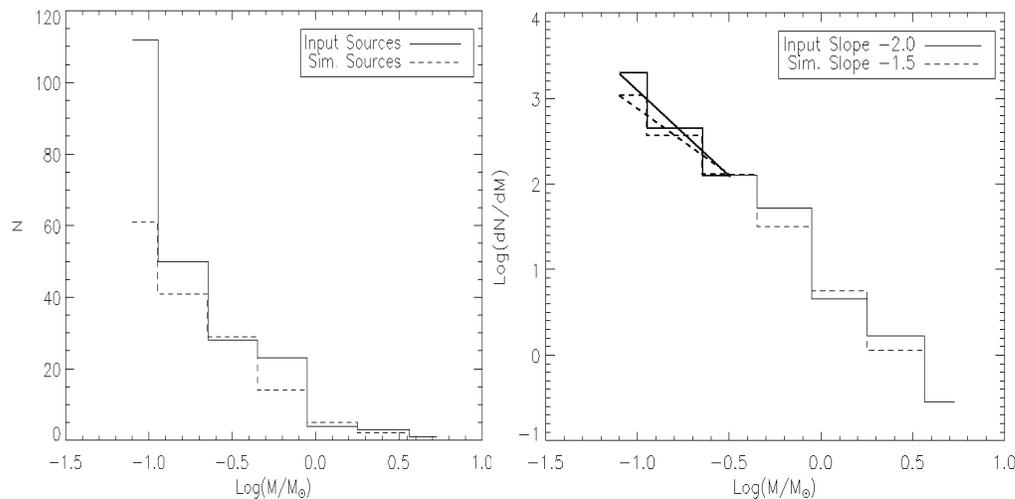
Fig.5: *On the left panel the input mass function, and the derived one (dot points) from the DAOFIND analysis, on the right, the effect on derived mass function.*

This very simple simulation shows that many more sources will be detected by HERSCHEL compared to ground based millimetre observations, allowing precise definition of temperatures and masses of the individual cluster members and producing large samples of high statistical significance. Given the field of view of SPIRE and PACS the entire Serpens protocluster can be mapped with few exposures with PACS and one exposure with SPIRE; *the time needed to survey this area at a sensitivity of 50 mJy at 5σ is of ~ 2 hours for the 60, 100, 170 μm bands (PACS), of only 0.4 hours for the 250, 350, 500 μm of SPIRE .* Since confusion will be the limiting factor, a high priority with HERSCHEL should be given to all the nearby (d<500 pc) clusters and protoclusters, where this problem is minimum.

4) A Survey of the Galactic Plane

The complete census of Galactic star forming regions in a wide range of masses and evolutionary stages is the new frontier that HERSCHEL will open for the first time. Indeed this task can be accomplished by surveying significant portions of the sky in the submillimeter continuum, where dust thermal emission is still substantially optically thin. The spatial resolution is also a critical parameter. It should not be too

high, to make the task realistic in terms of the time required to complete it. It should not be too low, to resolve the structure of the detected dense cores.
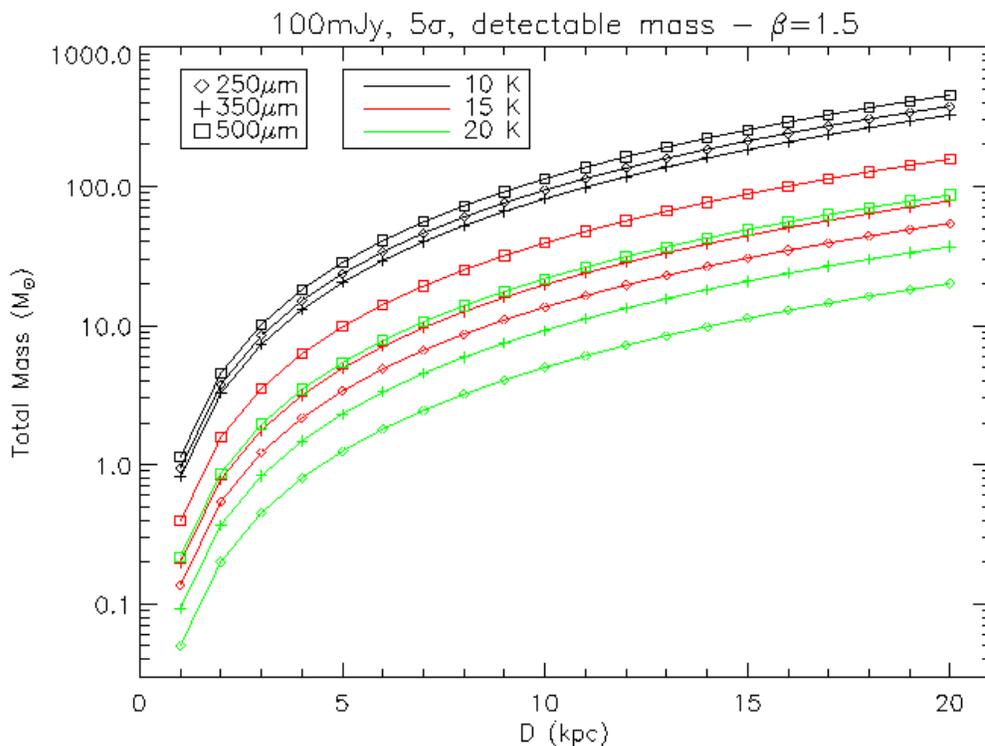
Fig.6: *Minimum (gas+dust) 5σ detectable mass as a function of distance from the Sun, for the 3 photometric bands (indicated by the three symbols), and for three dust temperatures. A β = 1.5 is assumed for the dust emissivity-frequency law.*

The instrument SPIRE offers a unique compromise in terms of wavelength coverage (250, 350 and 500μm simultaneous mapping), field of view (FOV=4'x8') and resolution (19", 27" and 38" HPBW). Star formation is mostly concentrated on the Galactic plane. The thickness of the molecular component of the Galactic disk is ~ 70 pc (Blitz 1990), corresponding to ~ 2.5^o at a distance of 1 kpc. A 5°-wide band centered on the Galactic plane should then contain all star forming regions at a distance greater than 1 kpc. We can cover this region with a set of strips obtained by scanning the telescope along the *b* Galactic axis. Using the latest SPIRE instrument performance figures and allowing for reasonable redundancy and scan overlaps, the full survey of the Galactic plane down to a 5σ sensitivity limit of 100 mJy in the three photometric bands can be completed in ~ 70 days. This estimate assumes 21 hours/day observing time and includes 10% observations overhead. The potentiality of this survey is summarised in Fig.6, where the flux sensitivity limit has been converted into total (gas+dust) detection limits at the 3 photometric bands for three

different dust temperatures as a function of distance from the Sun. Even a low-mass core like B335 (with a total mass of 3Mo and a dust temperature T~ 20 K) would be detectable at λ = 250μm up to a distance of 10 kpc from the Sun. PACS and HIFI follow-up investigations toward the most interesting areas of this survey are among its obvious outcomes. Physical and evolutionary characterization of the detected sources will be optimised by cross-analysing the survey database against similar databases in different wavebands but comparable spatial resolutions; obvious candidates for this type of study are the Midcourse Space Experiment (MSX, Shipman, Egan & Price 1996) 4.2-36μm survey of the Galactic plane, and the NRAO VLA Sky Survey (NVSS, Condon et al. 1998) at 6 cm for $\delta \geq$-40°.

BIBLIOGRAPHY

André, P., et al. 2000, "Protostars and Planets IV", 59

Blitz, L. 1990, Star Forming Giant Molecular Clouds, in: Bloemen H. (ed.) *The Interstellar Disk-Halo Connection in Galaxies*, IAU Symp. 144, p.41

Motte, F., André, P., and Neri R. 1998, A&A, 336, 150

Bernasconi, P.A., Maeder, A. 1996, A&A, 307, 829

Condon, J.J., et al.. 1998, AJ 115, 1693

D'Antona, F., Mazzitelli, I. 1994, ApJ, 90, 467

Franceschini, A., et al. 1991, A&AS, 89, 285

Gomez, M., et al. 1993, AJ, 105, 1927

Johnstone D, et al. 2001 ApJ 559, 307

Herbig, G.H., Terndrup, D.M. 1986. ApJ, 307, 609

Hildebrand R.H.Q. 1983, J.R.Astr.Soc., 24, 267

Lada, E.A., Bally, J., and Stark, A.A. 1991, ApJ, 368, 432

Mouskovias, T.C., and Ciolek, G.E. 1999, "The Origin of Stars and Planetary Systems", NATO Science Series, 540, 305

Nordh, L., et al. 1996, A&A, 315, L185

Palla, F., Stahler, S.W. 1999, ApJ, 525, 772

Pezzuto, S., et al. 2002, MNRAS in press

Saraceno, P., et al. 1996, A&A 309, 827

Saraceno, P., et al. 2000, Proceedings of "From darkness to Light: Origin and Evolution of Young Stellar Clusters", Cargese, April, 3-8, 2000

Saraceno, P, et al. 2001, *"The Promise of Herschel Space Observatory"* , ESA-SP-460, p.203

Shipman, R.F., Egan, M.P., Price, S.D. 1996, AAS Bull. 189, 6104

Shu, F.H., Adams, F.C., Lyzano, S. 1987, ARA&A, 25, 23

Stahler, W.S., 1988, ApJ, 332, 804

Testi, L., Palla, F., Natta, A. 1999, A&A, 342, 515

Testi, L., Sargent, A.I. 1998, ApJ, 508, L94

Wilking, B.A., Lada, C.J. 1985, "Protostars & Planets II",297

Zinnecker, H., et al. 1993, "Protostars & Planets III", 429

Star and early planet formation and the far-infrared/submillimeter regime

Rachel Akeson[1]

[1] Interferometry Science Center, California Institute of Technology, MS 100-22, Pasadena, CA, 91125, rla@ipac.caltech.edu

Abstract

This contribution was presented during the science panel discussion. The goal of this presentation is to describe why observations in the far-infrared and submillimeter are crucial to the study of star formation and the early stages of planet formation and to propose some specific areas that could be address by observations at these wavelengths.

1   Stages of star formation

A cartoon schematic of the stages of star formation in given in Figure 1, adapted from a figure in Andre (1994). Representative spectral energy distributions (SED) are also shown for each stage (neglecting line emission and absorption features). A classification scheme (Class I-III) for pre-main sequence stars based on the slope of the SED in the near and mid-infrared was proposed by Lada & Wilking (1984). This scheme was extended to the youngest, optically obscured stars by Andre et al. (1993). This sequence applies specifically to single, low mass stars, but intermediate mass stars (the Herbig AeBe's) are thought to have similar properties (see for example, Hillenbrand et al. 1992). Binary stars evolve through these stages as well, but are more complex due to interactions between the circumstellar/circumbinary material and the multiple stars.

In Class 0, the central protostar is completely obscured in the optical to mid-infrared by a massive envelope and the emission is dominated by a cold ($T \sim$ 10-30 K) dust component. This is the main accretion stage and there is also simultaneous outflow of material. In this stage, observations in the far-infrared to millimeter are the only method to locate these objects and study their characteristics.

In Class I, the envelope is somewhat dispersed, but accretion continues through the circumstellar disk. The stellar photosphere is visible (for some sources) in the optical, but long-ward of the near-infrared the SED is dominated by circumstellar emission from warm dust.

In Class II, the envelope is entirely dispersed, although the circumstellar disk is still optically thick. In the SED, the infrared excess is still present, but diminished. This class corresponds to the classical T Tauri stars.

In Class III, the circumstellar material is mostly or entirely dispersed. Any residual infrared excess is small. This class corresponds to the weak-line T Tauri stars.

~ 1000 AU

Class 0
– massive envelope
– high accretion
– outflow

cold BB

~ 100 AU

Class I
– residual envelope
– lower accretion rate

disk

star

Class II
– optically thick disk

star

disk

Class III
– little or no IR excess

star

disk?

λ (μm)
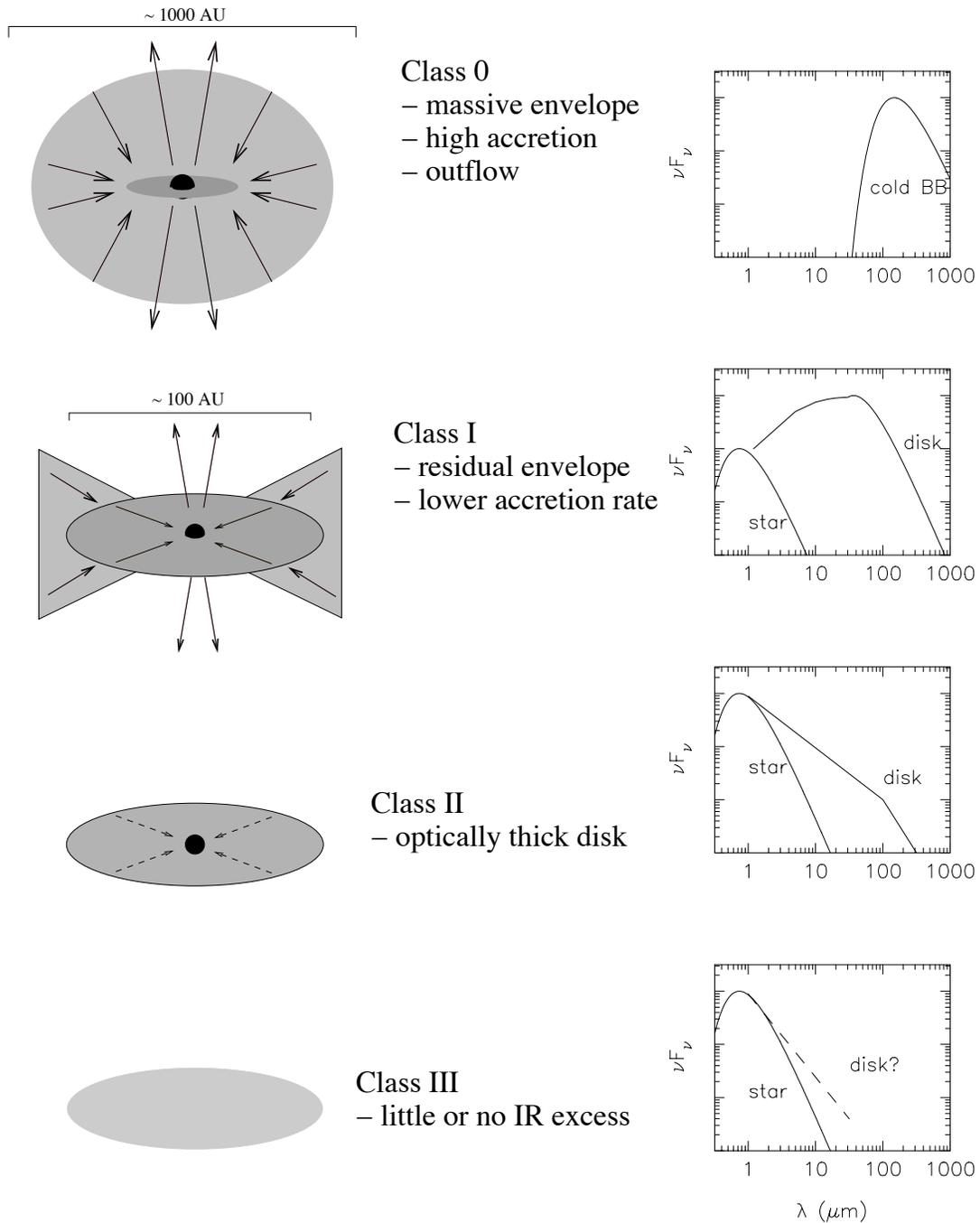
Figure 1: A schematic evolutionary star formation sequence

2 Why far-infrared and submillimeter observations?

The far-infrared and submillimeter wavelength regions are excellent probes of star formation because much of the dust continuum emission from the circumstellar material arises at these wavelengths. This is particularly true in the earliest stages of protostellar clump formation and collapse, which are completely obscured at optical and near-infrared wavelengths. The white papers discussing SAFIR (Rieke et al. 2002) and possible interferometry missions (Leisawitz et al. 2001) detail what these missions could contribute to the study of star formation. Other speakers in this workshop have also covered star formation (in particular, the contribution by Yorke).

Given the degeneracies in the physical model parameters (e.g. surface density and opacity) detailed observations at many wavelengths are necessary to fully study these objects. Thus, observations at far-infrared and submillimeter wavelengths are not only complimentary to the capabilities of such observatories as NGST and ALMA, observations from all these facilities are necessary to understand the processes involved in star and planet formation.

3 Some specific topics

Given below is a list of specific topics in star formation that could be addressed by future far-infrared/submillimeter observatories. This list is not intended to be complete, but to serve as a starting point for discussions.

A reminder to the reader on the physical size scales involved. In the Taurus cloud (the nearest well studied low mass star formation region) 1 arcsec corresponds to 140 AU, while 7 milliarcsec corresponds on 1 AU. In the Orion region (the nearest site of high mass star formation) 1 arcsec corresponds to 500 AU and 2 milliarcsec corresponds to 1 AU.

With resolution of a few arcsec

- Surveys for protostellar cores
- Resolving emission into envelope and disk components

  - Infall signatures
  - SED's are degenerate to several model parameters (e.g. debate over envelope vs. disk distributions for Herbig AeBe stars)

- Multiplicity / cluster formation

  - Intermediate and high mass stars often form in clumps
  - Test binary theories (capture vs. fragmentation)

With sub-arcsecond resolution

- Physical properties of circumstellar disks

    - Accretion rates, temperature distributions, lifetimes
    - Dust properties (size and composition)
    - Density distribution, gaps due to proto-planets
    - Particularly important for determining initial conditions for planet formation

- Jets/winds

    - Driving mechanism
    - Kinematics of ionized and molecular components

- High mass cores / ultra-compact HII regions

    - Lifetime "problem"
    - Disk photoevaporation
    - Kinematics and chemistry of hot molecular cores

Polarimetry

- Characterize magnetic field structure

This work was performed at the Infrared Processing and Analysis Center, California Institute of Technology.

References

Andre, P., 1994, in The Cold Universe, T. Montmerle, p.179

Andre, P., Ward-Thompson, D. & Barsony, M., 1993, Ap.J., 406, 122

Hillenbrand, L.A., Strom, S.E., Vrba, F.J. & Keene, J., 1992, Ap.J., 397, 613

Lada, C.J. & Wilking, B., 1984, Ap.J., 287, 610

Leisawitz, D., et al., 2001, Probing the Invisible Universe: The Case for Far-IR/Submillimeter Interferometry

Rieke, G., et al., 2002, Charting the Winds that Change the Universe II: The Single Aperture Far Infrared Observatory

INTERSTELLAR DUST MODELS CONSISTENT WITH EXTINCTION, EMISSION, AND ABUNDANCE CONSTRAINTS

Viktor Zubko[1], Eli Dwek, and Richard G. Arendt

Laboratory for Astronomy & Solar Physics
NASA Goddard Space Flight Center, Greenbelt, MD 20771

ABSTRACT

We present new interstellar dust models which have been derived by simultaneously fitting the far ultraviolet to near infrared extinction, the diffuse infrared emission, and, unlike previous models, the elemental abundances in dust for the diffuse interstellar medium. We found that dust models consisting of a mixture of spherical graphite and silicate grains, polycyclic aromatic hydrocarbon (PAH) molecules, in addition to porous composite particles containing silicate, organic refractory, and water ice, provide an improved fit to the UV-to-infrared extinction and infrared emission measurements, while consuming the amounts of elements well within the uncertainties of adopted interstellar abundances, including B star abundances. These models are a significant improvement over the recent Li & Draine (2001, ApJ, 554, 778) model which requires an excessive amount of silicon to be locked up in dust: 48 ppm (atoms per million of H atoms), considerably more than the solar abundance of 34 ppm or the B star abundance of 19 ppm.

---

[1]On leave from Department of Physics an Astronomy, University of Kentucky, Lexington, KY 40506–0055

Despite the considerable efforts of many researchers over past decades, we are still missing a thorough and reliable model of interstellar dust that would consistently expain a variety of available observed data. For example, even the composition of interstellar dust is still a matter of debate.

In the work presented here, we explore possible dust models that simultaneously comply with the three major observational constraints (see Figure 1): the average interstellar extinction, the thermal infrared emission from the diffuse ISM, and the interstellar abundance constraints. Other important constraints such as the interstellar polarization and X-ray halos are left for future research.

From a mathematical point of view, the problem of deriving the grain size distributions and composition reduces to a Fredholm integral equation of the first kind. To solve this typical ill-posed inverse problem, we implemented a special tool, the method of regularization (Tikhonov et al. 1995; Zubko 1997). Compared to other approaches, the method of regularization requires the minimum amount of information: the data to be fit and their uncertainty, without a need for default or template solutions.

Recently, Li & Draine (2001) proposed a dust model that consists of the polycyclic aromathic hydro-carbons (PAHs), graphite and silicate grains. Their model is consistent with the observed extinction and emission data, but requires too much silicon, magnesium, and iron to be in the dust: 48 ppm (atoms per million H atoms), significantly more than the maximum available solar abundance of 34 ppm.

Here we show that the Li & Draine model can be optimized by choosing different, more general size distributions (BARE models), thus producing good fits to the observational constraints without violating the abundance constraints. A more complex dust model (COMP) consisting of PAHs, bare graphite and silicate grains, and porous composite particles made up from silicates, organic refractory, and water ice, provides a somewhat improved fit to the observed constraints, including the interstellar abundances. This kind of dust model looks more realistic than the model in the light of the current view on the dust evolution and available IR data.

Our favored model is a COMP model with the solar abundances from Holweger (2001) assumed as the interstellar medium abundances (see Figure 2). It has minimum discrepancies in fitting the extinction and IR emission; it also looks better in explaining the observed infrared extinction and scattering properties: albedo and asymmetry parameter. Note that we also find equally good COMP models using B star and F and G stars abundance constraints.

One of the main results of the work is that no unique dust model can be chosen. In part, this can be explained by the paucity of observational constraints compared to the number of free model parameters.

REFERENCES

Arendt, R.G., et al. 1998, ApJ, 508, 74

Dwek, E., et al. 1997, ApJ, 475, 565

Fitzpatrick, E.L. 1999, PASP, 111, 63

Grevesse, N., & Sauval, A.J. 1998, Space Sci. Rev., 85, 161

Holweger, H. 2001, in Solar and Galactic Composition, ed. R.F. Wimmer-Schweingruber (Berlin: Springer), in press

Li, A., & Draine, B.T. 2001, ApJ, 554, 778

Sofia, U.J., & Meyer, D.M. 2001, ApJ, 554, L221

Tikhonov, A.N., Goncharsky, A.V., Stepanov, V.V., & Yagola, A. 1995, Numerical Methods for the Solution of Ill-Posed Problems (Dordrecht, the Netherlands: Kluwer Academic Publishers)

Zubko, V.G. 1997, MNRAS, 289, 305

Zubko, V., Dwek, E., & Arendt, R.G. 2002, in preparation
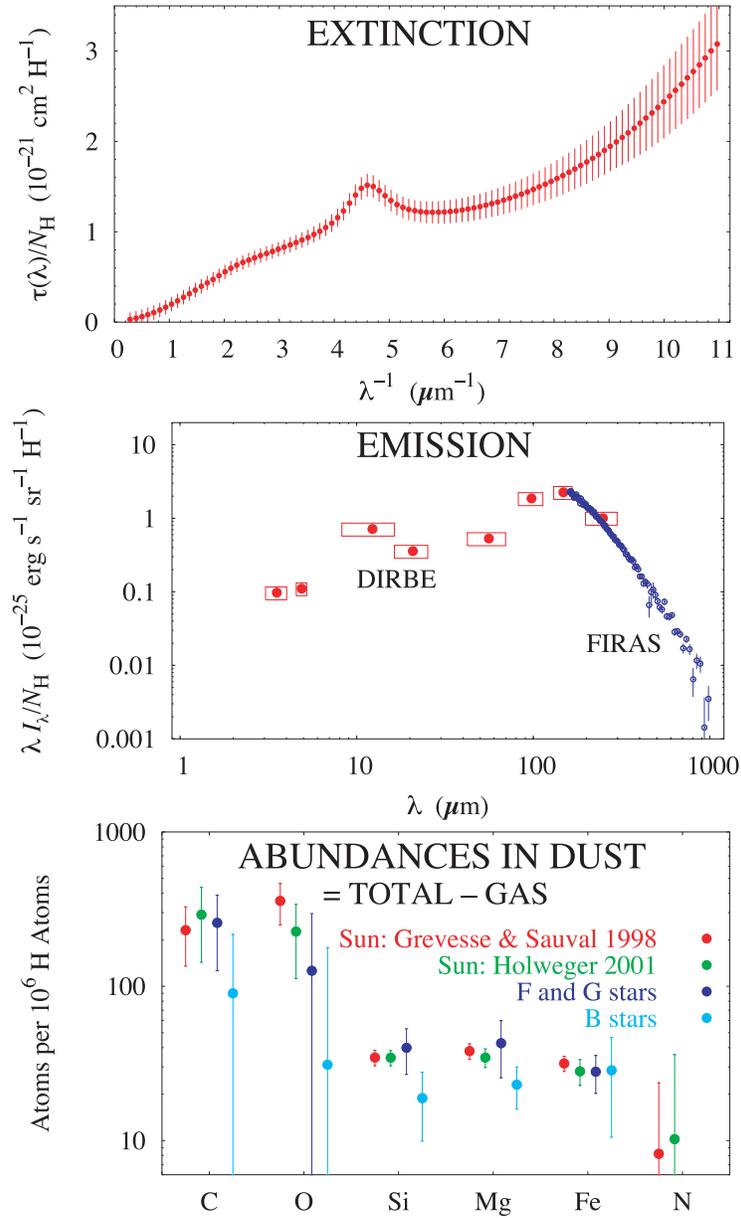
Fig. 1.— The mean Galactic extinction curve (Fitzpatrick 1999) (upper panel), infrared emission from high Galactic latitudes (Arendt et al. 1998; Dwek et al. 1997) (middle panel), and abundances in dust (lower panel).
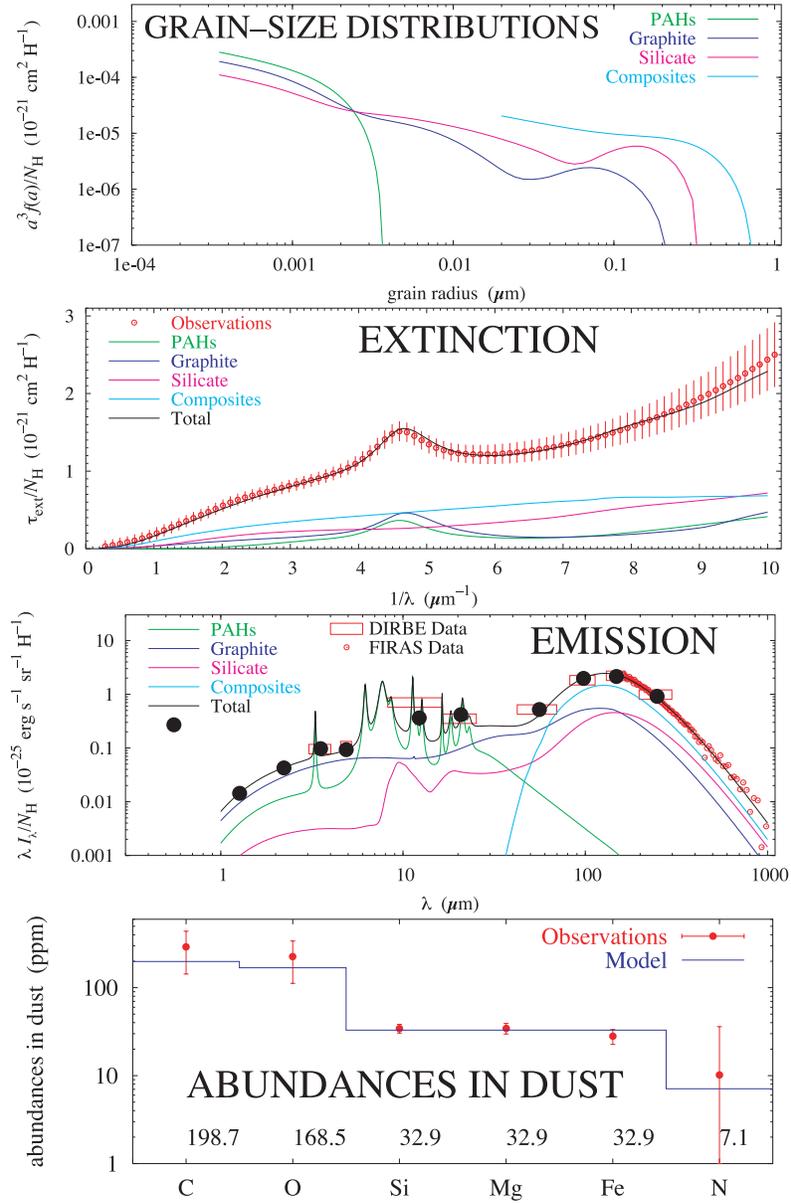
Fig. 2.— Our preferred dust model, COMP-S2: size distributions, extinction curve, emission spectrum, and elemental requirements. For abundances, solid line and numbers along the x-axis show model results

C^+ in ULIRGs and the Cosmos

C. C. Dudley,[1,2] J. Fischer,[1] M. Luhman,[3] M. Wolfire[4] & S. Satyapal[5]

`dudley@vivaldi.nrl.navy.mil`

ABSTRACT

SAFAIR will be able to measure the emission of 158 μm[C II] out to a redshift of 0.9 bridging an important gap in redshift space where the evolution of galaxies is a steep function of redshift. It will easily detect C^+ in normal galaxies of moderate luminosity and thus also in Ultraluminous Infrared Galaxies (ULIRGs) whose 158 μm[C II] line luminosities show an order of magnitude deficit compared with lower luminosity galaxies (Luhman et al. 1998, 2002). Present data on local ULIRGs, LIRGs and normal galaxies (more than three decades in luminosity) suggest that the onset of the deficit is abrupt (occurring in < a few dex) and centered near $L_{IR} = 10^{12}$ $L_\odot$.

Two scenarios which may have important consequences for observational cosmology are presented here. In the first scenario, if a high ionization parameter explains the deficit (Luhman et al. 2002) and enrichment is a gradual process then the apparently sharp break may migrate to higher luminosities as a function of redshift. In the second scenario, if the break is stable in luminosity, then the high surface density of SCUBA sources may allow a test of the isotropy of Λ. In the first scenario, SAFAIR will provide a ring-side-seat for observing galaxy evolution. In the second, SAFAIR will aid in the measurement of the geometry of the universe. Access to $z < 4$ may be possible if further observations of local sources confirm a possible 63 μm[O I] deficit of similar sharpness and strength.

[1]Naval Research Laboratory, Remote Sensing Division, Code 7213, 4555 Overlook Ave. SW, Washington, DC

[2]NRL/NRC Research Associate

[3]Institute for Defense Analysis, Alexandria, VA

[4]University of Maryland, College Park, MD

[5]George Mason University, Fairfax, VA

1.   Introduction

Advances in our technical abilities lead to advances in our knowledge which motivate further development of our experimental reach. Nowhere do space agencies do better service to astronomy than in providing observatories that defeat the opacity of the Earth's atmosphere. From the discovery of gamma ray bursts and X-ray transients, to the measurement of molecular hydrogen by *Copernicus* to the amazing results of IRAS on external galaxies, our view of our universe has been fundamentally transformed in each instance. The space age has opened new fields, and has brought a transformation of the whole field as well. Astronomers are more accustomed than any to thinking big, but now we are thinking big together as a result of the enormous planning and preparation needed to execute space missions successfully. As we work to plan ever more sophisticated missions, having delivered in passing the biggest bang-for-the-buck in existence (COBE), we in turn do service to space agencies by broadening their ability to carry out missions in non-astronomical endeavors.

As it turns out, the universe is probably somewhat blue (Hauser and Dwek 2001) but it has a lot of dust so we need to look at it longward of 2 μm or shortward of 2 keV to see it when it is busy making things for the most part. It is also expanding, so we need to think red. The *Infrared Space Observatory* (ISO) has provided us the opportunity to do some detailed thinking about some of the most powerful sources in the local universe and here we give a summary of one of our most important results from far infrared spectroscopy of ULIRGs and discuss possible implications of these results for more advanced missions and in particular SAFAIR.

2.   Results

Luhman et al. (1998) reported that the cooling line 158 μm[C II] was weak compared to FIR dust continuum emission in 5 of 6 ULIRGs by ~an order of magnitude compared to less luminous galaxies. Luhman et al. (2002) report that this result holds for 12 of 15 ULIRGs with the other 3 showing a more normal line strength. Fig. 1 reproduces the results of Luhman et al. (2002). Analyzing this line and other photo-dissociation region (PDR) lines and features, they conclude that the most likely explanation for the deficit in C^+ is actually an excess of continuum not associated with PDRs. They venture that a high ionization parameter such as may occur in the vicinity of AGN (Voit 1992) or in ultra compact H II regions (Bottorff et al. 1998; see Fischer et al. (2001) for detailed modeling and comparison of the effects of starburst and AGN input spectra) could provide the excess continuum emission.
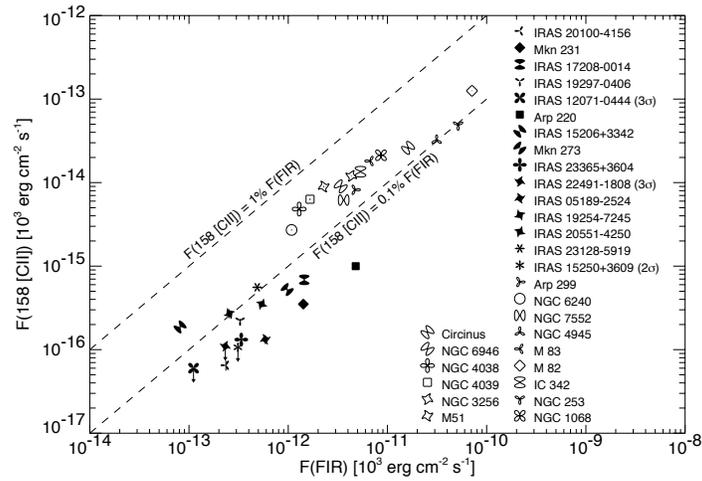
Fig. 1.— For details see Luhman et al. (2002) and references therein. ULIRGs are shown as filled symbols. The dashed lines mark the regime typical of normal and starburst galaxies.
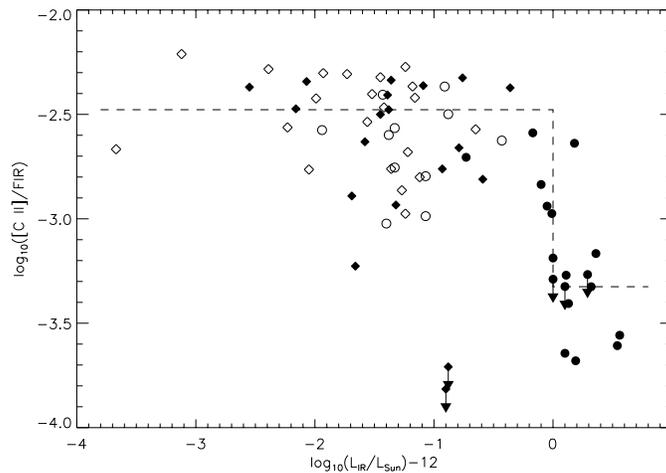
Fig. 2.— As given in Luhman et al. (2002), the ratio of C^+ intensity to FIR flux is plotted as a function of IR luminosity for the present sample (circles) as well as the US Key Project sample (Malhotra et al. 2001, diamonds). Sources plotted as open symbols may have both continuum and line emission that extend beyond the LWS beam. The rarity of C^+ deficit sources below 10^{12} $L_\odot$ and the ubiquity above is suggestive of a Heaviside step function (dashed line) changing from the median of the filled diamonds to the median of the filled circles at and above 10^{12} $L_\odot$ (ignoring upper limits).

Fig. 2 shows the C^+ deficit plotted against luminosity for sources plotted in Figure 1 and US Key Project sources appearing in the IRAS BGS Soifer et al. 1989; Sanders et al. 1995). The filled symbols should be entirely free of aperture effects. The figure shows that while the chances of finding deficit sources below 10^{12} $L_\odot$ is low (1:18 all sources) it is high at and above this luminosity (5:1). The change appears to be abrupt, and while further observations are needed to measure the sharpness of this change, we presume for most of the following that $\frac{\frac{\Delta[CII]}{FIR}}{\Delta log L_{IR}} \lesssim -7$ per 0.2 dex.

3.   Discussion

3.1.   The effects of high U

It has been proposed that either AGN or super starbursts account for the luminosities of ULIRGs (Sanders et al. 1988; Joseph & Wright 1985) and in either case a high ionization parameter ($U = \Phi(H)/cn_e$) can play a role in explaining the observed weakness of emission lines throughout the spectra of ULIRGs. When dust is present in an astrophysical plasma, it is generally thought to be heated via Lyα heating (Krishna Swamy & O'Dell 1968, however see Smith, Larson & Fink 1981 and Thronson 1983 for evidence of a molecular envelope associated with NGC 7027) which allows a sufficient path length through resonant scattering for dust to compete with two photon emission (Spitzer & Greenstein 1951) for these photons. Once this effect was understood, what is sometimes referred to as an infrared excess in the literature above the emission expected from Lyα heating needed explanation which is often provided by PDRs of $\sim$10 A_V (Tielens & Hollenbach 1985) at the edges of H II regions. The excess described by Luhman et al. (2002) is over and above the usual infrared excess, and thus needs further explanation. Direct heating of grains in plasmas can usually be neglected relative to Lyα heating, however, at high U, when dust is present, the path length from the ionizing source(s) to the edge of the plasma may approach an A_V or so and this leads to efficient competition by grains for ionizing photons. In such a case the plasma is dust bounded (Bottorff et al. 1998; Wood & Churchwell 1989) and infrared continuum rather than recombination lines comes to trace the intrinsic ionizing flux (and more). Such considerations led Voit (1992) to predict that ULIRGs might have weak emission lines in the infrared.

High U is generic with tens of pc of AGNs and should also occur when massive star clusters are formed. While AGNs probably don't form PDRs since they aren't ionization bounded (Ferland & Netzer 1983) dust bounded plasmas generated by star clusters may not form significant PDRs either since dust in the plasma will also

absorb non-ionizing radiation. Thus, PDR diagnostics and indeed plasma diagnostics may not easily distinguish the two dust bounded cases (Fischer at al. 2001 and this Workshop). However, U depends on n_e and not on the dust fraction. For very low metalicity systems it would be nearly impossible for a plasma to be dust bounded so that recombination lines would once again trace the ionizing flux in erstwhile ionization bounded plasmas regardless of U. Naturally, in this case, infrared continuum would be weak.

Pulling back from this extreme, if U can be taken as an increasing function of infrared luminosity (to explain the weakness of emission lines in ULIRGs, Fischer et al. this Workshop) and if ULIRGs form from gradually less and less enriched systems at earlier times, we might expect the break seen in Fig. 1 to migrate to higher luminosities as a function of redshift. Since we are discussing some of the most luminous sources, this is something that SAFAIR should be able to observe with ease.

All of this requires further detailed modeling but there is hope that this should be accomplished prior to launch.

3.2.   Could the break be stable in luminosity?

Obviously, ULIRGs require some dust to be ULIRGs and they also seem to be associated with major mergers, so there may be reason to think that ULIRGs occur in situations where the ISM is already enriched, perhaps at a minimum level. Could Cosmological Ultraluminous Infrared Sources (CULIRSs) be enriched at the same level as local ULIRGs? One senario where this might be the case would be if both ULIRGs and CULIRSs were powered by AGN (however see Trentham 2001 for an alternative accounting). Gebhardt et al. (2000) have associated the gravitational potential of local, dense, non-rotationally supported stellar systems with their central black hole masses. If ULIRG luminosities trace a portion of the accretion history of these black holes, where it is taken as given that major mergers may result is such stellar systems (eg Kormendy and Sanders 1992), then their environments are bound to be enriched owing to the systems from which they evolve. Note that Mihos & Hernquist (1994) propose that precursors to ULIRGs may require bulges to retain enough gas for a ULIRG phase based on dynamical modeling which, in light of the results of Gebhardt et al (2000), would lend support to the suggestion by Osterbrock (1993), in response to the evolutionary connection drawn by Sanders et al. (1988) between ULIRGs and quasars, that ULIRGs are built from pre-existing AGN.

At the same luminosity, CULIRSs would be tracing essentially the same step in

black hole accretion history as local ULIRGs and thus would very likely be associated with major mergers of well organized, relaxed and enriched systems.

Evidence that provides consistency with this view in terms of large scale structure formation is provided by (1) quasars at all redshifts being enriched, and (2) the suggestion by Stockton (2001) that morphologically elliptical sources demonstrably older that 3 Gyr at z=1.5 may select for a cluster environment with greater fidelity than quasars at the same redshift. Thus, the field environments where local ULIRGs are found may be provided by pre-cluster (over dense in the sense of Rees & Ostriker 1977) regions of earlier times which would have a head start in the formation of well organized (relaxed) precursor systems. In this senario one would expect $\frac{\Delta L_{\text{break}}}{\Delta z} \sim 0$ since the enrichment of CULIRSs would similar to that of local ULIRGs.

On the other hand, even granted the simplifying assumption (for this is in doubt) that ULIRGs are powered by AGN, it is not obvious that all CULIRSs need be as well, particularly as the probability of mergers of three or more systems increases at earlier (denser on average) times. The CULIRS senario discussed here might be diluted by sources of a different nature whose levels of enrichment would not be constrained by identifying similar phases of black hole accretion history locally and at high redshift.

3.3.   Implications of a stable sharp break

Before continuing, we should say that we don't actually know that the break shown in Fig 2. is sharp. One can discount NGC 4418 (lowest limit) as an oddball (modulo the unknown role of such source in the origin of the hard X-ray background) AGN (Roche et al. 1986, Dudley & Wynn-Williams 1997, Spoon et al. 2001) and the same may be true of NGC 1266 (Mouri et al. 1998) while further work is needed for IC 860 (three lowest filled diamonds) but observations of many more sources will be need to look carefully at the range $11.7 < log(L) < 12.3$ to ensure that the break is well defined. If it is, and if it is non-evolving, what are the consequences?

SCUBA sources are estimated to have a surface density of $\sim 10^4$ per square degree above 1 mJy at 850 microns and should have a redshift range of about 1–4 if their SEDs are similar to ULIRGs (Sanders 2000). Note that this puts them just past the capabilities of SAFAIR for C^+ observations in its presently conceived configuration though 63 μm[O I] would be available (see Fig. 3). Regardless, ULIRGs of $z < 1$ should be under represented in SCUBA surveys and strong evolution is suggested for ULIRGs starting with a mean $z = 0.15$ (Kim & Sanders 1998) so that the following estimates may also apply for $0.5 < z < 0.9$. This SCUBA source surface density

implies that in 30 square degrees, a redshift bin of width $\Delta z = 0.01$ should contain about about 1000 sources up to the vagaries of large scale structure. If a third of these fall within 0.3 dex of the break then it may be possible to determine the break position as a function of FIR flux to 0.03 dex (7%). If so, the break would provide a fairly precise standard candle for a given redshift bin over a solid angle scale that allows coverage in ~ 2 Ms (500 s pointings with a $3'$ field of view based on $1\sigma(300$ km/s FWHM) $=10^{-22}$ W/m^2 in 1 h Leisawitz private communication; for $z = 0.9$, Arp 220 at $> 10\sigma$, Arp 299 at $> 30\sigma$ in C$^+$). If it is a true standard candle, it is obvious that Λ could be well and easily (Riess et al. 2001) measured in this way. On the other hand, other standard candle methods, even if they rely on sources with a lower (per year) surface density but which provide a distance estimate on each occurance, could be used to calibrate the method described here.

Strong evidence for $\Lambda \neq 0$ is of recent origin and has spurred some thought on its relation to inflation. In one model, that if subject to anthropic selection (Peebles & Vilenkin 1999), angular anisotropy in a quintessence field would not be expected to be measurable. Tracker models (eg Steinhardt, Wang & Zlatev 1999) are conceived to describe acceleration that is variable in time. If the acceleration is squeezed out of structure formation (Wetterich 2001) then angular anisotropy might result though the angular scale is uncertain. The observations reported by Stockton (2001) are suggestive that $d\Lambda/dz \sim 0$ (if we may be permitted this tautology) nearly out to redshifts where the big bang background would inhibit sufficient baryon cooling to allow star formation.

It seems likely that even an uncalibrated differential experiment over regions of 30 square degrees might set interesting limits on any angular anisotropy, while a calibrated experiment might provide a dense sampling of time variability if present, thus constraining the nature of the acceleration and related models of of its origin.

The CMB dipole measured by COBE is thought to arise from the Milky Way's motion with respect to the universe on very large scales. Evidence consistent with this interpretation has been provided by number counts of non-local radio sources (Blake & Wall 2002). Observations of the break should provide a measure of the actual velocity of the Milky Way relative to a similarly distant background just so long as the break is persistent an non-evolving or homogeneously evolving.

4.   Conclusion

The recently discovered and now confirmed C^+ deficit in ULIRGs may indicate a quite sharp break in luminosity space which, if stable, suggests a potentially useful standard candle for future far infrared and submillimeter observatories of sufficient sensitivity such as SAFAIR. Alternatively, the migration and/or persistence of the break may probe the enrichment history of the universe.
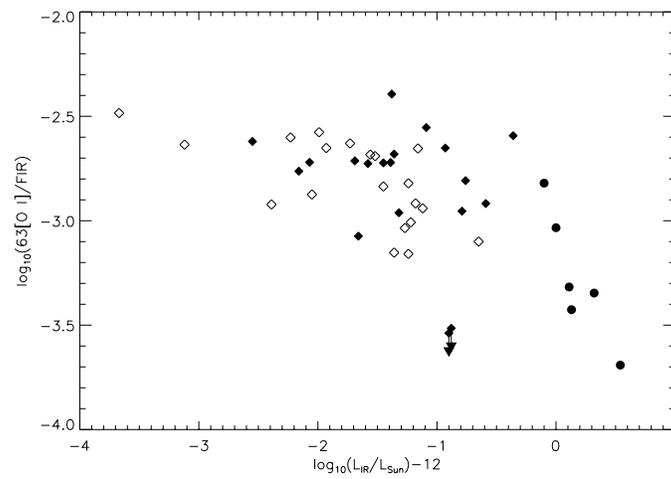
Fig. 3.— It is possible that 63μm[O I] may also have a sharp break (symbols the same as Fig. 2; ULIRG data from Luhman et al. (2002)) which may extend the reach of SAFAIR in redshift.

REFERENCES

Blake, C. & Wall, J. 2002, Nature 416 150

Bottorff, M., Lamothe, J., Momjian, E., Verner, E., Vinkovic, D. & Ferland, G. 1998, PASP 110 1040

Dudley, C. C. & Wynn-Williams, C. G. 1997, ApJ 488 720

Ferland, G. J. & Netzer, H. 1983, ApJ 264 105

Fischer, J. et al. 2001, in "Spectroscopic Challenges of Photoionized Plasmas" ASP Conference Series, eds. G. Ferland & S. Wolf Savin 247 499

Gebhardt, R. et al. 2000, ApJ, 539, 13

Hauser, M. G. & Dwek, E. 2001, ARA&A 39 249

Joseph, R. D. & Wright, G. S. 1985, MNRAS 214 87

Kim, D.-C. & Sanders, D. B. 1998, ApJS 119 41

Kormendy, J. & Sanders, D. B. 1992, ApJ 390 L53

Krishna Swamy, K. S. & O'Dell, C. R. 1968, ApJ 151 L61

Luhman, M. L. et al. 1998, ApJ 504 L11

Luhman, M. L., Satyapal, S., Fischer, J., Wolfire, M., Sturm, E., Dudley, C. C., Lutz, D., & Genzel, R. 2002, in prep

Malhotra, S., et al. 2001, ApJ 561 766

Mihos, J. C. & Hernquist, L. 1994, ApJ 431 L9

Mouri, H. Tanaguchi, Y., Sato, Y. & Kawara, K. 1998 A&A 334 482

Osterbrock, D. E. 1993, ApJ 404 551

Peebles, P. J. E. & Vilenkin, A. 1999, Phys Rev D59 063505

Rees, M. J. & Ostriker, J. P. 1977, MNRAS 179 541

Riess, A. G., et al. 2001, ApJ 560 49

Roche, P.F., Aitken, D. K., Smith, C. H. & James, S. D. 1986, MNRAS 218 19P

Sanders, D. B., Egami, E., Lipari, S., Mirabel, I. F. & Soifer, B. T. 1995, AJ, 110, 1993

Sanders, D. B., Soifer, B. T., Elias, J. H., Madore, B. F., Matthews, K., Neugebauer, G. & Scoville, N. Z. 1988, ApJ, 325, 74

Sanders, D. B. 2000, AdSpR 25 2251

Smith, H. A, Larson, H. P. & Fink, U. 1981 ApJ 244 835

Soifer, B.T., Boehmer, L., Neugebaur, G. & Sanders, D. B. 1989, AJ, 98, 766

Spitzer, L. & Greenstein, J. L. 1951, ApJ 114 407

Spoon, H. W. W., Keane, J. V., Tielens, A. G. G. M., Lutz, D. & Moorwood, A. F. M. 2001, A&A 365 L353

Steinhardt, P. J, Wang, L. & Zlatev, I. 1999, Phys Rev D59 123504

Stockton, A. 2001, in Astrophysical Ages and Time Scales, ASP Conference Series, Von Hippel, Sipmson & Manset eds., 245, 517 (astro-ph/0104191)

Tielens, A. G. G. M. & Hollenbach, D. 1985, ApJ 291 747

Thronson, H. 1983, ApJ 264 599

Trentham, N. 2001, MNRAS 323 542

Voit, G. M. 1992, ApJ 399 495

Wetterich, C. 2001, astro-ph/0111166

Wood, O. D. E. & Churchwell, E. 1989, ApJS 69 831

Far-Infrared and Submillimeter Spectroscopy of
Low to Moderate Redshift Normal and Active Galaxies

Jacqueline Fischer

Naval Research Laboratory
Remote Sensing Division
4555 Overlook Ave SW, Washington DC 20375 USA,
Jackie.Fischer@nrl.navy.mil

Abstract. Broadband surveys with planned observatories such as SIRTF, ASTRO-F, NGST, Herschel, SMA, and ALMA will provide large samples of dust-enshrouded galaxies undergoing varying proportions of star formation and accretion onto supermassive black holes. Recent evidence suggests that the mass of supermassive central black holes are proportional to central velocity dispersions of their host galaxy bulges, but it is not yet known how these properties are evolutionarily connected. To trace the evolution of these components during the important luminous, dusty merger episodes as a function of redshift, sensitive mid- to far-infrared spectroscopic diagnostic observations from space will be necessary. The high sensitivity and spatial resolution of a large aperture, natural background limited far-infrared/submillimeter telescope such as SAFAIR will allow detection and identification of the far-infrared/submillimeter components of galaxies and of a multitude of rest-frame mid- to far-infrared recombination, fine-structure, and molecular lines toward dusty galactic nuclei. The line diagnostics beginning to be developed from ISO work will help determine important parameters such as the intrinsic spectral energy distribution, ionization parameters, metallicities, densities, extinction to the embedded energy sources, and ultimately the star-formation and accretion rates in these galaxies. Ultimately, the higher spatial resolution provided by a cold, far-infrared submillimeter interferometer is needed to view the distribution of structure in galaxies in order to unravel the as yet untold story of the evolution of galaxies over the lifetime of the Universe.

1. Science goals of Far-IR/Submillimeter Large Aperture and Interferometer Missions

Broadband imaging surveys in optical, near-infrared, and high energy bands will tell us how the number counts of different galaxy types vary as a function of photometric redshift and thus give us quantitative information on galaxy evolution. The missions of interest in this workshop must have the sensitivity and spatial resolution to uniquely detect and identify far-infrared and submillimeter galaxies (or protogalaxies) and the far-infrared and submillimeter emission components associated with the galaxies discovered in other bands. SAFAIR was conceived to be able to do this. Further, it will allow us to characterize the broadband distribution of the far-infrared and submillimeter emission

and to carry out sensitive spectroscopy to determine in which types of cosmological sources these emission components trace accretion onto supermassive black holes and when they trace powerful bursts of star formation. These observations will help determine the relationship of the evolution of black holes to the evolution of galaxies. For example, do supermassive black holes evolve throughout normal galaxy evolution or is their evolution mediated mainly through collisions and mergers of galaxies?

2. Current understanding and puzzles

Interpretation of cosmological broadband and spectroscopic signatures is dependent on the thorough understanding of local ones. ISO observations have brought a clearer understanding of some infrared signatures, while adding further puzzles in the cases of others. Here, I highlight some of the ISO results on galaxies and the questions that they have raised.

Moorwood et al. (1996) were able to use the plethora of mid-infrared lines from highly ionized species from the visually obscured, Compton-thick active nucleus in the nearby Circinus galaxy to determine the shape of the ionizing spectrum including a strong UV bump, while Clavel et al. (1999) have found evidence for large differences in obscuration by dust that explain the different mid-infrared signatures in Seyfert 1 and 2 galaxies, consistent with unification theories of AGN. Mid-infrared diagnostics have been used to quantitatively ascertain the starburst and AGN contributions to the infrared luminosity of ultraluminous infrared galaxies (Genzel et al. 1998; Laurent et al. 2000). Based on the ISO determined Milky Way galactic center extinction law, Genzel et al. find that the ratio of ionizing photons to bolometric luminosity in ultraluminous galaxies is similar to that of starburst galaxies. Lutz (1999) has found that the extinction laws in these galaxies may differ systematically from each other, so the sensitivity of future missions must allow us to *measure* the extinction law towards these galaxies. If confirmed, these differences may shed light on the varied conditions in the interstellar media of these galaxies and will need to be used to derive the important diagnostic ratio of ionizing flux to bolometric luminosity.

Luhman et al. (1998; 2002) and Malhotra et al. (1997; 2001) have found unexpected weakness of the [CII]158μm and [OI]63μm luminosities of ultraluminous and lower luminosity warm galaxies, while the ratio of the [CII] and PAH feature luminosities is constant (Helou et al. 2001; Luhman et al. 2002), except in the important case of low metallicity galaxies (Madden 2001). Helou et al. suggest that this may indicate that both high radiation density and increased importance of large grains in warm galaxies may explain these phenomena, while Luhman et al. (2002) posit that the presence of dust-bounded ionized regions and the resulting excess infrared emission may be contaminating the PDR emission diagnostics. These explanations need not be in conflict! Fischer et al. (2001; these proceedings) have modeled fine-structure line diagnostics of high ionization parameter regions illuminated by AGN and starbursts that may be responsible for producing dust-bounded ionized regions and the observed weakness of fine-structure lines from ionized regions. Lastly, atomic and molecular absorption in lines of OH, H_2O,

and [OI] can be used to constrain the intervening column densities towards and physical conditions in the far-infrared emission regions in these galaxies (for discussion on OH, see Smith, these proceedings).

The higher sensitivity and spatial resolution of observatories such as SIRTF, SOFIA, NGST, Herschel, SMA, and ALMA will propel these endeavors forward, paving the way for interpretation of these spectroscopic diagnostics in cosmological sources.

3. The need for natural-background limited sensitivity and high spatial resolution

Bock et al. (these proceedings) illustrate the dramatic improvement in sensitivity produced by cold, natural-background limited, space-borne spectrometers, while Bradford et al. (these proceedings) discuss a novel, lightweight, stackable spectrometer concept with no moving parts that can be used for redshift identification and other spectroscopic studies. A good illustration of the need for the high sensitivity that SAFAIR will provide, is the large scale spectroscopic survey suggested by Dudley et al. (these proceedings) to study the [CII]158μm deficiency break at $\sim 10^{12}$ $L_\odot$ in luminous infrared galaxies. Such a survey is capable of statistically tracing metallicity evolution with redshift, if it occurs in infrared luminous galaxies, with a novel alternative possibility of its use as a *statistical* standard candle if the metallicity in such objects is similar to that found in the local universe.

The need for high spatial resolution cannot be overemphasized. The high spatial resolution of SAFAIR will be necessary to correctly associate far-infrared and submillimeter cosmological sources with those in other spectral bands. Further in the future, direct measurement constraints on the physical sizes of objects such as obscured AGN (10 mas ~ 3.7 pc at the distance of the nearest ultraluminous infrared galaxy) can be conclusive proof of their nature. Associated interferometric spectroscopic signatures can then be used as diagnostics of these types of objects at cosmological distances as long as the spatial resolution is high enough to prevent swamping of these features by emission from other structures within the same galaxy. Thus the high spatial resolution provided by a cold, far-infrared/submillimeter interferometer equipped with spectroscopic capability will be necessary in order to produce a detailed account of the evolution of galaxies over the lifetime of the Universe.

References

Bock, J.J., Bradford, C.M., Earle, L., et al., these proceedings
Bradford, C.M., Naylor, B., Zmuidzinas, J., et al., these proceedings
Clavel, J., Schulz, B., Altieri, B., et al. 2000, A&A, 357, 839
Dudley, C.C., Fischer, J., Luhman, M.L. et al. these proceedings
Fischer, J., Allen, R., Dudley, C.C, et al. 2001, in *Spectroscopic Challenges of Photoionized Plasmas*, ASP Conf. Ser., v. 247, eds. G. Ferland & D. Wolf Savin, 499

Fischer, J., Allen, R., Dudley, C.C., et al., these proceedings

Genzel, R., Lutz, D., Sturm, E., et al. 1998, ApJ, 498, 579

Helou, G., Malhotra, S., Hollenbach, D. J., Dale, D. et al. 2001, ApJ, 548, 73

Laurent, O., Mirabel, I.F., Charmandaris, V., et al. 2000, A & A, 459, 887

Luhman, M.L., Satyapal, S., Fischer, J., et al. 1998, ApJ, 504, L11

Luhman, M.L., Satyapal, S., Fischer, J., et al. 2002, ApJ, submitted

Lutz, D. 1999, in *Universe as seen by ISO,* eds. P. Cox, M. Kessler, ESA-SP 427, p. 623

Madden, S.C. 2001, in *The Promise of the Herschel Space Observatory*, eds. G.L. Pilbratt, J. Cernicharo, A.M. Heras, T. Prusti, & R. Harris. ESA-SP 460, p. 159

Malhotra, S., Helou, G., Stacey, G., et al. 1997, ApJ, 491, L27

Malhotra, S., Kaufman, M. J., Hollenbach, D., et al. 2001, ApJ, 561, 766

Moorwood, A.F.M., Lutz, D., Oliva,E. et al. 1996, A&A, 315, L109

Smith, H.A., these proceedings

SAFIR AND INTERSTELLAR MEDIUM STUDIES:

Gary Melnick (SAO)

The questions of greatest scientific interest addressable at far-infrared and submillimeter wavelengths at the time SAFIR would fly will have been shaped by the results from the Space Infrared Telescope Facility (SIRTF) and the Herschel Space Observatory (HSO) as well as ongoing studies using the Stratospheric Observatory for Infrared Astronomy (SOFIA) and the Atacama Large Millimeter Array (ALMA). In light of these other efforts, SAFIR must offer unique and compelling capabilities. For investigations of the interstellar medium, it is important that SAFIR include the following features:

Wavelength coverage between about 50 and 540 μm:

The choice of the short- and long-wavelength cutoffs is driven by the desire to study species key to understanding the oxygen chemistry as well as to determine the importance of chemical fractionation effects. Because oxygen is the most abundant heavy element in the Universe, the chemical evolution involving oxygen controls the abundance of many other important species, including those involving carbon, nitrogen, sulfur, and silicon. Both atomic oxygen and ortho-water are indispensable diagnostics of the oxygen chemistry and their ground-state lines at 63 and 538 μm, respectively, should be encompassed by SAFIR's wavelength coverage. At the same time, the abundance of deuterium can provide important and direct insights into the effects of chemical fractionation. For these reasons, it is important that SAFIR also possess sufficient spectral coverage, on the short wavelength end, to include the two lowest rotational lines of HD at 112 and 56 μm.

Since SAFIR will offer higher spatial resolution than any other observatory at wavelengths between 30 and 350 μm, one of its clear strengths will be the study of gas that displays interesting structure on small scales. Examples include: (1) hot (T ≥ 300K) post-shocked gas in which a large number of rapid endothermic reactions are favored thus creating a chemical composition that is distinct from the surrounding cold gas; and, (2) gas within proto-planetary disks.

Shocks

Collisions involving gas with relative velocities as small as 10 km/s are sufficient to generate shocks behind which temperatures exceed 300 K. As shown in Figure 1, spectral coverage between 50 μm (6000 GHz) and 540 μm (~550 GHz) would include a large number of transitions within important oxygen and carbon-bearing species.

Proto-Planetary Disks

Temperature and density gradients across a proto-planetary disk are predicted to lead to significant chemical variations as a function of distance from the central star. Since this gas will be incorporated into planets as they form, understanding the composition of this gas is necessary to an understanding of the composition of these planets. The ability of SAFIR to
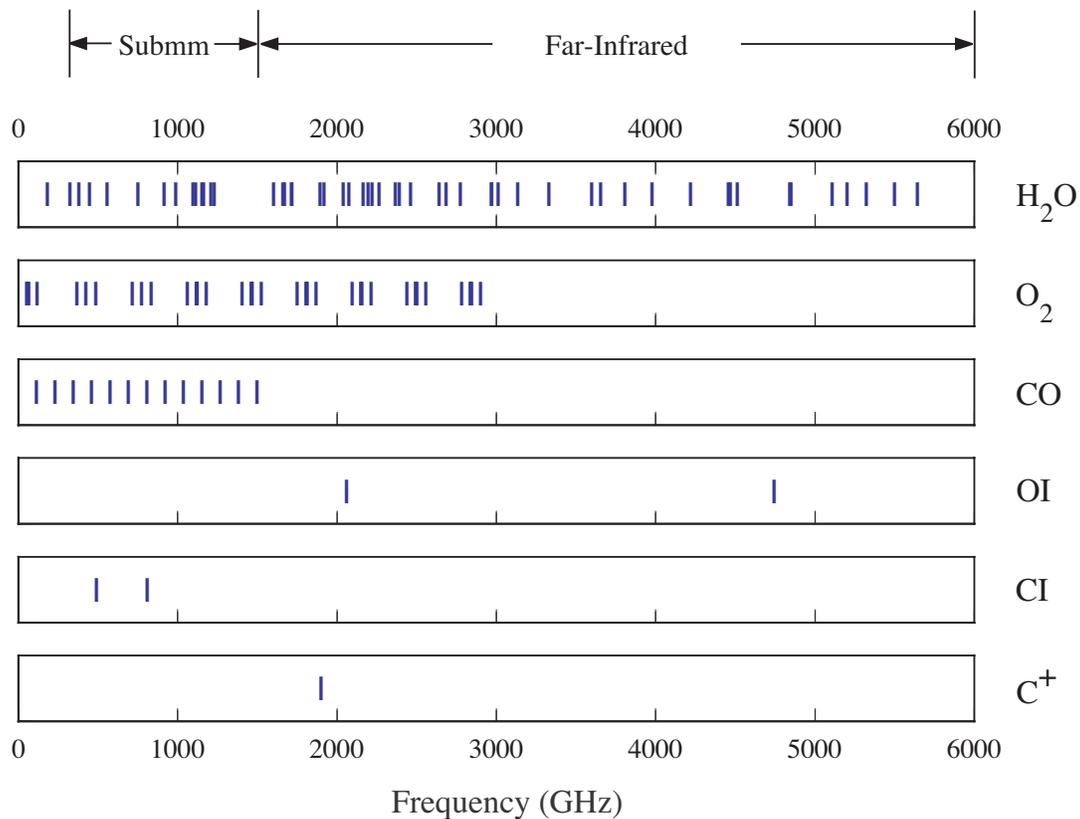
Figure 1. Line frequencies of the lowest-lying (T<500 K) transitions among the most important oxygen-and carbon-bearing species.

sample a large number of water transitions will help in modeling the water emission and can, in part, compensate for a lack of even higher spatial resolution.

Spectral Resolution of 1 km/s (or Better) at All Wavelengths:

Quite often, a single line-of-sight through a molecular cloud contains several distinct components: cold intervening gas, a warm cloud surface, a cold cloud interior characterized by internal motions of a few km/s and, sometimes, embedded outflows that exhibit p-Cygni profiles and line wings of a few tens of km/s. Only observations using high spectral resolution can separate these components and allow the observer to retrieve useful diagnostic information. As shown in Figure 2, a comparison of the SWAS water spectrum of the galactic star forming region NGC 6334 obtained at the 1 km/s resolution of SWAS and a co-average of spectral elements meant to simulate the same spectrum at a resolution of 30 km/s demonstrates that considerable information is lost when the velocity resolution is much less than the intrinsic widths of spectral features. Similarly, the high spectral resolution of SWAS is used to isolate cold absorbing gas in galactic spiral arms along the line-of-sight to W49, as shown in Figure 3.

The HIFI instrument aboard HSO will provide ≤ 1 km/s spectral resolution between 150 and 600 μm; however, between 57 and 150 μm, the highest spectral resolution available on HSO corresponds to about 300 km/s. To build upon the results of HSO as well as to allow useful
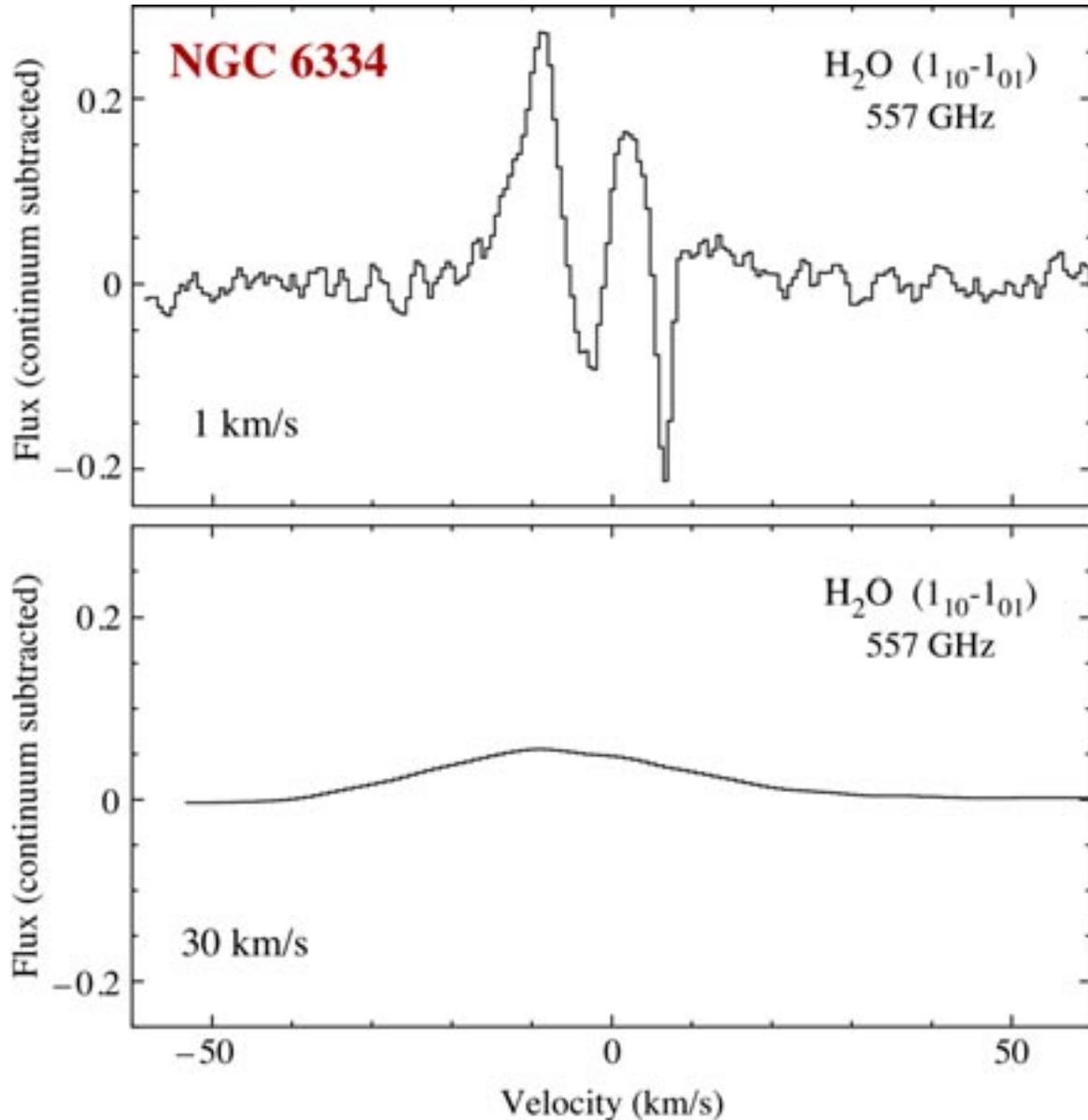
Figure 2. SWAS ortho-H_2O 557 GHz spectrum of the galactic star forming region NGC 6334 (*top*). This spectrum illustrates the important details, such as self-absorption, that would be lost at a spectral resolution of 30 km/s (*bottom*).

comparisons with high velocity resolution measurements that will be routinely made using ALMA and SOFIA, it is important that SAFIR offer similarly high spectral resolution.

Spatial Resolution at Least 2x Better than HSO:

Many galactic regions are expected to show marked changes in their physical conditions and chemical composition over small spatial scales. As mentioned above, shocks and proto-planetary disks are but two examples. Unfortunately, the ability of far-infrared telescopes to resolve structure on small scales has been extremely limited; because far-infrared
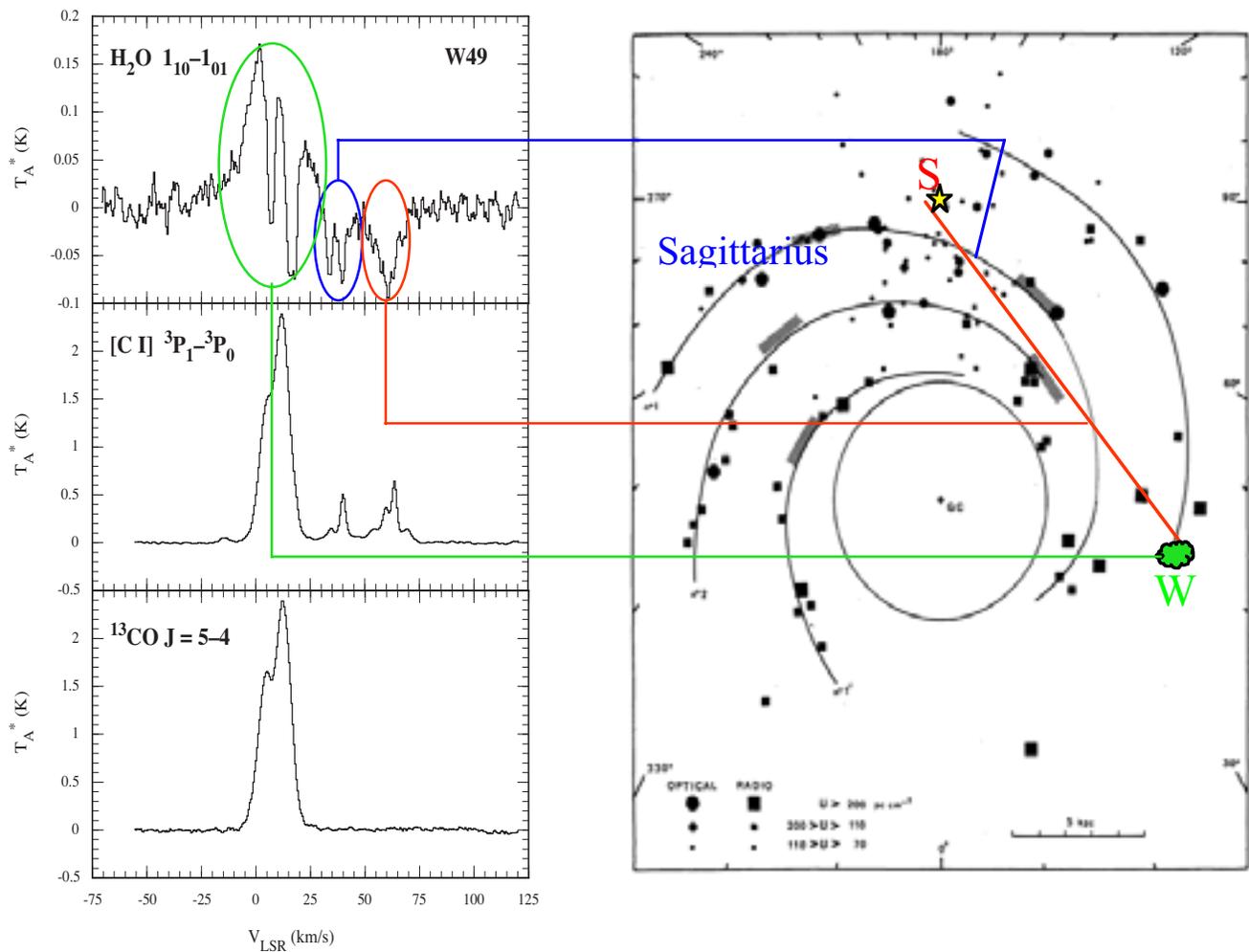
Figure 3. SWAS spectra of H_2O (top left), CI (middle left), and ^{13}CO J=5-4 (bottom left) toward the galactic star forming complex W49. The 1 km/s spectral resolution of SWAS allows the 35 and 60 km/s absorption features associated with two line-of-sight passages through the Sagittarius spiral arm to be easily separated from the emission intrinsic to W49 at lower velocities.

telescopes must operate above most of the atmosphere, by necessity, these telescopes have been substantially smaller than their ground-based counterparts operating at both shorter and longer wavelengths. This has resulted in a conspicuous drop in the spatial resolution of far-infrared maps when compared to both near-infrared and submillimeter and millimeter wavelength maps. HSO and SOFIA will improve upon the spatial resolution obtained by previous facilities, such as IRAS, ISO, and the Kuiper Airborne Observatory, but even these newer facilities will not be able to produce maps at the sub-arcsecond spatial resolutions offered by the new generation of large optical/infrared telescopes and by ALMA. (ALMA will process sufficient spatial resolution (0.02 arcsec x λ_{mm}) to measure compositional variations across many proto-planetary disks; however, ALMA will not have access to many important species, such as H_2O.) Though it is highly unlikely that a single aperture far-

infrared space telescope sufficiently large to produce sub-arcsecond images can be flown –
this capability will have to await the development of space interferometric techniques –
SAFIR must strive to improve upon its largest predecessor, HSO, by at least a factor of two.
Figures 4 and 5 illustrate how the diffraction-limited beamsize produced by
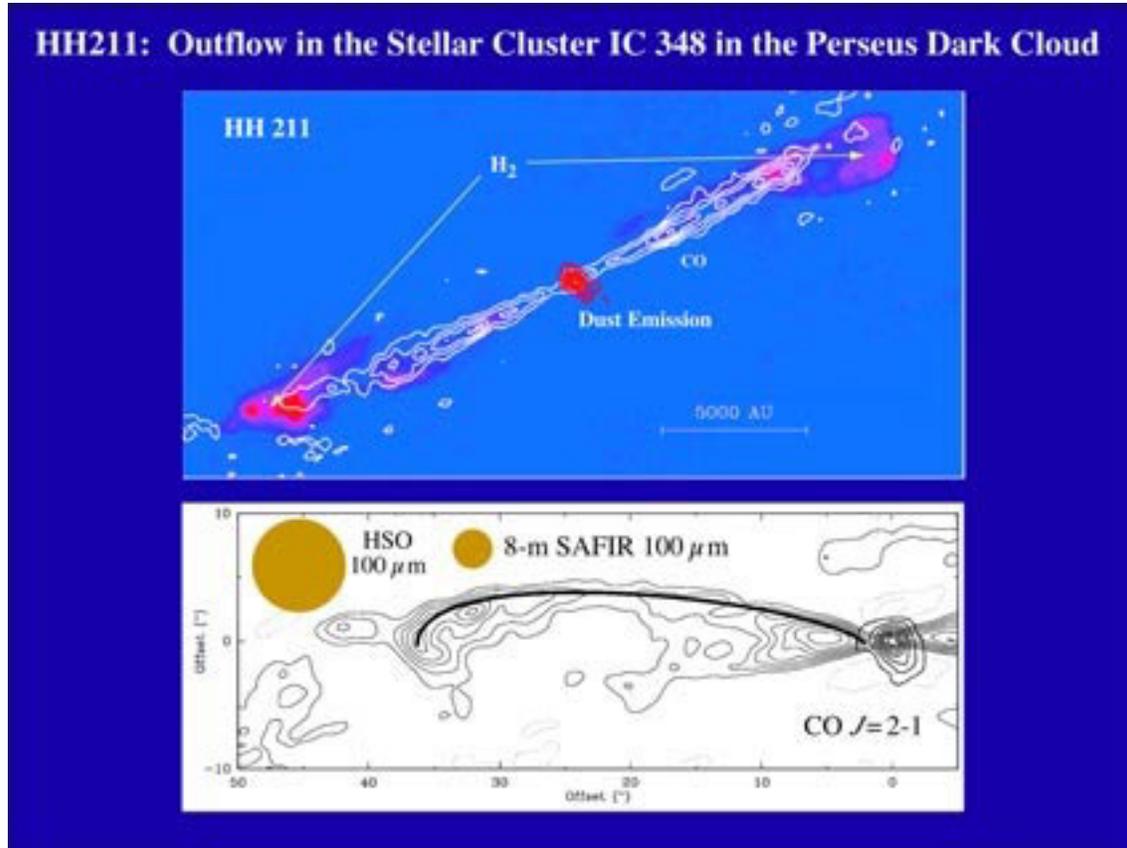
Figure 4. The Herbig-Haro jet HH211, as observed in the H_2 v=1-0 S(1) line by
McCaughrean et al. 1994, *ApJ*, 436, L189 (*top*), and CO *J*=2-1 by Gueth and Guilloteau
1999, *A&A*, 343, 571 (*bottom*). Also indicated are the projected beamsizes of HSO and
an 8-meter SAFIR at 100 μm. The smaller beamsize would allow SAFIR to better
resolve and study the sharp compositional changes predicted to occur across and along
the shock-excited surface of the cavity created by the outflow. Such an understanding
is important since many of the molecules formed behind shocks are injected into the
surrounding quiescent material.

an 8-meter diameter SAFIR compares with HSO at 100 μm for the study of shocked
outflows and proto-planetary disks.

Scientific Emphasis on Lines Accessible Only to a Space Telescope:

For SAFIR to promise truly unique scientific results (i.e., results unobtainable by any other

Beta Pictoris   Fomalhaut
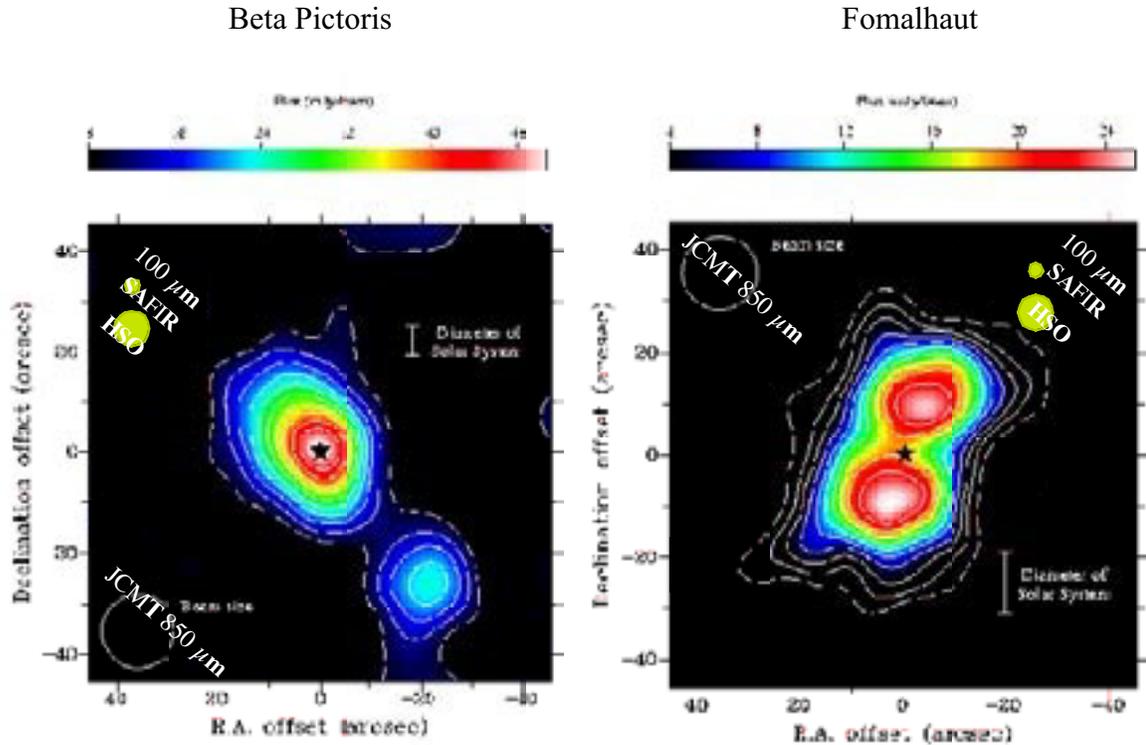
Figure 5. Comparison of the spatial resolutions of the JCMT at 850 μm, HSO at 100 μm, and an 8-meter diameter SAFIR at 100 μm relative to the size of two well-known disk sources.

facility) it should highlight its ability to study important atomic, ionic, and molecular species possessing detectable transitions only from space. One such example of a molecule observable only from space is $H_2^{16}O$; the atmosphere remains sufficiently opaque at even airplane altitudes (e.g., 14 km) that studies of galactic sources of water emission are impossible. The scientific case for SAFIR must emphasize such areas of uniqueness.

SESSION III

MISSION CONCEPTS

Charting the Winds that Change the Universe, II
The Single Aperture Far Infrared Observatory (*SAFIR*)

G. H. Rieke, D. J. Benford, P. M. Harvey, C. R. Lawrence, D. T. Leisawitz, D. F. Lester,
J. C. Mather, G. J. Stacey, M. W. Werner, and H. W. Yorke

Abstract: "*SAFIR*** will study the birth and evolution of stars and planetary systems so young that they are invisible to optical and near-infrared telescopes such as NGST. Not only does the far-infrared radiation penetrate the obscuring dust clouds that surround these systems, but the protoplanetary disks also emit much of their radiation in the far infrared. Furthermore, the dust reprocesses much of the optical emission from the newly forming stars into this wavelength band. Similarly, the obscured central regions of galaxies, which harbor massive black holes and huge bursts of star formation, can be seen and analyzed in the far infrared. *SAFIR* will have the sensitivity to see the first dusty galaxies in the universe. For studies of both star-forming regions in our galaxy and dusty galaxies at high redshifts, *SAFIR* will be essential in tying together information that NGST will obtain on these systems at shorter wavelengths and that ALMA will obtain at longer wavelengths." – page 110, Astronomy and Astrophysics in the New Millenium, National Research Council - AANM, National Academy Press, 2001.

1. The 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 until 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.

Regardless of the original emission process, cosmic energy sources glow in the far infrared and submm. The continuum emission is due to the incredible efficiency of interstellar dust in absorbing visible and ultraviolet photons and reemitting their energy. The appearance of the early Universe, of AGNs and starbursting galaxies, and of star forming regions is transformed through suppression of the visible and ultraviolet and augmentation of the far infrared and submm. Low-lying far infrared fine structure lines are the major coolants for interstellar gas. Molecular transitions in this spectral range carry the signature of conditions in warm and dense interstellar clouds where stars form. Thus, we must look in the far infrared and submm for clues to the underlying processes shaping the origin, structure, and evolution of our Universe.

Accessible advances in technology can produce huge advances in our capabilities for far infrared and submm astronomy. Consequently, the consensus-based Astronomy and Astrophysics Survey Committee recommended *SAFIR*, an 8-m class, space-borne far infrared telescope (operating from $\sim 20\mu m$ to $\sim 1mm$), as a high priority to be started in this decade (page 10, AANM). The scientific importance of this mission spans both the Origins and Structure and Evolution of the Universe science themes within OSS. The

Origins roadmap with the same goals and community endorsement as *SAFIR*. As an observatory-class instrument, *SAFIR* (and FAIR) would serve a broad range of science needs.

With a start near the end of the decade, *SAFIR* follows several missions that will lay its scientific and technological foundation. SIRTF, whose mission will be complete by the time *SAFIR* would go to Phase B definition, will leave a Great Observatory legacy of mid and far infrared observations along with the technical demonstration to make a large cold telescope a compelling next generation facility. SOFIA will have operated for a number of years, providing both a testbed for detector technology and a high level of flexibility and accessibility in its explorations of the far infrared sky. Herschel will be completing its examination of the submm range, both for continuum sources and with powerful spectroscopic instrumentation. However, none of these missions will approach the potential for this spectral range, SIRTF because of its small aperture (85cm) and SOFIA and Herschel because of the thermal background due to their relatively high operating temperatures (~ 220K and ~80K respectively). Yet *SAFIR* is well in reach technically, given the modest requirements on its optics and pointing, the infrastructure under development within NASA for large telescopes and large satellites in general, and the potential for dramatic advances in detector technology at modest cost. *SAFIR* will lay the foundation for the next steps in exploring this spectral region, such as the SPECS space-borne interferometer (see "Probing the Missing Universe" by D. Leisawitz et al.).

NGST and ALMA will begin operation toward the end of the decade. The timely ability of *SAFIR* to bridge the spectral gap between them will complement well the capabilities of both facilities, maximizing the productivity of two other major investments in astronomical capability.

2. *SAFIR* and the Formation and Evolution of AGNs

It appears that central supermassive black holes are a universal component of galactic bulges. Do the central black holes form first and serve as condensates for galaxies? Or do they build up as galaxies grow and merge? The low lying H_2 lines at 17 and 28.2μm are one of the few conceivable ways to study warm molecular gas condensations prior to the formation of metals, for example molecular gas around primordial massive black holes. A number of processes, such as formation of a small number of stars, can heat molecular clouds above the ~100K threshold for high visibility of these lines. The lines are undetectable together from the ground until $z > 35$ (both must be detected to confirm the identification). *SAFIR* will be well suited to searching for them. Line widths and profiles would indicate whether the central mass is highly compact (suggesting a black hole), or if the molecular cloud is just in a mild state of turbulence (as expected if it is self gravitating without a central black hole).

At the current epoch, galaxy mergers produce huge far infrared fluxes through a combination of violent starbursts and of AGNs associated with their cental black holes. "Distinguishing starbursts from supermassive black holes is complicated by the fact that AGNs are often shrouded in dust, so that much of the direct emission is hidden from view. Long wavelengths penetrate the dust more readily, so .. *SAFIR* and NGST with an extension into the thermal infrared are .. suitable for separating the two phenomena" (page 85, AANM).

What happens during the much more common mergers that build galaxies in the early Universe? COBE showed that the far infrared/submm energy density in the early

Universe is comparable to that in the visible/near infrared. What are the relative roles of dust embedded AGNs and starbursts in producing this luminosity? Do AGNs at high redshift differ in basic properties from nearby ones? Models of the cosmic X-ray background indicate that the great majority of AGNs at high redshift are heavily absorbed (Gilli et al.; Comastri et al.). Thus, these answers must be sought in the far infrared where optical depths are low (ISM optical depths are similar at 20μm and 6keV and rapidly decrease at longer infrared wavelengths and higher X-ray energies).

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. Figure 1, based on work by Voit and Spinoglio and Malkan, is a demonstration. 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 8-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, allowing *SAFIR* to constrain the UV spectra of AGNs and extending work with the Infrared Space Observatory (ISO) on a few nearby Seyfert galaxies to large lookback times. 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.
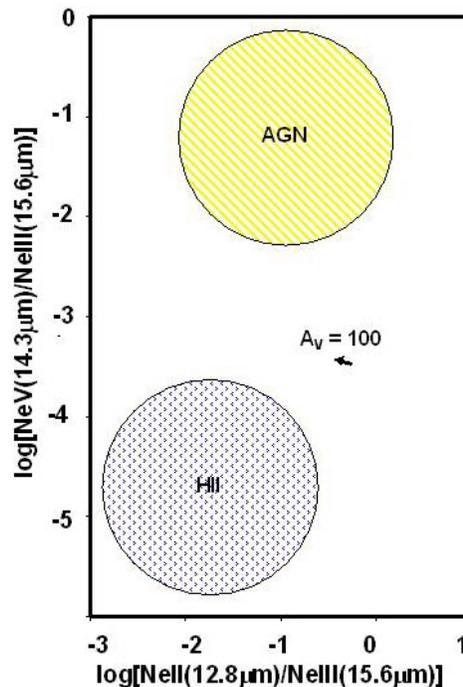
Figure 1. Ne V is not excited appreciably by hot stellar spectra, but is produced by the hard UV spectra of AGNs. Hence, the line ratios plotted distinguish the two types of ionizing source unambiguously. The plotted ratios are virtually extinction independent; the short arrow shows the effect of 100 magnitudes of visible extinction on the line ratios. Even this level of extinction has little effect compared with the separation of

The angular resolution of *SAFIR* is a critical contribution to these studies. Figure 2 shows the Hubble Deep Field and the X-ray sources discovered in a deep Chandra exposure. A portion of the HDF is degraded to 1" resolution, the beam diameter of a 8-m telescope operating at 40μm. Individual galaxies are well isolated for detailed study.

3. *SAFIR* and the Emergence of Stars and Galaxies

The history of star formation determines the evolution of galaxies and the generation rate for heavy elements. It has been traced by deep Hubble Space Telescope (HST) imaging followed up with large groundbased telescopes. However, even at modest

Figure 2. Deep Chandra detections of X-ray sources in the HDF. In the degraded portion, the resolution has been reduced to the diffraction limit of a 8m telescope operating at 40μm. The individual galaxies, including the X-ray identifications, are readily isolated at this
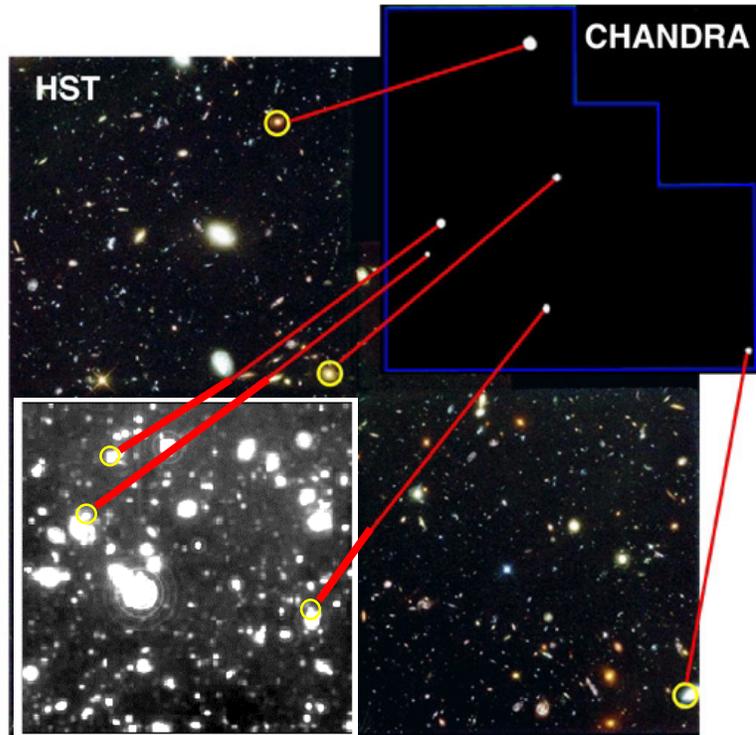
redshifts, these techniques only probe the rest frame ultraviolet. Interstellar dust can absorb nearly all the UV in star forming galaxies. In the best-studied starburst galaxies such as M82, a debate raged for more than a decade regarding how to correct even the near infrared emission for interstellar 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 can only 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 large cosmic far infrared/submm energy density discovered by COBE. This background has been partially resolved by ISO in the very far infrared and is thought to arise from starburst galaxies at z up to 3. A 8-m telescope would 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. A full understanding of star formation in the early Universe requires that we extend far FIR/submm measurements 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. Combining *SAFIR* and ALMA measurements of SEDs can give photometric redshift estimates, and *SAFIR* spectroscopy can measure redshifts using the strong PAH features in galaxy spectra near 8μm. The sensitivity of *SAFIR* will allow us to measure ordinary galaxy luminosities to below L_*, even out to z ~ 5. Equally important is the resolution afforded by an 8-m aperture; previous space far infrared missions like SIRTF will be limited in this science objective by confusion noise.

4. *SAFIR* and the 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? Once even traces of metals have formed, the C^+ line at 158μm 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. N^+ lines at 122 and 205μm also play important roles in cloud cooling. Study of the molecular hydrogen and these emission lines in the early Universe and as a function of redshift promises to reveal many of the processes occurring in the gas clouds that build early galaxies. Space-borne observations in the FIR/submm must be a major component of this study.

The far infrared fine structure lines also control the cooling of molecular clouds in the Milky Way. Understanding this process and related ones revealed by far infrared spectroscopy is a key to advancing our knowledge of how these clouds begin their collapse into stars and planets (see below).

5. *SAFIR* and the Birth 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 birth of the star 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. Imaging with the resolution provided by a far infrared 8-m telescope ($\leq$ 100 AU at 40μm for the nearest star forming regions) can probe the density, dynamics, and temperature structure of these ~1000 AU collapsing cores on critical physical scales. 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.

As it collapses, 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, can probe the physical and dynamical conditions in the collapse. The spectrum predicted for a collapsing cloud core is shown in Figure 3. 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 birth thoroughly.
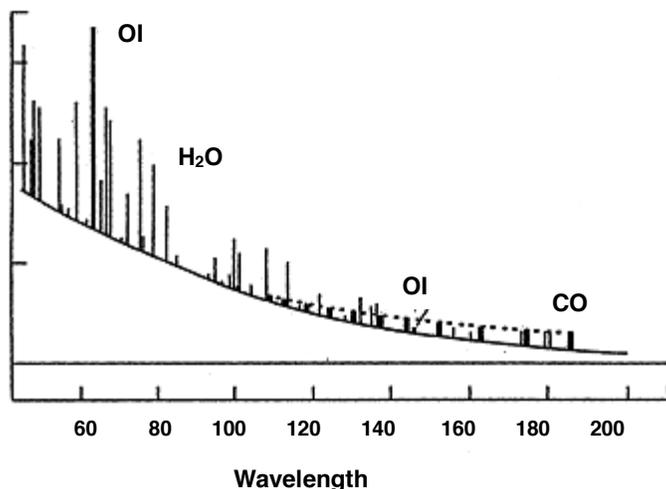
Figure 3. Predicted far infrared spectrum of a collapsing cold cloud core, from Ceccarelli, Hollenbach, and Tielens. The spectrum is dominated by [OI] and complexes of CO and H O.

6. *SAFIR* and the Evolution of Planetary Systems and the Origin of Life

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 in the Kuiper Belt outside the orbit of Neptune gives access to objects where accretion proceeded slowly and its products should be primitive and still reflect conditions in the early solar nebula.

There is a large population of Kuiper Belt Objects (KBOs), including objects of size rivaling the largest asteroids. They have a broad variety of surface characteristics. To interpret the clues they provide for evolution 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 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.

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. Many more will be discovered by SIRTF. The Kuiper Belt is 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 (e.g., Beust et al.). 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 have been produced in

Figure 4. Planetary debris disks of differing ages. All systems are illustrated at the same scale on the sky. The ISO spectrum shows minerological spectral features expected in such systems, and the beam sizes for NGST (for the PAH feature) and SAFIR (for Fosterite and water ice) are shown projected onto the β Pic disk.
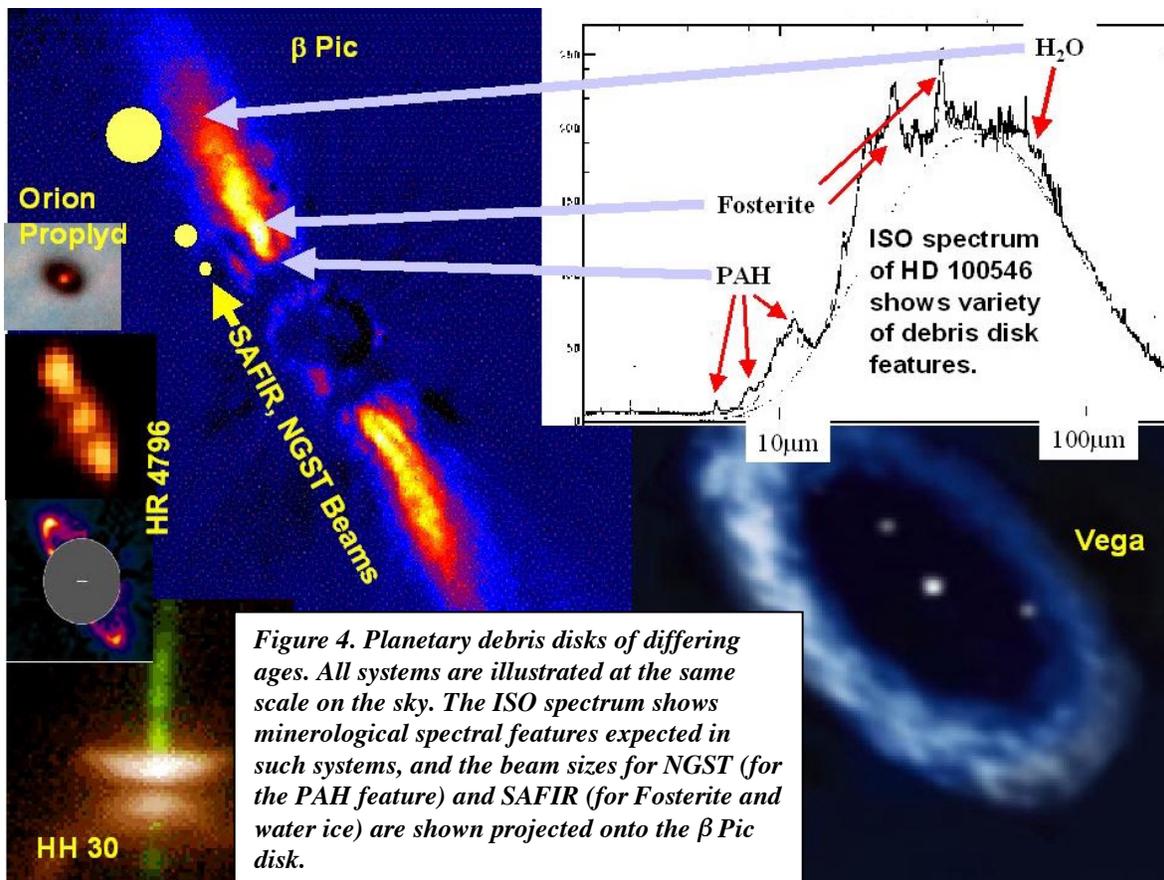
recent collisions between planetesimals. Thus, this system and others like it demonstrate the potential of examining the early, violent evolution of debris disks and the infall of comets.

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. Figure 4 illustrates the potential advances with **SAFIR**. 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.

7. *SAFIR* and the Discovery of New Phenomena

Technological advances enable astronomical discoveries. Harwit tried to quantify this relation in "Cosmic Discovery." In the 25 years preceding publication of the book, important discoveries were made within 5 years of the development of new technology making them possible. The exceptional discovery potential in the FIR/Submm region arises because the sensors are still substantially short of fundamental performance limits and the telescopes available to date have been very modest in aperture (less than 1 meter!).

The previous decadal survey developed a parameter to describe the discovery potential of new missions, which they called astronomical capability. This parameter is proportional to the time required to obtain a given number of image elements to a given sensitivity limit. **SAFIR** will have astronomical capability exceeding that of past far

infrared facilities by a factor of about 10^{10}, and will still offer a gain of about 10^5 after SIRTF and Herschel have flown. A gain of 10^5 is similar to the gain from the initial use of the Hooker 100-Inch Telescope on Mt. Wilson to the Hubble Space Telescope.

8. Mission Development

8.1 Telescope

With the imminent selection of the NGST prime contractor, it is timely to begin mission concept studies for *SAFIR*. There are two general possibilities, as indicated in Figure 5. The development of the NGST telescope may result in approaches that can be readily adapted to the far infrared, with the differing requirements of (1) colder operating temperature; (2) relaxed image quality; and (3) larger aperture (now that NGST has decreased in size to 6m). However, these three important differences may lead to unique architectures for the far infrared telescope. This basic decision must be made as soon as possible to guide further development of the mission. Although autonomous deployment is a possibility, opportunities for reduced cost and risk through in-space assembly should also be explored. Finally, there is a high level of interest in the far infrared and submm both in Europe and Japan, making it timely to consider possible international collaborations. Such cooperation has already been fruitful in the infrared and submm for Herschel, Planck, and ALMA.
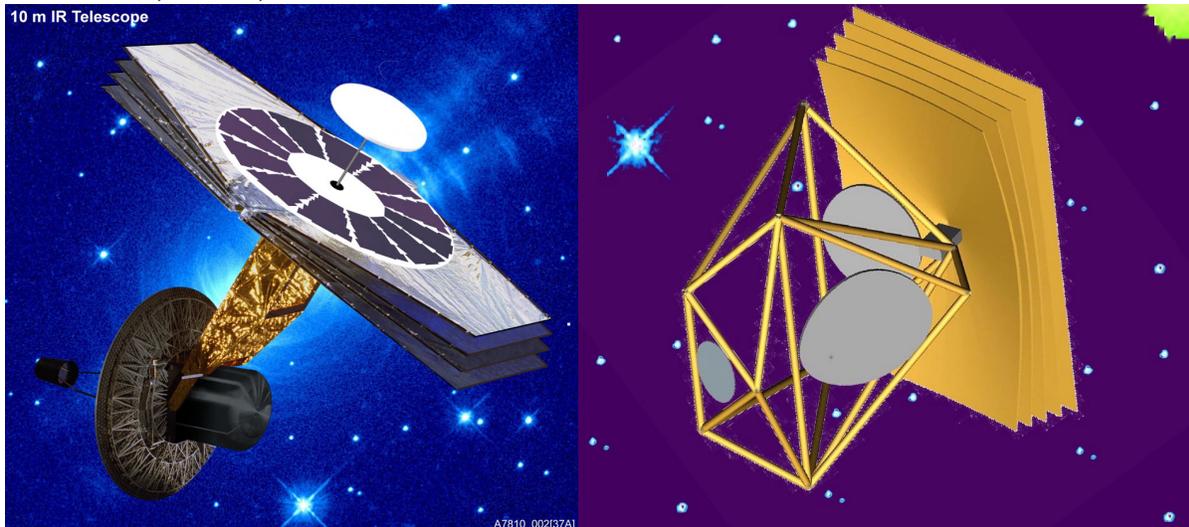
Figure 5. Two possible development paths for SAFIR. The figure to the left illustrates the potential for a telescope based on NGST developments, in this case placed at about 4AU to obtain greater radiative cooling (courtesy Ball Aerospace). The figure on the right illustrates that focused developments for the far infrared may also be promising. In this case, the telescope uses a stretched membrane approach that may offer a lower construction cost than NGST-based telescopes (courtesy M. Dragovan).

8.2 Detector Technology

The far infrared and submillimeter ranges have benefited relatively little from investments in detector technology by non-astronomical pursuits. In this regard, they differ dramatically from the visible, near and mid-infrared, and radio regions. Detectors in those spectral regions closely approach theoretical performance limits. For example, in the visible, CCDs have quantum efficiencies greater than 90%, read noises of about two electrons, and formats including many millions of pixels. In the far infrared, the much smaller prior investment has left the possibility for orders of magnitude further progress

toward fundamental limits. NASA missions are the best customers for this technology, and an augmented NASA investment will return substantial benefits to *SAFIR* and other far infrared and submillimeter missions such as Explorers. These investments will guarantee our nation's leadership in this important technology.

Figure 6 illustrates the three major detector technologies. Each has current strengths and weaknesses. Far infrared photoconductors are the most advanced in array construction, as shown by the space qualified SIRTF array in the figure, and require relatively modest cooling. However, they fall somewhat short of theoretical limits in potential performance and respond only up to the excitation energy. Development should address larger arrays, at least 128x128. Bolometers have broad spectral response and are the most advanced submm continuum detectors. They require extremely low operating temperatures. Development needs to emphasize improved array technology, such as SQUID-based multiplexing, and superconducting-thermometer bolometers that interface well to SQUID electronics. Hot electron bolometer mixers provide the best heterodyne operation above the superconducting gap frequency of NbTiN, around 1200 GHz. They can have large advantages for spectroscopy over photoconductors and bolometers. Development needs to address reducing noise temperatures and developing support electronics to allow large scale spatial arrays.
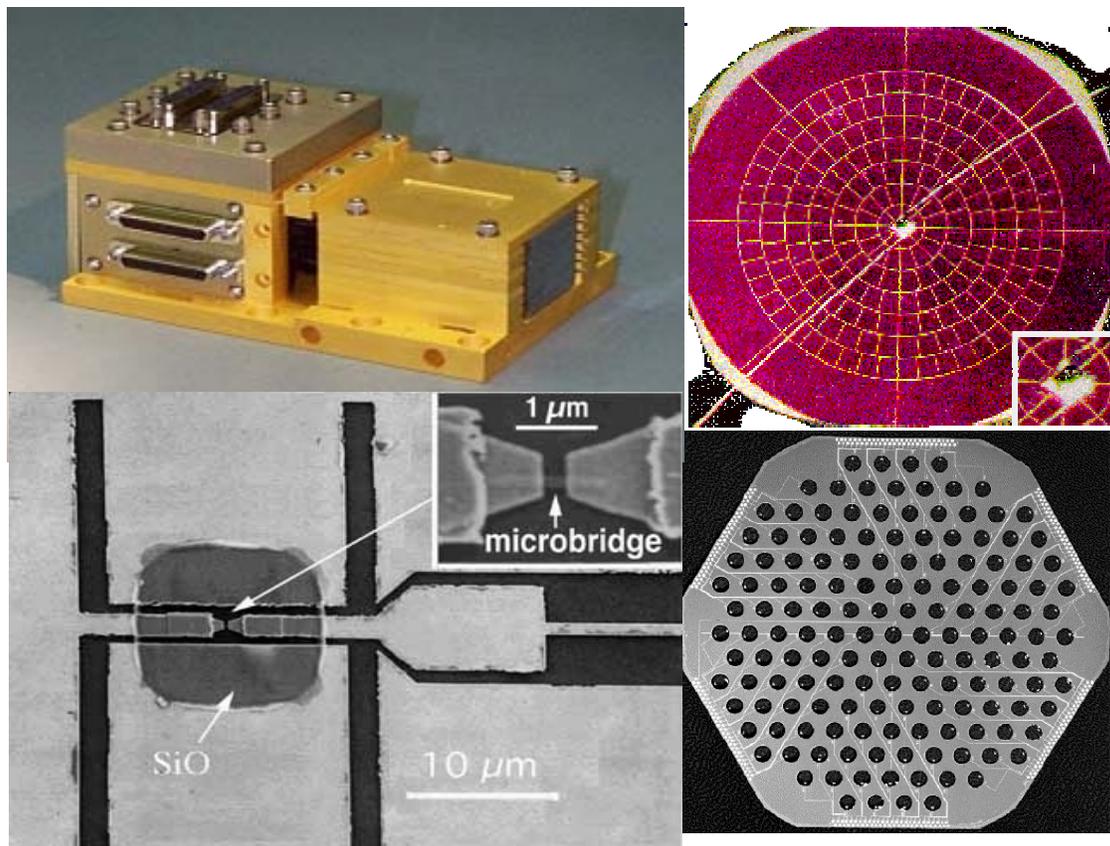
Figure 6. Far Infrared and Submillimeter Detector Approaches. Clockwise from upper left: (1) the SIRTF 32x32 Ge:Ga far infrared photoconductor array; (2) a spiderweb bolometer element; (3) an array of spiderweb bolometers; and (4) a hot electron bolometer mixer.

8.3 Budget

Goddard Space Flight Center carried out an estimate of the budget for *SAFIR* for the UVOIR panel of the decadal survey. They drew on their experience estimating the cost of NGST, so the comparison of the two missions is also pertinent. Their results are in Table 1. They assumed that no additional development would be required beyond that for NGST, although the report indicated that this was probably not entirely correct. We allow here for a significant development program, perhaps even departing significantly from the NGST telescope architecture. In the spirit of the above estimate, we take this program to be half that for NGST, or an additional $125M, for a total cost of $620M. For comparison, the estimate of the UVOIR panel for NGST is $1114M.

Construction (assumes an ESA instrument)	$310M
Launch (using a new mid-sized EELV)	$85M
Science and Mission Operations (5 yrs)	$100M
Total	**$495M**

Table 1. Cost estimate for SAFIR from the decadal UVOIR panel report.

Augmentation to budget for far infrared detector arrays	$10M
Very low temperature refrigerators for space environments	$50M
Large, lightweight optics	$80M

Table 2. Technology investments relevant to SAFIR and other far infrared/submillimeter projects.

The decadal survey committee also recommended a budget over the decade for the technology development that would support *SAFIR* and other projects in the far infrared and submillimeter, as shown in Table 2.

9. Summary

"*SAFIR*…will study the relatively unexplored region of the spectrum between 30 and 300μm. It will investigate the earliest stage of star formation and galaxy formation by revealing regions too shrouded by dust to be studied by NGST, and too warm to be studied effectively with ALMA…. It will be more than 100 times as sensitive as SIRTF or the European [Herschel] mission….To take the next step in exploring this important part of the spectrum, the committee recommends *SAFIR*. The combination of its size, low temperature, and detector capability makes its astronomical capability about 100,000 times that of other missions and gives it tremendous potential to uncover new phenomena in the universe." – pages 39, 110 Astronomy and Astrophysics in the New Millenium, National Research Council, National Academy Press, 2001.

SAFIR can contribute substantially to both the Structure and Evolution of the Universe and the Origins themes of NASA space science, through realizable technology developments of a moderate scale. With clear science priorities and exciting science goals that are intellectually accessible to the greater public, SAFIR is a mission that would help engage the nation in shared exploration. It will also showcase new technology in dramatic images capturing events at the dawn of time and at the birth of stars and planets.

Probing the Invisible Universe: The Case for Far-IR / Submillimeter Interferometry

D. Leisawitz, T. Armstrong, D.J. Benford, A. Blain, K. Borne W. Danchi, N. Evans, J. Gardner, D. Gezari, M. Harwit, A. Kashlinsky, W. Langer, C. Lawrence, P. Lawson, D. Lester, J. Mather, S. H. Moseley, L. Mundy, G. Rieke, S. Rinehart, M. Shao, R. Silverberg, D. Spergel, J. Staguhn, M. Swain, W. Traub, S. Unwin, E. Wright, and H. Yorke

Abstract

The question "How did we get here and what will the future bring?" captures the human imagination and the attention of the National Academy of Science's Astronomy and Astrophysics Survey Commitee (AASC). Fulfillment of this "fundamental goal" requires astronomers to have sensitive, high angular and spectral resolution observations in the far-infrared/submillimeter (far-IR/sub-mm) spectral region. With half the luminosity of the universe and vital information about galaxy, star and planet formation, observations in this spectral region require capabilities similar to those currently available or planned at shorter wavelengths. In this paper we summarize the scientific motivation, some mission concepts and technology requirements for far-IR/sub-mm space interferometers that can be developed in the 2010-2020 timeframe.

1.   Science goals

The Decade Report posed a number of "theory challenges," two of the most compelling of which are that astrophysicists should strive to: (a) *develop an "integrated theory of the formation and evolution of [cosmic] structure";* and (b) *"develop models of star and planet formation, concentrating on the long-term dynamical co-evolution of disks, infalling interstellar material, and outflowing winds and jets."* (Decade Report, p. 106)

Rieke et al. (2002; hereafter the "SAFIR white paper") explain the vital role that will be played by future far-IR/sub-mm observations in confronting these challenges and the need for a 10-m class Single Aperture Far Infrared Observatory (SAFIR). SAFIR will represent a factor 10^5 gain in astronomical capability relative to the next-generation missions SIRTF and Herschel, yet it will have the visual acuity of Galileo's telescope. An additional hundred-fold increase in angular resolution can be achieved with interferometry after SAFIR and within the NASA Roadmap time horizon. In this section we explore the science potential of sub-arcsecond resolution in the far-IR/sub-mm, picking up where the SAFIR white paper leaves off. In particular, we don't bother to explain why the far-IR spectrum (line and continuum radiation) is rich in information content, as doing so would only restate facts already eloquently presented in the SAFIR white paper.

1.1 The heritage and destiny of cosmic structure

After we locate in space and time the first generations of stars, galactic bulges, galactic disks, and galaxy clusters, we will want to relate these early structures to the "seeds" of structure seen in the cosmic microwave background fluctuations and learn how they formed. We will need measurements that show us how the cosmic structures changed over time to the present day. We will want to lift the veil of dust that conceals galactic nuclei, including our own, from view at visible wavelengths. How did the Milky Way form, and why is there a black hole at its center? What happens to the interstellar medium when galaxies collide, and how does a starburst work?

Did bulges form first and disks form later, or did disks merge to form bulges? What accounts for the diversity of galaxy types? How might the universe and its constituents look when it is twice or ten times its current age?

It will take a telescope much bigger than 10 m to see structure in galaxies at redshift z ~ 1 or greater in the far-IR/sub-mm. These objects subtend angles of ~1 arcsec. SAFIR will measure far-IR spectra of huge numbers of high-z galaxies, and they will be analyzed statistically and with the aid of models and complementary NGST and ALMA observations. However, to study the astrophysics of distant galaxies it will be important to resolve them in the far-IR/sub-mm, where they emit half or more of their light (Trentham et al. 1999). As noted by Adelberger & Steidel (2000), high-z galaxies "are undeniably dusty…. Large corrections for dust extinction will be necessary in the interpretation of UV-selected surveys, and only IR observations can show whether the currently adopted corrections are valid or suggest alternatives if they are not." The far-IR spectrum tells us the amount of dust present, but says little about how the dust is distributed. The dust distribution, which will be seen directly when the galaxies are resolved in the far-IR/sub-mm, strongly influences the extinction (Calzetti 2001). The galaxy assembly process could be studied via high spatial resolution spectral line maps. For example, a C^+ 158 μm line map at $\lambda/\Delta\lambda \sim 10^4$ would provide vital information about the gas dynamics in merging and interacting systems and reveal the rotation speeds and velocity dispersions within and among galaxies and protogalaxy fragments.

Figure 1. A far-IR/sub-mm interferometer that provides HST-class resolution would resolve as much detail in a galaxy at z = 10 as ISO did in M31. These images illustrate that it is impossible to deduce the far-IR appearance of a galaxy from an optical image. The far-IR image reveals the sites of star formation and the reservoir of interstellar matter available for new star formation (Haas et al. 1998).

M31 in the visible.

ISO map of M31.
(wavelength of 175 microns)
North is up, east is left.

As noted in the SAFIR white paper, far-IR continuum and line emissions are excellent indicators of the star formation rate and the physical conditions in star forming molecular clouds. At 10 Mpc, the distance of a nearby galaxy, a giant H II region subtends about 1 arcsec, and the typical spacing between neighboring regions is about 10 arcsec. SAFIR could be used to study individual sites of star formation spectroscopically. Later, with HST-class resolution, we could make similarly detailed observations of objects at much greater distances to learn how star formation works in protogalaxies and systems having very low heavy element abundance. With the same resolution we could study star forming regions in Virgo cluster galaxies at the linear scales sampled by past IR missions (IRAS and ISO) in the Milky Way. This would help us to understand the chemical and energetic effects of star formation on the interstellar and intergalactic medium and better interpret measurements of the high-z universe.

"The central regions of galaxies were likely heavily dust enshrouded during their formation epoch. Future far-IR observations can provide a window into this formation process and help determine the relationship between bulge formation and black hole formation." (Spergel, 2001) Black hole masses could be routinely measured with high spatial resolution spectral line mapping in the far-IR/sub-mm. High angular resolution submillimeter timing observations of the black hole at the Galactic center have the potential to enable a measurement of its spin (Melia et
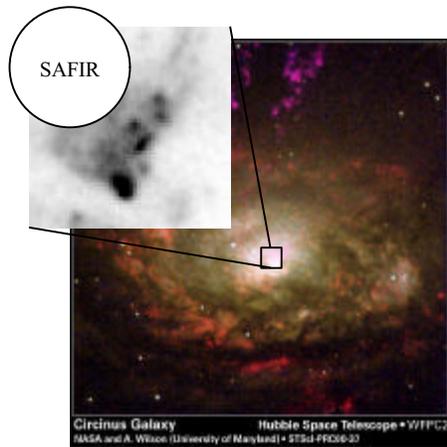
al. 2001). Such a measurement could substantially advance our understanding of the role played by supermassive black holes in galaxy formation and evolution (Elvis et al. 2002), and could yield new insight into fundamental physics, perhaps with cosmological implications.

Figure 2. Multiple optical emission line sources were seen near the central active nucleus of Circinus with HST (Wilson et al. 2000). These sources would lie within a single SAFIR beam (circle). A far-IR/sub-mm interferometer could produce high-resolution images and spectral line maps and provide valuable information about the physical conditions, gas dynamics, and star formation in Active Galactic Nuclei, unequivocally testing AGN emission and orientation hypotheses.

High resolution is important for another, subtler reason. At wavelengths $\lambda > 200$ μm, the sensitivity of SAFIR will be confusion limited to ~10 μJy (Blain 2000); at these long wavelengths it could detect starburst galaxies out to z > 5, but L_* galaxies only out to z ~ 2. However, an observatory with only three times better angular resolution than SAFIR would have a significantly lower confusion limit and could detect even a galaxy like the Milky Way out to z ~ 10. At far-IR/sub-mm wavelengths galaxies do not decrease in brightness with increasing redshift as $(1+z)^{-2}$, as one might expect, because an increasing portion of the emission is shifting *into* the observed wavelength band. At sub-mm wavelengths this so-called "negative K correction" compensates cosmological dimming out to z ~ 10. While a single aperture telescope larger than SAFIR may be possible, a factor of three seems very challenging. However, the nature of the far-IR sources is such that adequate sensitivity can be achieved with smaller apertures, and hence the spatial resolution can better be provided with interferometry. Thus, by beating confusion, far-IR/sub-mm interferometers could follow up on *all* the galaxies and proto-galaxies seen by HST, NGST, and ALMA. An important observational goal is to sample a representative volume of the high-z universe in the far-IR/sub-mm with HST-class angular resolution and spectral resolution sufficient to resolve the velocity structure in distant objects.

Figure 3. A far-IR/sub-mm interferometer with 10^{-20} W/m^2 sensitivity (equivalent bolometric magnitude 31.2) would slice through the Milky Way and between nearby galaxies to image galaxies and protogalactic objects out to z ~ 10. (Credit: A. Benson near-IR simulation for NGST)
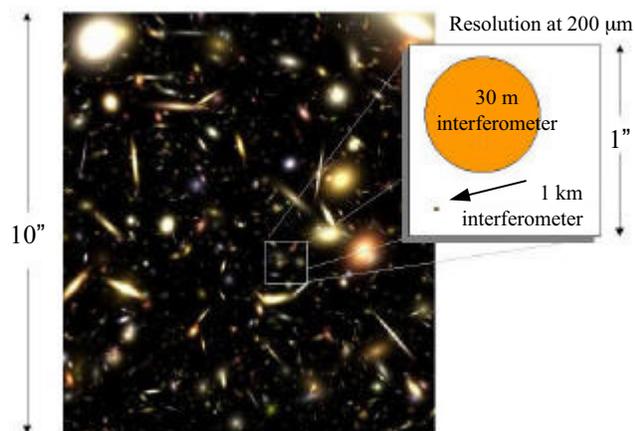
It would be a great scientific achievement to image the pristine molecular hydrogen that must have allowed primordial gas clouds to cool, collapse, and give birth to the first generation of stars, before any heavier elements existed (Haiman et al. 1996). The most likely signature is a pair of H_2 cooling lines (rest wavelengths 17 and 28 μm) redshifted to z > 10 (Abel et al. 2002). SAFIR could detect this emission if it arises at z < 10, but its discovery may have to await a far-

IR/sub-mm interferometer if it comes from higher redshifts and is concentrated in discrete objects. An interferometer would resolve out confusing background emission and could have sufficient sensitivity to make the measurement.

1.2 The formation of small structures: stars, planets, and their inhabitants

How did the solar system and the Earth form? What are the various possible outcomes of the star and planet formation process, and how does the process work? How are the initial conditions in protostellar disks reflected in the properties of planetary systems? What chemical processes occur during star and planet formation?

Star and planet formation are parts of a single process that involves the movement of matter from envelopes extended over about 10,000 AU to disks on scales of 1 - 100 AU, and ultimately into stars and planets on much smaller scales (Evans 2001). The nearest protostellar objects are at 140 pc, where 1 AU subtends an angle of 7 mas. Future astrophysicists will need high spatial and spectral resolution measurements that reveal the bulk flows of material and the physical conditions (density, temperature, magnetic field strength, and chemical abundances) in dense molecular cores, protostars, protoplanetary systems, and debris disks (Evans 1999). The SAFIR white paper explains the essential need for far-IR continuum and spectral line measurements of these systems and the capability of SAFIR to resolve protostars down to the 100 AU scale.

With a far-IR/sub-mm interferometer we will be able to probe much smaller physical scales, particularly the scales relevant to studies of planet formation. Far-IR interferometric studies of circumstellar disks will reveal dust concentrations that represent the early stages of planet formation, and measurements of exozodiacal debris disks will show gaps and structures produced by resonances with already-existing planets (Ozernoy et al. 2000). By observing planetary systems in a wide range of evolutionary states and following individual systems over a period of years we could learn how protostellar material migrates and coalesces to form planets. Observations such as these would almost surely revolutionize our understanding of how the solar system formed. The rich far-IR line spectrum would be exploited to "follow the evolution of chemical abundances and locate reservoirs of biogenic materials" (Evans 2001). We will want to understand chemical evolution from molecular cores to planets in a unified way. The spectra of gas giant planets, which emit most of their light in the far-IR, could be measured with an interferometer.
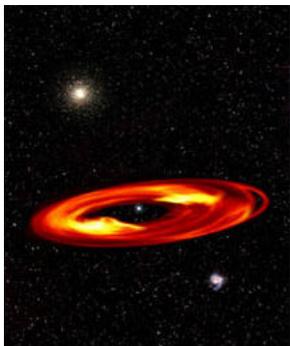
Figure 4. This artist's concept of the Vega debris disk illustrates the resonance features that could be studied with far-IR/sub-mm interferometers. At Vega's distance 1 AU subtends an angle of 128 mas. Studies of the structure of protoplanetary and debris disks will go a long way toward advancing our understanding of planet formation. (Credit: D. Wilner, M. Holman, P. Ho, and M. Kuchner; CfA Press release http://cfa-www.harvard.edu/newtop/previous/011802.html).

The next generation of far-IR and submillimeter observatories – SIRTF, SOFIA, and Herschel – will be very sensitive, but will have insufficient spatial resolution to achieve the observational goals outlined above. The key to making further progress will be to increase angular resolution

by many orders of magnitude without sacrificing sensitivity or spectral resolution. About a decade from now ALMA will provide unprecedented spatial and spectral resolution at millimeter and submillimeter wavelengths, far out on the Rayleigh-Jeans tail, and probe regions with very high dust column densities. SAFIR will look where protostars are most luminous, in the far-IR, image them at ~100 AU scales for the first time, and chart the velocity structure to provide definitive evidence of envelope collapse. An additional factor of 10 - 100x improvement in spatial resolution will be needed to image protoplanetary and planetary debris disk structure in the spectral region where these objects emit the bulk of their energy.

2. Desired measurement capabilities

Table 1 summarizes the measurement capabilities needed to achieve the science goals outlined in sections 1.1 and 1.2 and shows that the desired capabilities are similar for the two applications. Table 2 summarizes the capabilities of the next-generation observatories NGST, SAFIR, and ALMA. SAFIR will pry open the door to the "invisible" far-IR universe and leave the astrophysics community desiring the next critical capability: better angular resolution. As can be seen by comparing the SAFIR column of Table 2 with Table 1 an improvement by two orders of magnitude in angular resolution is desired.

Table 1. Desired Measurement Capabilities for the Mid-IR to Millimeter Spectral Range

Science goal	Formation and evolution of cosmic structure	Formation of stars and planetary systems
Sample targets	Hubble Deep Fields, gravitational lens sources, interacting galaxies	Nearest protostars, Orion prolyds, Vega, HH 30, and other disks
Wavelength range (peak emission) (μm)	40 – 1000	30 – 300
Angular resolution (mas)	20	10
Spectral resolution ($\lambda/\Delta\lambda$)	$>10^4$	3×10^5
Point source sensitivity, νS_ν (W/m^2)	10^{-20}	10^{-20}
Field of view (arcmin)	4	4

Table 2. Measurement Capabilities of Next-Generation Observatories

Observatory	NGST	SAFIR (10 m)	ALMA
Wavelength range (μm)	0.6 – 30	30 – 300	850 –10,000 plus windows at 350, 450
Angular resolution (mas)	50 at 2 μm	2500 at 100 μm	10 at 1 mm
Spectral resolution ($\lambda/\Delta\lambda$)	10^3	10^6	$>10^6$
Point source sensitivity, νS_ν (W/m^2)	$10^{-21} – 10^{-20}$	10^{-20}	10^{-19} at 1 mm
Field of view (arcmin)	4	4	0.3 at 1 mm, bigger field with mosaicing

3. Mission concepts

How will we satisfy the inevitable desire for detailed far-IR/sub-mm views of the high-redshift universe and protoplanetary disks? An interferometer with total aperture comparable to that of SAFIR (78 m^2) would have the desired sensitivity *and* could provide the desired angular resolution (Table 1). The resolution of an interferometer with maximum baseline b$_{max}$ is $\Delta\theta$ = 10 mas (λ/100 μm)(b$_{max}$/1 km)$^{-1}$. Thus, a 1 km maximum baseline is needed to provide the angular

resolution ultimately desired in the far-IR/sub-mm. To obtain excellent image quality all spatial frequencies would have to be sampled in two dimensions; in other words, measurements would have to be made on many baselines $b < b_{max}$, and at many baseline position angles. This so-called "u-v plane" filling is accomplished with ground-based interferometers by deploying many apertures, allowing for array reconfiguration, and relying on Earth rotation. In space there is more freedom to move apertures to desired locations, so one can tailor the u-v coverage to the problem at hand. There is a substantial cost advantage to limiting the number of apertures, particularly because to achieve background-limited performance and the desired sensitivity, the mirrors would have to be very cold (~5 K). However, in space, where there is no atmosphere to distort the wavefront, 2 or 3 apertures would suffice. The preferred location for the interferometer is the Sun-Earth L2 point, as it is distant enough to help with cooling and pointing, yet near enough to handle a large data rate.

A Michelson interferometer, in which parallel beams are combined using a half-silvered mirror or the equivalent, offers several advantages. First, a relatively modest number of detectors would be required. In a conventional Michelson interferometer a single-pixel detector is needed for each baseline, or two such detectors can be used because there are two "output ports." Detector arrays would provide a multiplex advantage that could be used either to widen the field of view or improve signal-to-noise by spectrally dispersing. The field of view could be as large as 5 arcmin $(N_{pix}/100)$ $(\lambda/100 \ \mu m)$ $(d/4 \ m)^{-1}$, where d is the diameter of the individual aperture mirrors and N_{pix} is the pixel count in one array dimension. A 100 x 100 pixel array would provide the desired field size. Second, a Michelson interferometer can be operated in "double Fourier" mode (Mariotti & Ridgway 1988), so it naturally provides high spectral as well as high spatial resolution. The spectral resolution $R = 10^4$ $(2\Delta/1 \ m)$ $(\lambda/100 \ \mu m)^{-1}$, where Δ is the length of the delay line stroke (i.e., 2Δ is the optical delay), so a 0.5 m stroke would yield the desired spectral resolution in every spatial resolution element. A small additional optical delay would be needed to compensate for geometric delay associated with the off-axis angles in the wide field.

Mather et al. (1999) first suggested the possibility of a 1 km maximum baseline far-IR/sub-mm imaging and spectral interferometer space mission called SPECS (Submillimeter Probe of the Evolution of Cosmic Structure). The concept and a technology roadmap were further developed with science and engineering expertise provided through the February 1999 community workshop on "Submillimeter Space Astronomy in the Next Millennium" (http://space.gsfc.nasa.gov/astro/smm_workshop/). The concept of a science and technology pathfinder mission called SPIRIT (Space IR Interferometric Telescope) originated at the workshop. SPIRIT is much like SPECS, except that the interferometer would be built on a boom and have $b_{max} \sim 30$ m ($\Delta\theta \sim 0.34$ arcsec at 100 μm). SPECS, like the original concept for TPF, would use formation flying to maneuver the interferometer apertures. For more information on the SPIRIT and SPECS concepts see Leisawitz et al. (2000).
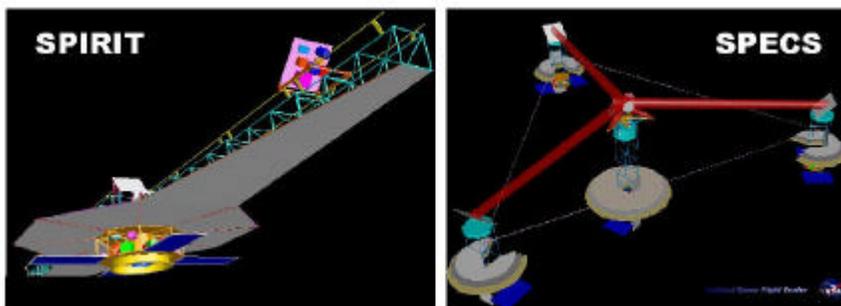
Figure 5. SPIRIT, a scientific and technology pathfinder for SPECS, could achieve the spatial resolution of a 60 m filled aperture telescope on a 30 m boom.
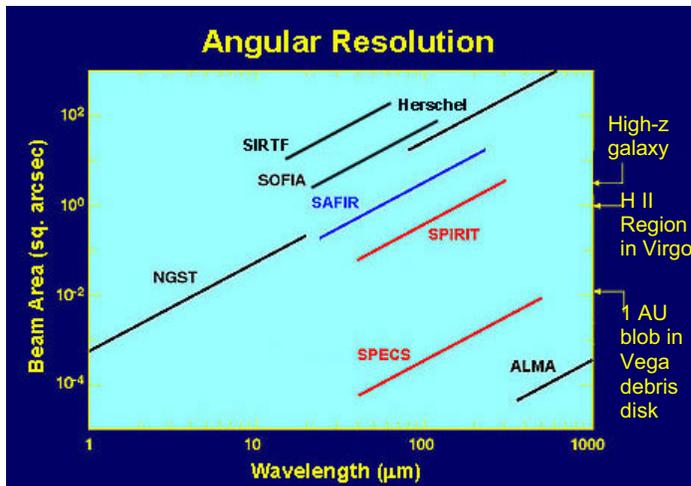
Figure 6. After SAFIR an additional hundred-fold improvement in angular resolution will be needed to achieve the science goals outlined in section 1 and attain resolution comparable to that of NGST and ALMA in the spectral regions that bracket the far-IR/sub-mm. The resolution gain could be accomplished in two steps, first with the pathfinder mission SPIRIT, and then with SPECS. N.B. A protoplanetary disk at 140 pc is about the same size as an H II region in the Virgo cluster.

4. Enabling technologies

"The [Radio and Submillimeter Wave Astronomy] panel supports the recommendation of the Panel on Ultraviolet, Optical, and Infrared Astronomy from Space that NASA pursue technology development leading toward a far-infrared/submillimeter interferometer in space." – Decade Report, Panel Reports, p. 170.

New technology will be needed in four areas: 1) detectors, 2) cooling, 3) optics and interferometry, and 4) large structures and formation flying. In this section we summarize the requirements in each area and cite possible solutions, then we conclude with a brief discussion of technology validation on space missions due to launch in the coming decade. More information on the enabling technologies for far-IR/sub-mm interferometry is given by Shao et al. (2000).

4.1 Detectors

The detector goal is to provide noise equivalent power less than 10^{-20} W Hz$^{-1/2}$ over the 40 – 850 μm wavelength range in a 100 x 100 pixel detector array, with low-power dissipation array readout electronics. This low noise level is a prerequisite for background-limited telescope performance. The ideal detector would count individual photons and provide some energy discrimination, which would enable more sensitive measurements. Among the encouraging recent developments in detector technology are superconducting transition edge sensor (TES) bolometers (Benford et al. 2002), SQUID multiplexers for array readout (Chervenak et al. 1999), and single quasi-particle counters built out of antenna-coupled superconducting tunnel junctions and Rf-single electron transistors (Schoelkopf et al. 1999).

4.2 Cooling

The cooling requirements for space-based far-IR/sub-mm interferometry are similar to those for a large single-aperture telescope like SAFIR. To take full advantage of the space environment, the mirrors will have to be very cold (~5 K) and the detectors even colder (<0.1 K). Active coolers will have to operate continuously and not cause significant vibrations of the optical

components. The coolers should be light in weight. Cooling power will have to be distributed over large mirror surfaces. Thermal transport devices will likely have to be flexible and deployable. Large, deployable sunshades will be needed, and they will have to provide protection without seriously compromising sky visibility. Since several stages of cooling must be used to reach the required temperatures, the devices that operate in each temperature range must be able to interface with each other both mechanically and thermally. The Astro E-2 mission will use a three-stage cooling system for its X-ray microcalorimeters, which operate at 65 mK (Breon et al. 1999; Shirron et al. 2000), and important advances in cooler technology will be made for NGST. Cryogenic capillary pumped loops, which have already been tested in space, have the potential to distribute cooling power over long distances (Bugby et al. 1998).

4.3 Optics and Interferometry

The mirrors needed for far-IR/sub-mm space interferometry are similar to those needed for SAFIR, only smaller. The mirrors must: (a) be light in weight ($1 - 3$ kg m^{-2}), (b) have a surface roughness not exceeding ~0.5 μm rms, (c) be able to be cooled to <10 K, and (d) maintain their shape to a small fraction of a wavelength when subjected to cooling or mechanical stress. Flat mirrors, perhaps stretched membranes (Dragovan 2000), could be used for the light collecting elements of the interferometer. The additional requirements for interferometry are beamsplitters that can operate at ~4 K and over the far-IR/sub-mm wavelength range, and long-stroke cryogenic delay lines. For a 5-year SPIRIT mission the delay line would have to be able to stroke (full amplitude) at ~10^{-2} Hz and survive at least 10^6 cycles; for SPECS the ideal delay line would move 100x faster and survive a proportionately greater number of cycles. (These numbers are based on the assumption that the mirror movement is fast enough to completely sample the synthetic aperture plane in the time required to build up the typical desired sensitivity.) The delay line would have to impart minimal disturbance on the metering structure. Finally, mosaicing techniques and algorithms for wide-field interferometry will have to be developed. Research on cryogenic delay lines and beam combiners (Swain et al. 2001; Lawson et al. 2002) and wide-field imaging interferometry (Leisawitz et al. 2002; Rinehart et al. 2002) is now underway.

4.4 Large Structures and Formation Flying

A variety of architectures are possible for SPIRIT, but all of them depend on the availability of a lightweight, deployable truss structure measuring at least 30 m in length when fully expanded. Any parts of the truss that will be seen by or in thermal contact with the mirrors must be cryogenic. One possible design requires the deployed structure to be controllable in length. Another requires tracks and a mirror moving mechanism. A third design solution uses a series of mirrors along the structure to provide non-redundant baseline coverage. In all cases the boom would spin to sample different baseline orientations. Any repeating mirror movements will have to be smooth and rely on a mechanism that is robust enough to survive at least 10,000 cycles. Structures designed to meet the challenges of space-based optical interferometry have been under study for a long time for SIM (Laskin & San Martin 1989), which has far more demanding control and metrology requirements than those of SPIRIT because SIM will operate at much shorter wavelengths.

Free-flying spacecraft will be needed to accomplish imaging interferometry with maximum baseline lengths in the 1 km range. The requirement is to sample the u-v plane completely, yet avoid the need for an unaffordable amount of propellant for formation flying. It may be necessary to combine tethers with formation flying to form a long-baseline observatory that maintains symmetry while rotating. The system will have to be deployable, stable, and capable of being pointed at a variety of targets. A modeling effort is now underway, and early results suggest that tethered formation flying is feasible (Farley & Quinn 2001).

4.5 Technology Validation

"SAFIR, the [UVOIR from Space] panel's top-priority moderate-size mission, … will enable a distributed array in the decade 2010 to 2020…The single most important requirement is improved angular resolution. The logical build path is to develop a large, single-element (8-m class) telescope leveraging NGST technology on time scales set by NGST's pace of development. A later generation of interferometric arrays of far-infrared telecopes could then be leveraged on SIM or TPF technologies …." – Decade Report, Panel Reports, p. 329.

Table 3 shows that there could be a rich heritage in space-validated technologies for far-IR/sub-mm interferometry by the beginning of the next decade. Ground-based laboratory or field research and testbed experiments are already underway, and more such research will be proposed to advance the technology readiness of components (e.g., detectors and array readout devices), systems (e.g., cryogenic delay line), or techniques (e.g., wide-field imaging interferometry) this decade.

Table 3. Technology Heritage for Long-baseline Far-IR/Sub-mm Interferometry [a]

Technology	SIRTF 2002	SOFIA 2004	Astro-E2 2005	StarLight 2006	NGST 2009	SIM 2009	SPIRIT 2010	ST-? [b] ~2010	SAFIR 2012
Detectors	X	*					**		**
Coolers	X		*		*		**		**
Optics & Interferometry	X			* **	*	*	**		**
Large structures & Formation Flying				**	*	*	X	**	X

NOTE: X denotes mission contributing to technology development; * denotes mission critical to success of SPIRIT (similar to technology inheritance for SAFIR); ** denotes mission critical to success of SPECS
[a] TPF will contribute substantially to the technology heritage if an interferometric solution is selected from among several concepts under consideration
[b] A hypothetical New Millennium Mission designed to validate tethered formation flying

5. Recommendations

"A rational coordinated program for space optical and infrared astronomy would build on the experience gained with NGST to construct SAFIR, and then ultimately, in the decade 2010 to 2020, build on the SAFIR, TPF, and SIM experience to assemble a space-based, far-infrared interferometer." – Decade Report, p. 110.

A coordinated, intensive technology program this decade is the key to success on this timescale. The critical technology areas outlined in section 4 – detectors, cooling systems and components, large optics, interferometric techniques, cryogenic delay lines, deployable structures, and formation flying – deserve particular attention. Much of this investment will apply to SAFIR as well as far-IR/sub-mm interferometry.

A study program for far-IR/sub-mm space astronomy should be initiated as soon as possible. To ensure that the technology funds will be wisely invested it is essential to take a system-level look at the scientific, technical, and design tradeoffs. SAFIR and far-IR/sub-mm interferometry concepts could be studied together to ensure that each mission takes the best advantage of its architecture type, and to explore the possibility that overall cost savings could accrue through, for example, reuse of test facilities, hardware, design solutions, and coordinated technology validation. The study might identify presently unplanned but necessary technology demonstration experiments.

References

Abel, T., Bryan, G.L., and Norman, M.L. 2002 Science, 295, 93

Adelberger, K.L., and Steidel, C.C. 2000, ApJ, 544, 218

Astronomy and Astrophysics in the New Millenium 2001, National Research Council, (National Academy Press: Washington, DC) [Decade Report]

Astronomy and Astrophysics in the New Millenium Panel Reports 2001, National Research Council, (National Academy Press: Washington, DC) [Decade Report, Panel Reports]

Benford, D.J., Ames, T.A., Chervenak, J.A., Grossman, E.N., Irwin, K.D., Khan, S.A., Maffei, B., Moseley, S.H., Pajot, F., Phillips, T.G., Renault, J.-C., Reintsema, C.D., Rioux, C., Shafer, R.A., Staguhn, J.G., Vastel, C., and Voellmer, G.M. 2002, AIP Proc., in press.

Blain, A. 2000, astro-ph/0011479

Breon, S., Shirron, P., Boyle, R., DiPirro, M., Serlemitsos, A., Tuttle, J., and Whitehouse, P. 1999, Cryogenics, 39, 677

Bugby, D., Nguyen, T., Kroliczek, E., Ku, J., Swanson, T., Cullimore, B., and Baumann, J. 1998, IECEC-98-I288

Calzetti, D. 2001, PASP, 113, 1449

Chervenak, J. et al. 1999, Appl. Phys. Letters, 74, 4043

Dragovan, M. 2000, astro-ph/0001241

Elvis, M., Risaliti, G., and Zamorani, G. 2002, ApJL, in press (astro-ph/0112413)

Evans, N.J., II 2001, BAAS, 198, 8205 (see http://space.gsfc.nasa.gov/astro/aasbeyond/evans)

Evans, N.J., II 1999, ARAA, 37, 311

Farley, R.E., and Quinn, D.A. 2001, Proc. AIAA, 2001-4770

Haas, M., Lemke, D., Stickel, M., Hippelein, H., Kunkel, M., Herbstmeier, U., and Mattila, K. 1998, A&A, 338, L33

Haiman, Z., Rees, M.J., and Loeb, A. 1996, ApJ, 467, 522

Laskin, R.A., and San Martin, A.M. 1989, Proc. AAS/AIAA Astrodynamics Specialist Conf., Stowe, VT

Lawson, P.R., Swain, M.R., Dumont, P.J., Moore, J. D., and Smythe, R. F. 2002, BAAS, 199, 4503

Leisawitz, D., and Hauser, M.G. 1988, ApJ, 332, 954

Leisawitz, D., Danchi, W., DiPirro, M., Feinberg, L.D., Gezari, D., Hagopian, M., Langer, W.D., Mather, J.C., Moseley, S.H., Jr., Shao, M., Silverberg, R.F., Staguhn, J., Swain, M.R., Yorke, H.W., and Zhang, X. 2000, Proc. SPIE 4013, 36

Leisawitz, D., Leviton, D., Martino, A., Maynard, W., Mundy, L.G., Rinehart, S.A., and Zhang, X. 2002, Proc. SPIE, in prep (see http://space.gsfc.nasa.gov/astro/aasbeyond/contrib/wiit)

Mariotti, J.-M., and Ridgway, S.T. 1988, A&A, 195, 350

Mather, J.C., Moseley, S.H., Jr., Leisawitz, D., Dwek, E., Hacking, P., Harwit, M., Mundy, L.G., Mushotzky, R.F., Neufeld, D., Spergel, D., and Wright, E.L. 1999, Rev. Sci. Instr., submitted (astro-ph/9812454)

Melia, F., Bromley, B.C., Liu, S., and Walker, C.K. 2001, ApJ, 554, L37

Ozernoy, L.M., Gorkavyi, N.N., Mather, J.C., and Taidakova, T.A. 2000, ApJ, 537, L147

Rieke, G.H., Harvey, P.M., Lawrence, C.R., Leisawitz, D.T., Lester, D.F., Mather, J.C., Stacey, G.J., Werner, M.W., and Yorke, H.W. 2002, SAFIR white paper

Rinehart, S.A., Leisawitz, D., Leviton, D., Martino, A., Maynard, W., Mundy, L.G., and Zhang, X. 2002, Proc. SPIE, in preparation.

Schoelkopf, R.J., Moseley, S.H., Jr., Stahle, C.M., Wahlgren, P., and Delsing, P. 1999, IEEE Trans. Appl. Supercond., 9, 2935

Shao, M., Danchi, W., DiPirro, M., Dragovan, M., Feinberg, L.D., Hagopian, M., Langer, W.D., Lawson, P.R., Leisawitz, D., Mather, J.C., Moseley, S.H., Jr., Swain, M.R., Yorke, H.W., and Zhang, X. 2000, Proc. SPIE, 4006, 772

Shirron, P., Canavan, R., DiPirro, M. J., Tuttle, J.G., and Yeager, C.J. 2000, Adv. Cryo. Eng. 45,

Spergel, D. 2001, BAAS, 198, 8203 (see http://space.gsfc.nasa.gov/astro/aasbeyond/spergel)

Swain, M.R., Lawson, P.R., Moore, J.D., and Jennings, D. 2001, Proc. 36th LAIC (Liege International Astrophysical Colloquium: From Optical to Millimetric Interferometry; Scientific and Technical Challenges), Liege, Beligum

Trentham, N., Kormendy, J., and Sanders, D.B. 1999, AJ, 117, 2152

Wilson, A.S., Shopbell, P.L., Simpson, C., Storchi-Bergmann, T., Barbosa, F.K.B., and Ward, M.J. 2000, AJ, 120, 1325

Acronyms

ALMA – Atacama Large Millimeter Array
HST – Hubble Space Telescope
NGST – Next Generation Space Telescope
ISO – Infrared Space Observatory
SAFIR – Single Aperture Far-IR Telescope
SIM – Space Interferometry Mission
SIRTF – Space Infrared Telescope Facility
SOFIA – Stratospheric Observatory for IR Astronomy
SPECS – Submillimeter Probe of the Evolution of Cosmic Structure
SPIRIT – Space Infrared Interferometric Telescope
TPF – Terrestrial Planet Finder

Complementarity of NGST, ALMA, and far IR space observatories

John C. Mather

NASA Goddard Space Flight Center, Code 685, Greenbelt, MD 20771
john.mather@gsfc.nasa.gov

Abstract: The Next Generation Space Telescope (NGST) and the Atacama Large Millimeter Array (ALMA) will both start operations long before a new far IR observatory to follow SIRTF into space can be launched. What will be unknown even after they are operational, and what will a far IR space observatory be able to add? I will compare the telescope design concepts and capabilities and the advertised scientific programs for the projects and attempt to forecast the research topics that will be at the forefront in 2010.

1. Introduction

In a strategic planning sense, the most ambitious space observatories must follow after the most ambitious ground-based and space-based observatories of the previous generation. Almost by definition, they look impossibly difficult when they are conceived. By the time the consensus is built for such a space observatory, it must be clear that the other tools have been nearly exhausted and that there are important mysteries that can never be resolved without new tools.

When can we anticipate that other tools will be exhausted, so that a series of far infrared space observatories becomes mandatory? The crystal ball is cloudy, but we do anticipate several new observatories that have been strongly endorsed by the National Academy, and may be online within the next 10 years. The SIRTF (Space Infrared Telescope Facility) is coming within a year, pending resolution of certain flight software issues, and will provide an extraordinary leap in sensitivity, reaching the confusion limits for its size at all wavelengths. The ALMA (Atacama Large Millimeter Array) and the NGST (Next Generation Space Telescope) are the two most powerful, approaching the far infrared region from both ends. Some kind of 25 m GSMT (Giant Segmented Mirror Telescope) visible – near IR telescope will be built.   By 2015, all three will have been in operation for about 5 years, and the cream of the new discoveries may already have been found. What might those be, and what will they imply for far IR observations in space?

The technical capabilities are easily summarized. NGST imaging sensitivity will be photon background limited out to about 12 μm, with an aperture of 6 or 7 meters, and there will be moderate resolution spectroscopy (R ~ 1000 out to 5 μm, R ~ 2000 from 5 to 28 μm). ALMA will have about 64 12 m aperture dishes but detector technology and atmospheric interference limit its sensitivity to several times the quantum noise limit in

those submillimeter atmospheric windows where observations can be made at all. ALMA will have high spectral resolution, as high as wanted to resolve any thermal motions. The GSMT will presumably be outfitted with a full array of imaging and spectroscopic capabilities, but its longer wavelength infrared performance might not be optimized since the prime driver will be the possibility for good imaging.

The case for a far IR observatory in space has been very ably presented in this conference (Rieke, 2002), and includes many wonderful goals and objectives. There is much to be learned from SIRTF and from a future successor, e.g. the SAFIR (Single Aperture Far IR) telescope. However, I am concerned that we have not let our imaginations run sufficiently far into the future. Such a mission will not fly today or tomorrow, but more than a decade into the future. So what will be discovered by 2010 or 2015, and what will be left for far IR astronomy? This is an important exercise, because most of the arguments advanced today for the need for future missions are written from today's perspective, and many will need revision in another decade. As we explain the need for SAFIR, let us consider where our colleagues will be in the future.

2. Key Scientific Questions and Forecast

The National Academy has identified the following key scientific questions in the Decadal Survey (McKee and Taylor, 2000), many of which will be rather well addressed by the ALMA and NGST. What will be left open? What will the competition be? The UV astronomy community does not have a major mission planned after HST, but a committee is meeting shortly in Chicago to outline new strategies for UV astronomy in space. Neither far IR astronomy nor UV astronomy is ripe for commitment to a new mission, but either or both could be ready soon and I believe both should appear in the NASA strategic plan.

The Decadal Survey (p. 52) identified 5 key topics that will set the direction for astronomy and astrophysics for many years to come:

1. "How did the universe begin, how did it evolve from the soup of elementary particles into the structures seen today, and what is its destiny?
2. How do galaxies form and evolve?
3. How do stars form and evolve?
4. How do planets form and evolve?
5. Is there life elsewhere in the universe?"

They identified 5 goals for the new initiatives of this coming decade as well (p. 53):

1. "Determining large scale properties of the universe
2. Studying the dawn of the modern universe
3. Understanding black holes
4. Studying star formation and planets
5. Understanding the effects of the astronomical environment on Earth"

These goals are stated in very broad terms, and most of the new initiatives have something to say about most of them.

The NGST science working groups (SWG's) have defined a set of 23 large observing programs that might be carried out, and a subset of 7 of them that should be used to drive the observatory design. These 23 are called the Design Reference Mission, and are not meant as a substitute for the competitive process that will eventually select observers and their programs. The 7 core programs are:

1. Formation & Evolution of Galaxies - Imaging
2. Formation & Evolution of Galaxies - Spectroscopy
3. Mapping Dark Matter
4. Search for Reionization Epoch
5. Measuring Cosmological Parameters
6. Formation & Evolution of Galaxies - Obscured Stars & AGN
7. Physics of Star Formation: Protostars

These 7 programs lead to the SWG's recommendation for 3 core instruments. There would be a near IR camera covering 0.6 to 5 μm with about 48 Mpixels across about 16 square arcminutes, Nyquist sampled at 2 μm, with an R ~ 100 grism mode, suitable for studying the first light, galaxy formation, dark matter, supernovae, young stars, Kuiper Belt Objects (KBO), and stellar populations. There will be a near IR (1-5 μm) multiobject spectrometer covering about 9 square arcminutes, capable of observing > 100 spectra simultaneously, with resolution of about 1000, suitable for studying Galaxy formation/diagnostics (clustering, abundance, star formation., kinematics), Active Galactic Nuclei, and young stellar clusters (Initial Mass Function (IMF)/stellar populations). There will also be a mid IR camera and spectrometer covering 5 to ~28 μm with about 4 square arcminute field, with a spectral resolution up to about 3000, suitable for studying the physics of old stars at high redshift, z~5 obscured star formation and Active Galactic Nuclei to z ~ 5, PAHs to z~5, Hα to z~15, the cool stellar IMF, protostars and disks, KBO sizes, and comets.

The predicted sensitivity of NGST is extraordinary, significantly less than 1 nJy for long exposures at near IR wavelengths. The sensitivity is worse at wavelengths > 5 μm where zodiacal light becomes bright, and worse again for wavelengths > 12 μm where the stray light from the warm parts of the sunshield becomes brighter than the zodiacal light.

Nevertheless, NGST will still be the most sensitive telescope over its entire wavelength range. NGST will be capable of seeing individual young globular clusters and galactic nuclei all the way back to their hypothetical formation time at z > 10, and individual Type Ia supernovae back to z > 5.

The ALMA objectives are similarly ambitious, and the ALMA web site gives many pages of quantitative goals. For the Distant Universe topic, they offer the following:

- "Image thermal dust emission in evolving galaxies at epochs as early as z=10.
- Yield kiloparsec-resolution images of dust emission in active galaxies and QSOs.
- Detect CO, CI, CII emission lines from Galaxies and QSOs. A line luminosity of 10,000,000 solar luminosities at z=0.5 will be detectable in under 10 minutes.
- Image the microwave decrement in galaxy clusters; together with AXAF observations, this will provide an accurate determination of the Hubble constant.
- Resolve regions of particle acceleration in the jets and lobes of radio galaxies.
- The MMA, with the world's largest aperture at wavelengths around 1 mm, will be crucial to millimeter-wave VLBI. Accretion disks in galaxies as distant as the Virgo Cluster will be seen at 20 microarcseconds resolution."

It is clear that this project has enormous potential for detailed analysis of complex phenomena, and for discovery of completely unexpected types of objects.

I will now engage in a little crystal ball gazing. I think it likely that these and other major instruments will achieve some significant portion of the following discoveries in the next 15 years:

1. Measurement of the cosmic background radiation anisotropy on all scales, and of its polarization on small scales, leading to consensus on the development of large scale structure, and of the amounts and types of dark matter at z~1000
2. Measurement of the cosmic dark matter clustering by weak lensing at z ~ 1-5
3. Measurement of cosmic dark energy by its effect on the acceleration of the universe and supernova brightness
4. Measurement of neutrino masses in terrestrial labs deep underground, confirming that they form an appreciable part of the cosmic dark matter
5. Small chance: detection of non-neutrino dark matter in the lab
6. Development of a new theory of elementary particles, perhaps based on strings or m-branes, including a first apparently successful theory of quantum gravity, with "predictions" explaining the inflationary period
7. Measurements of frame dragging in general relativity, by GP-B (around Earth) and Con-X (around black holes)
8. Detection of gravity waves with LIGO and LISA (later) due to binary collapsed objects, and maybe asymmetric supernovae
9. Discovery of the first luminous objects, including star clusters, galactic nuclei, ordinary supernovae, hypernovae, and supermassive primordial metal-free stars

10. Discovery of highly obscured luminous infrared galaxies at high redshift, showing that the first galaxies immediately produced dust
11. Discovery of the objects that re-ionized the early universe at redshift around 6
12. Beginnings of mapping the intergalactic hydrogen distribution through its influence on QSO absorption, its Lyman alpha emission, and its 21 cm emission, using advanced imaging spectroscopic equipment to map much deeper than is now possible
13. Discovery of metal-poor Population III and completely metal-free Population IV stars from the very early universe, still orbiting within or around our Galaxy
14. Agreement on the formation sequence of galaxies, probably from small objects building up by mergers, both by observations of distant objects, and from kinematic reconstruction of the history of the Milky Way and nearby galaxies
15. Discovery of black holes (quasars) at extremely high redshifts in the first luminous objects, fully equipped with dusty torii
16. Theoretical understanding of early (cool) stages of star formation, confirmed by detailed analysis of atomic and molecular lines seen with ALMA
17. Theoretical understanding of late stages of star formation, confirmed by detailed analysis of spectra taken with NGST and ALMA
18. Direct detection and characterization of planets in orbit around nearby stars, through clues from Doppler and astrometric measurements, transits, analysis of structure in dust clouds, deployment of a coronagraph on NGST or on a new special purpose planet finding space or ground-based telescope, and application of new search strategies. Determination of atmospheric properties from transit spectroscopy.
19. Discovery of large numbers of loose planets in star clusters and in the Earth's neighborhood
20. Understanding some of the effects of past astrophysical events on Earth, including passage through dust clouds, and nearby supernovae
21. Understanding the formation and stability of planetary systems through migration of bodies and interpretation of chaotic dynamics
22. Discovery of large numbers of Kuiper belt objects, and the beginning of mineralogical interpretation from their colors
23. Recovery of many more Antarctic meteorites from Mars, including one that has startling and maybe convincing evidence of water and life
24. Imminent return of Mars surface samples; when they are brought back they will be disappointing because they won't come from the hypothetical wet spots.
25. Interpretation of solar system dynamics to understand the formation history of the planets and the asteroid belt, including the event that broke up the iron meteorites from parent bodies
26. Theoretical analysis and computer modeling for all categories of mysterious objects, based on the continuation of Moore's law, and the continued development of efficient 3-D hydro and gravity codes

3. "What Ain't We Got? You Know Darned Well!" (R. Rodgers, O. Hammerstein, 1949)

We'll have sunlight on the sand and moonlight on the sea, but (at least) one subject will still be mysterious. Half of the light in the universe, the part that's converted from starlight to far infrared, will still be observed only with the small SIRTF and limited ground-based facilities. After all the progress made with other instruments, astronomers will still be concerned that they are studying their elephant blindfolded, touching the toenails to measure the mass and luminosity and origin of the elephant. The essentially unobservable objects and mysteries will include:

1. Some object classes we already know about that emit primarily in the far IR. We'll still wonder what they're doing there, whether they harbor marvelous surprises.
2. Some bright and rare object classes. Without a full sky survey at the relevant wavelengths we won't know they exist. NGST and ALMA are not survey instruments, but the LSST (Large Synoptic Survey Telescope) will do a wide field survey at visible wavelengths. However, this is certainly far from the far IR.
3. Some objects that emit primarily in the far IR will be too faint to discover at any other wavelength. To elude discovery with the NGST, which is background limited out to about 12 µm, they have to be either highly obscured, or quite cold, or highly redshifted.
4. To elude discovery and detailed study with ALMA, they need to be either faint or camouflaged or rare. To be camouflaged, they need only have featureless spectra that do not distinguish highly redshifted warm dust from local cool dust at the long wavelengths that penetrate the atmosphere enough for ALMA to see them. To have featureless far IR spectra, they need to have significant far IR optical depth, so that the dust outshines the line emission, or sufficient ionizing radiation to destroy atomic and molecular emitters at far IR wavelengths.

As there are about 2 orders of magnitude in wavelength between the bands where NGST (12 µm) and ALMA (1 mm) are at their best sensitivity, I think that there are many possibilities of objects that fit these categories. It is straightforward to plot a blackbody or a thermal emitter with an emissivity proportional to frequency on the sensitivity plots for NGST, SIRTF, and ALMA. An emitter with a flux of $\nu I_\nu = 5 \times 10^8$ Hz-Jy at 100 µm and a temperature of about 70 K in the observing frame could easily escape detection by all these instruments. I translate these numbers into luminosities of 2.35×10^{-4} L_{sun} at 1 kpc, 2.35×10^2 L_{sun} at 1 Mpc, 2.35×10^8 L_{sun} at 1 Gpc, and 6.1×10^9 L_{sun} at 3 Gpc. (I've made an approximate correction for redshift at 3 Gpc). If the temperature is a different, the undetectable luminosities are lower.

Even now with the short list of objects seen with SCUBA there are many that are not known at other wavelengths (Smail et al., 2002). I suggest the following possibilities:

1. Redshift 10-20 objects with ordinary late type stellar colors, due to some first generation galaxies that have already lost their gas supply because of supernovae blowing them apart, and therefore have only old stars. These could be a new category of object, and could change our view of the early universe and the formation of galaxies.
2. Redshift 10-20 objects that are so compact that dust clouds surround them and hide all the hot, bright stars within; these may be AGN or star-forming regions. It is generally thought that the first objects were UV bright and dust-free, but we don't know how quickly dust may be released. If the lifetime of the objects is 10^9 years but dust is copiously produced after 10^7 years, then only 1% of the objects will be found with redshifted UV radiation. It's more likely that most objects have some UV stars outside the dust cocoons even after dust is released, but we already know that some of the most luminous objects are completely obscured.
3. Molecular hydrogen cooling lines at rest frame 17 and 28 µm from the formation of the first galaxies. These lines could be quite faint, and covering large areas of sky, so we'll need a wide field very efficient spectrometer of modest resolution to find them. Maybe this will be an imaging Michelson spectrometer.
4. Redshifted nebular 5-20 µm lines of highly ionized neon, etc., which distinguish starburst galaxies from AGN. These occur in the far IR out to very high redshift, there are no known equivalents at radio wavelengths, and the UV is often obscured so is not sufficient.
5. Redshifted C^+ (157 µm) and N^+ (122, 205 µm) cooling lines from the interstellar medium in galaxies in the range out to redshift 5 or so; at higher redshift these lines come into view for ALMA at 1 mm. At lower redshift ALMA could see them if they're bright and they fall in the atmospheric windows.
6. Cool dust clouds around most stars, due to comets and asteroids as in the zodiacal light. These might be best detected as far IR excesses, and with sufficient angular resolution to recognize structure, they could lead to many inferences about the presence of planets.
7. Molecular and atomic line cooling systems for interstellar gas where the dust has been vaporized or has never formed.
8. The large scale polarization of the CMBR. This will still not be measured yet at the levels interesting to Big Bang theorists, as it will require very large and sensitive far IR arrays, as well as very good understanding of the astrophysical foregrounds and methods for removing them. Far IR observations (i.e. at wavelengths shorter than the CMBR) may be required in support of the CMB polarization measurements.

4. Other Possible Advances, Scientific and Technical

Some possible technical advances will also make far IR technology ready for space and for the astronomical challenges of extreme sensitivity. I think the following are likely developments in the next 10 years, particularly given support from NASA.

1. Far IR detector arrays will continue to improve, with improved coolers routinely available below 0.1 K, with superconducting readout electronics, and the possibility of far IR photon counters. When they reach sufficient maturity, a new generation of telescopes will become an obvious next step. This may take another decade, depending on funding and commitment levels.

2. Coherent receivers will improve, so that the ALMA looks much more attractive at 350 μm when the weather is good. There may also be a call for a warm 15 m space telescope operating at shorter wavelengths, to follow the Herschel with coherent receivers and improved spectroscopic back ends.

3. A low-power correlator will be invented that could enable a space version of ALMA, with several or many dishes flying in formation around a correlator hub spacecraft. With warm dishes and a reduced requirement for precise formation flying, this will look very attractive for observations of many ALMA and Herschel types of targets.

4. UV detectors will also continue to improve, and the UV astronomy community will correctly argue that there are important scientific goals that are ripe for their approach.

5. Innovative new spectrometer designs will enable the use of far IR arrays for efficient high-resolution spectrometry, combining dispersion with Fourier or Fabry-Perot interferometry. These will be tested first on SOFIA and then will serve as the basis for a new observatory.

6. Earth scientists will fly an imaging FTS (Fourier Transform Spectrometer) in geosynchronous orbit for the 10 μm region in the NASA EO3 technology demonstration program, demonstrating the principles and developing the signal processing technology.

7. NGST will demonstrate deep passive radiative cooling, and engineers will develop concepts for two-stage radiative coolers that reach 7 K, the zodiacal light limit at Earth

8. Active coolers will be developed for NGST and TPF, capable of large capacity at moderate power consumption, providing cooling for the telescope and for the detectors without stored cryogens.

9. Someone will develop formation flying for constellations of many spacecraft looking down at the Earth.

10. Ground-based interferometry will develop quite well, using AO on large telescopes to boost sensitivity by orders of magnitude over present capabilities. Astronomers will develop an appreciation for the possibility of high-resolution imaging of stellar surfaces, cores of AGN, dust in close proximity to young and old stars, planetary surfaces in the Solar System, and star-forming regions.

5. Implications for Far IR Space Mission Planning

It's clear from all this that the crystal ball is very cloudy. The next conclusion is that all areas of astronomy benefit from improved detector technology, so that should be the top priority for long term funding. Larger scale technologies like coolers, radiative shields, mirrors, and formation flying may have other customers who can develop them without specific requirements from the far IR astronomy community.

Then, based on the principle that we must do everything simple and cheap before we can do anything complicated and costly, there's a logical progression of far IR projects to do.

1. A wide field sky far IR survey, ideally of the whole sky, whenever the detector arrays are ready. There are many previously unknown objects that could be found with even short exposures and a small mirror, and some will represent new classes that are not expected. This mission is too large to fit the existing Explorer competition guidelines, but is probably not so large that it requires the full NASA strategic planning process. It might be suitable for a competition in an Explorer-Plus category if this could be made available. If my arguments made above are sound, it could be a winner.

2. At about the same time, a Cosmic Background Polarization mission could be built. It would like be an all-sky survey mission as well, and depends on much the same technology but requires a very different mode of operation. It might be in the same price range as the wide field far IR survey. Given the scientific backing of the community this could also be a winner, but only if the predicted astrophysical foregrounds are sufficiently well understood at the time.

3. At about the same time, a single aperture far IR telescope, SAFIR, as advocated by the Decadal Survey. This would be a follow-on to NGST, with as large an aperture as is affordable given the NASA budget possibilities. It would have to be a strategic mission, and it would have to compete with other such missions at the advisory committee and National Academy level. This could have a budget less than the NGST, if the NGST aperture is sufficient, or larger, if the scientific goals are sufficiently compelling to justify a larger telescope. It could have a variety of instruments like NGST, optimized for imaging and spectroscopy over the whole range from the NGST zodiacal light limit of 12 μm out to maybe 500 μm, depending on detailed comparison with capabilities on the ground.

4. After SAFIR, I think it will be clear that we need more angular resolution to study the objects we have discovered, and to beat the confusion limit at wavelengths > 100 μm. To do this we will need either an imaging interferometer or a new telescope 3 times as large; a smaller gain in aperture would not be scientifically compelling, and a larger one could be seen as too great a leap. An interferometer span a few times the aperture of SAFIR will be enough to make a major advance, so I think some form of single-spacecraft interferometer will be the next step. On the other hand, by that time, it may be possible that formation flying technology has been fully developed by other fund sources, and a separated spacecraft

mission like SPECS will be competitively priced. It might also be that really large aperture telescope mirrors will be possible by then, based on stretched membranes of some sort.

6. Conclusions

The short-term actions for NASA should be to 1) support far IR detector development, 2) start small study efforts to develop concepts for the first three possible missions on this list. The time scale to develop a scientific consensus and a mission concept is of the order of 4 years, just in time to get ready to do the next big thing after NGST, ALMA, and GSMT. It is important to start this journey of a thousand miles with a single step.

7. Acknowledgments

The Next Generation Space Telescope project is supported directly by NASA Headquarters. The GSFC work on far IR interferometry is supported by NASA Headquarters through grants for this conference and for technology development of the Wide Field Interferometric Imaging Testbed. We are grateful for the recognition by the scientific community and NASA management for far infrared studies in general, and to G. Rieke for his articulate advocacy of far IR astronomy.

8. References

ALMA web site, http://www.alma.nrao.edu/, 2002.

Leisawitz, D., et al., "Probing the Invisible Universe: The Case for Far-IR/Submillimeter Interferometry," (SPECS white paper), this conference, and http://arXiv.org/abs/astro-ph/0202085, 2001

McKee, C., and Taylor, J., *Astronomy and Astrophysics in the New Millennium* (the Decadal Survey), National Research Council, National Academy Press, 2001

NGST web site, http://www.ngst.nasa.gov; Design Reference Mission, http://www.ngst.nasa.gov/science/drm.html, 1999.

Rieke, G., et al., "Charting the Winds that Change the Universe - Far Infrared and Submm Astronomy," http://mips.as.arizona.edu/MIPS/fircase3.doc, 1999

Rieke, G., et al., "Charting the Winds that Change the Universe: The Single Aperture Far Infrared Observatory," this conference, 2002

Smail, Ian; Ivison, R. J.; Blain, A. W.; Kneib, J.-P., "The nature of faint submillimetre-selected galaxies," MNRAS, 331, 495, 2002.

SURVEYING GALAXY EVOLUTION IN THE FAR-INFRARED
A FAR-INFRARED ALL-SKY SURVEY CONCEPT

D.J. Benford, M.J. Amato, E. Dwek, M.M. Freund, J.P. Gardner, A. Kashlinsky[†],
D.T. Leisawitz, J.C. Mather, S.H. Moseley, R.A. Shafer, J.G. Staguhn[†]

NASA / Goddard Space Flight Center   †SSAI

ABSTRACT

Half of the total luminosity in the Universe is emitted at rest wavelengths ~80–100μm. At the highest known galaxy redshifts ($z \geq 6$) this energy is redshifted to ~600μm. Quantifying the evolution of galaxies at these wavelengths is crucial to our understanding of the formation of structure in the Universe following the big bang. Surveying the whole sky will find the rare and unique objects, enabling follow-up observations. SIRCE, the Survey of Infrared Cosmic Evolution, is such a mission concept under study at NASA's Goddard Space Flight Center. A helium-cooled telescope with ultrasensitive detectors can image the whole sky to the confusion limit in 6 months. Multiple wavelength bands permit the extraction of photometric redshifts, while a large telescope yields a low confusion limit. We discuss the implications of such a survey for galaxy formation and evolution, large-scale structure, star formation, and the structure of interstellar dust.

JUSTIFICATION

The Cosmic Infrared Background (CIRB) is the integral of the light from all sources at all distances. Much of this light comes from ultraluminous infrared galaxies, but some fraction arises in AGN and from normal galaxies. The energy released by the formation of stars and in regions around AGNs is absorbed and reemitted by dust. Half the total luminosity in the Universe is emitted at infrared wavelengths, much of it at ~100μm (Figure 1). The fraction of dust emission was higher in the past than it is today, implying that dusty galaxies produce a greater portion of the luminosity at high redshifts.
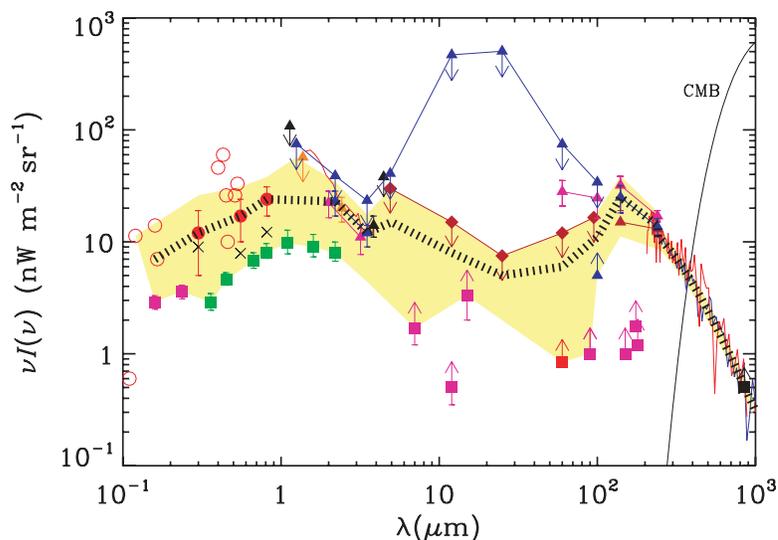
Figure 1: Extragalactic background light (Hauser & Dwek 2001). DIRBE & FIRAS measured this at long wavelengths; two peaks are known, at ~1μm and ~100μm.

A complete picture of star formation and AGN activity in the Universe can be obtained only when far infrared observations reach the ability to probe to high redshifts comparable to that at shorter wavelengths. Determination of the cosmic star formation rate history, the growth of cosmic structure, and the accompanying energy release requires direct observations of the sources that dominate the luminosity of the early Universe, which were previously seen only as a component of the CIRB. Surveying the history of the Universe – particularly for star formation – becomes harder in the optical as we look to higher redshifts and greater visual extinctions. The strong inverse-K correction at far-infrared and submillimeter wavelengths makes it possible to see galaxies to great distances (Figure 2).
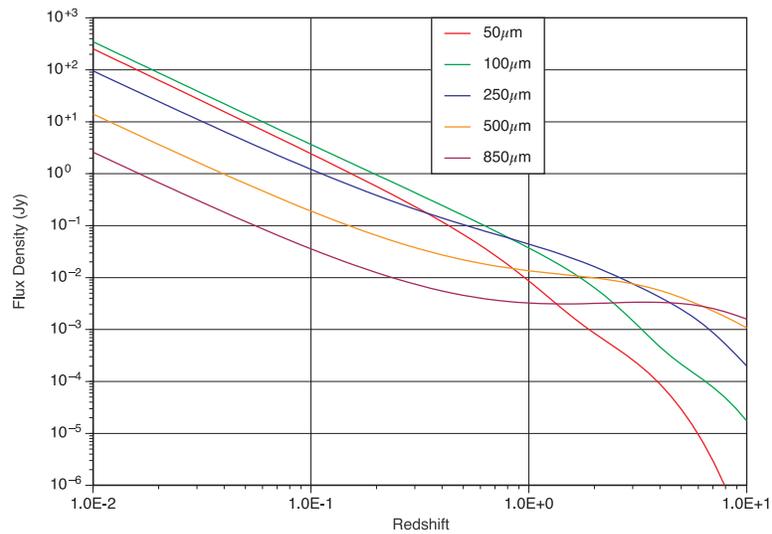
Figure 2: Flux density of an ultraluminous infrared galaxy (ULIRG) as a function of redshift illustrates the sensitivity of submillimeter wavelengths to high redshift galaxies.

At achievable flux density limits (1mJy at 100μm, roughly 1000 times more sensitive than IRAS), thousands of dusty sources at z>7 can be discovered – if they exist. Shown at right are the fluxes of a set of template galaxies (Figure 3). If we redshift these galaxies until they become too faint to resolve (Figure 4), their fluxes at three fixed observed wavelengths will trace out paths in a color-color-color space. The paths are generally well separated, enabling reliable photometric estimates of redshift and hence luminosity.
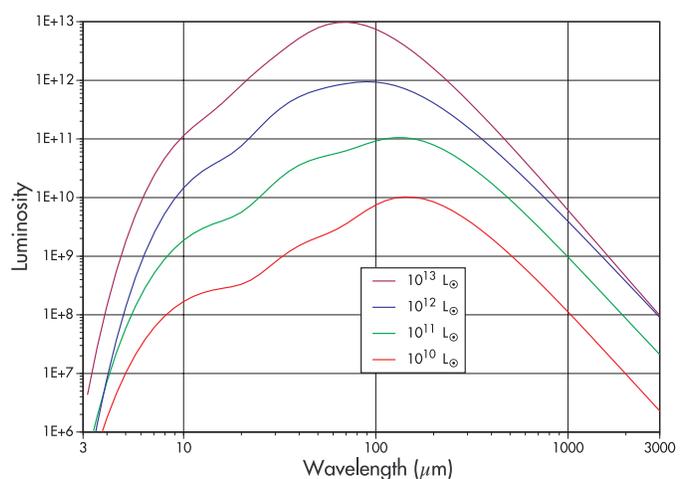
Figure 3: Template spectra of galaxies showing the strong dust emission peak near 100μm.

189
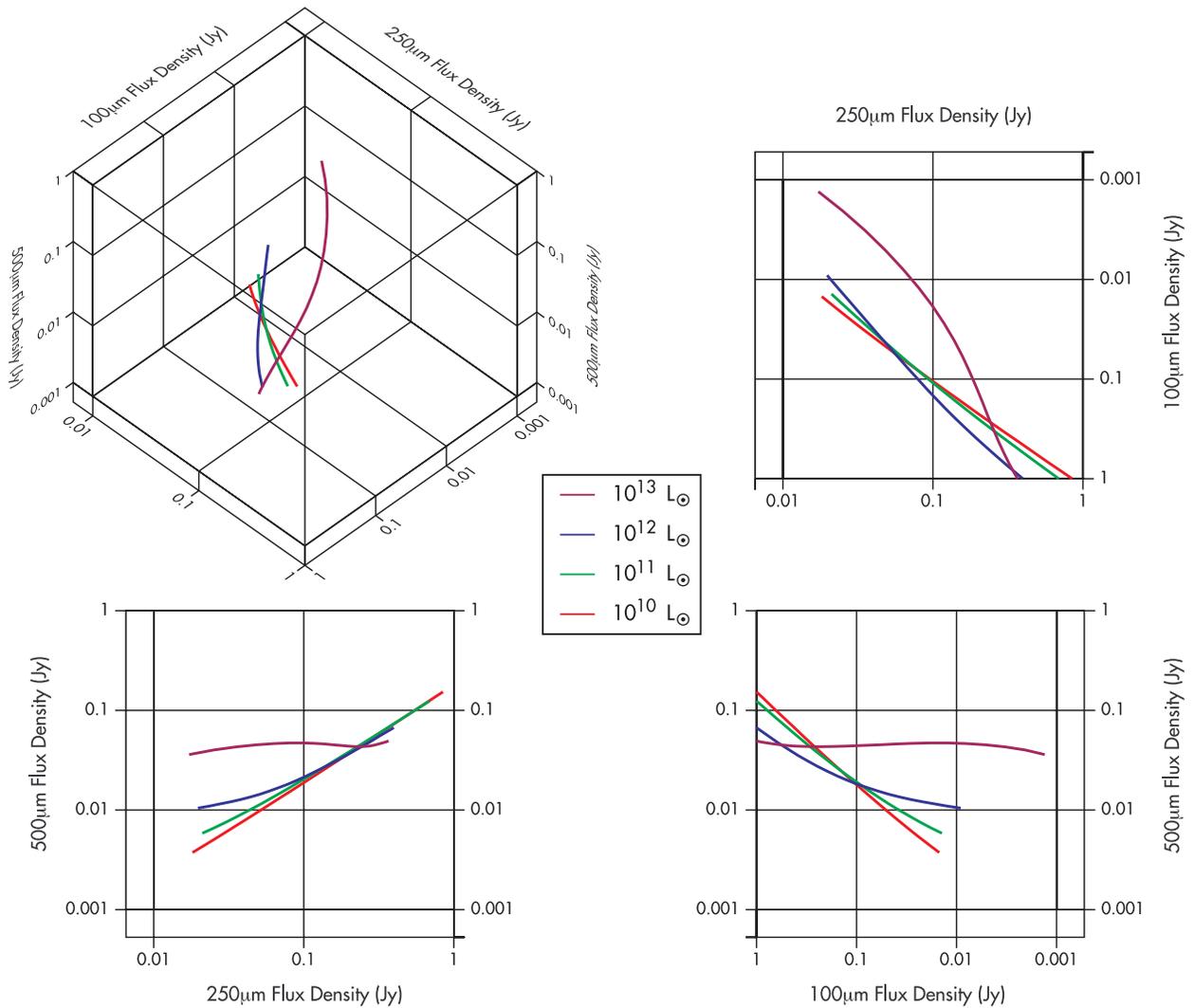
Figure 4: Paths of template galaxies observed in three far-infrared bands as they redshift until becoming fainter than the confusion limit.

The key scientific questions to be answered include:

- History of energy release in the Universe
- Evolution of the different source types with redshift
- Relative importance of nuclear and gravitational energy release
- Nature, redshift distribution and spatial distribution of the CIRB sources
- Epoch of initial star formation in the Universe

MISSION CONCEPT

How can we learn more about the sources which make up the CIRB, and how do we learn about the history of energy release in the Universe?

A space-based observatory can be built with existing technology which will:

- Map the sky at high spatial resolution at far infrared wavelengths
  Requires large telescope; large format detector arrays
- Achieve sensitivity high enough to detect very distant objects
  Telescope must be cryogenically cooled and operate at wavelengths up to 500µm
- Cover enough sky to enable statistically significant discoveries
  All sky survey a necessity
- Integrate until images are confused
  Very high detector sensitivity is required to complete survey in finite lifetime

This observatory, called **SIRCE** (Survey of Infrared Cosmic Evolution), is shown in a cutaway view in Figure 5.
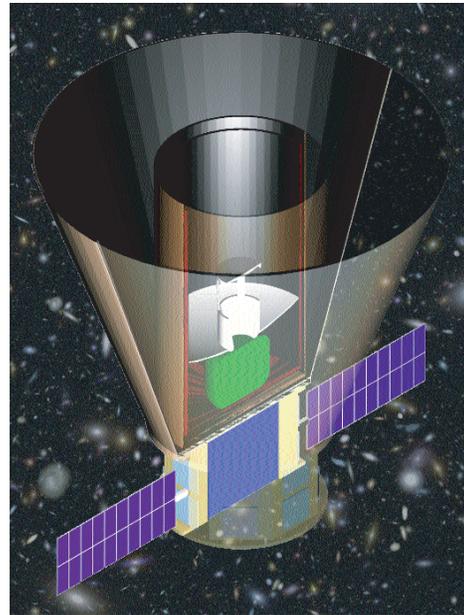
Figure 5: Cutaway view of a cryogenically cooled telescope for imaging the far-infrared sky to the confusion limit.

Why be confusion limited? For the extraction of point sources, confusion-limited images have the faintest detection limits available for a given telescope size. Furthermore, since the statistics of confusion-limited images are not noise-dominated, they are useful for constraining the distribution of sources below robust detection limits. The confusion limit is shown below in Figure 6; a representative confused image is shown below in Figure 7.
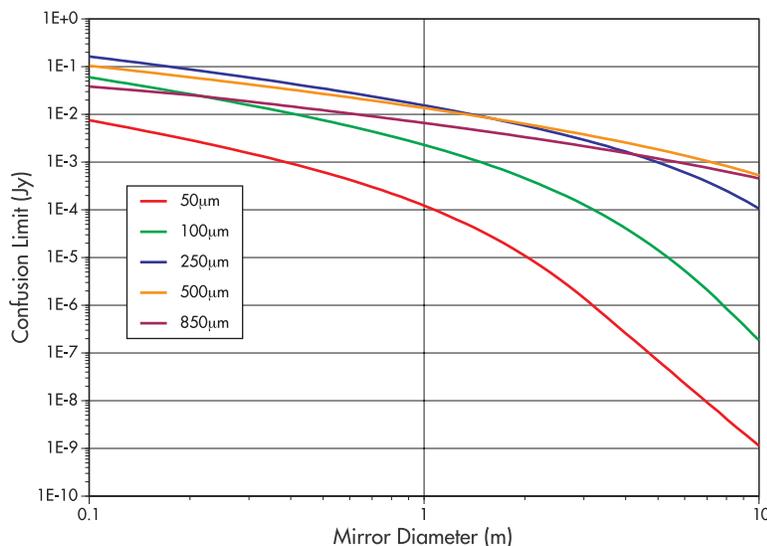
Figure 6: Confusion limit vs. mirror diameter for relevant wavelengths (from Blain 1999).
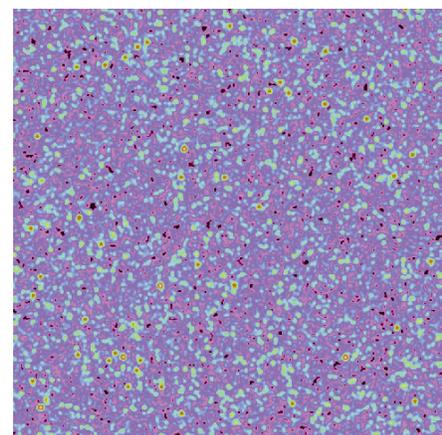
Figure 7: Simulated confused image of a 1° square region of the sky. Every source is a galaxy; very little dark sky remains.

Several questions need to be answered in the development of a far-infrared survey mission:

- How big does the telescope have to be to detect sources at great distances?
- What wavelengths must be covered?
- How cold must the telescope be?
- How sensitive do the detectors have to be?

TELESCOPE SIZE:

At left (Figure 8) is a comparison of the Galactic vs. extragalactic confusion; in order to determine the distribution of galaxies, a 1–2 m mirror will minimize Galactic confusion. At right (Figure 9) is an estimate of the differential source counts for a 2m telescope. Such a telescope can find tens of thousands of z>7 galaxies, measuring star formation activity back to an era unreachable by existing telescopes.
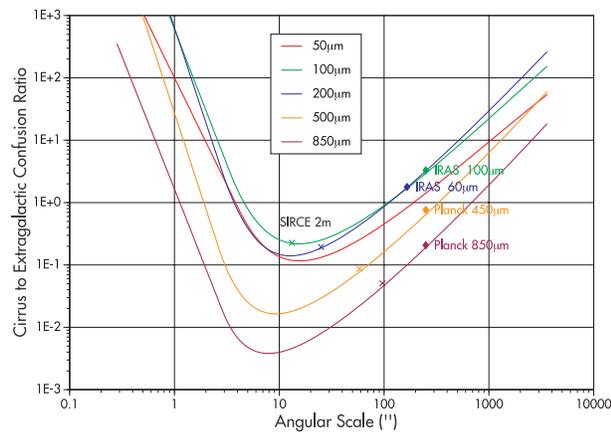
Figure 8: Galactic-to-extragalactic confusion limit ratio.
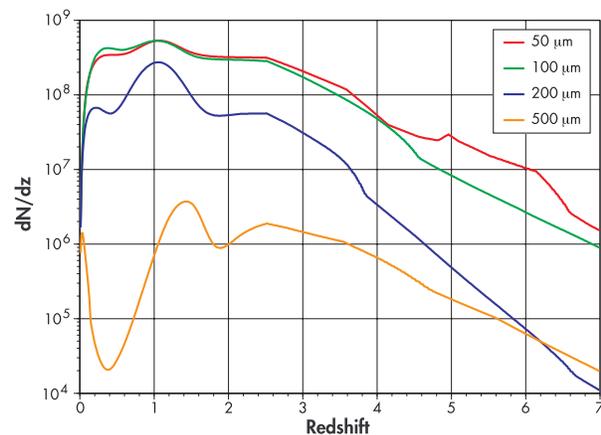
Figure 9: Differential source counts vs. redshift and wavelength for a 2m telescope (from Pearson 2001).

WAVELENGTH RANGE:

In order to determine photometric redshift accurately, measuring fluxes at wavelengths around a feature such as the dust emission peak at ~80-100μm is required. The short wavelength requirement is that the short wavelength side of this peak be seen in nearby galaxies; therefore ~50μm is a good choice. The long wavelength cutoff of the observatory should permit the peak to be seen in z~5 galaxies of the higher luminosities (those with ~80μm rest frame peaks). This sets the wavelength at about 500μm. Clearly, a longer wavelength would be beneficial, but the difficulty of overcoming confusion requires ever larger apertures. Fortunately, very high redshift (z>5) objects can be selected using a 500μm dropout technique.

DETECTOR SENSITIVITY:

Large format detector arrays with extremely high sensitivity (i.e., bolometers with NEP$\approx$6·10^{-19} W/$\sqrt{}$Hz) can integrate down to the confusion limit as they scan at the orbital rate for

a survey with a helium-cooled telescope; therefore the telescope sees everything it can see as fast as it can observe!

TELESCOPE TEMPERATURE:
Maximizing sensitivity implies that the dominant photon background must come from the distant sky, not from the telescope. Figure 10 shows the relative contributions of background power from the sky in space compared to the emission of a 5% emissivity telescope. The telescope must be cooled to around 4K to suppress self emission below the natural sky backgrounds.
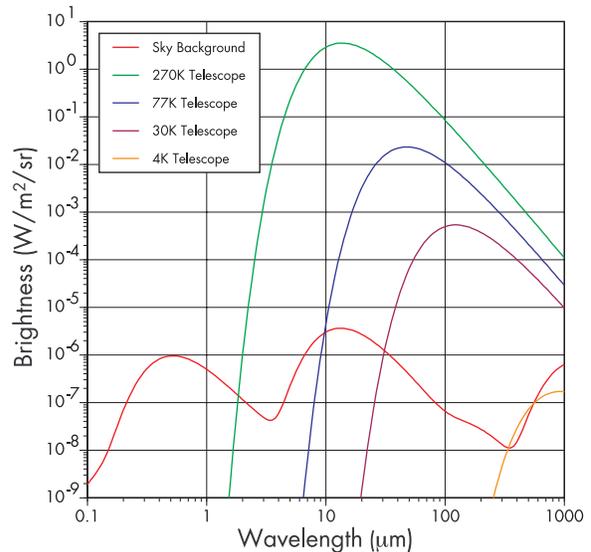
Figure 10. Photon background from the space environment compared with telescope emission.

CONCLUSION

We have developed a far-infrared all-sky survey mission concept with the goal of answering a specific set of important questions about the evolution of galaxies over cosmic time. A helium-cooled telescope with ultrasensitive detectors can image the whole sky to the confusion limit in 6 months. SIRCE has been developed by the Goddard Space Flight Center for launch within the decade. It is an ideal candidate for the proposed expanded Explorer mission line.

REFERENCES

Blain, A.W., 1999, *"Future millimeter, submillimeter and far-infrared surveys and their successful follow up"*, in ASP Conference.Series #191 `Photometric Redshifts and the detection of high-redshift galaxies', p. 255-264

Hauser, M.G. & Dwek, E. 2001, *"The Cosmic Infrared Background: Measurements and Implications"*, ARAA #39, p.249-307

Pearson, C.P. 2001, *"Evolutionary constraints from infrared source counts"*, MNRAS #325 (4), p.1511-1526

H2EX a space mission dedicated to the observation of molecular hydrogen

François Boulanger[1] and Edith Falgarone[2]

1: Institut d'Astrophysique Spatiale, Orsay, France,

2: Laboratoire de Radio-Astronomie, Ecole Normale Supérieure, Paris, France

Abstract

We present a proposal for a space mission, H2EX, that would be dedicated to the mapping of H_2 emission in its four lowest rotational transitions, at 28.2, 17.0, 12.3 and 9.7 micron with a spectral resolution $\sim 10^4$ sufficient to provide kinematical distances in galaxies. The proposed instrument will be extremely sensitive to H_2 with temperatures higher than 80 K. Colder H_2, which may contribute most of the baryonic dark matter in galaxies, will be traced by the emission of the warm H_2 heated, throughout the medium, by the dissipation of omnipresent turbulence. The main scientific objectives are to (i) directly measure the mass and temperature distribution of the warm H_2, far from star forming regions, in particularl in low metallicity environments where the traditional tracers of H_2 (CO, dust emission) fail, (ii) trace the dissipation of turbulence in the perspective of building a global view of the star formation process in galaxies and (iii) provide mass constraints on baryonic dark matter in the form of H_2 in a large sample of galaxies along the Hubble sequence.

1. Scientific Rationale for an H_2 Explorer

Unlike atomic hydrogen, which has been detected at large in the universe, the hydrogen molecule is very difficult to observe in emission because its rotational transitions in the mid-IR are blocked by the Earth atmosphere. Further, since the first rotational lines of para and ortho H_2 arise from levels with large energies ($E_{J=2} = 510$ K, $E_{J=3} = 1020$ K, see Fig. 1) collisional excitation of H_2 is effective only for warm molecular gas.

ISO-SWS and FUSE have opened a new perspective on the interstellar medium by detecting large amounts of warm H_2 (i.e. in rotational levels $J \geq 2$, hereafter H_2^*) far from star forming regions. These results suggest the existence of previously unseen components of warm ($T_k \sim 10^3$K) and lukewarm ($T_k \sim$ 100-200 K) H_2 distributed throughout the cold interstellar medium for which UV photons are not the sole heating source; in low metallicity environments, H_2 is expected to be lukewarm because of the drop of the C^+ cooling. These components could thus be a novel tracer of the cold H_2 (i.e. in rotational levels $J < 2$), traced so far by the CO lines with large uncertainties

and exclusively in metal-rich environments. The H2EX experiment is designed to fully exploit this new perspective on the interstellar medium.

The only heating source for the warm H_2 observed far from star forming regions is the dissipation of the ubiquitous turbulence fed by differential rotation (galactic shear). Dissipation is the process which drives the progressive infall of gas toward the center of galaxies and star formation at all galactocentric distances. Mapping warm H_2 in galaxies have therefore bearings on star formation at large scale and galaxy evolution, on star formation in low metallicity environments and eventually in the early universe, and last on the amount of cold dark matter in the form of H_2.

H2EX will detect H_2 warmer than $T_k \sim 80$ K. Observation of the four lowest rotational transitions is mandatory to disentangle the warm and lukewarm components and determine the mass of H_2^*. The central goal of the mission will be to determine a conversion factor H_2^*/H_2, and thereby to provide a tracer of H_2 molecular gas in galaxies independent of metallicity.

A survey will be conducted in dedicated regions of the Milky Way and in galaxies along the Hubble sequence, including spirals with different (rising, flat or falling) rotation curves and thus different shears and need for dark matter. The measurement of the power radiated by H_2 will provide the first census of the sinks of the turbulent energy, widely accepted to be the stabilizing agent against collapse, at all scales. Together with accurate mass determinations of molecular gas, in regions where other tracers (CO lines, dust continuum) are unavailable, this will provide the first global view of the processes which control star formation a critical issue in the evolution of galaxies. In the Milky Way, H2EX will greatly extend the range of environments where the emission from H_2 is detectable and will probe physical processes such as H_2 formation rate, excitation following formation, collisional excitation, ortho-to-para conversion and gas heating.

In the following sections we detail the scientific rationale of H2EX and present an outline of a preliminary mission concept as it has been recently submitted to the french Space Agency, CNES, in response to a call for ideas for future space missions.

2. Scientific Context

2.1 H_2 Tracers

UV Observations

In the 1970's, the Copernicus satellite observed molecular hydrogen electronic lines in absorption towards UV bright sources, essentially stars (Spitzer et al. 1974). This absorption experiment yielded column densities for diffuse clouds ($A_V < 1$) along a large number of lines of sight. With its greater sentivity FUSE (the Far Ultraviolet Spectroscopic Explorer satellite) extends the work done with Copernicus on diffuse clouds to translucent clouds with visual extinction up to about 5 mag, or H nuclei column densities of $\leq 10^{22}$ cm^{-2}.

The absorption measurements provided the first measurements of the temperature of

H_2 in diffuse clouds, $\sim 80K$ on average, as well as an estimate of the formation rate of H_2 on interstellar dust grains (Jura 1974). Unfortunately, UV observations only allow the study of molecular gas in the Solar Neighborhood and provide no information on the spatial distribution of H_2. Moreover, UV absorption lines are most often optically thick. The interpretation of UV data thus suffers from large error bars related to opacity corrections.

CO emitting gas

In the absence of direct measurements of H_2, many studies focussed on the CO molecule. Due to its dipole moment, CO emits line emission easily detectable from the ground. The CO molecule is generally used as a tracer of H_2 in spite of a number of highly uncertain assumptions. The abundance of CO with respect to H_2 is small (of the order of 10^{-4}) and depends critically on metallicity and radiation field. In addition, CO is invariably optically thick. The CO to H_2 conversion factor, X_{CO}, has been estimated through the comparison with other mass tracers and using the virial theorem. Models and observations show X_{CO} to be dependent on the metallicity and radiation field. Clearly X_{CO}, which is a product of the local star formation history, will be different for different types of galaxies and is expected to be rather different in gealactic nuclei and the outer regions of galaxies. This leads to large uncertainties of the H_2 mass estimates in spiral galaxies, ranging from about 20% to about 10 times the mass of H I. This last figure is actually approaching the amount of dark matter required to explain the flat rotation curves of galaxies. The overall sub-mm Galaxy spectrum obtained with the COBE far-infrared absolute spectrophotometer (FIRAS) shows, from the CO excitation, that the bulk of the CO emitting gas in the Galaxy is cold ($T < 20$ K) (Fixsen et al. 1999).

Near-infrared observations

Up until the advent of ISO, the only routine detections of H_2 were made in the near-IR region towards galactic shocks and photo-dissociation regions (PDRs), and in some instances the nuclei of active galaxies. Such detections were of ro-vibrational transitions, most notably the v=1–0 J=3–1 (or 1–0 S(1)) line at 2.12 μm. In PDRs, these lines result from the fluorescence cascade following pumping of electronic transitions in the far-UV. In sufficiently powerful shocks, a narrow gas layer becomes warm enough to have vibrationally excited levels populated by collisions. Therefore, in both PDRs and shocks, only a small fraction of the H_2 is vibrationally excited, and the near-IR lines do not give information on the temperature or mass of the bulk of the gas. A further disadvantage of the near-IR observations is the appreciable extinction by dust, more than 10 times larger than at 17.0 and 28.2 μm.

Mid-IR lines: ISO

The most stringent tests so far of our current understanding of the physics of H_2 in space, were provided by the Infrared Space Observatory (ISO) from observations of the lowest ro-vibrational and pure rotational lines, the latter providing a thermometer for the bulk of the gas above about 80 K (Figure 1).

Several ISO-SWS results on the rotational lines of H_2 do challenge the traditional view of the interstellar medium (ISM). In several environments, the intensities of the S(0) and S(1) lines are greater than predicted by models. For instance, in a PDR illuminated

by a late type star, models fall short of the observed fluxes (Thi et al. 1999). In the edge-on galaxy NGC 891, the same lines observed far from the centre of the galaxy are only marginally explained in terms of extended diffuse PDRs in cloud envelopes, under the assumption of an ortho/para ratio of unity (Valentijn & van der Werf 1999). Observations of the five lowest rotational transitions of H_2 along a line of sight close to the galactic plane, carefully avoiding star forming regions (Verstraete et al. 1999; Falgarone et al. 1999) show that the intensities of the S(2) and S(3) lines are about one order of magnitude above those predicted by PDR models of the medium sampled (Figure 2) which involve gas at a few 100 K. These findings may be put in perspective with the results of Copernicus, and FUSE which show that the temperatures derived from the lowest levels could not account for the excitation of the higher levels (Spitzer & Jenkins 1975). These results suggest the existence of a previously unseen component of warm H_2 ($T_k > 100$ K) distributed throughout the cold interstellar medium and for which UV photons cannot be the sole heating source. An alternative excitation process is collisional excitation in small regions transiently heated to high temperatures by the deposition of mechanical energy in shocks or in coherent vortices, responsible for the viscous dissipation of turbulence (Falgarone et al. 2001). Interestingly, observations of H_2 absorption lines by FUSE, in the direction of late B stars probe a similar fraction of warm H_2 embedded in the cold diffuse interstellar medium, as did ISO-SWS (Gry et al. 2001).

2.2 Open issues

Turbulence dissipation in the interstellar medium

The dissipation of the kinetic energy of turbulent motions is a source of heating and excitation of H_2 relevant to all the components of the interstellar medium which, unlike radiative excitation is not necessarily confined to the vicinity of massive stars. The sources of turbulent energy are stellar (supernova explosions, HII regions, molecular
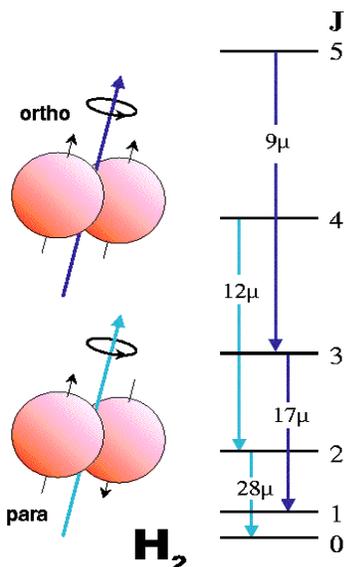
Figure 1: *H2EX will measure the four lowest rotational lines of the hydrogen molecule (H_2) in the Universe. The 12.3 and 28.2 μm lines probe the para (anti-parallel spins) and the 9.7 and 17.0 μm lines probe the ortho (parallel spins) form of H_2. They originate from levels only a few hundred to a few thousand Kelvin above ground and thus can be emitted from the bulk of the warm molecular gas ($T \geq 50\,K$ up to $1000\,K$). The lines are exceedingly difficult, if not impossible, to observe from ground and airborne observatories due to a combination of large thermal backgrounds, poor atmospheric transmission, and of low decay rates (longer than 100 years).*
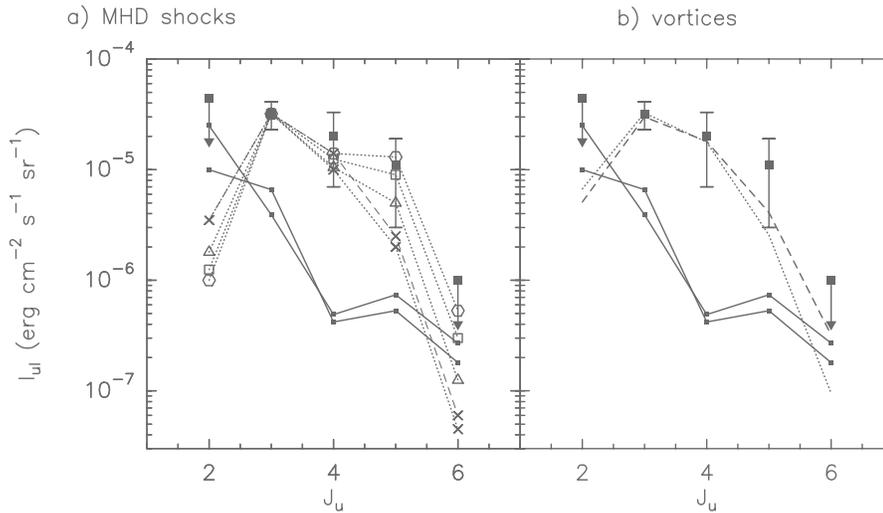
Figure 2: *Model predictions compared with the ISO-SWS observed line intensities (solid squares) along a line of sight across the galactic plane, avoiding star forming regions:* (a) H_2 *line emission of magneto-hydrodynamical (MHD) shocks with velocities of 8 km s^{-1} (crosses), 10 km s^{-1} (triangles), 12 km s^{-1} (empty squares) and of 14 km s^{-1} (hexagons) travelling in a preshock gas of density $n_H = 50$ cm^{-3}, magnetic field intensity $B_0 = 7\mu G$ and shielding $A_v = 0.1$ mag. The small influence of the density is illustrated by the dashed curve for which $n_H = 100$ cm^{-3}.* (b) H_2 *line emission of coherent vortices for $n_H = 30$ cm^{-3}, $B = 10\mu G$, $A_v = 0.1$ mag, $v_D = 3.5$ km s^{-1}. The dotted line corresponds to the case where the populations of the H_2 rotational levels have reached statistical equilibrium, while for the dashed line this assumption is relaxed. The number of shocks and vortices on the line of sight are adjusted to the observed S(1) line intensity. In each panel, the solid lines show two different PDR models of the 20 magnitudes of gas sampled by the line of sight. PDR models fail by more than one order of magnitude to reproduce the observed S(2) and S(3) intensities.*

outflows) but also the galactic shear associated with differential rotation.

The dissipation of supersonic turbulence occurs in bursts, either in shocks (Flower & Pineau des Forêts 1998) or in the regions of large shear at the boundary of coherent vortices (Joulain et al. 1998). Both processes generate localized and transient warm regions in the cold medium. These hot spots have so far escaped direct detection but observations and modeling severely constrain their phase-space structure i.e. they must have a small volume filling factor (a few %), and average surface filling factors close to unity, with large fluctuations about average. Indirect evidence includes the large abundances of typical shock produced molecular species (e.g. CH^+, HCO^+) in the cold diffuse medium, the formation of which require much more energy than is available in the bulk of the ISM volume and the pioneering ISO-SWS and FUSE observations mentioned

above.

Turbulence dissipation is a major step in the star formation process itself since non-thermal energy is recognized as the main support of dense gas against gravitational collapse. It is also a major step in galactic evolution since this dissipation contributes to the progressive infall of the interstellar gas toward the nuclei of galaxies. Yet, turbulence dissipation has always eluded direct detection. The pure rotational lines of H_2 are the major coolant of the regions heated by dissipation of turbulence. It is noteworthy that the power per unit surface in the Galaxy, $\frac{1}{2}\Sigma v_{rot}^2/\tau$, provided by the dissipation of the kinetic energy available in the galactic rotation (where $\Sigma = 240$ M$_\odot$/pc^2 is the characteristic surface density of giant molecular clouds and v_{rot}=250 km s^{-1}) over $\tau \sim 3$ Gyr (the time scale for gas depletion through star formation, Kennicut, 1983, Kennicut et al. 1994) corresponds to an intensity $I = 1.5 \times 10^{-5}$ erg s^{-1} cm^{-2} sr^{-1}. This value is of the order of what has been detected at large scale in the H_2 lines in the Milky Way and in NGC 891 Observations of the H_2 lines will allow a direct measurement of the dissipation rate of turbulent energy in galaxies. Interestingly, the spatial distribution of the energy sources being known (supernova remnants (SNR), HII regions, differential rotation), tracing the H_2 line emission across galaxies should determine the contribution of each kind of source as a function of galactocentric distance and provide estimates of how far from injection, turbulence propagates and dissipation takes place.

Hidden molecular gas in galaxies

Our increasing knowledge about the evolution of galaxies along the Hubble sequence suggests that some of the baryonic dark matter might be hidden molecular gas, concentrated in the outer part of galactic discs. The evolution of galaxies along the Hubble sequence from late to early types is thought to result from the dynamical action of spiral waves and bars leading to a progressive concentration of their mass towards their nuclei (Pfenniger et al. 94). Besides, H I rotation curves show that the fraction of dark matter in discs is decreasing along this sequence. This seems possible only if some of the baryonic dark matter is interstellar gas which can be transformed into stars. An additional argument for a gas contribution to dark matter comes from the fact that the surface density of dark matter, deduced from the rotation curves, follows that of the atomic gas. More precisely, the rotation curves can be reproduced with the visible matter (stars and gas), by multiplying the H I mass by a single factor between 7 and 10 (Hoekstra et al. 2001). This factor is constant with radius in a given galaxy and varies slightly from galaxy to galaxy, being larger in early types.

Star formation thresholds in the external parts of galaxies

Star formation is found to plummet abruptly beyond the optical disk of many galaxies. In most of the spirals, the transition occurs when the the gas surface density Σ_{gas} drops below the critical surface density Σ_{crit}, (the Toomre criterion), which involves the epicyclic frequency and a more local quantity, the gas velocity dispersion. Yet, large star forming regions exist in the outer disks of spirals in spite of subcritical gas surface densities. Other factors than the ratio $\Sigma_{gas}/\Sigma_{crit}$ may define the threshold for the growth of gravitational instability. In particular, the link (or the absence of link) between gravitationally bound clouds and molecular clouds should be explored in environments with very different pressure and metallicity from the solar neighborhood where the thresholds

for the formation of bound, and molecular, clouds are similar. Although the mass of the cold H_2 will have to be inferred through the tracer H_2^*/H_2, the low metallicity and low pressure environments (outer regions of disks of spirals and dwarf galaxies) can be explored through direct detection of H_2 only.

3. Payload and Mission Concept

The science goals lead to the following mission and instrumental requirements for the H2EX projects.

- A telescope system passively cooled to $\leq 50\,\mathrm{K}$, to suppress thermal system noise.

- sensitive detectors not affected by cosmic radiation operating at a temperature of $6\,\mathrm{K}$

- a primary mirror with a diameter $\sim 1m$ to get the angular resolution ($\sim 10''$) necessary to well resolve local Universe galaxies.

- an imaging spectrometer without moving parts and with a spectral resolving power $\sim 10^4$. An instrument with a long input slit, fixed gratings and an infrared detector array for each spectral line is foreseen.
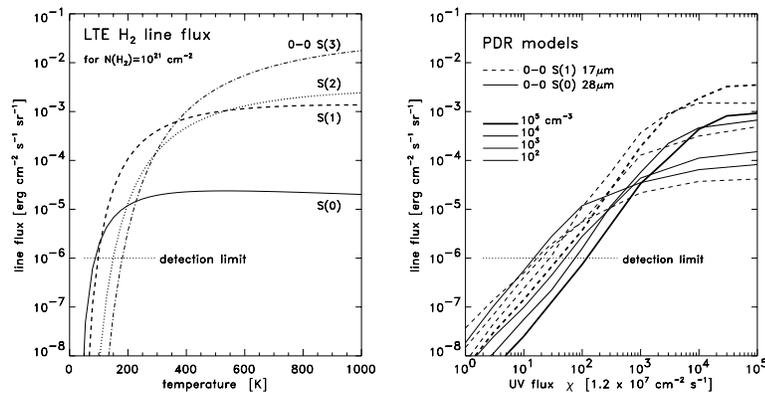
Figure 3: **Left** *Predicted intensities of the four lowest pure rotational H_2 lines (0–0 S(0), S(1), S(2) and S(3) at 28.2, 17.0, 12.3 and 9.66 µm respectively) for a layer of gas with an H_2 column density of 10^{21} cm^{-2} in local thermodynamic equilibrium (LTE). Such a column density is typical of small galactic clouds. The required detection limit of H2EX for an exposure time of 100 s is indicated.* **Right** *Model predictions of the 0-0 S(0) and S(1) line intensities observed face-on toward a PDR, as a function of the UV intensity incident on the PDR, for several values of the total H nuclei density. H2EX will detect all low density PDRs in regions where the UV flux is ten times the average diffuse background of the solar neighborhood, and in even lower UV environments with limb brightening.*

The technologies necessary to H2EX have already been developed for space-borne applications. The science requirements described above are technically feasible due to recent developments in thermal design, optical systems, and infrared detector technology. In particular, this project will benefit from the mid-IR detector developments carried out for NGST. For passive cooling, the initial plan was to launch H2EX to L2 as Planck and Herschel. But recent modeling carried out by JPL indicates that currently available passive cooling techniques can achieve H2EX telescope temperatures in the range of 45K to 55K in a low earth polar orbit similar to those used by the IRAS and COBE infrared observatories. For the cooling of the detectors, Planck H and He-sorption coolers should be considered but the option of a cryostat should also be studied.

A preliminary version of the H2EX proposal has been presented to the french Space Agency CNES in response to a recent survey for future projects. Our plan is to get a more formal proposal ready by the end of 2002. H2EX includes a collaboration with a team of JPL scientists.

Table 1: *Assumed combined properties of the H2EX spectrograph using detectors based on the present SIRTF detectors, the resulting sensitivity limits are listed in Table 2. Larger arrays are being developed for MIRI on NGST.*

parameter	baseline mission
size arrays in pixels	256×256
detector operating temperature	$6\,\mathrm{K}$
read noise	$20\ \mathrm{e}^-$
dark current	$10\ \mathrm{e}^-$
projected size of pixel	$2.5'' \times 2.5''$
system total throughput	0.2

Table 2: *Grating characteristics and sensitivity of the H2EX spectrometer. The sensitivity is for a 5σ detection in a $5'' \times 5''$ resolution element of 2×2 pixels assuming 100 s integration and $R = 10^4$*

Transition	Wavelength (μm)	Blaze angle degrees	lines/mm	Resolution	Sensitivity $\mathrm{erg\,cm^{-2}s^{-1}sr^{-1}}$
S(0)	28.2	44.83	50	9600	$4.1 \cdot 10^{-7}$
S(1)	17.0	42.84	80	10200	$3.3 \cdot 10^{-7}$
S(2)	12.3	40.22	105	10900	$2.5 \cdot 10^{-7}$
S(3)	9.7	40.90	135	10400	$1.5 \cdot 10^{-8}$

4. H2EX Surveys

The observational concept is that H2EX will carry out dedicated sky surveys, to address the main science goals of the mission. With an integration time of 100 s per sky position

and rastering with steps of $5''$ (for a 2.5" pixel scale the slit length will range from 10 to 42' for 256 to 1024 long arrays), hundreds of square degrees can be surveyed within a 2 years mission. The sensitivity achieved in this survey (see Table 2) would enable mapping of lukewarm gas with temperature $T_k \geq 80\,\mathrm{K}$ and column densities $N(\mathrm{H_2}) > 10^{21}\ \mathrm{cm}^{-2}$. This survey will be sensitive to column densities of warm gas ($T_k \sim 10^3\,\mathrm{K}$) as small as $N(\mathrm{H_2}) \sim 10^{18}$ cm^{-2} (see Figure 3). In addition, deep integrations in some dedicated fields, will provide maps of unprecedented sensitivity, with $3\,\sigma$ detection limits down to $10^{-7}\mathrm{erg\,s^{-1}cm^{-2}sr^{-1}}$. The spectral resolution allows us to resolve large scale motions in galaxies (~ 30 km s^{-1}). We consider that this is essential for a comparison with CO and H I observations.

The main observing program comprise (i) a Solar Neighborhood interstellar medium and star forming regions survey dedicated to study of the physics of $\mathrm{H_2}$ in space over the widest range of environments, (ii) a Milky Way survey sampling the star forming molecular ring and quiescent molecular gas in the outer Galaxy and (iii) a survey of large number of local Universe galaxies (~ 200) sampling the Hubble sequence and including the Magellanic Clouds and the largest nearby galaxies (e.g. M31, M33, ...).

These observations will allow to address the following questions:

- Physical processes relevant to $\mathrm{H_2}$: formation, excitation, ortho to para conversion, thermal budget in the ISM ...

- The radiative and dynamic interaction of stars with their ambient medium

- What is the distribution and spatial structure of the warm and lukewarm $\mathrm{H_2}$ components? How closely related are they to star forming regions?

- Does the warm $\mathrm{H_2}$ follow the cold material traced by CO? By H I? How does this depend on metallicity?

- What are the heating sources of $\mathrm{H_2}$ far from star forming regions?

- How far out of the optical disc are the $\mathrm{H_2}$ lines detectable?

- How good are the rotational $\mathrm{H_2}$ lines as a tracer of the bulk of cold $\mathrm{H_2}$ and can we determine an $\mathrm{H_2}^*/\mathrm{H_2}$ conversion factor?

- How does the warm/lukewarm $\mathrm{H_2}$ content in a galaxy correlate with galactic type, the shape of the rotation curve, the galactic shear, the metallicity?

- What is the contribution of star forming regions to the excited $\mathrm{H_2}$ content (shocks from HII regions, SNRs,...)?

- What are the $\mathrm{H_2}$ properties in galaxies with active nuclei? With starbursts?

5. H2EX in the light of other missions

Several infrared observatories will become operational before or around the timeframe of H2EX, including SIRTF, ASTRO-F, NGST, and SOFIA. It is instructive to compare them all to H2EX.

H2EX is dedicted to surveying the 4 lowest pure rotational transitions of H_2 simultaneously over large regions of sky. The proposed observing programme for H2EX requires deep, high-resolution spectroscopic observations. The spectrograph concept with a long slit will permit to survey large areas efficiently. High resolution ($\sim 30 km s^{-1}$) is a main requirement to determine kinematical distances in galaxies and to detect narrow emission lines that would have a weak contrast with respect to the dust continuum otherwise.

None of the other observatories has this combination of capabilities. H_2 observations with the infrared spectrometer of SIRTF will be limited by the low $600\ km\,s^{-1}$ resolution of the IRS instrument. H2EX will benefit of new detector technology developed for the MIRI instrument of NGST but only from its slit length, H2EX will be more efficient at mapping H_2 emission than IRS on SIRTF by more than one order of magnitude.

The AIRES instrument on SOFIA has a similar resolution and sensitivity to our proposed instrument. It can, however, only observe the 17 and 28 μm lines and its slit length is approximately ten times less. While the EXES instrument has a mid-IR spectroscopic capability that can observe all four lines (not simultaneously), it is not as sensitive as AIRES and has an even smaller slit length. Given the limited number of hours per year for observing (approximately 1000 hours, or 40 days) from SOFIA and the need to use more time to observe the same sky area leads us to doubt if such a survey project would ever be considered.

NGST will have a mid-IR instrument (MIRI) with a spectro-imaging mode operating at the relevant wavelengths with a spectral resolution, $R = \lambda/\Delta\lambda$, a factor of a few smaller than H2EX. For H_2 observations H2EX and MIRI which will use the same detector technology are providing complementary observing capabilities. Thanks to its much larger mirror aperture will be much more sensitive for detecting point sources and imaging compact sources but H2EX will be as sensitive and more efficient at mapping emission extended on scales larger than the field of view MIRI $10 - 20''$.

In summary none of the above missions can duplicate the survey capabilities of H2EX. H2EX will complement these observatories by making large scale surveys from which interesting objects/areas can be selected for NGST and/or SOFIA follow up.

References

Bertoldi F. et al.: 1999 A&A 346, 267
Falgarone E., Verstraete L., Hily-Blant P., Pineau des Forêts G., 2000 in "Molecular Hydrogen in Space", eds. F. Combes and G. Pineau des Forêts.
Falgarone E., Verstraete L., Pineau des Forêts G., Hily-Blant P.: 2001 A&A submitted
Fixsen D.J., Bennett C.L., Mather J.C.: 1999 ApJ 526 207

Flower D., Pineau des Forêts G.: 1998, MNRAS 297, 1182

Gry C., Boulanger F., Nehmé C., Pineau des Forêts G., Habart E., Falgarone E.: 2001 A&A submitted

Hoekstra H., van Albada T.S., Sancisi R.:2001 MNRAS 323 453

Kennicut R.C.: 1983 ApJ 272, 54

Kennicut R.C., Tamblyn P., Congdon C.E.: 1994 ApJ 435, 22

Jura, M.: 1974, ApJ 191, 375

Nakagawa et al., 1998, SPIE Vol 3356

NGST-DSS-WHRP-004, DaimlerChrysler Aerospace, 1 Oct 1999

Pfenniger D., Combes F., Martinet L.: 1994, A&A 285, 79

Pfenniger D., Combes F.: 1994, A&A 285, 94

Spitzer, L., Cochran W.D., Hirshfeld A.: 1974, ApJS 28, 373

Spitzer, L., Jenkins E.B.: 1975 ARA&A 13, 133

Thi W.F. et al.: 1999, in "The Universe as seen by ISO", ESA SP-427, Eds. M.F. Kessler and P. Cox

Valentijn E. A., van der Werf, P.P., de Graauw, T., de Jong, T.: 1996, A&A 315, L145

Valentijn E.A., van der Werf, P.: 1999, ApJ 522, L29

Valentijn E.A., Thi, W.F.: 1998, "ISO SWS - Ultimate sensitivity", IDC report in http://www.iso.vilspa.esa.es/science/pub/1999/detpaper.ps

Verstraete, L., Falgarone, E., Pineau des Forêts, Flower, D., Puget, J.-L, 1999, in "The Universe as seen by ISO", ESA SP-427, Eds. M.F. Kessler and P. Cox, p779

Wright C.M., et al.: 1999 ApJ 515 L29

THE CASE FOR SPACE-BORNE FAR-INFRARED LINE SURVEYS

J. J. Bock[1], C. M. Bradford[2], M. Dragovan[1], L. Earle[3], J. Glenn[3], B. Naylor[2], H.T. Nguyen[1], and J. Zmuidzinas[2]

[1]Jet Propulsion Laboratory, Pasadena, CA, 91109
[2]California Institute of Technology, Pasadena, CA, 91125
[3]University of Colorado, Boulder, CO, 80309

James J. Bock
M/S 169-327
Jet Propulsion Laboratory
Pasadena, CA, 91109
Email: jjb@astro.caltech.edu, phone (818) 354-0715

ABSTRACT

The combination of sensitive direct detectors and a cooled aperture promises orders of magnitude improvement in the sensitivity and survey time for far-infrared and sub-millimeter spectroscopy compared to existing or planned capabilities. Continuing advances in direct detector technology enable spectroscopy that approaches the background limit available only from space at these wavelengths. Because the spectral confusion limit is significantly lower than the more familiar spatial confusion limit encountered in imaging applications, spectroscopy can be carried out to comparable depth with a significantly smaller aperture. We are developing a novel waveguide-coupled grating spectrometer that disperses radiation into a wide instantaneous bandwidth with moderate resolution (R ~ 1000) in a compact 2-dimensional format.

A line survey instrument coupled to a modest cooled single aperture provides an attractive scientific application for spectroscopy with direct detectors. Using a suite of waveguide spectrometers, we can obtain complete coverage over the entire far-infrared and sub-millimeter. This concept requires no moving parts to modulate the optical signal. Such an instrument would be able to conduct a far-infrared line survey 10^6 times faster than planned capabilities, assuming existing detector technology. However, if historical improvements in bolometer sensitivity continue, so that photon-limited sensitivity is obtained, the integration time can be further reduced by 2 to 4 orders of magnitude, depending on wavelength. The line flux sensitivity would be comparable to ALMA, but at shorter wavelengths and with the continuous coverage needed to extract line fluxes for sources at unknown redshifts. For example, this capability would break the current spectroscopic bottleneck in the study of far-infrared galaxies, the recently discovered, rapidly evolving objects abundant at cosmological distances. The role for line survey instrumentation will become acute as the detection rate of far-infrared galaxies dramatically increases with the next generation of space-borne and ground-based

bolometer cameras. A space-borne spectrometer can rapidly follow up Herschel/SIRTF far-infrared galaxies at wavelengths inaccessible from the ground, conduct deep line surveys, and even search for redshifted H_2 line emission from the first luminous objects.

INTRODUCTION

Direct detectors, unlike hetrodyne systems, are not limited by fundamental quantum noise and can achieve background-limited sensitivity even under the very low backgrounds encountered in spectroscopy. Background-limited direct detectors combined with a cooled aperture promise a significant improvement in our capability to detect far-infrared spectral lines from galaxies as cosmological distances. Far-infrared lines can provide the redshifts and reveal the energetics of distant far-infrared galaxies. Furthermore, line emission from molecular hydrogen may be detectable at high redshifts in primordial galaxies undergoing the first episode of star formation. Because these distant objects may be at undetermined redshifts, we propose that moderate resolution spectroscopy with direct detectors over a large instantaneous bandwidth is attractive for future space-borne instrumentation. We shall show that the line survey speed possible in space is significantly higher than currently planned capabilities.

Unlike photometry, which encounters a confusion limit that varies as a function of angular resolution on the sky, spectroscopy offers the ability to distinguish multiple objects in a single beam by virtue of spectral discrimination. Thus the spectral confusion limit, set by the density of spectral lines, is much lower than the more familiar photometric confusion limit. If sufficient sensitivity can be achieved, spectroscopic line surveys may even be deeper than a photometric survey for a given aperture diameter. Several authors have even promoted deep spectral line surveys in blank regions of sky (e.g. Blain *et al*. 2000), but current sensitivities are far too low to make such surveys practical.

ACHIEVABLE SENSITIVITY

Surveys for spectral lines are best carried out with a dispersive spectrometer, such as a grating. We are developing a compact 2-D waveguide spectrometer (WaFIRS) which operates in one polarization over a wide bandwidth (see C.M. Bradford *et al*. in these proceedings). We calculate the sensitivity achievable with a spectrometer with direct detectors as follows. We combine both detector and photon noise equivalent powers in quadrature,

$$NEP_{tot}^2 = NEP_{bol}^2 + 2hnQ.$$

The optical power absorber by the detector is

$$Q = \lambda I_\lambda \, R^{-1} \, \lambda^2 \, \eta_{opt} \, (N_{pol}/2),$$

where λI_λ is the specific intensity of the sky, including emission from the cosmic microwave background, interstellar dust, zodiacal dust, and thermal emission from the telescope. The total optical efficiency is η_{opt}, the number of polarizations detected is N_{pol}, the spectral resolution is $R = \lambda/\Delta\lambda$ and we assume single-mode throughput λ^2. For

WaFIRS we assume R = 1000, η_{opt} = 0.5 and N_{pol} = 1. The staring sensitivity to integrated line intensity in one Hz of audio bandwidth is thus

$$NEF = (A\ \eta_{opt}\ (N_{pol}/2))^{-1}\ NEP_{tot}\quad [W\ m^2\ Hz^{-1/2}],$$

where we have neglected the small factor of the choice of detector sampling of the spectrometer focal plane. The time to survey the available spectral bandwidth of an instrument to a given line sensitivity δF is thus

$$T = 0.5\ (NEF/\delta F)^2\ (\Delta\nu_{instr}/\Delta\nu_i),$$

where $\Delta\nu_{instr}$ is the available spectral bandwidth of the instrument, and $\Delta\nu_i$ is the instantaneous spectral bandwidth available at a given moment during the observation.

We compare the achievable sensitivity with a grating spectrometer on a cooled telescope with planned instruments, SIRTF/IRS (http://sirtf.caltech.edu/SSC/IRS), Herschel (http://astro.estec.esa.nl/herschel/key_pubs.html), and ALMA (http://www.alma.nrao.edu/info/sensitivities) in Fig. 1. For ALMA, we assume broad 300 km s^{-1} linewidths appropriate for galaxies, and 4 GHz of instantaneous bandwidth. With a 12 GHz instantaneous bandwidth, the ALMA survey times in Fig. 1 will be reduced by a factor of 3.

The purpose of this comparison is to motivate the need for wideband space-borne spectroscopy. We hasten to emphasize that the WaFIRS sensitivity only represents what is achievable, which is optimistic by definition, whereas the other instruments are designed with full noise budgets, and are generally in a high state of engineering development. In the case of IRS in particular, the full instrument has been assembled and tested and awaits launch. Furthermore WaFIRS requires large numbers of high sensitivity detectors, in quantities and background-limited sensitivities that exceed the current state of the art. The scientific role of a space-borne spectrometer with moderate spatial and spectral resolution, but with complete spectral coverage, is very complementary to ALMA with high spectral and spatial resolution, but with spectral coverage limited to atmospheric windows.

Nevertheless, the improvement in line survey speed in Fig. 1 is very large, more than 6 orders of magnitude for detectors with NEP = 1e-18 W/√Hz, and up to 9 orders of magnitude in the case of background-limited detectors. An additional 2 orders of magnitude of improvement would be obtained with a 10 m aperture such as the proposed SAFIR mission.

SCIENTIFIC APPLICATIONS

Our proposed technology enables moderate resolution $(R = \lambda/\Delta\lambda \sim 1000)$ spectroscopy of galaxies to determine redshifts and energetics. A distribution of redshifts determines the history of deeply embedded star formation and the earliest epoch of galaxy formation. Furthermore, with complete spectra, important questions about the nature of the galaxies can be resolved, such as the cooling rates, excitation conditions, and AGN fractions.
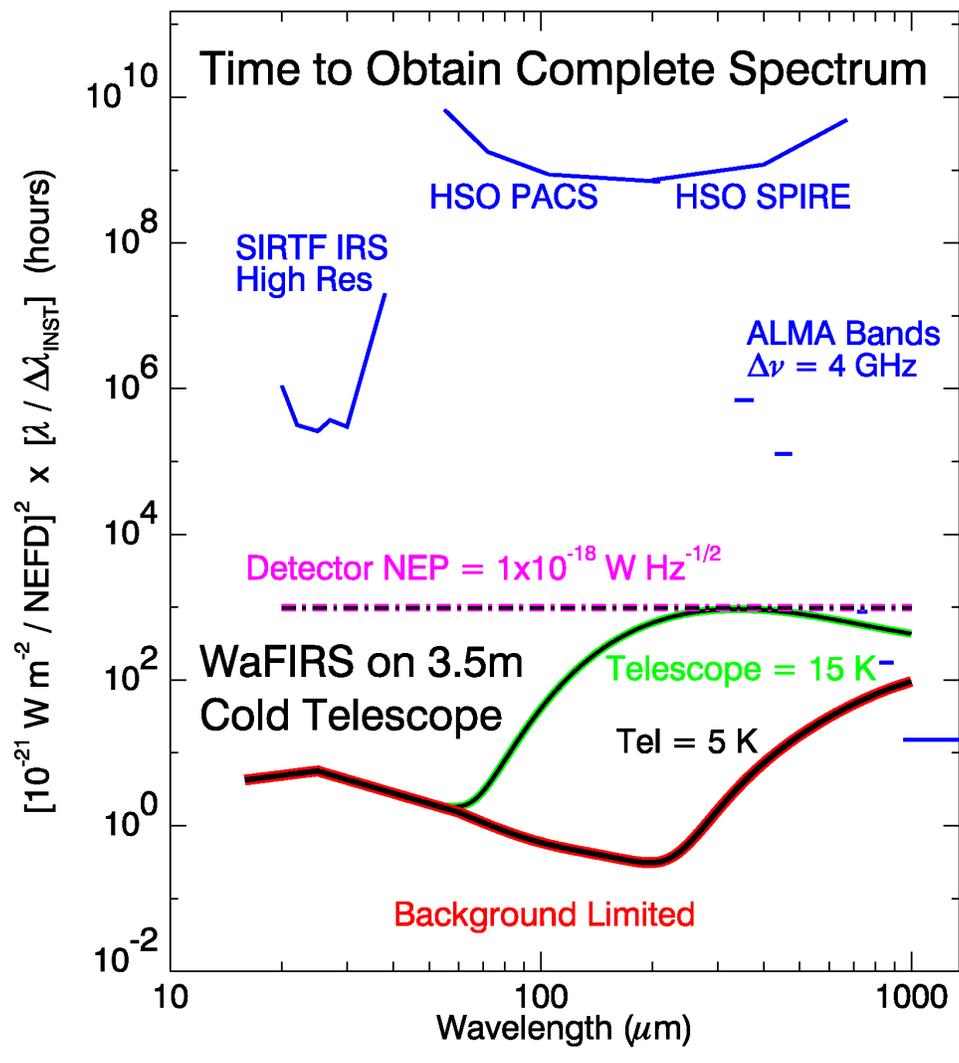
Figure 1: *Spectral spectral line survey speed plotted for several instruments. The figure of merit, [1e-21 W m^{-2}/NEFD(5σ)]^2 [λ/Δλ$_{inst}$], gives the time required to survey for a line in an octave of bandwidth, to a 5σ sensitivity of 1e-21 W/m^2. This is the appropriate figure of merit for detecting lines at unknown redshift. We estimate the observation time required from a cooled 3.5 m space-borne telescope (WaFIRS) with existing NEP = 1e-18 W/√Hz detectors (dashed line) and background-limited detectors (green and red curves) for different telescope temperatures. This projected observing time is compared with the times required for SIRTF/IRS, Herschel, and ALMA.*

Follow-up of SIRTF and Herschel Far-Infrared Galaxies

The redshift and far-infrared line strengths of virtually every Herschel/SIRTF galaxy should be detectable spectroscopically with WaFIRS on a 3.5 m telescope, as shown in Fig. 2. We estimate the minimum galaxy flux from the confusion limit of Herschel (Blain, Ivison, & Smail, 1998) and SIRTF/MIPS (Rieke *et al*, 1996). The lines in the figure are the expected [CII] 158 μm fluxes as a function of redshift for sources with fluxes corresponding the 5σ confusion noise level for each instrument. The FIR spectral energy distributions are assumed to be thermal dust emission greybodies (T_{dust} =

40 K, β =1.6, λ_0 = 25 μm). The line luminosities were derived by multiplying the bolometric FIR luminosities by $L_{[CII]}/L_{FIR}$ = 1 x 10^{-3} (e.g., Fisher 2000), where the observed FIR flux is equal to the 5σ confusion noise level. The bold line is the WaFIRS sensitivity assuming background-limited detectors. The predicted line flux initially drops with increasing redshift because the peak of the galaxy spectral energy distributions is being redshifted into the FIR bands. Therefore, fainter galaxies will be detected at the photometric confusion levels. The curves turn up steeply again beyond a redshift of a few because the luminosity of the observed galaxies at the confusion noise level is larger.
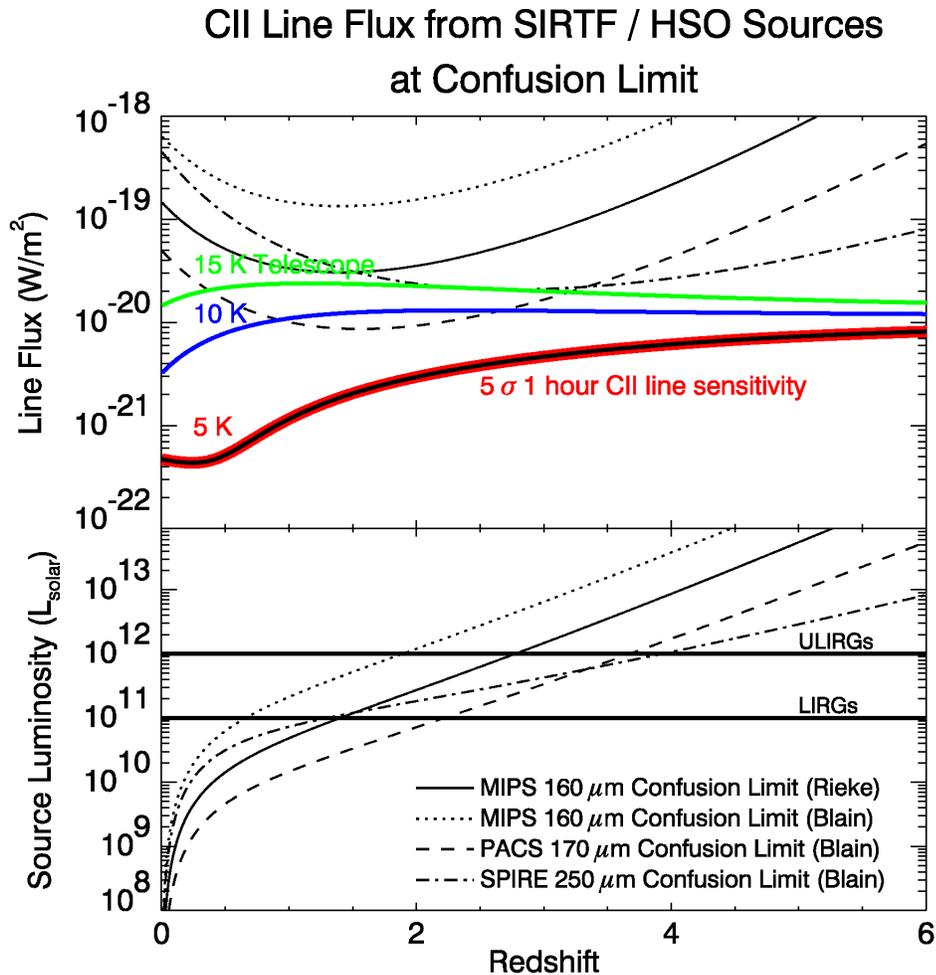
Figure 2: *[CII] line emission from any object detected photometrically by SIRTF or Herschel is readily detected by the space-borne spectrometer. We assume a line-to-continuum ratio L(line)/L(bol) = 1 x 10^{-3}.*

This analysis shows that for redshifts where the [CII] line is observable (z < 5), WaFIRS will be able to detect the [CII] line for all galaxies at the confusion noise levels of SIRTF and Herschel. For galaxies with fluxes at the 5σ confusion noise level, it will take a few minutes to detect the [CII] line with a signal-to-noise ratio of five at z = 2, and less time at the other redshifts. Galaxies at large distances, z > 5, must either be extremely luminous or lensed to be detected by HSO or SIRTF. They should be detectable in 1 hour of integration in faint atomic lines (e.g. NII, NIII, OI, OIII), with $L_{line}/L_{FIR} > 10^{-6}$. The redshift and far-infrared line strengths of virtually every Herschel or
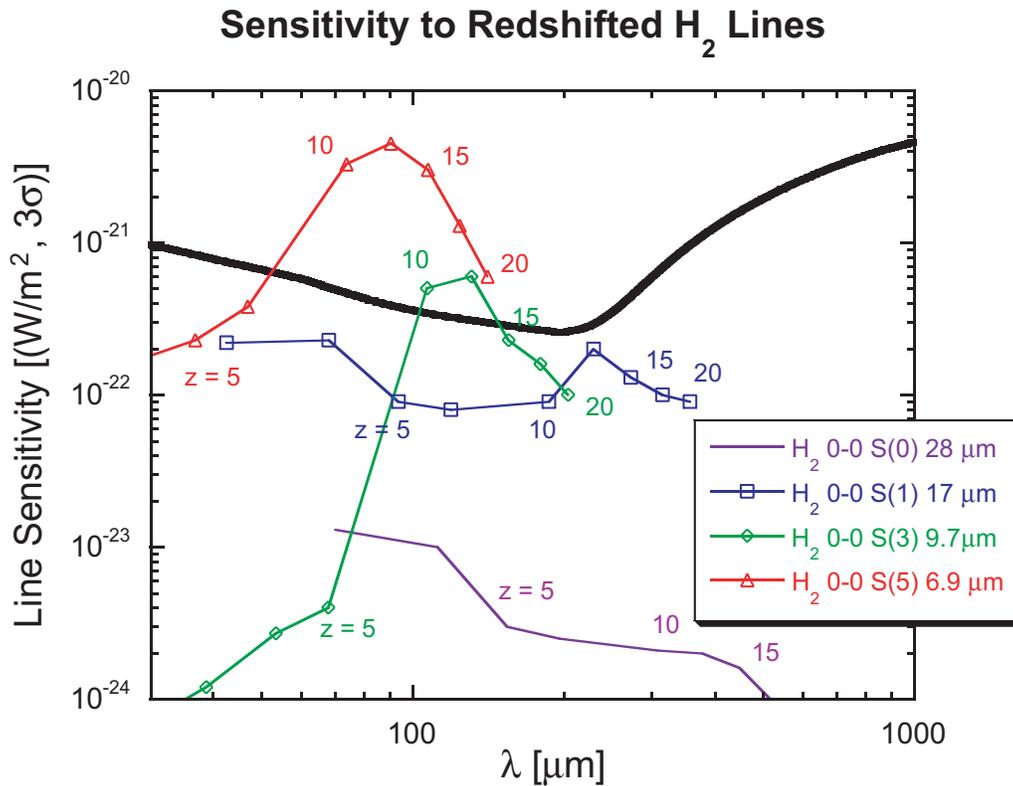
Figure 3: *Molecular hydrogen line emission driven by the first generation of star formation, calculated according to the model of Ciardi and Ferrara (2000). Line emission is detectable out to large redshift (z = 20) in moderate integration time.*

SIRTF galaxy should be measurable. The high line survey speed makes it possible to survey a large fraction of the ~10^5 galaxies expected from a key-program survey with Herschel.

Blind Spectral Line Surveys

The high sensitivity of WaFIRS allows for such a deep spectral line survey for the first time. Although the spectral confusion limit is highly uncertain, we estimate that ~10^2 hours are required to reach line sensitivities of 1e-22 W/m^2 (5σ), a rough estimate of the confusion limit (Blain *et al.* 2000). Such surveys could be combined with high-resolution images from ALMA to reveal the far-infrared line cooling of the faintest and most distant objects.

Molecular Hydrogen Line Emission from Primordial Galaxies

The first objects to form in the early universe gravitationally collapsed from over-dense regions. Energy from these structures was predominantly radiated by molecular hydrogen, because metals were not present to provide more efficient cooling. Molecular hydrogen thus determined the physical size of the earliest objects (M = $10^4 - 10^6$ solar masses), and their epoch of formation (z = 15 – 100) (Tegmark et al. 1997). The line emission from any one of these objects is faint, but it is possible that a large agglomeration of collapsing clouds might be detectable (Kamaya & Silk 2001).

Table 1: Measured Line Intensities for M82 and Arp 220

			M82		Arp 220	
L/L(solar)				3.30e10		1.10e12
Species	**Transition**	**L [um]**	**L(line)L(bol)**	**Ref**	**L(line)/L(bolo)**	**Ref**
C I	3P 1-0	609.1	1.60E-06	stu	2.30E-06	gp
C I	3P 2-1	370.4	7.20E-06	stu	1.20E-05	gp
N II	3P 1-0	205	8.80E-05	pet		
C II	2P 3/2-1/2	157.7	1.60E-03	col	1.26E-04	luh
O I	3P 0-1	145.5	1.50E-04	col	< 1.2E-05	jf
NII	3P 2-1	121.9	2.08E-04	col	< 3.6E-05	jf
O III	3P 1-0	88.4	1.05E-03	col	< 7.2E-05	jf
O I	3P 1-2	63.2	2.20E-03	col	-6.80E-05	jf
N III	2P 3/2-1/2	57.3	4.17E-04	col	< 2.4E-05	jf
O III	3P 2-1	51.8	1.26E-03	col	< 4.8E-05	jf
Si II	2P 3/2-1/2	34.8	1.12E-03	fs	7.73E-05	gen
S III	3P 1-0	33.5	8.40E-04	fs	7.30E-05	gen
O IV	2P 3/2-1/2	25.9			< 5.50E-6	gen
Ne II	2P 1/2-3/2	12.8	1.20E-03	fs	7.45E-05	gen
H2	0-0 S(0)	28.3	9.70E-06	fs		
	0-0 S(1)	17.0	1.80E-05	fs	3.30E-05	sturm
	0-0 S(2)	12.3	1.94E-05	fs		
	0-0 S(3)	9.7	1.20E-05	fs		
	0-0 S(4)	8.0	8.42E-06	fs		
	0-0 S(5)	6.9	1.38E-05	fs	3.30E-05	sturm
CO	1-0	2600				
	2-1	1299	7.00E-07	wild	1.20E-06	gp
	3-2	866	2.80E-06	wild	2.50E-06	gp
	4-3	649				
	5-4	520				
	6-5	433	1.00E-06	wild		
	7-6	371				
	8-7	324				
	9-8	289				

Notes: col = Colbert, J.W., et al. 1999, ApJ 511, 721.
fs = Foester-Screiber, N.M., et al 2001, ApJ 552, 544.
gen = Genzel, R. et al., 1998, ApJ 498, 579.
gp = Gerin, M. & Phillips, T.G., 1998, ApJ 509, L17.
jf = Fisher, J., private communication.
luh = Luhman et al., 1998, ApJ 504, L11.

pet = Petuchowski, S.J. et al., 1994, ApJ 427, 17.
sch = Schilke, J. et al., 1993, ApJ 417, 67.
stu = Stutzki, J. et al., 1997, ApJ 477, 33.
sturm = Sturm, E. et al., 1996, A&A 315, L133.
wild = Wild, W. et al., 1992, A&A 265, 447.

The onset of the first star formation heats the surrounding gas, eventually dissociating H_2 and ionizing the IGM by $z = 5$. During the first phase of star formation and their supernovae, H_2 line emission may be bright. For $5 < z < 20$, Ciardi & Ferrera (2001) calculate that the redshifted H_2 6.9 μm (rest wavelength) line should have an observed brightness 10^{-21} to 10^{-20} W m^{-2}, and the 9.7 μm and 17 μm lines should have an observed brightness $10^{-22} - 10^{-21}$ W m^{-2} (see Fig. 3).

Finally, H_2 lines may be detected in massive systems in the foreground of a FIR luminous distant galaxy in absorption. For example, using a 10 mJy background source, a column density of $N(H_2) = 10^{23}$ cm^{-2}, corresponding to a 10^{11} solar masses of gas in a 10 kpc diameter, H_2 may be detectable (Shibai et al. 2001).
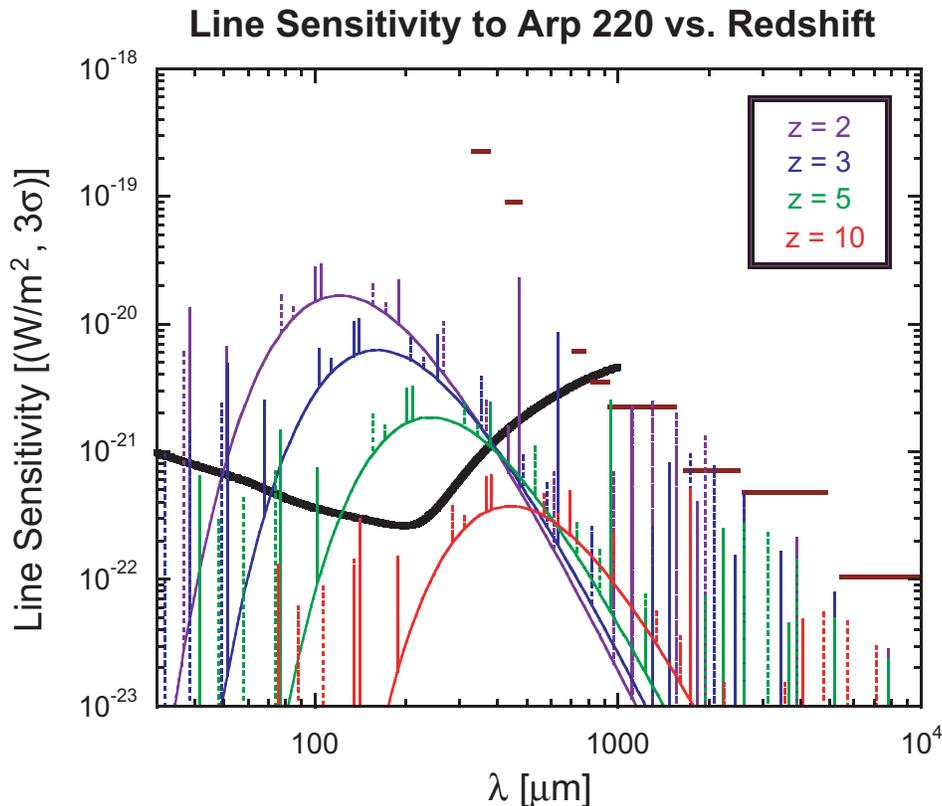
Figure 4: *The ULIRG (L = 1.1e12 L(solar)) Arp 220 is detectable to z = 10, based on measured (solid) and estimated (dashed) line strengths. We have reduced the continuum emission by 10,000 to emphasize the line intensities. We assume ALMA surveys the atmospheric windows for 300 km/s lines with an instantaneous bandwidth of 4 GHz. Arp 220 is notable for having extremely weak atomic lines, and represents a difficult object for line detection.*

Local Case-Studies: M82 and Arp 220

As examples, we take two nearby, well-studied far-infrared luminous galaxies, M82 and Arp 220, and place them at cosmological distance assuming a ΛCDM cosmology ($\Omega_\Lambda = 0.7$, $\Omega_m = 0.3$). In fact, both of these galaxies represent difficult objects for spectroscopy. M82 is a starburst galaxy with prominent atomic and molecular lines. By the standards of the luminous objects detected by SCUBA, M82 is not very luminous. Arp 220 is a ULIRG with very faint atomic line emission. It has the faintest measured $L_{[CII]}/L_{FIR}$ ratio of any ULIRG. The measured line strengths are listed in Table 1.

CONCLUSIONS

Space-borne far-infrared spectroscopy with direct detectors and a cooled aperture can improve on the sensitivities of planned space missions, but requires significant improvement in the sensitivity and format of direct detectors. Space-borne spectroscopic surveys are complementary to the ground-based ALMA interferometer.
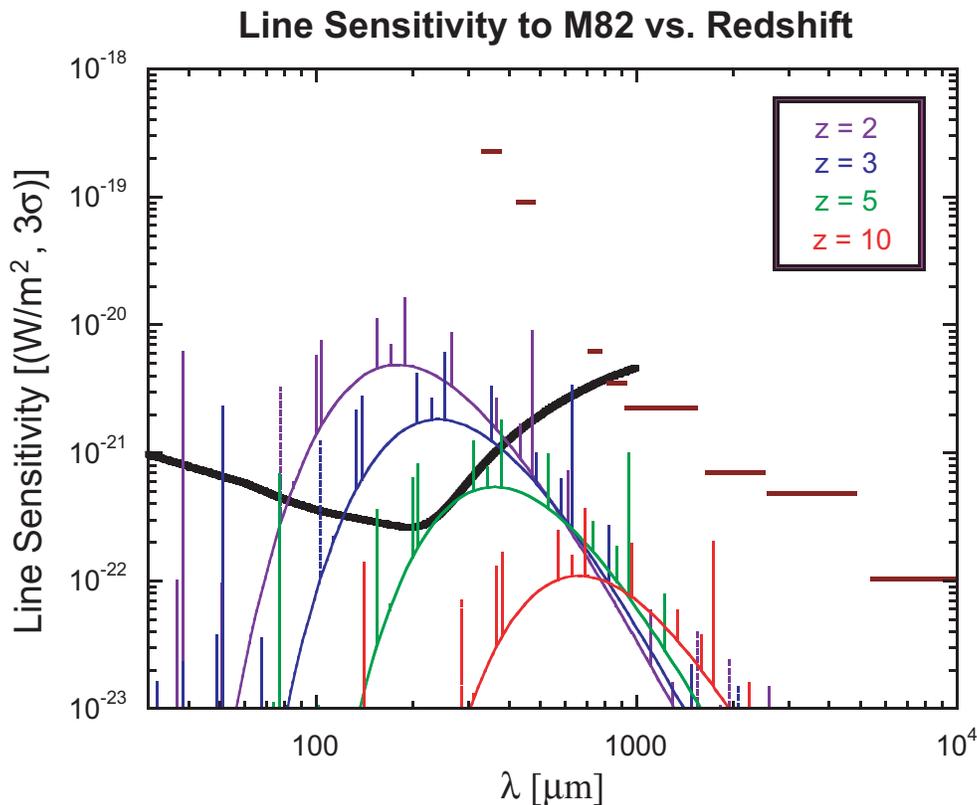
Figure 5: *The starburst galaxy M82 is detectable at z = 5, based on measured (solid) and estimated (dashed) line strengths. We have divided the continuum emission by 1000 to emphasize the line intensities. We assume ALMA surveys the atmospheric windows for 300 km/s lines with an instantaneous bandwidth of 4 GHz. M82 is not an extremely luminous (L = 3.3e10 L(solar)) far-infrared galaxy compared to objects currently being detected by SCUBA at cosmological distances.*

REFERENCES

Blain, A.W. *et al.*, 2000, MNRAS, 313, 559.
Blain, A.W., Ivison, R.J., & Smail, I., 1998, MNRAS, 296, 29.
Bradford, C.M. *et al.*, in these proceedings.
Fischer, J., 2000, astro-ph 0009395.
Kamaya, H. & Silk, J., 2002, MNRAS, 332, 251.
Rieke, G.H. *et al.*, 1996, AAS 189, 309.
Ciardi, B. & Ferrara, A. 2001, MNRAS, 324, 648.
Shibai, H. et al., 2001, PASJ 53, 589.
Tegmark, M. *et al.*, 1997, ApJ 474, 1.

An Infrared Telescope for Planet Detection and General Astrophysics

C. F. Lillie, C. B. Atkinson, L. S. Casement, M. R. Flannery, K. V. Kroening, S. L. Moses
TRW Space & Electronics

ABSTRACT

NASA plans to launch a Terrestrial Planet Finder (TPF) mission in 2014 to detect and characterize Earth-like planets around nearby stars, perform comparative planetology studies, and obtain general astrophysics observations. During our recently completed a TPF Mission Architecture study for NASA/JPL we developed the conceptual design for a 28-meter telescope with an IR Coronagraph that meets these mission objectives. This telescope and the technology it embodies are directly applicable to future Far-IR and Submillimeter space missions.

The detection of a 30th magnitude planet located within 50 milli-arcseconds of a 5th (Visual) magnitude star is an exceptionally challenging objective. Observations in the thermal infrared (7-17 μm) are somewhat easier since the planet is 'only' 15m fainter than the star at these wavelengths, but many severe challenges must still be overcome.

These challenges include:

- Designing a coronagraph for star:planet separations less than or equal to λ/D
- Developing the deployment scheme for a 28m space telescope that can fit in an existing launch vehicle payload fairing
- Generating configuration layouts for the IR telescope, coronagraph, spacecraft bus, sunshade, solar array, and high-gain antenna.
- Providing:
  - Structural stability to within 10 microns to support the optics
  - Thermal control to achieve the necessary structural stability, as well as providing a stable (~30K) thermal environment for the optics
  - Dynamics isolation from potential jitter sources
- Minimizing launch mass to provide the maximum payload for the science mission
- Interfacing to an EELV Heavy launch vehicle, including acoustic and stress loads for the launch environment
- Identifying the key technologies (which can be developed by 2009) that will enable TPF mission to be performed
- Generating a manufacturing plan that will permit TPF to be developed at a reasonable cost and schedule.

Many of these design challenges result in inherently conflicting requirements on the design of TPF. Drawing on our experience with large space telescopes such as the Chandra X-ray Observatory and the Next Generation Space Telescope, we have created a conceptual design for TPF that successfully meets these challenging requirements. This paper describes our solution to these challenges.

1.0 Introduction

This paper describes our conceptual design for a large aperture telescope with an IR Coronagraph that we developed during our mission architecture study for the Terrestrial Planet Finder (TPF) Mission. In addition to its capabilities for planet detection, characterization and comparative planetology, this observatory also has a significant capability for general astrophysics in the 3 to 28-micron spectral region. This architecture is readily scaleable to both larger and smaller apertures, and with minor modifications the present design could also be used at Far Infrared and Sub-millimeter wavelengths.

2.0 The Terrestrial Planet Finder Mission

The TPF Mission is a key element of NASA's Origins program. Now planned for launch in ~2014, TPF is designed to detect and characterize the properties of Earth-like planets in the habitable zones around solar type stars. It will also carry out a program of comparative planetology in a large number of solar systems, studying gas giant and terrestrial planet and debris disks. TPF observing time will also be used to collect important new data of general astrophysics interest.

During its five year mission TPF will search over >60% of the sky for solid bodies with ~ 1 Earth radius and a temperature of ~270K around >150 stars of spectral type F5 to K5. Spectra will be obtained for >5 of the detected objects in the 7-17 micron region, at a resolution of ~20, to look for features due to CO_2, H_2O, CH_4 and ozone that could indicate the presence of life. TPF's observing time will be split 50:50 between planet detection and characterization and general imaging and spectroscopy.

3.0 Mission Architectures

During the first phase of our study we examined several different architectures for the TPF mission. These architectures included (1) a >100-meter baseline IR nulling interferometer with a linear array of four 4-meter cryogenic telescopes; (2) a 30-meter cryogenic telescope with excellent mid-spatial frequency figure and a coronagraph with deformable optics; (3) a 30-meter Fresnel telescope with free flying spacecraft for the primary mirror and modules, separated by ~ 6 km; (4) a 100-meter sparse aperture IR telescope with ~100 randomly distributed 2 to 4-meter sub-apertures and a separate spacecraft with the correction optics/coronagraph/sensors located ~500 meters away; and (5) a 70-meter apodized occulter flown in formation with an ~8-meter diffraction-limited visible telescope ~100,000 km away.

From this effort we concluded that (1) the contrast ratio was too severe and the technology development was too challenging for direct detection of Earth-like planets with a visible interferometer; (2) that the contrast ratio in the visible was too severe for the sparse aperture telescope; and (3) that an IR occulter was not practical since the occulter must be very large and very distant from the "camera" telescope.

During the second phase of the study we elected to perform a more detailed study of the IR Coronagraph described herein, while the Ball Aerospace team studied visible

coronagraphs, the Lockheed Martin team studied IR nulling interferometers, and the Boeing-SVS studied a "hyper telescope" and an apodized square aperture telescope.

4.0 IR Coronagraph Design Concept

Our conceptual design for a large aperture telescope with an IR Coronagraph is shown in Figure 1. It draws heavily from our previous work for the Next Generation Space Telescope, with a large multi-layer sunshield that allows the segmented deployable telescope and science instrument module to be passively cooled to less than 30K.

The primary mirror consists of 36 hexagonal panels measuring ~4-meters flat-to-flat, arranged in 3 rings around a central opening. Each panel has a thin, gold-coated composite membrane mirror mounted attached to a composite backing structure by 6 rigid body actuators for tip-tilt-piston control, and 7 figure control actuators for control of low-order figure errors. The mirrors are produced with a low-cost replica optics process. The panels' areal density is ~5 kg/m^2.
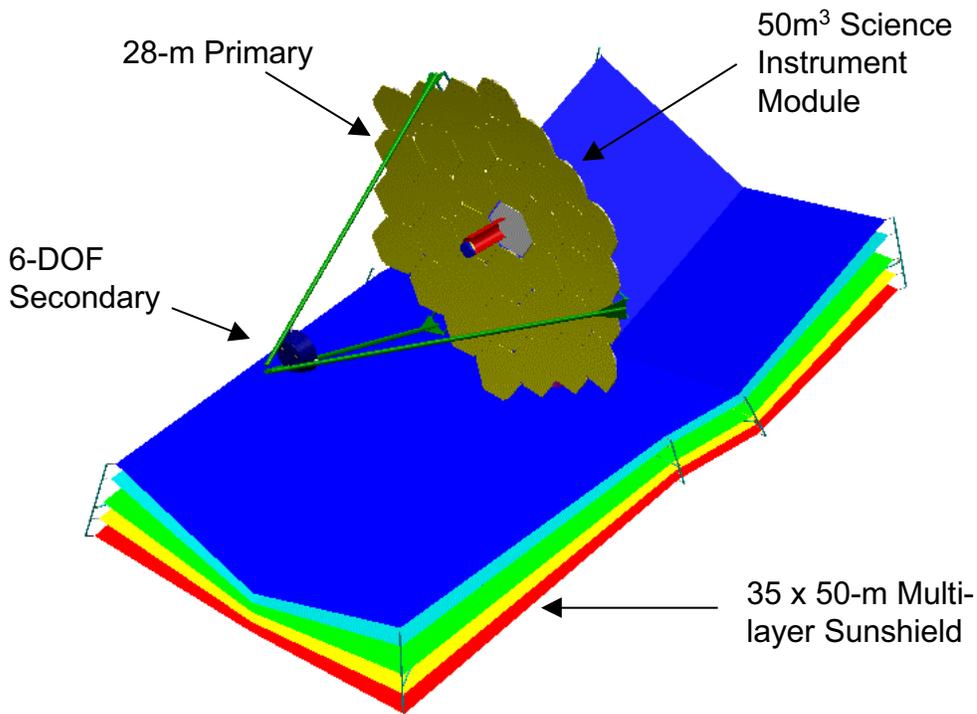
Figure 1. Conceptual Design

A science instrument module (Figure 2) behind the primary houses a coronagraph with an IR imager for planet detection and an IR spectrometer for planet characterization. The coronagraph occupies ~1/3 of the instrument module's 50 m^3 volume, leaving room for other instruments, such as imagers and spectrometers for general astrophysics observations.

This science payload (telescope plus the SIM) is attached to the sunshade and the spacecraft bus by a deployable mast that also provides thermal and vibration isolation from the ~300K spacecraft with its rapidly rotating gyros and reaction wheels.
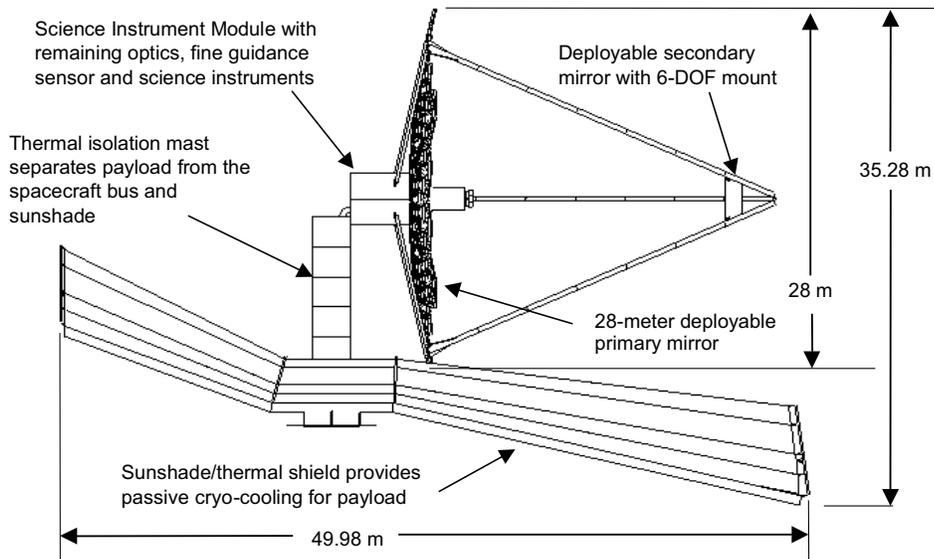
Figure 2. Deployed Observatory (Side View)

The Spacecraft equipment compartment (Figure 3) houses all of the avionics and warm payload electronics. A deployable, non-articulating solar array provides constant electrical power while low gain omni antennas and a high gain dish antenna provide communications with the ground system. Bi-propellant Secondary Combustion Augmented Thrusters (SCAT)s provide propulsion for orbit insertion and station keeping, while hydrazine thrusters provide momentum unloading and backup attitude control. The equipment module has dedicated panels for parallel integration and test. The thermal isolation mast stows in the equipment module's central cylinder.
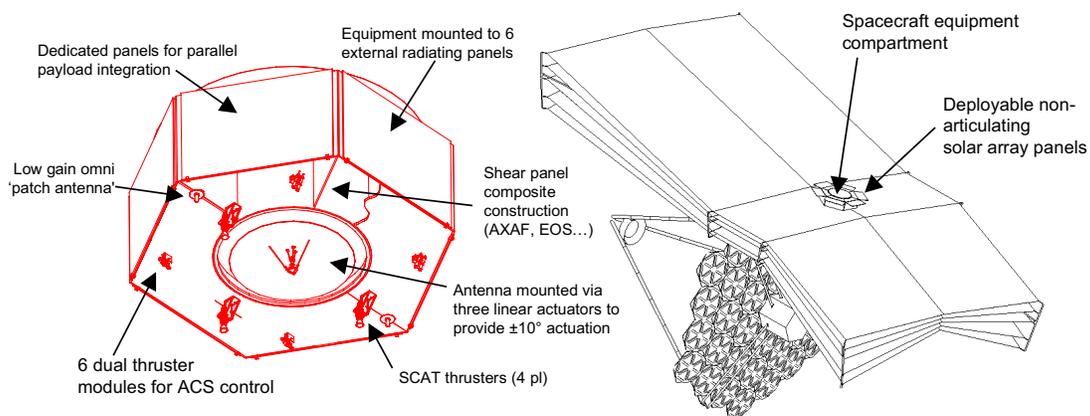
Figure 3. Spacecraft Bus

The entire TPF observatory can be packaged to fit in the fairing of the Delta IV Heavy launch vehicle (Figure 4), which has sufficient lift capability to place it in a transfer orbit to its operational orbit around the L2 point.

replica optics
36 x 0.205 m
7.41 m total

16.5 m

Instrument
4.00 m

sunshade stowage 0.92 m
Spacecraft bus 1.06 m
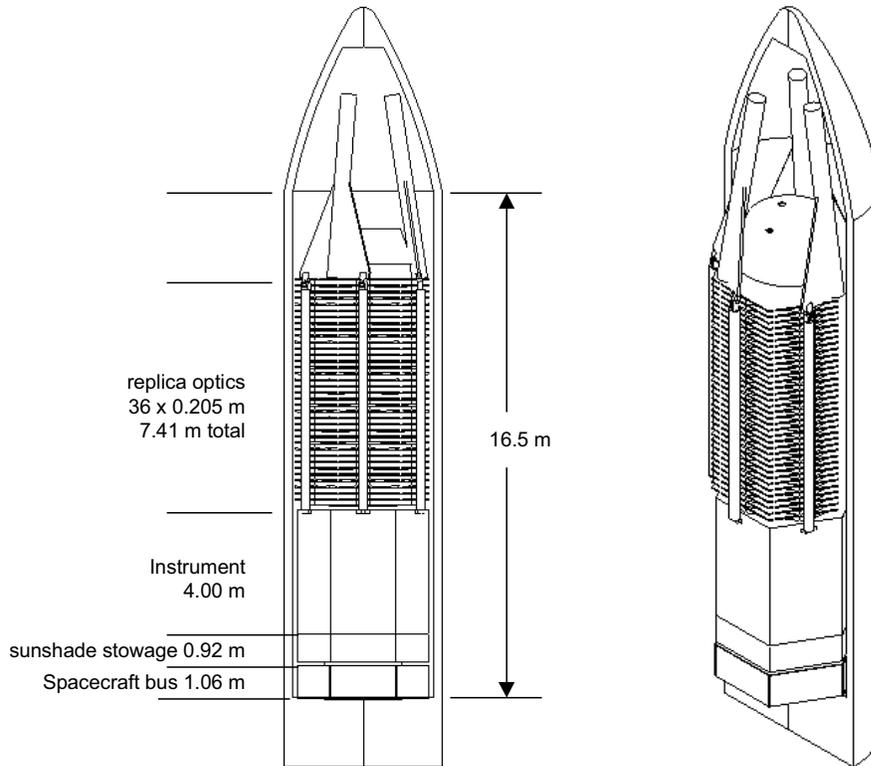
Figure 4. Observatory Stowed in 73.5 foot Delta IV Heavy Fairing

5.0 Deployment

After separation from the launch vehicle's upper stage the spacecraft's solar array panels are deployed, followed by the Optical Telescope Assembly (OTA) and the Thermal Isolation Mast (TIM). The OTA deployment begins with the extension of the telescoping secondary mirror support struts. The struts then hinge in the middle and rotate outward at their point of attachment to the Science Instrument Module (SIM), providing clearance for deployment of the primary mirror. The primary mirror panels are deployed using the approach proven by TRW's highly successful High Accuracy Reflector Demonstration (HARD) Program in the early 1990's. Figure 5 shows how HARD reflector panel stack is raised, rotated, and lowered so that the bottom panel can be latched into place. For TPF this process is repeated until all three rings (36 panels) have been deployed. The secondary support struts are then straightened and latched to the periphery of the primary mirror (see Figure 2). The tertiary mirror, deformable mirror and central baffle that have been stowed in the instrument compartment are then deployed into position with a telescoping mechanism. The final step in the deployment is deployment of the sunshade, using proven mechanical deployment technology developed for large RF antennas.

TPF's 11 deployment systems with 68 elements are not much more complex than those on our Tracking and Data Relay Satellites (TDRS) with 8 systems and 45 elements that deployed perfectly on all 6 flight systems. [In our over 40 years of experience in the design, integration, verification and flight operation of spacecraft deployments we have deployed 672 systems with 1920 elements with a 100% mission success rate].
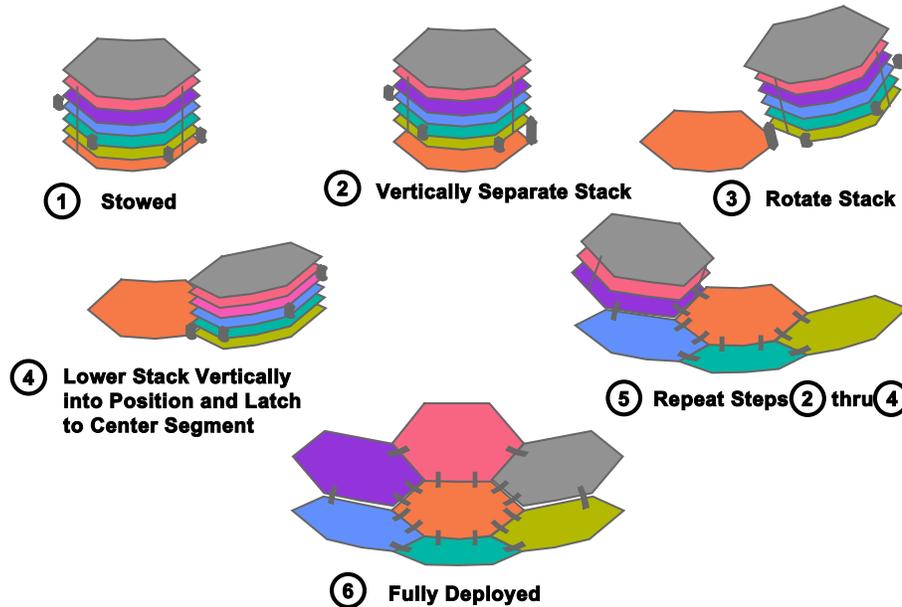
Figure 5. HARD Reflector Deployment Sequence

6.0 Telescope Design

The OTA (Figure 6) is a three-mirror anastigmat (TMA) with a fourth (steerable and deformable) mirror that provides fine pointing and wavefront error correction. The off-axis in field, on-axis in aperture design provides excellent stray light control and well corrected aberrations in the focal plane with residual spot sizes <0.038 x diffraction limit at 7 μm over a 2.4 x 7.2 arcminute field of view. The F/20 optical system provides a plate scale of 0.367 arcsec/mm at the telescope focal plane.

The OTA image quality requirement is diffraction limited at 7 μm (with a goal of 3 μm). In order to minimize scattered light the surface roughness requirement is <10 nm rms (with a goal of 3 nm). The IR coronagraph requires correction of spatial frequencies from ~0.8 to 98 cycles/diameter. Our actuator density study show 7 figure actuators are ample for correcting low order deformations such as RoC, astigmatism and trefoil, and we can use global influence functions to control low-spatial frequencies from 0.5 to 10 cycles/dia. The residual is corrected by the DM with ~200 actuators/dia.

The TMA system wavefront error due to actuator residual errors, static wavefront errors, launch induced errors and on-orbit thermal/structural errors (with contingency) is ~980 nm rms (0.14λ at 7 μm). After correction with the DM the mid-spatial frequency wavefront error is predicted to be 3.69 nm rms.
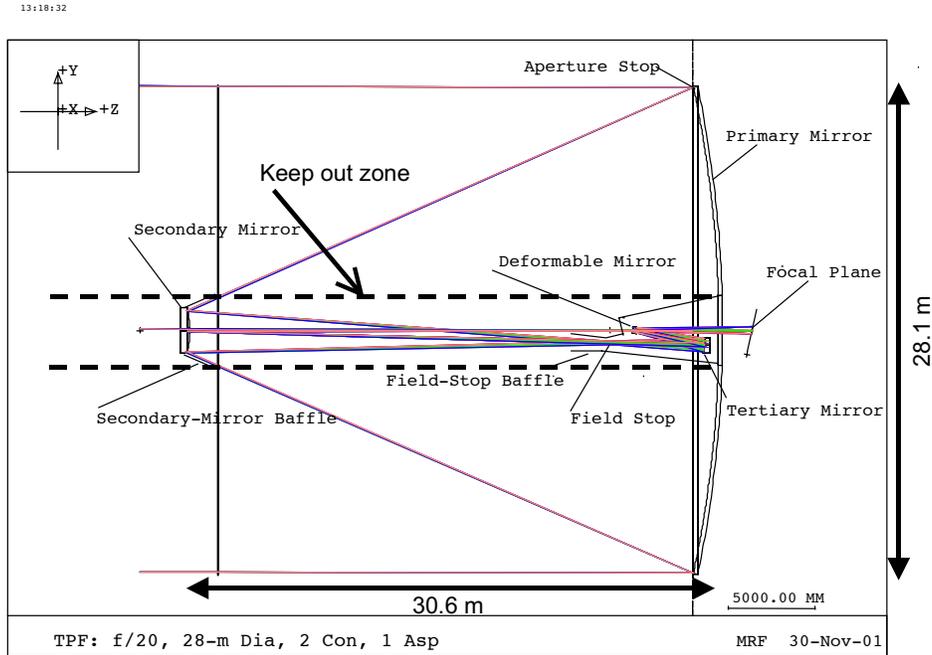
Figure 6. Optical Telescope Assembly Layout

7.0 Science Instruments

Figure 7 lists the science instrument characteristics, including the coronagraph, guide star sensor and the (optional) mid-IR camera/spectrometer. Figure 8 shows the coronagraph's optical layout, including the occulting spot, Lyot stop and the imaging camera. The Fast Steering Mirror is used to correct for jitter-induced line-of-sight pointing errors. These errors are sensed with a quad cell that receives the starlight reflected off the back of the occulting spot. A filter wheel selects the bandpass that reaches the imaging system focal plane and other wheels hold a variety of occulting

	λ (μm)	IFOV (mas)	FPA size (pixels^2)	FOV (as^2)	Comments
Guide Camera (Guide *and* Wide Field / Acquisition modes)	3-5	10 *and* 60	1024	10 *and* 60	HgCdTe likely. Run fast enough to drive the mirror for guiding. Baseline a magnification Guiding Mode. Could be separate units for Mode and Wide Field Acquisition Mode. Based NGST science instrument and guider.
IR Coronagraph	7-17 (5-28)	20	512	10	Base on Eclipse testbed heritage. Filter wheel spot wheel to optimize detection based on type & expected separation; dispersive gratings for low resolution spectra for
IR Imaging Camera / Spectrometer	5-28	60 or 30	1024 or 2048	60 or 120	Largest available Si:As detector. Dual filter CVF, grisms, spectrometer pickoff are all options. Base on NGST MIR camera, and spectrometers.
Optional: Vis/NIR Imager	0.5-5	15	4096 (NGST based)	60	InSb or HgCdTe detector as used for NGST. driven by figure errors of primary mirror. wavelength cutoff driven by mirror

Figure 7. Science Instrument Characteristics

spots and Lyot Stops that enable the coronagraph to optimize its characteristics for each target. Not shown here is the folding flat that feeds a spectrograph similar to the SIRTF Infrared Spectrometer short wavelength, low-resolution module with a micro-shutter device to select the portion of the image that contains the object to be observed.
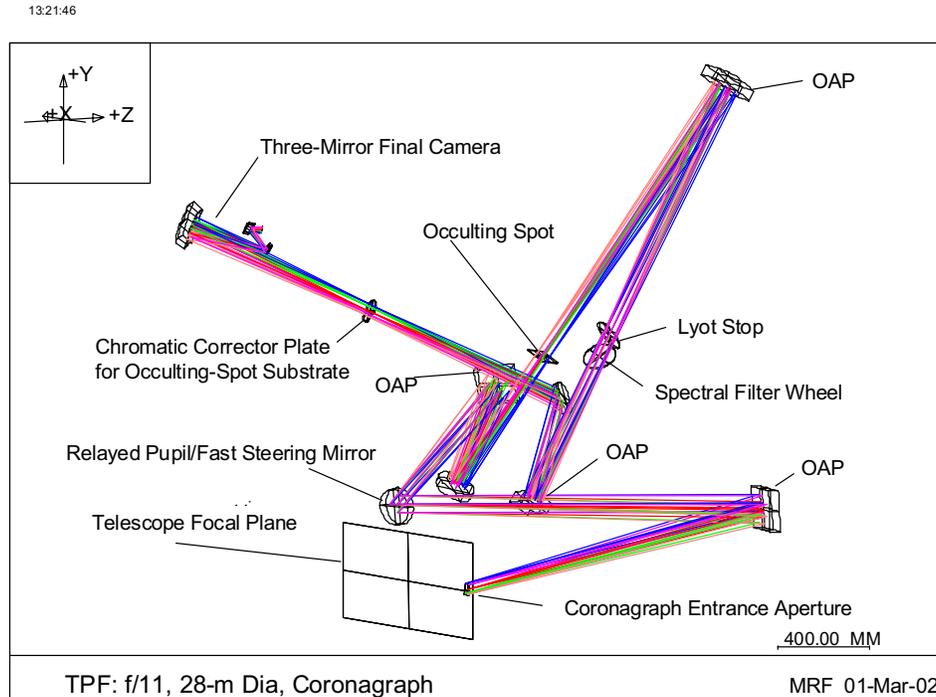
Figure 8. Coronagraph Layout

8.0 Planet Detection and Characterization Performance

Figure 9 summarizes the expected performance of our IR coronagraph for the planet detection, including the input parameters to our performance model and the values

Parameter	Value	Units
Primary Mirror Diameter	28	m
Number of Hexes	36	
Wavelength	10	microns
Integration time	2	hours
Pixels in FOV	512	
IFOV (per pixel)	0.0250	arcseconds
Distance of Star	24.2	parsecs
Overall optics thruput	0.6150	Read from optics
Spectral Resolution	5	λ / Δλ
Maximum single t_exp	901	seconds (NGST)
Pixel size	30	microns
Instrument temperature	10	K
Optics Temperature	50	K
Exo-zody (factor * Earth)	1	0.1 - 10x

Detector Parameters (from Detector worksheet)		
Dark current 1	30	electrons/sec
Read noise 1	100	electrons
QE 1	0.407	Si:As
Optics emissivity 1	0.1	

Occulting Spot Parameters		
Spot FWHM	80	mas
Attenuation	8	10^x
Lyot stop atten	1.00E-02	on diffraction
Apodized Lyot?	0	1=yes, 0=no

System Parameters		
Mirror panel vibration alloc.	3	mas
LOS error allocation	3	mas

Case	Time to SNR = 5 for R=5, λ=10 μm
Earth @ 10 pc	2.4 hours
Phi2Pav @ 24.2 pc	71.2
DYEri @ 5 pc	0.4
HIP 48113 @ 18.4 pc	12
71 Ori @ 21.1 pc	73.5
HIP 92549 @ 26.1 pc	783 hours, or 101 hrs @ 8 μm, & 17.6 hrs @ 8 μm w/ 60 mas occulting spot

Multiple occulting spots and filters required for detection mission

"Average" integration time per target is estimated to be ~20 hours

Figure 9. System Performance for Planet Detection

assigned to them. We assumed the occulting spot has a log-Gaussian form with its transmission increasing from 10^{-8} at the center to 10^{-4} at a radial distance of 0.04 arcsec (40 mas) and unity at 120 mas. This sharp increase in transmission enables the instrument to detect planets at radial distances of order $\sim\lambda/D$, i.e.: at $\sim$50 mas at 7 μm and $\sim$70 mas at 10 μm. As shown in Figure 10, at 10 μm, for a solar-type star at 10 parsecs, the background signal from scattered and diffracted starlight is still $\sim10^4$ greater than the signal from an Earth-like planet. Nevertheless, by PSF fitting and/or subtraction of images with different roll angles, it should be possible to detect the planet at a radial distance of 100 mas with a SNR= 5 and $\lambda/\Delta\lambda$=5 in $\sim$0.8 hours. This integration time increases to $\sim$2 hours at a radial distance of 80 mas and $\sim$5 hours at 70 mas. The integration times to obtain spectra with $\lambda/\Delta\lambda$=20 and SNR=10 for planet characterization are a factor of $\sim$18 longer.
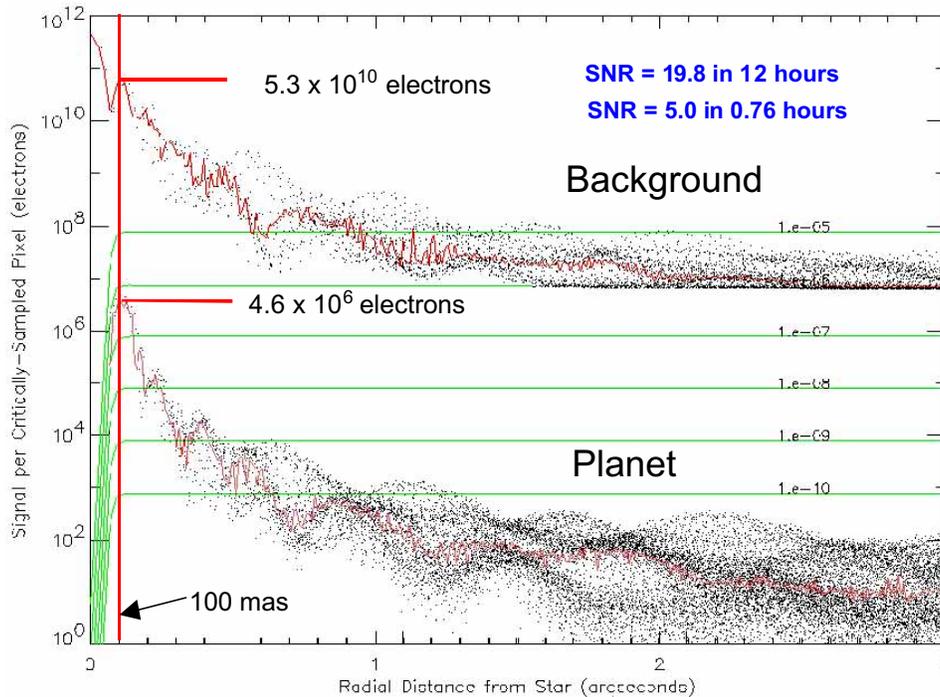
Figure 10. Observed counts per Pixel for a 12 hour Integration

9.0 General Astrophysics Capability

TPF also has a great power for general astrophysics. With an angular resolution of 63 mas at 7 μm, a collecting area of $\sim$490 m^2, a 2.4 x 7.2 arc minute field-of-view, and an average mirror temperature of 21K (and a maximum of 35K) it is well suited for observations in 3 to $\sim$50 μm region. [Since thermal radiation from the spacecraft is the dominant heat source, additional thermal shielding could lower the optics temperature into the 10-15K range]. As shown in Figure 11, TPF's point source sensitivity would be $\sim10^2$ times greater than NGST, $\sim10^4$ times greater than SIRTF, and 10^7 times greater than SOFIA. In particular, TPF could obtain a 5σ observation of a 25-day period Cepheid at J-band at 750 Mpc in $\sim$6 hours; measure IR surface brightness fluctuations at several gigaparsecs; get light curves and spectra of Type Ia supernovae at Z=3; and image super

starclusters at Z=19 at an SNR of 10 in 6 hours; study disks and outflows around protostars and disks jets in active galactic nuclei; and obtain spectra of faint galaxies and other faint SIRTF and NGST discoveries.

- TPF point source sensitivity compared with other IR Observatories for
  - Integration time = 10,000 sec
  - Resolution $(\lambda/\Delta\lambda) = 5$
  - Signal to Noise = 5σ
- The 4-m and 8-m NGST and the 28-m TPF telescopes are assumed to have optics with an equilibrium temperature of 35K and emissivity of 0.05
- The zodiacal light background limits NGST and TPF sensitivity for $\lambda < 15\mu m$
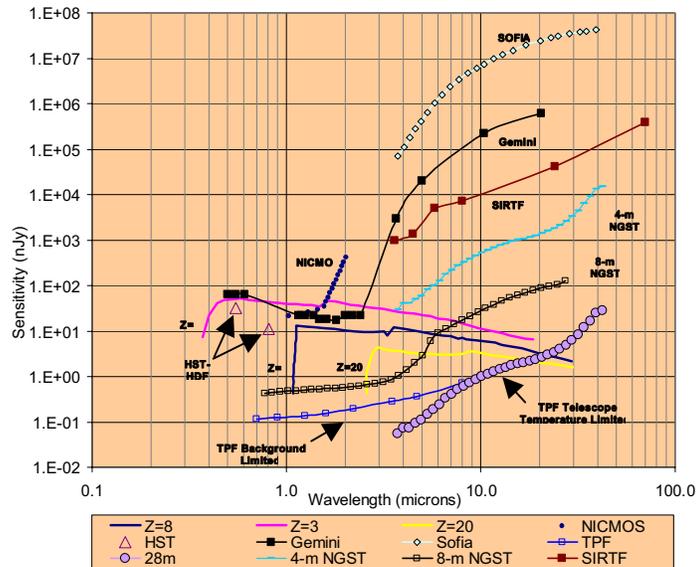
Figure 11. TPF Point Source Sensitivity Compared to Other Observatories

10.0 Key Technologies

The key enabling technologies for our IR Coronagraph are large lightweight optics, high-contrast imaging technologies, large format Si:As detectors, large IR filters and broadband transmissive substrates. Most of these technologies are already being developed for other missions such as NGST, but additional effort is required. Many other NASA programs require similar technologies, including FAIR, SPIRIT, SUVO, Life Finder and Planet Imager. These technologies are currently at TRL 2 or 3. No breakthroughs are required to develop them for TPF, however, just a good engineering effort with reasonable funding.

11.0 Cost and Schedule

The next 4 years of the TPF program will be devoted to technology development, precursor mission(s) development, and pre-phase A studies of two architecture classes leading to selection of a single architecture class in 2006. Two 24-month phase A studies initiated in 2007 will precede downselection to a single prime contractor, followed by a 24-month Phase B and 48-month Phase C/D development effort. We currently envision launch in December 2014. At the TPF Final Architecture Review in December 2001 the four study contractors presented life cycle cost estimates that ranged from ~$1.2B to $1.9B in FY'02 dollars.

12.0 Acknowledgements

This work was performed under JPL Contract No. 1217284. We thank our Science Team members and Industry Partners for their support, and Dr. John Trauger for his Performance Model Calculations.

An Overview of the StarLight Mission
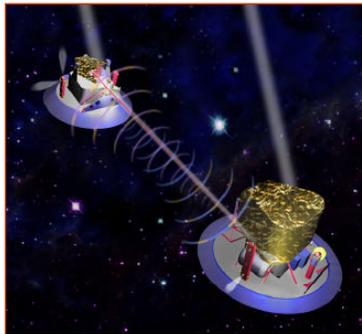
Oliver Lay, Gary Blackwood, Serge Dubovitsky, Riley Duren

Jet Propulsion Laboratory M/S 301-451
4800 Oak Grove Drive
Pasadena CA 91109
Tel: (818) 354-2521
Email: oliver.p.lay@jpl.nasa.gov

StarLight and Formation Telescopes

StarLight is the precursor to a new class of Formation Telescopes, led by TPF

DARWIN

Life Finder

StarLight

TPF

Planet Imager

SPECS

Detector Spacecraft

MAXIM Pathfinder

Stellar Imager

...MAXIM

Optics Spacecraft
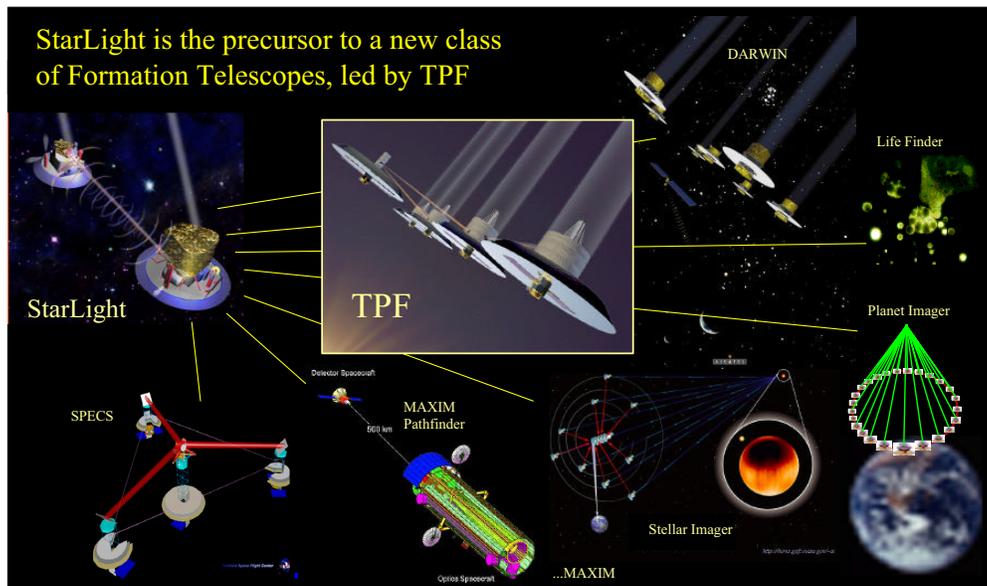
2   Far IR 3/7/02

StarLight mission summary

- June 2006 launch to heliocentric orbit

- Nominal 6 month mission with option of additional 6 month extension

- Validate autonomous formation flying system
  - range control to 10 cm
  - bearing control to 4 arcmin

- Demonstrate formation flying optical interferometry
  - wavelength 600 - 1000 nm
  - baselines 30 - 125 m

1. Spacecraft launched together on Delta 2925

2. Checkout and formation flying experiments

3. Combiner-mode (single spacecraft) interferometry with collector parked in safe stand-off position

4. Formation interferometry

3 Far IR 3/7/02

---

2 spacecraft parabolic geometry

- Original 3 spacecraft design did not fit the budget
- 2 spacecraft concept demonstrates all key areas of formation flying interferometry
- Collector flown on the surface of a virtual paraboloid, with combiner at the focus
- Gives a baseline of 125 m with a fixed delay of only 14 m

Collector

F

V

125 m

Synthetic Aperture Plane

Baseline	Separation	Bearing angle
30 m	40 m	46 deg
125 m	600 m	12 deg

4 Far IR 3/7/02

225
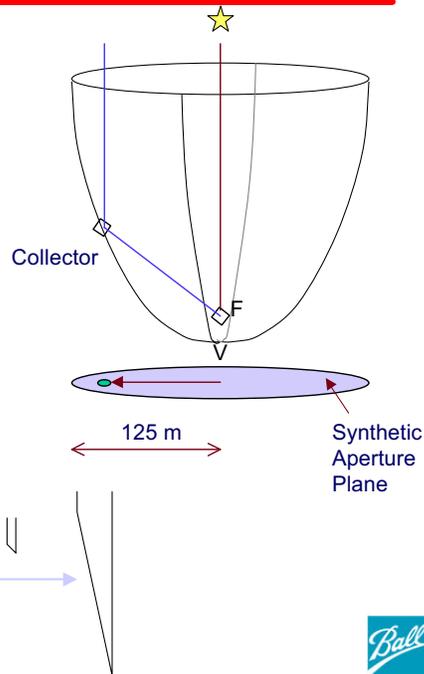
Observing configuration

Fold mirror

Articulating siderostat

Starlight field of regard

Incoming starlight direction for 30 m baseline

Collector

Fixed sunshade

Starlight field of regard

Inter-Spacecraft Line of Sight

Thruster stalk (8 nozzles per stalk)

Combiner

Instrument envelope

Deployed sunshade

3.5 m

18º x 18º Star Tracker FOV (x2 Shown)
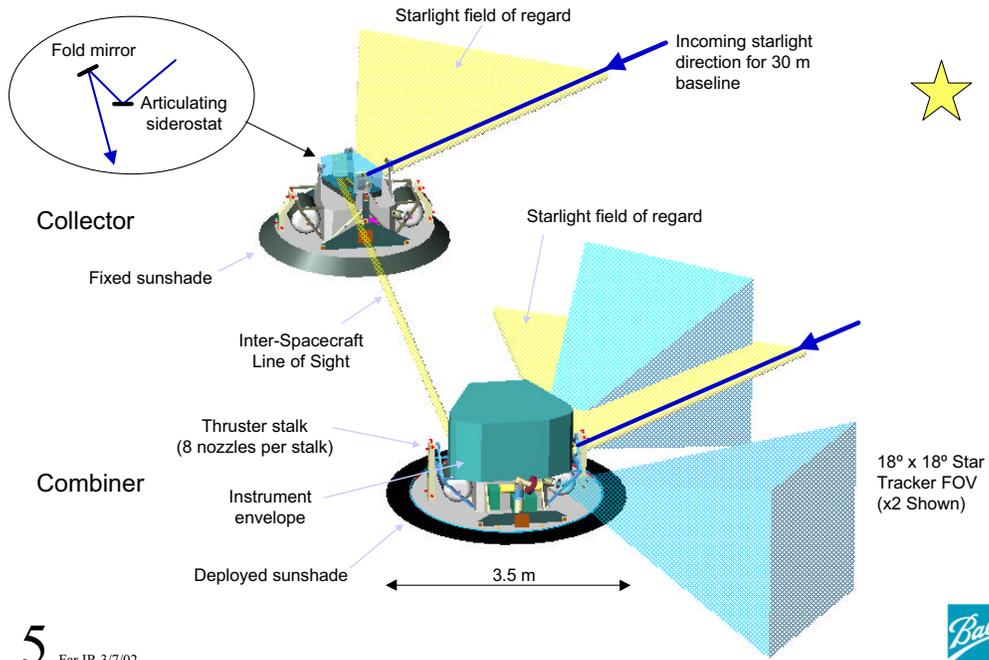
5 Far IR 3/7/02

Combiner instrument optics (unfolded)

Compressor Primary

Left starlight (via collector)

Right starlight (direct from star)

~1 m

Siderostat

Compressor secondary

Active delay

Combiner sub-bench

Fixed delay primary (14 m)

Siderostat (12 cm clear aperture)

Compressor secondary
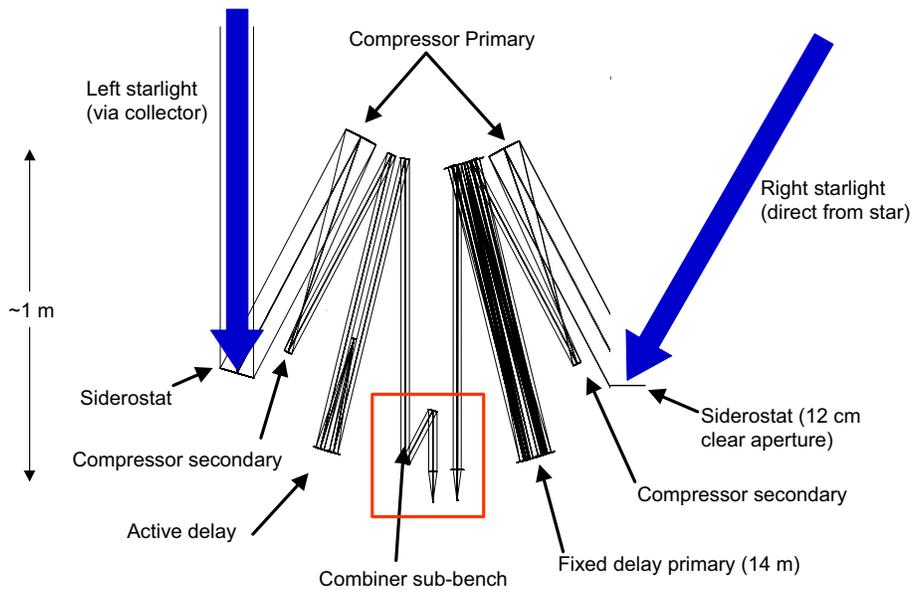
6 Far IR 3/7/02

Combiner instrument layout

1.3 m

Thermal blanket standoffs

left siderostat

right siderostat

active delay

14 m fixed delay

compressor

beam combiner

Bottom side of bench

Top side of bench
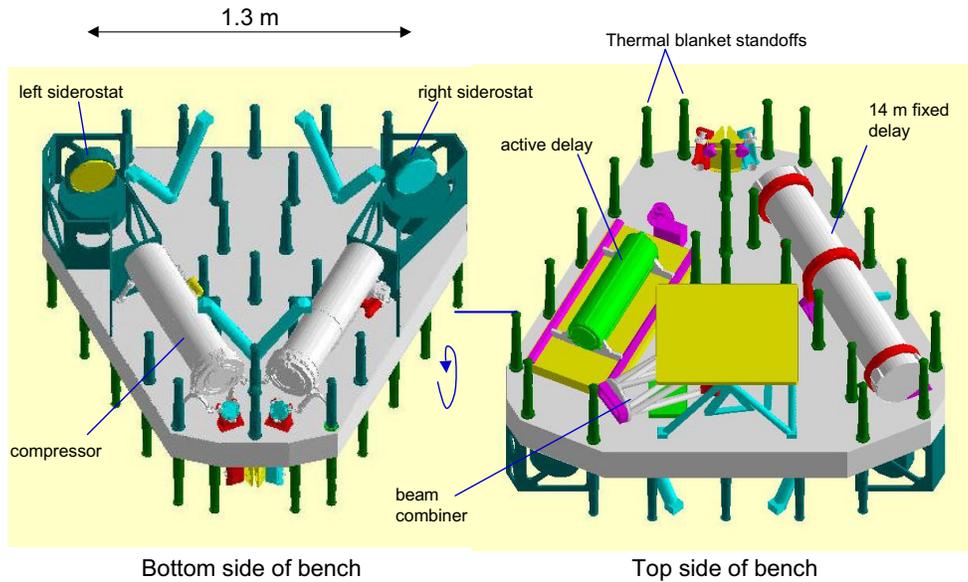
7 Far IR 3/7/02

Acquisition and Observation Sequence

1 Move spacecraft

2 Acquire right starlight

3 Acquire metrology

4 Acquire left starlight

5 Estimate delay & delay rate; trim formation

6 Find fringe

7 Track & Measure fringe

Target star

• Initial state

Collector

Combiner

8 Far IR 3/7/02

227
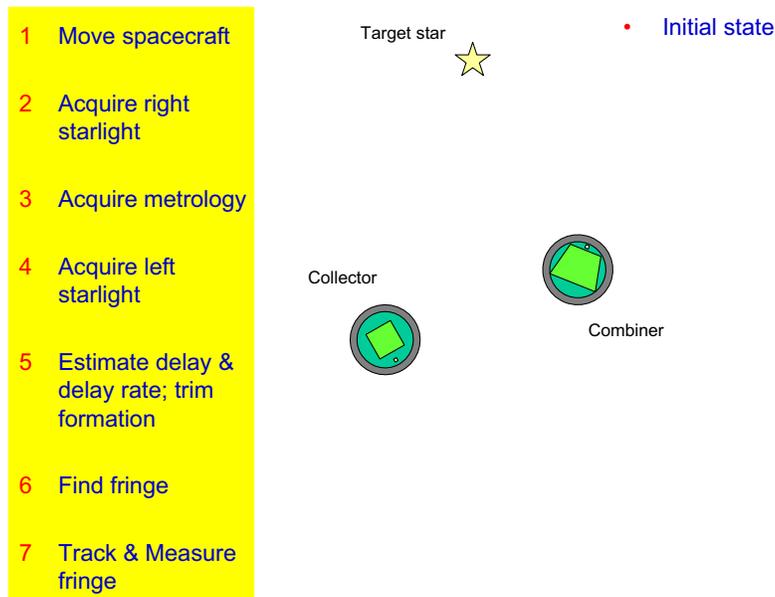
Move spacecraft

1 **Move spacecraft**

2 Acquire right starlight

3 Acquire metrology

4 Acquire left starlight

5 Estimate delay & delay rate; trim formation

6 Find fringe

7 Track & Measure fringe

Collector

Virtual paraboloid of equal delay

Combiner
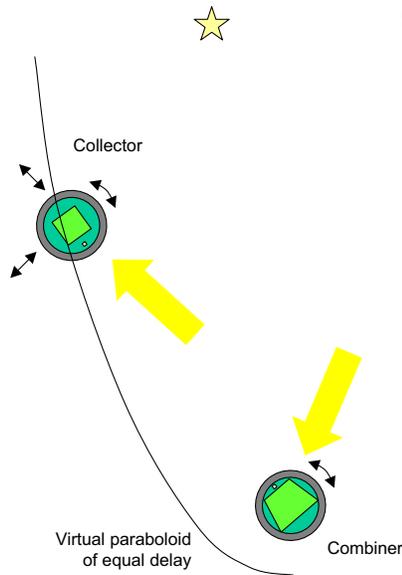
- Formation & Attitude Control System (FACS)
  - Fusion of sensors inputs:
    - startrackers (~ 5 arcsec)
    - gyros
    - AFF sensor
    - (laser metrology)
  - Actuators
    - Reaction wheels (not used for interferometry)
    - Cold gas thrusters
      - 16 per spacecraft
      - 7 mN thrust
      - 74 μN-s minimum impulse
  - Master/Slave architecture
  - Functions:
    - collision avoidance
    - sun avoidance
    - fuel balancing
    - minimize thruster plume impingement
  - Performance:
    - +/- 10 cm range
    - +/- 4 arcmin bearing
    - +/- 3 arcmin spacecraft attitude

9 Far IR 3/7/02

Autonomous Formation Flying sensor

AFF sensor rear equipment

AFF sensor front mounting plate

RF signal from Collector to Combiner portion of AFF sensor
RF signal from Combiner to Collector portion of AFF sensor
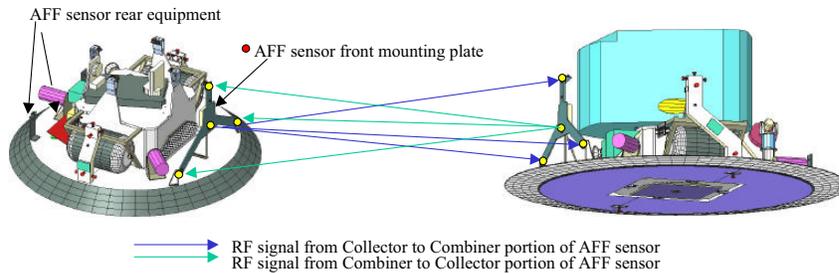
- Ka band system (30 GHz)
- 2 Tx and 4 Rx per spacecraft
- Ranging codes based on NAVSTAR GPS
- Relative bearing from carrier phase difference
- Symmetric between spacecraft
- 2 cm range, 1 arcmin bearing (1 σ)

Prototype Ka-band Antenna with choke rings

Electronics mounted on back of mounting plate
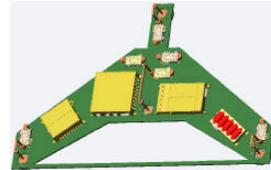
10 Far IR 3/7/02

Acquire right starlight

1 Move spacecraft

2 **Acquire right starlight**

3 Acquire metrology

4 Acquire left starlight

5 Estimate delay & delay rate; trim formation

6 Find fringe

7 Track & Measure fringe
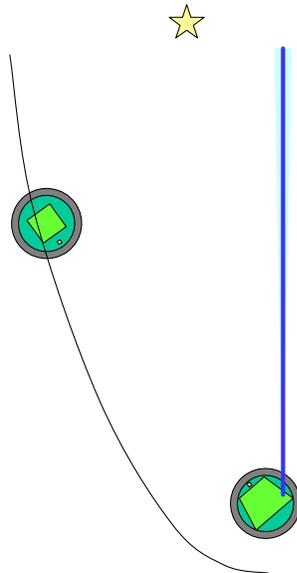
- Instrument has single CCD for acquisition, angle tracking and fringes
- Acquisition field of view = 1 arcminute
- Once acquired, control loop tracks out changes of combiner inertial attitude

11 Far IR 3/7/02

Acquire metrology

1 Move spacecraft

2 Acquire right starlight

3 **Acquire metrology**

4 Acquire left starlight

5 Estimate delay & delay rate; trim formation

6 Find fringe

7 Track & Measure fringe
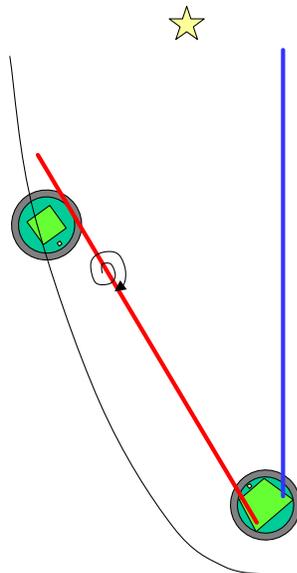
- Outgoing 1.3 μm metrology beam is coincident with left combiner boresight
- Collector fold mirror houses 4-diode Metrology Pointing Sensor
- AFF sensor gives combiner - collector bearing angle
- Left combiner siderostat performs spiral search to acquire laser signal
- With metrology pointing loop locked, system can resolve 10 μm/s transverse motion @ 600 m
- Dual target linear metrology monitors external and internal paths simultaneously with one beam to 10 nm

20 mm
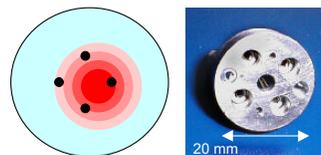
Metrology Pointing Sensor

12 Far IR 3/7/02

229

Acquire left starlight

1. Move spacecraft

2. Acquire right starlight

3. Acquire metrology

4. **Acquire left starlight**

5. Estimate delay & delay rate; trim formation

6. Find fringe

7. Track & Measure fringe
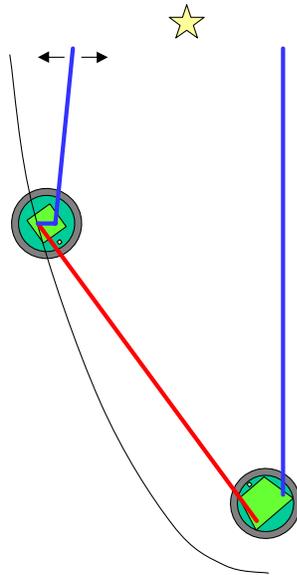
- Left starlight boresight is locked to center of collector optics
- Collector siderostat executes a small search until star appears in combiner focal plane
- Left angle tracking control and metrology pointing loop form a coupled control system, distributed across both spacecraft
- Stray light from collector sunshade is an issue:
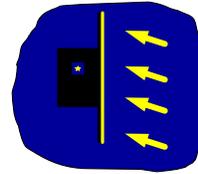
13 Far IR 3/7/02

Estimate delay & delay rate

1. Move spacecraft

2. Acquire right starlight

3. Acquire metrology

4. Acquire left starlight

5. **Estimate delay & delay rate; trim formation**

6. Find fringe

7. Track & Measure fringe
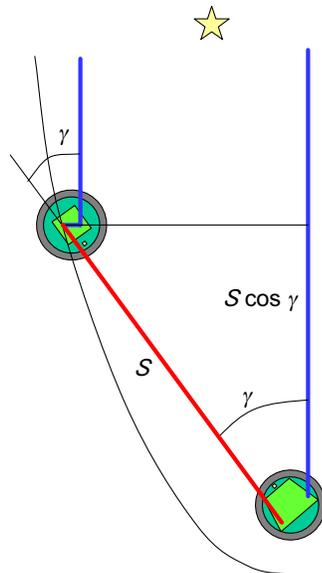
Delay offset = Left path − Right path

$$\approx S - \left(S\cos\gamma + D_{fixed} \right)$$

$$\approx S\left(1 - \cos\gamma\right) - D_{fixed}$$

Delay (10 mm)

S γ

AFF sensor (20 mm)   Angular metrology (10')

Delay rate (20 μm s^{-1})

S γ

Linear metrology (25 μm s^{-1})   Angular metrology (15 mas s^{-1})

γ

$S\cos\gamma$

S

γ
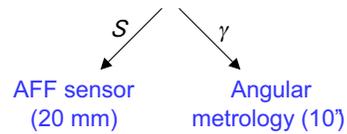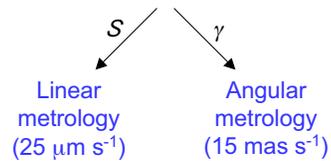
14 Far IR 3/7/02

Trim formation

1	Move spacecraft
2	Acquire right starlight
3	Acquire metrology
4	Acquire left starlight
5	**Estimate delay & delay rate; trim formation**
6	Find fringe
7	Track & Measure fringe
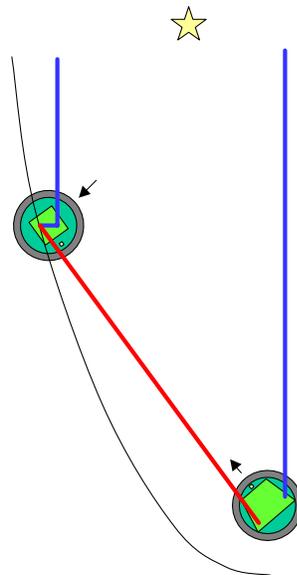
- Need to get within the range and rate capabilities of the active delay line
- Instrument laser metrology provides precision range rate and bearing data to FACS
- Small thrusts to trim the delay and delay rate
- Iterative process until requirements met
- When complete we have a stabilized optical structure:
  - Tip/tilt stabilization:
    - 3 pointing loops locked (left & right stellar angle tracking, metrology beam to collector)
  - Path stabilization:
    - active delay line tracks out jitter sensed by dual target linear metrology

15 Far IR 3/7/02

Find fringe

1	Move spacecraft
2	Acquire right starlight
3	Acquire metrology
4	Acquire left starlight
5	Estimate delay & delay rate; trim formation
6	**Find fringe**
7	Track & Measure fringe
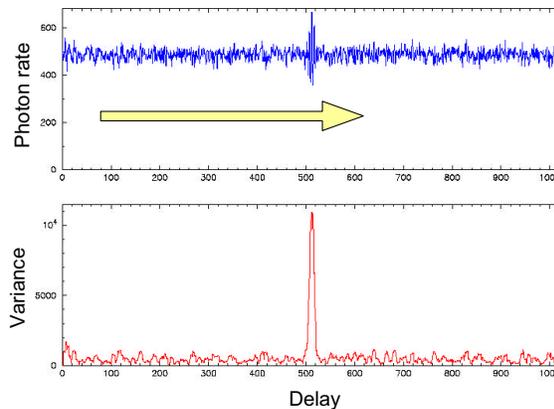
- Delay uncert ~ 10 mm (1 σ)
- Delay rate uncert ~ 20 μm s^{-1} (1σ)
- Search 50 mm in delay at 100 μm s^{-1}
- Scan the delay line across the search range
- Fringes give increased variance in detected photon rate

16 Far IR 3/7/02

231

Track & measure fringe

1	Move spacecraft
2	Acquire right starlight
3	Acquire metrology
4	Acquire left starlight
5	Estimate delay & delay rate; trim formation
6	Find fringe
7	**Track & Measure fringe**

- StarLight measures visibility amplitude at each of 5 baselines (30 - 125 m)
- White light and 4-channel dispersed fringe outputs, sampled at 500 Hz
- Inherit algorithms from Keck Interferometer
- The interferometer performance will be characterized by observations of ~ 20 known stars of different size and brightness, down to about 5th magnitude
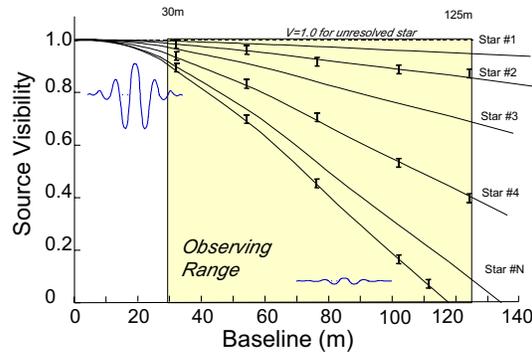
17 Far IR 3/7/02

Technology development

Formation Interferometry Testbed

Formation Flying Simulations
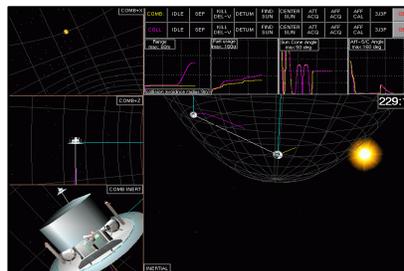
Autonomous Formation Flying Sensor tests
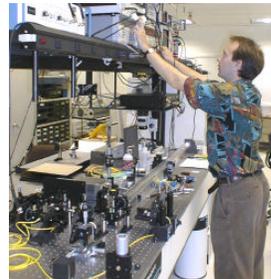
Linear and angular metrology

18 Far IR 3/7/02

Formation Interferometer Testbed

Psuedostar   Combiner

Collector

Located at JPL

fringe data, 01/25/02
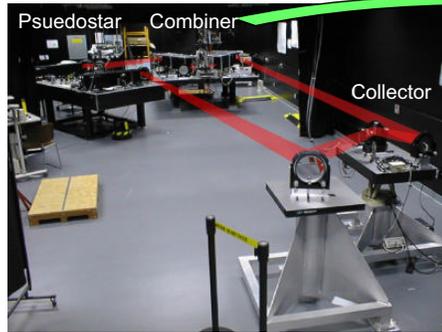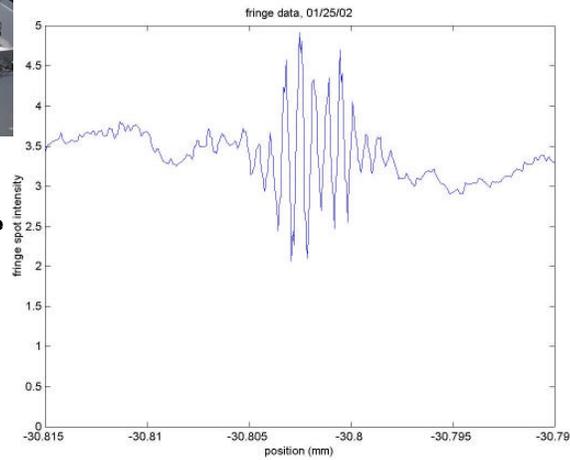
- **Recently achieved white light fringe tracking**
  - **Instrument visibility ~40% (matches the predicted value)**
  - **All control loops operating simultaneously**
  - **Performance limited by:**
    - **20m airpath each arm**
    - **50 Hz camera**

19 Far IR 3/7/02

Key performance metrics

Performance metric	StarLight	TPF Planet-finding	TPF Astrophysics	StarLight Technology	TPF need
1 Wavelength band (fringes)	600 - 1000 nm	7 - 20 μm	3 - 30 μm		
2 Baselines	30 - 125 m *	75 - 200 m	75 - 1000 m	2-spacecraft parabolic geometry	angular resolution for astrophysics
3 Separation	40 - 600 m *	25 - 70 m	25 - 330 m		
4 Range control (+/-)	10 cm *	5 cm	5 cm	AFF sensor, angular metrology, formation-flying algorithms, low impulse thrusters	sizes delay lines
5 Bearing control (+/-)	4 arcmin *	1 arcmin	0.2 arcmin		
6 Range knowledge (1 σ)	2 cm	1 cm	1 cm	AFF sensor	formation control ; fringe search limiting magnitude; delay line range
7 Range rate knowledge (1 σ)	< 1 μm / s	35 μm / s	1 μm / s	Dual target linear metrology system	
8 Inertial Bearing knowledge (1 σ)	10 arcsec	10 arcsec	2 arcsec	Angular metrology system	
9 Bearing rate knowledge (1 σ)	30 milliarcsec / s	40 milliarcsec / s	0.2 milliarcsec / s	Angular metrology system	
10 Path length stabilization (1 σ)	35 nm	3.5 - 70 nm	70 nm	Dual target linear metrology system	nulling
11 Tip/tilt stabilization (1 σ)	0.1 λ/D	0.025 λ/D	0.05 λ/D	Angular metrology system, distributed control loops	nulling

* nominal performance; limits to be pushed on orbit

20 Far IR 3/7/02

233

SESSION IV

MISSION ISSUES

FUTURE FAR-IR AND SUBMM MISSIONS: JAPANESE PERSPECTIVE

Takao Nakagawa[1]

Institute of Space and Astronautical Science (ISAS), 3-1-1 Yoshinodai, Sagamihara, Kanagawa 229-8510, Japan

ABSTRACT

We present the concept of the SPICA (Space Infrared Telescope for Cosmology and Astrophysics) mission, which incorporates a 3.5 m telescope cooled to 4.5 K. SPICA will focus on high-resolution mid- to far-infrared observations with unprecedented sensitivity. It will make great contributions to our understanding of important astronomical questions, such as the history of star-formation in the universe, the birth and evolution of AGN, the formation of planets in extrasolar systems, and the history of our solar system.

In order to reduce the total weight dramatically, we propose a "warm launch" cooled telescope concept; the telescope is to be launched at ambient temperature and is to be cooled in orbit to 4.5 K by a modest mechanical cooler system with the assistance of effective radiative cooling. SPICA is proposed to be launched into a halo orbit around S-E L2 in 2010.

Key words: Galaxies: formation – Stars: formation – Planets: formation – Missions: SPICA

Figure 1. Conceptual design of the SPICA mission with a 3.5 m primary mirror.

1. INTRODUCTION

Since the success of the all-sky survey by the Infrared Astronomical Satellite (IRAS, 1983), infrared observations from space have been one of the essential tools in many fields of astronomy. The first observatory-type infrared mission, the Infrared Space Observatory (ISO), also demonstrated the effectiveness of infrared observations from space.

Following these pioneering missions, two more missions are to be launched within the next four years. One is SIRTF (Gallagher & Simmons 2000), which is the last mission of NASA's great observatory series; it has an 85 cm cooled telescope, and is to be launched in 2002. The other is ASTRO-F (Murakami 1998, Nakagawa 2001), a survey mission of ISAS with a 70 cm telescope, to be launched in 2004.

Although these missions are very powerful, their mirror sizes are relatively small (60 – 85 cm), and the spatial resolution is not so high. Hence high-resolution missions with much larger telescopes have been long-awaited.

Two missions with large telescopes in the infrared and the sub-mm regions are now proposed to be launched within a decade. One is the "Next Generation Space Telescope" (NGST) (Mather & Stockman 2000) and the other is the "Far InfraRed and Submillimetre Telescope" (FIRST) (Pilbratt 2000). These will be powerful observatories, but their telescopes are only passively cooled and the thermal radiation from the telescopes degrades their sensitivity especially in the mid- and the far-infrared.

We propose a new space observatory, the SPICA mission (formerly called HII/L2 mission), which incorporates a large-aperture (3.5 m) "actively cooled" (4.5 K) telescope, and is optimized for mid- and far-infrared astronomy.

2. SCIENTIFIC OBJECTIVES

There are a variety of scientific goals that can uniquely be explored by SPICA. Here we list some representative goals.

2.1. STAR-FORMATION HISTORY OF OUR UNIVERSE

One of the most important questions in astronomy is the history of star-formation in our universe. Most of the previous studies were based on optical observations (e.g. Madau et al. 1996), which traced ultraviolet radiation in the rest-

frame of high-redshift galaxies and left large uncertainties due to dust extinction.

Far-infrared observations, on the other hand, are free from these uncertainties and can reveal the star-formation history reliably even at high-redshift. The high spatial resolution of SPICA is essential to detect faint galaxies at high-redshift in the confusion-limited far-infrared sky.

Moreover, the mid- and far-infrared region is rich with many bright lines useful for the estimates of the redshift of each galaxy. In particular PAH features can be powerful tools, since these features can be measured even with low-resolution spectroscopic observations. SPICA covers most of the PAH features at any redshift, and will enable high-sensitivity determination of the redshift of galaxies at high z with high efficiency.

2.2. BIRTH AND EVOLUTION OF AGN

AGN are one of the important energy release mechanisms in our universe and are believed to be the result of mass accretion on to massive black holes. Their origins are still uncertain, although massive black holes seem to be ubiquitous in galactic bulges.

ISO demonstrated the effectiveness of spectroscopic observations in the mid-infrared to reveal the origin of the luminosity (AGN vs starbursts) even in very dusty galaxies (e.g. Genzel et al. 1998), but these observations were limited to the local universe. The high-sensitivity of SPICA will enable this type of observation even for high-redshift galaxies and will reveal how and when AGN were formed in the early universe.

2.3. FORMATION OF PLANETARY SYSTEMS

Discoveries of "Vega-like stars" evoked much discussion on the formation process of proto-planetary systems. The high-resolution capability of SPICA in the mid-infrared will reveal the density and temperature profiles of the dusty disks of "Vega-like stars", which are essential to understand the formation process of planetary systems.

One more important capability of SPICA is that it is expected to detect extrasolar planets directly. We now have $\sim$50 giant planet candidates in extrasolar systems discovered through "indirect" methods. SPICA will enable the first "direct" detection of extrasolar planets beyond $\sim$2 AU around nearby ($\sim$ 5pc) stars (see Tamura 2000 for details). The nature of these extrasolar planets will be revealed by mid-infrared spectroscopy of planetary atmosphere also by SPICA. These observations will show us a general picture of the formation process of planets.

2.4. DISCOVERY POTENTIAL

The capability of the 3.5 m telescope of SPICA is a big jump from those of previous missions with smaller telescopes ($<$ 1 m). Moreover, SPICA is a very efficient obser-

Table 1. Summary of current specifications of SPICA.

Parameter	Value
Mirror Size	3.5 m
T_{Mirror} in Space	4.5 K
T_{Mirror} at Launch	300 K
Core Wavelength Range	5-200 μm
	(Diffraction Limit at 5 μm)
Orbit	S-E L2 Halo
Cooling	Radiative Cooling and
	Mechanical Coolers
Total Mass	2,600 kg
Telemetry Rate	30 G bytes day^{-1}
Launch Vehicle	H-IIA Rocket
Launch Year	2010

vatory with its wide field of view (6′). Hence, SPICA has a great potential to discover interesting objects serendipitously. Since SPICA can make both photometric and spectroscopic observations, we can make follow-up observations of "serendipitously found objects" with SPICA itself.

3. OUTLINE OF THE SPICA MISSION

3.1. DESIGN CONCEPT

To achieve high sensitivity in the mid- and the far-infrared, we have to cool the whole telescope and the focal plane instruments. All of the infrared astronomical satellites flown so far carried liquid helium for cooling; this made the satellites big and heavy and reduced the sizes of the telescopes themselves significantly. Moreover, their mission lives were limited by the hold time of liquid helium.

To overcome these difficulties, we propose a "warm-launch, cooled telescope" design concept, i.e. the telescope and focal plane instruments are "warm" at launch but are cooled in orbit. The cryogenic system, which enables this "warm launch" concept, is a key issue and will be discussed later.

3.2. OUTLINE OF THE MISSION

Figure 1 shows a conceptual design of the SPICA mission based on the above concept, and Table 1 summarizes its specifications. The "warm launch" reduces the total size significantly and enables the payload fairing of the Japanese H-IIA rocket to accommodate a telescope with a 3.5 m primary mirror.

We do not employ a deployable mirror design for the telescope of SPICA and use a conventional "monolithic mirror" design, in order to make the mission technically feasible and reliable.
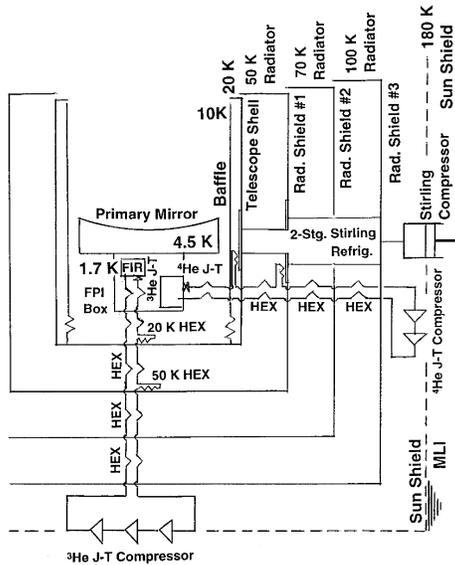
Figure 2. Schematic drawing of the cryogenic system of the SPICA mission. The telescope is enshrouded with multi-layered radiators. One set of a JT cooler (with ^{4}He) and a two-stage Stirling cycle is used to cool the telescope and some of focal plane instruments to 4.5 K. Another set of a JT cooler (with ^{3}He) and a two-stage Stirling cycle is used to cool the far-infrared detectors to 1.7 K.

Figure 3. Heat flow (unit: mW) diagram of the SPICA mission cooled by radiation and cryogenic coolers. Black arrows show the heat flow by conduction and shaded arrows show the heat flow by radiation. We assume a structure of the conceptual design (Figure 1)

4. CRYOGENIC SYSTEM

4.1. CONFIGURATION

The biggest technical challenge of the SPICA mission is the cryogenic system. In order to cool the observing system, we plan to use (1) radiative cooling and (2) mechanical cryocoolers. Figure 2 shows a schematic drawing of the cryogenic system and Figure 3 shows the heat flow diagram of the current design.

In order to make radiative cooling most effective, we propose a halo orbit around one of the Sun-Earth libration points (L2) (hereafter S-E L2, the point at the opposite side of the sun from the earth) for SPICA. In this orbit, heat sources (Sun and Earth) are almost in the same direction and radiative shielding can be simplified. Hence we can make radiative cooling very effective at S-E L2. Radiative cooling alone can cool the telescope below 30 K, which is low enough for near-infrared observations.

To cool the telescope further, we propose to use mechanical cryocoolers: the combination of a two-stage Stirling cycle cooler and a Joule-Thomson (JT) cooler with ^{4}He.

As the heat flow diagram (Figure 3) shows, the dominant cooling process of the whole system is through radiation at various temperature stages, and the cooling power required for mechanical coolers is only 30 mW at 4.5 K,

i.e. we can cool the whole telescope and the focal plane instruments down to 4.5 K by a modest cryocooler system.

4.2. MECHANICAL CRYOCOOLERS

As discussed in the previous subsection, mechanical cryocoolers are key elements in the cryogenic system of SPICA. We have been developing two types of mechanical cryocoolers for space applications.

One is a two-stage Stirling cycle. This has been developed to be onboard ASTRO-F and it is now under extensive tests. The cooling power is about 200 mW at 20 K. We plan to use this type of cooler as pre-coolers for the JT coolers.

The other type is a 4 K JT cryocooler (Narasaki & Tsunematsu 2000) for SMILES on the international space station. Typical cooling power of the JT cooler, together with the two-stage Stirling cooler, is about 30 mW at 4.85 K, which is sufficient for SPICA. Typical total input power for this system is about 180 W at room temperature.

Both systems are now being tested extensively and are to be flight-proven in 2004. We can use basically the same system for SPICA to cool the telescope to 4.5 K.

Some focal plane instruments require lower temperatures. For example, stressed Ge:Ga detectors for far-infrared have to be cooled to 1.7 K. Hence we plan to use another Stirling-JT system but with ^{3}He for the JT to achieve lower temperature as shown in Figure 3. We are now making a proto model of this system and will start extensive tests soon.

5. FOCAL PLANE INSTRUMENTS

The core wavelength range of SPICA will be 5–200 μm and we plan to cover this wavelength range with two focal plane instruments.

One is the Mid-infrared Camera and Spectrometer, which covers 5–25 μm with two channels. We propose three modes of observations. One is diffraction-limited imaging. The pixels size is $0.''18$ (shorter channel) and $0.''36$ (longer channel) with a common field of view of $6'$. The second is the mid-resolution ($\lambda/\Delta\lambda \sim 10^3$) spectroscopy mode. The third is the coronagraphic mode for the direct detection of planets in extrasolar systems (Tamura 2000). The single-segment, cold optics of the SPICA telescope will be an ideal platform for the coronagraphic observations.

The second instrument is the Far-infrared Camera and Spectrometer, which covers the 50–200 μm range with two channels. This instrument also has two modes: one is diffraction-limited imaging and the other is mid-resolution imaging spectroscopy with a Fourier-transform spectrometer.

6. COMPARISONS WITH OTHER MISSIONS

Figure 4 shows the photometric sensitivity of SPICA for point sources as a function of wavelength. Since SPICA has a cooled telescope, it can achieve superior sensitivity throughout the infrared wave band.

Figure 4 also shows the sensitivity of two other large missions (NGST and FIRST) in the infrared and sub-mm regions.

Each of the three missions has its own unique capability. NGST is geared for near-infrared (core wavelength range of 1-5 μm) observations and can achieve very deep observations in near-infrared with high spatial resolution. FIRST concentrates on longer wavelengths and can make high-resolution spectroscopic observations as well as photometric observations.

However, both NGST and FIRST will have only moderately cooled telescopes, and the thermal radiation from the telescopes will degrade the sensitivity at mid- and far-infrared wavelengths (see dotted lines in Figure 4). On the other hand, SPICA has good sensitivity throughout the infrared region and achieves excellent sensitivity especially in the mid- to far-infrared region, since its telescope is cold enough. In this sense, SPICA is complementary to NGST and FIRST. Since these three missions are unique and complementary to each other, we need these three missions to cover the whole infrared region with good sensitivity and high spatial resolution. International collaboration is essential to make these important missions possible and also to make well-organized observations with these three unique missions.
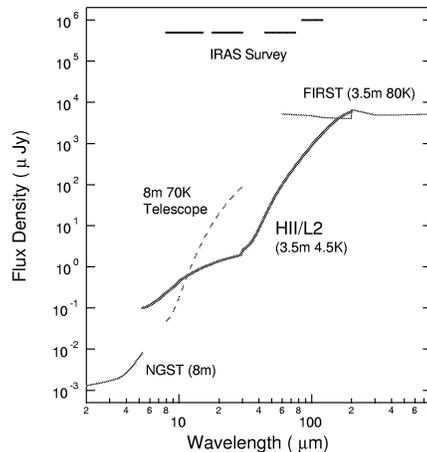
Figure 4. Comparison of point-source sensitivity (5 σ flux density at 3,600 s integration) of various infrared missions. Diffraction-limited observations with $\lambda/\Delta\lambda = 5$ is assumed for SPICA.

REFERENCES

Gallagher, D. B. & Simmons, L. L. 2000, Proc. SPIE 4013, 80

Genzel, R. et al. 1998, ApJ 498, 579

Madau, P. et al. 1996, MNRAS 283, 1388

Mather, J. C., & Stockman, H. S. 2000, Proc. SPIE 4013, 2

Murakami, H. 1998, Proc. SPIE 3356, 471

Nakagawa, T. 2001, this volume

Narasaki, K. & Tsunematsu, S. 2000, in ISAS Report SP 14 Mid- and Far-Infrared Astronomy and Future Space Missions, ed. T. Matsumoto & H. Shibai (Sagamihara, ISAS), 331

Pilbratt, G. L. 2000, Proc. SPIE 4013, 142

Tamura, M. 2000, in ISAS Report SP 14 Mid- and Far-Infrared Astronomy and Future Space Missions, ed. T. Matsumoto & H. Shibai (Sagamihara, ISAS), 3

Future Far-IR and Submillimetre Space Missions: the European Perspective

Matt Griffin

Department of Physics and Astronomy, University of Wales, Cardiff,
5, The Parade, Cardiff, CF24 3YB, UK

Abstract: The European astronomy and instrumentation community has a long history of participation in far-IR space astronomy satellites including IRAS, ISO, Herschel and Planck. While there are no specific far-IR/submillimetre missions in ESA's current long-term plan after Herschel and Planck, technology and design studies are currently underway, in connection with the IRSI-DARWIN infrared interferometer mission, which are relevant to any future FIR space observatory. In addition, following work on the Herschel, Planck, and NGST satellites, as well as many ground-based, balloon-borne and air-borne instruments, the European instrumentation community will have much to contribute at an international level to the next generation of far-IR/submillimetre missions. This includes the design and construction of large cryogenic focal plane units, optical design and stray light control, filter and quasi-optical component technology, cryogenic mechanisms, and large-format detector arrays for imaging and spectroscopy. Instrumentation heritage and expertise will be complemented by a strong scientific enthusiasm amongst European astronomers to exploit the great astronomical potential of the far-IR and submillimetre range.

1. The European far-IR instrumentation and astronomy community

European astronomers and laboratories have a long record of participation in and far-IR/submillimetre astronomy, both Earth-based and space-borne. This has included a strong partnership role in IRAS and a leading role in ISO. In addition, European scientists and agencies have successfully devoted considerable attention and resources to ground-based submillimetre astronomy through major facilities including SEST, JCMT and IRAM. The field continues to expand with projects currently in development including ESA's Herschel and Planck satellites and major participation by European scientists and institutes in NGST, ALMA, and SOFIA. These major facilities are complemented by many recent and current instrument projects and international collaborations including the SCUBA and MAMBO bolometer cameras, SCUBA-2, the next-generation superconducting bolometer camera for the JCMT, the QUEST cosmic background polarimetric experiment, and existing and planned balloon programmes such as BOOMERanG, Maxima, Archeops, Pronaos, and Elisa. This vigorous instrumentation programme is matched by a substantial and growing community of European astronomers who are active in the field and recognise the growing importance of the far-IR/submillimetre range.

2. The ESA programme

ESA's strategic plan for astrophysics and space science is set out in its *Horizons-2000* programme. This includes the successfully completed ISO mission, the major satellites Herschel and Planck currently in development for launch in 2007, and a substantial involvement in instrumentation for NGST. There are plans for IRSI-Darwin, a major thermal-IR interferometry mission to detect and spectrally characterise Earth-like planets, which would be launched in the 2013 - 15 time-frame. No far IR missions are explicitly included in the present programme after Herschel and Planck. However, there are currently un-allocated (Flexible) mission slots in the programme which are expected to be the subject of Announcements of Opportunity in the future. While these would not be compatible with major European-only far-IR missions, they would offer the prospect of significant participation in international collaborations. The ESA programme is currently being reviewed and redefined, and a clearer picture of possible future mission opportunities is expected to emerge over the course of the next year.

2.1 Herschel and Planck

The Herschel Space Observatory, due to be launched with Planck in 2007, will be the fourth cornerstone mission in ESA's science programme. It will carry out imaging photometry and spectroscopy in 60 - 670 μm range. The key science goals are the detection and investigation of galaxies at high redshift, and the study of star formation and the interstellar medium in our own and nearby galaxies. Herschel will operate in the stable thermal environment of the L2 Lagrange point of the Sun-Earth system, carrying a 3.5-m passively cooled (~ 80 K), low emissivity (< 4%) telescope, with three cold focal plane instruments installed in a superfluid helium cryostat based on the flight-proven ISO design. PACS and SPIRE are direct detection instruments with imaging photometry and spectroscopy. HIFI is a high-resolution heterodyne spectrometer. The operational lifetime of the cryostat will be at least three years. Scientifically and technically, Herschel, together with SIRTF and Astro-F, will make great advances in surveying galaxies out to high redshifts; but, being limited by aperture size and thermal background, will lack the sensitivity and angular resolution needed to study their detailed physics and chemistry in the same way that we can currently do for nearby galaxies. That will be the task of the next generation of FIR missions for which Herschel will pave the way both scientifically and technically.

The Planck satellite will also operate at L2, and is designed to make exquisitely sensitive measurements of the temperature and polarisation anisotropies of the cosmic background radiation. Two instruments, LFI (based on 20-K HEMT radiometers) and HFI (based on 100-mK NTD Ge bolometers) will map the entire sky in ten bands covering the 350 μm - 10 mm range. The mission lifetime will be at least 14 months, during which time two complete surveys of the sky will be carried out.

The Herschel and Planck payload instruments incorporate a number of technologies which will be highly relevant for future missions.

Submillimetre telescopes and associated technology: Herschel's 3.5-m diameter antenna will be fabricated in SiC, and will be an important intermediate step between the 0.5-1-m class telescopes flown so far and the larger apertures needed for future missions. Optical alignment

expertise from Herschel and also the Pronaos balloon experiment will be relevant to NGST and future missions.

Large cold focal plane units: the Herschel instruments have large focal plane units with mass up to 75 kg and dimensions of nearly 1 metre. Multiple temperature stages (at approx. 10, 4, 2 and 0.3 K) are provided with strict constraints on the thermal and mechanical design.

Cryogen-free cooling systems: Planck will be the first satellite to employ a cryogen-free cooling system capable of reaching 4 K, incorporating passive cooling to approximately 50 K, a hydrogen sorption cooler to achieve approximately 18 K, a closed-cycle mechanical cooler based on Joule-Thompson expansion of ^{4}He to provide the 4-K stage. Cooling to 100 mK for the HFI bolometric detectors is provided by a zero-gravity open-cycle ^{3}He-^{4}He dilution refrigerator with another Joule-Thompson expansion to provide an intermediate 1.6-K stage.

Complex opt-mechanical design: The Herschel instruments all incorporate sophisticated optics to maximise their scientific capability in terms of spectral and spatial coverage. The same will be true of the mid-infrared instrument for NGST, for which the optics module will be built by a European consortium.

Cryogenic mechanisms: As in the case of ISO, the Herschel instruments will use low-power cryogenic mechanisms to provide a wide and flexible range of capabilities.

Large-format far-IR/submillimetre arrays: The Herschel direct detection instruments will employ large arrays of ^{3}He bolometers (feedhorn-coupled NTD Ge bolometers for SPIRE; filled arrays of micromachined Si bolometers for PACS), and PACS will also have Ge:Ga photoconductor arrays. Planck HFI will use feedhorn-coupled NTD-Ge bolometers operating at 100 mK with 36 individually optimised channels, 12 of which will employ pairs of polarisation-sensitive bolometers. Whilst none of these technologies are likely to be appropriate for future ultra-low background space-borne instruments (which will probably be based on superconducting detectors), the Herschel and Planck instruments will provide an invaluable base of experience from which to develop the system-level technologies needed for future array instruments in space.

Heterodyne mixers and receiver technology: The HIFI instrument for Herschel incorporates state-of-the art mixers and back-end spectrometers developed in both Europe and the USA.

Stray light control: The sensitivities of Herschel PACS and SPIRE, and of Planck HFI, will be determined by the thermal background from the telescope and any stray light from within or without the cryostat. The instruments, and telescope optical system designs have been carefully optimised and studied to minimise or eliminate potential stray light contributions to the radiation reaching the detectors.

FIR filters and quasi-optical components: The Herschel instruments and Planck HFI will employ state-of-the art filters, dichroics and beam dividers designed to provide high in-band transmission over well-defined pass-bands and high rejection of out-of-band radiation extending

to UV wavelengths. In addition to defining the spectral characteristics of the instrument, these filters serve to control the thermal balance of the FPUs and their various temperature stages.

Submillimetre direct detection spectroscopy: Herschel PACS will carry a cold grating spectrometer and SPIRE will have an FTS survey spectrometer.

Experience gained in building and using these instruments is likely to prove invaluable in designing spectroscopic and imaging instruments for future cold-aperture space telescopes, and this is therefore an area in which European laboratories will be well placed to contribute. It is important to note that four of the five Herschel and Planck instruments involve significant participation from US groups under NASA support, also paving the way for future satellite instrumentation collaborations. Indeed, this is already occurring with Europe planning to provide the optics module for the NGST mid-infrared instrument, MIRI.

2.2 Other European far-IR/submillimetre satellites

ODIN, a mainly Swedish satellite for astronomy and atmospheric research, was launched in 2001. It carries a 1.1-m CFRP telescope and heterodyne radiometers covering 118.25 - 119.25 GHz, 486 - 504 GHz, and 541 - 580 GHz, bands which include key transitions of water and molecular oxygen. The ESA comet rendezvous mission Rosetta, to be launched in 2003, carries a mm/submm radiometer, MIRO, which is being provided by a US/European consortium.

2.3 IRSI-Darwin and SMART-2

IRSI-Darwin is a thermal-IR interferometer mission designed to detect Earth-like planets and carry out spectroscopic observations to search for signatures of biological activity. The Darwin baseline design involves a flotilla of six 1.5-metre diameter cold (~5 K) free-flying telescopes. The signals from the six telescopes are brought together for interference in a central beam-combining spacecraft. An eighth spacecraft located behind the array plane is responsible for control and communications and telemetry to Earth. The Darwin concept is based on a Bracewell nulling interferometer whereby the light from the parent star is nulled to a high degree ($\sim 10^5$) through destructive interference while the light from the planet is not cancelled.

Although not firmly approved at this stage, the mission has been the subject of several internal and industrial studies. Work in preparation for Darwin involves proving a number of key technologies that are relevant to any future FIR interferometer mission. These include:

Precise formation flying and associated techniques: metrology (RF and laser-based), fringe sensing, control systems for deployment, pointing, array re-sizing, path-difference control, and micro-Newton thrusters.

Optics: Achromatic phase shifters, optical delay lines, wavefront filters, amplitude and polarisation matching

Detectors: Quantum well Far IR detector arrays and associated read-out electronics

Design studies are being done by ESA and European industry in preparation for Darwin and the gravitational wave interferometer mission LISA, which also relies on some of the same

technology developments. A two-spacecraft technology demonstrator mission, SMART-2 is currently in development with an envisaged launch date of 2006. The main goals of SMART-2 are demonstration of formation flying (deployment, rotation, slew, resizing), inertial sensor, laser interferometer, and micro-Newton thruster technologies.

LISA is already a NASA-ESA collaboration, and it is conceivable that Darwin and the similar NASA mission TPF may at some stage be combined. Much of the technology needed for interferometer-based missions is generic, whether they be based on free-flying or tethered systems or designed for infrared or gravitational-wave astronomy. Given the mutual interest and specific plans of the two agencies, there is a good rationale for joint ESA-NASA initiatives in the future.

3. European far-IR/submillimetre instrumentation and technology

In addition to the space instruments mentioned in Section 2, European groups are active in the deployment of advanced instrumentation of ground-based, balloon-borne and airborne facilities, with much of this work involving collaboration with US groups. These include NTD-Ge submillimetre bolometer cameras SCUBA (UK), MAMBO (Germany), and BOLOCAM (UK); balloon instruments such as BOOMERanG (Italy, UK), Archeops (France, UK), BLAST (UK), and Elisa (France, UK, ESA-ESTEC); and SOFIA instruments FIFI-LS, GREAT (Germany), and SAFIRE (UK). A key technology for future low-background missions will be the use of large-format arrays of sensitive superconducting detectors. An important step along this road will be SCUBA-2, a two-band submillimetre camera for the JCMT, planned be built by a UK/Canadian/US collaboration with micromachined TES detector arrays and SQUID multiplexers. SCUBA-2 will address many of the array implementation, multiplexing, and system-level issues that must be thoroughly understood before such technology can be deployed in space instruments. The development of 80 x 80 pixel arrays for SCUBA 2 is currently the most ambitious superconducting detector program anywhere.

European laboratories are also continuing to pursue the development of heterodyne components and receivers, including high-frequency hot-electron bolometer mixers (France, Germany, Netherlands, UK, Sweden), integrated receivers with antenna, mixer and LO all integrated on the same chip (Netherlands, Russia, Denmark), and local oscillators (Germany, Ireland), and back-end spectrometers (Sweden, Germany, France).

The technique of millimetre bolometric interferometry, (important for a future long-baseline FIR interferometer mission such as SPECS or for a possible fourth-generation cosmic background anisotropy satellite optimised for B-mode polarisation detection) is being pioneered in the UK. In addition, through operation of the IRAM interferometer and in the future, ALMA, European experience in mm/submm interferometry will be valuable in defining future space interferometry missions.

4. Conclusions

There is a large community of European astronomers active in FIR/submillimetre-based research, and there is intense interest in the single-dish and interferometer mission concepts which are being studied in the US (as is reflected by the large number of European attendees at this workshop). This community has access to, and has flourished through use of, ground-based facilities including JCMT, SEST, and IRAM; satellite missions IRAS and ISO - with Herschel, and Planck, and involvement in Astro-F coming next; and a variety of balloon- and air-borne projects. It is certain that it will regard strong participation by Europe in the further development of far-IR/submillimetre astronomy as essential. European scientists will have much to contribute to technology developments and design studies that will be taking place in the coming years in preparation for future missions. Indeed the European and US communities are already working in a highly integrated and mutually beneficial manner on today's projects. We can expect this to continue and hopefully to be matched by a comparable level of activity and collaboration at space agency level to help realise the full potential of the far-IR/submillimetre band for astrophysics.

The ATACAMA MILLIMETER ARRAY (ALMA)
IMAGING COSMIC DAWN AT Near-THz FREQUENCIES

Alwyn Wootten[1]
email: awootten@nrao.edu
National Radio Astronomy Observatory[2], 520 Edgemont Road
Charlottesville, Virginia, United States of America

The Atacama Large Millimeter Array, or ALMA, is an international telescope project which will be built over the coming decade in Northern Chile. With over 7000 m^2 of collecting area comprised of 64 12m antennas arrayed over baselines up to 14 km in extent, ALMA will provide images of unprecedented clarity and detail. One revolutionary feature of ALMA will be its ability to combine interferometric and single telescope data, providing complete flux recovery. ALMA will cover a spectral wavelength range from 7mm to 0.3 mm or shorter wavelengths, providing astronomy with its first detailed look at the structures which emit Terahertz and subTerahertz photons, the most abundant photons in the Universe.

1. The Atacama Large Millimeter Array

The Atacama Large Millimeter Array, or ALMA, is an international telescope project which will be built over the coming decade in Northern Chile. With over 7000 m^2 of collecting area comprised of 64 12m antennas arrayed over baselines up to 14 km in extent, ALMA will provide images of unprecedented clarity and detail. One revolutionary feature of ALMA will be its ability to combine interferometric and single telescope data, providing complete flux recovery. ALMA will cover a spectral wavelength range from 7mm to 0.3 mm or shorter wavelengths, providing astronomy with its first detailed look at the structures which emit Terahertz and subTerahertz photons, the most abundant photons in the Universe. In this paper we review the history of ALMA, its politics and goals, the chosen site at Chajnantor, the scope of the project, and a few of the technical hurdles and proposed solutions, as well as the ALMA schedule.

1.1. *Goals: Origins of ALMA*

ALMA had its roots in two projects: (1) the Millimeter Array (MMA) project of the NRAO, and (2) the Large Southern Array (LSA) project imagined in Europe.

---

[1]U. S. Project Scientist, ALMA

[2]The National Radio Astronomy Observatory (NRAO) is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

The scientific capabilities required in the U. S. MMA were refined in community science workshops sponsored by the NRAO throughout the decade of the 1980s beginning with the 1985 Millimeter Array Science Workshop, held in Green Bank W. Va.[1]. This process culminated in the 1995 Workshop held in Tucson, Arizona. Five areas of interest were identified as target areas for the MMA:

(i)   Cosmology and Extragalactic
(ii) Star Formation and Stellar Evolution
(iii) Galactic Molecular Clouds and Astrochemistry
(iv) Solar System
(v)   Sun and Stellar

Reports from these workshops may be found with other materials on the ALMA web pages at http://www.alma.nrao.edu. These very different scientific areas all require precision imaging over the subTHz frequency bands observable from beneath the Earth's atmosphere at resolutions from an arcsecond to several hundredths of an arcsecond, well matched to that provided by large optical telescopes (e.g. Gemini, Keck, the Very Large Telescope (VLT) or the Giant Segmented Mirror Telescope GSMT) on the ground, or future space telescopes (e.g. the Hubble Space Telescope, the Next Generation Space Telescope, or SPECS). The specifications for the MMA were chosen to enable it to produce very precise images of the sky at frequencies below 0.35 THz, where atmospheric windows are especially transparent from high sites. However, interest increased in frequencies up to 1 THz as groundbased submillimeter observatories (e.g. the Caltech Submillimeter Telescope, the James Clerk Maxwell Telescope and the Heinrich Hertz Telescope) pioneered investigation into this spectral region. After the 1995 meetings the frequency range of the MMA was expanded to cover the range 0.03 through 0.95 THz.

The scientific requirements for the LSA were similarly discussed and summarized in documents describing "Science at high Z or the youth of the Universe" and "Planetary Formation or the youth of the Solar System." The arguments were summarized at a 1995 meeting in Garching. The LSA science requirements were issued, and in the 1998 April *LSA/MMA Feasibility Study* options for combining the arrays were explored. A primary difference in the concepts for the two arrays was the antenna diameter, with the MMA employing 8m diameter primary antennas and the LSA employing 15m primaries. Early discussions in the U. S. had contemplated complementing the MMA observations with an array of smaller antennas, but for 8m primary diameter and operation below 0.35 THz it was shown that this was not necessary. In November 1998 the Millimeter Array Advisory Committee (MAC) discussed an array with inhomogeneous antenna diameters and the project adopted a primary of 12m in December 1998, driven by a desire for sensitivity (which requires a large collecting area most straightforwardly achieved with larger primaries) but also because of simplicity of an array with a uniform design. There was general realization that wide-field imaging at submillimeter wavelengths would not be optimal but detailed studies were not done to quantify this owing to time pressure.

At a meeting in 1999 February in Tucson, the science requirements were joined. The two projects merged into the ALMA project in 1999 June, resulting in a merger of the similar science needs. The capabilities of the merged array were discussed at the 1999 October Workshop "Science with the Atacama Large Millimeter Array[2]". The scope of the merged array includes 64 antennas of 12m diameter arrayed in configurations with baselines reaching about 14km in extent and including an initial four of ten prospective receiver bands.

In the meantime, Japanese astronomers were discussing plans for a Large Millimeter/Submillimeter Array, also to be located in northern Chile. Discussions on how ALMA might be enhanced through the entry of another major partner continue; for the present Japan pursues an independent design and development phase. The ALMA Science Advisory Committee identified three desirable enhancements at their meeting in September 2001. These include (1) an Atacama Compact Array of about twelve 7m telescopes to improve the high frequency performance of ALMA; (2) a 2nd Generation Correlator of increased capacity; and (3) additional receiver bands to bring the complement of receivers to at least eight. The smaller 7m telescopes might be more accurately figured and provide access to frequencies above 1 THz.

1.2. *Political: International Partners*

North American ALMA partner institutions include the United States National Science Foundation, through its National Radio Astronomy facility operated by Associated Universities, Inc. and the Canadian National Research Council. NRAO works in cooperation with a University consortium including the Owens Valley Radio Observatory of Caltech and the Berkeley Illinois Maryland Association. European partners include European Southern Observatory, The Centre National de la Recherche Scientifique (CNRS), France; The Max Planck Gesellschaft (MPG), Germany; The Netherlands Foundation for Research in Astronomy, (NFRA); Nederlandse Onderzoekschool Voor Astronomie, (NOVA); The United Kingdom Particle Physics and Astronomy Research Council, (PPARC); The Swedish Natural Science Research Council, (NFR); and the Oficina de Ciencia y Tecnologia and Instituto Geografico Nacional (IGN,)(Spain). The National Astronomical Observatory of Japan may join the ALMA consortium soon; Chile, as host nation for the ALMA project, participates through its presence on the ALMA Coordinating Committee, the ALMA Science Advisory Committee and by providing the excellent site high in the Andean Altiplano and support for it.

ALMA has gained widespread support as an instrument astronomy needs to develop astronomical ideas through observation in the twenty-first century. The project has been endorsed by both the 1991 and 2000 Astronomy and Astrophysics Survey Committees of the National Academy of Sciences (US) as among the highest priority items for astronomical facilities to be constructed. In France, at a colloquium at Arcachon in March 1998 organized by CNRS/INSU, French astronomers put construction of ALMA as highest priority for future instruments. In Canada, ALMA was identified as top priority by the National Research Council's Long Range Planning Panel on Canada's Future in Astronomy until 2015. Dutch astronomers likewise established ALMA as the top priority for instrumentation in the coming decade. In the U.K., the Astronomy Vision Panel of PPARC identified ALMA as the highest priority medium-term project.

1.3. *Goals: Science Aims and Drivers*

In the complete electromagnetic spectrum of the Universe, there are three primary peaks. One, the biggest, is the peak from the 3 K blackbody radiation relic of the Big Bang. That peak occurs in the millimeter range of the spectrum, as expected for any black body radiating at such a low temperature. Another peak occurs near one micron–this is the accumulated light from all of the stars and planets in the Universe. But it is the third strongest peak. The second strongest occurs at about 1.5 THz or 200 microns wavelength.

Light of these wavelengths cannot penetrate the atmosphere, as it is absorbed by water and other molecules in the atmosphere–this peak was identified only recently through satellite observations. Alas, with a satellite one is limited as to the size of telescope one can observe with and anything we can currently put into space is far too small to give good images of what it is which is producing this second peak. We know that it must be something at a temperature of about 30 K from its peak frequency. This leads us to suspect that it is emission from all of the stars and planets which ever formed in the Universe, and the cold molecular clouds from which they formed, and from the young galaxies full of dust which hosted those molecular clouds. Recently, some telescopes have made progress in identifying the source of this unknown radiation–much of it–some would claim all of it–appears to come from tremendous episodes of star formation in galaxies at the earliest stage of their creation. The only way to tell for sure is to get clear images of these sources. ALMA provides the only currently practicable way to provide these images.

ALMA will be able to detect thermal emission from nearby large planets directly. ALMA will be able to detect wobbling motion in stars caused by planets orbiting them. ALMA will not directly image earthlike planets around nearby stars–no telescope currently existing or conceived would be able to do that, though many are designing telescopes to approach this goal. ALMA will be able to detect planets in the process of formation. It will be able to detect the chemicals which might have been available for the nourishment of early life (see http://www.cv.nrao.edu/~awootten/allmols.html for a list of the 122 molecules now known).
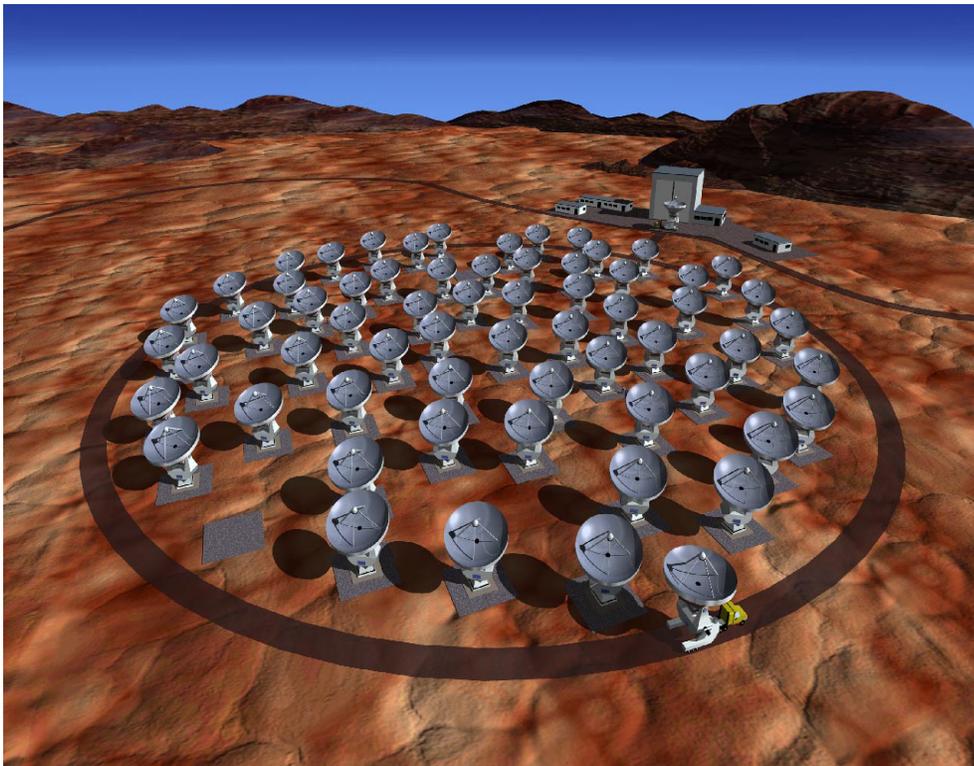
Fig. 1. Artist's rendition of ALMA in its compact configuration. Image courtesy European Southern Observatory.

1.4. *The instrument: the steel, silicon and carbon fiber*

With construction funding beginning in FY2002, ALMA will be built over the coming decade in Northern Chile. ALMA will be a revolutionary telescope, operating at the entirety of the millimeter and submillimeter wavelength band observable from the Earth's most lofty regions. Table 1 summarizes the specifications for ALMA.

Table 1.   Summary of ALMA Specifications

Parameter	Specification
Number of Antennas	64
Antenna Diameter	12m
Antenna Surface Precision	$< 25\ \mu$m rss
Antenna Pointing Accuracy	$< 0.$"6 rss
Total Collecting Area	>7000 m^2
Angular Resolution	0".02 λ (mm)
Configuration Extent	150 m to 14 km
Correlator Bandwidth	16 GHz per baseline
Spectral Channels	4096 per IF
Number of IFs	8

The ALMA specifications are described in more detail in the ALMA Construction Project Book, which like other project details may be found on the Worldwide Web at www.alma.nrao.edu. Table 2 summarizes the sensitivity of ALMA in various modes and at various frequencies.

Table 2.   Summary of ALMA Sensitivities

Band no.	Frequency Range (GHz)^a	Continuum (mJy; 60s)	Line (1 km s^{-1}) (mJy; 60s)
1	31.3–45	0.02	5.1
3^b	84–116	0.027	4.4
4	125–163	0.039	5.1
6^b	211–275	0.071	7.2
7^b	275–370	0.120	10
9^b	602–720	0.849	51
10	787–950	1.26	66

^{a} Frequency ranges as given in *Specifications for the ALMA Front End Assembly*, Draft 1.4, August 31, 2000.
^{b} These 4 bands are the highest priority bands, the others are the second priority bands.

1.5. *Location: the Site at Chajnantor*

In May 1998 NRAO recommended construction of the MMA on a site in Region II of northern Chile which lies on a plain at the foot of three volcanic peaks, Cerro Toco, Cerro Chajnantor and Cerro Chascon. The site (longitude 67d 45m W, latitude -23d 01m S)lies near the Tropic of Capricorn, about 50 km east of the historic village of San Pedro de Atacama, 130 km southeast of the mining town of Calama, and about 275 km ENE of the coastal port of Antofagasta. It lies close to the border with Argentina and Bolivia beside the Paso de Jama road into Argentina. The mean elevation is about 5000m (16400 ft). Several sites had been tested but as scientific interest increased for the highest frequencies, sites at the highest altitudes became favored. Testing of the Chajnantor site began in April 1995 and continues to the present, a joint effort of NRAO, the European Southern Observatory (ESO). The Nobeyama Radio Observatory (NRO) operates a similar testing facility nearby at Pampa la Bola. The testing operations continue with the involvement of the Chilean university community.

Up-to-date details of the monitoring of a number of parameters critical to ALMA's success continue and current details from the instrumentation may be found at the ALMA web site. Some salient characteristics include: the median annual temperature is -2.5^{o}C with annual 50th percentile winds of 10.4 m/s. The average barometric pressure is only 55 percent of the value at sea level. Humidity averages 39% and ultraviolet radiation is about 170% that at sea level. Transparency at .225 THz has been monitored for several years; the 50th percentile zenith optical depth at this frequency is 0.061 corresponding to a column of precipitable water of a little more than 1mm^3. With such a low water column, observations are possible at the atmospheric windows covered by ALMA Bands 9 and 10 (see Table 2) for roughly half of the time. Direct observations of atmospheric transparency including the super THz windows at 1.035, 1.3 and 1.5 THz have been published[4] showing transmission of up to 10% in the highest frequency window.

2. Technical challenges and proposed solutions, schedule

2.1. *The Antenna: Mechanical Engineering at the Heart of the Array*

Each of the two major partners plus Japan have contracted for construction of a prototype antenna meeting the specifications imposed by the scientific goals ALMA will attain. A prototype design for the ALMA antennas will be delivered in July to the NRAO's Very Large Array (VLA) site near Socorro, N.M., for tests to ensure it meets the very demanding specification of this project. VertexRSI of Santa Clara, California is the U.S. contractor for the antenna. A second prototype design will be delivered to the same site in April 2003 by a European consortium including Alcatel, Costamasagna and EIE. A third prototype is expected to be delivered through contract with the National Astronomical Observatories of Japan in Spring 2003. A series of tests will determine the best design; all three designs are the property of the ALMA project. The successful design will become the production design, to be built by the winner of a bidding process.

Fig. 2. An engineering rendering of the 12 meter diameter prototype antenna that is being fabricated by VertexRSI to the very demanding ALMA specifications. The prototype antenna will be delivered to the NRAO Very Large Array site for testing, as shown in this drawing.

The ALMA prototype antenna makes extensive use of carbon fiber reinforced plastic (CFRP) technology in order for the antenna to maintain a stable parabolic shape in the harsh thermal and wind environment characteristic of the ALMA site at the 16,500 feet. The antenna surface accuracy must be better than 25 microns to enable efficient observations at the very highest frequency. Much of the time, ALMA will image fields larger than the primary beam, which means that multiple pointings of the antenna will be fed into a single image. This requires that the antennas maintain a pointing accuracy better than 0.6 arcseconds, despite breezy conditions (50 percentile 6.5 m/s). The Chajnantor site affords no vegetative cover of consequence, so windblown grit and dust will occur and must not degrade the performance of the antenna. At such an altitude, temperature extremes occur. The annual median temperature is below freezing and can range ± 20 C. The ultraviolet

radiation at this altitude is 170% of that at sea level. All these factors provide a challenge to modern antenna design.

As the array is reconfigurable, the antennas must be transportable on an occasional basis by a transporter. The maximum baselines available are on the order of 14km. Current plans are to move antennas on a few-per-day basis along a self-similar configuration of roughly spiral geometry out to the largest configuration, where topography heavily constrains the design. This will provide images with a range of detail.

2.2. *The receivers: The world's largest superconducing array.*

Many laboratories worldwide contribute to the ALMA receiver effort. The receivers will cover the entire observable submillimeter/millimeter spectrum observable from ALMA's superb site. Ten receiver bands will span this region, employing superconducting tunnel junction mixers at 4K with transistor amplifiers operating over a 4-12 GHz range, for frequencies above $\sim$.08 THz. Orthogonal polarizations will be received, to total over one thousand receivers–the most extensive superconducting electronic receiving system in astronomy. Currently, receivers for the lowest two bands are planned to be transistor amplifiers. The complement is rounded out by a system of radiometers operating at .183 THz which will monitor water in the atmosphere for path length correction of the incoming signals.

To cover the observable spectrum in ten bands, radiofrequency bandwidths up to 30% will need to be covered by receivers at some bands. Below .35 THz, the receiver sidebands will be separated to lessen atmospheric noise. Mixers will be balanced to reduce the requirements on local oscillator power.

To achieve this performance, the receivers will be housed in cartridges of a modular design,with all ten bands enclosed within a dewar of roughly 1m diameter and stages cooled to 70, 15 and 4K. For each band, a modular cartridge will be developed which fits into the dewar from the bottom and which is held in place by flexible thermal links. Thus, all bands share the focal plane, and optics and calibration devices will sit atop the dewar.

2.3. *The Correlator: Achieving 1.7 x 10^{16} multiply/add operations per second*

The ALMA Correlator will be located at the Chajnantor site. The analog input from the receivers on 64 antennas, 8 IFs, each of 2GHz bandwidth will be digitized and transmitted at a rate of 96 Gigabits per second from each antenna. The signal travels over fiber optic cables to digital filters, then to the correlator. The correlator must achieve 1.7 x 10^{16} multiply and add operations per second. It cross-correlates signals from 32*63=2016 pairs of antennas on 16 msec timescales; it also autocorrelates signals from 64 antennas on 1 msec timescales, producing 32 Gbyte/s output. The correlator is currently under construction at the Central Development Laboratory of the NRAO on the University of Virginia grounds. Its design offers flexibility of selection of bandwidth, spectral window placement and a reasonably low power requirement of 150 kW.

Correlator design and construction require a particularly long lead time; given the pace of technology prudence requires that ALMA consider the next generation correlator as the current generation correlator is constructed. Therefore, ALMA has fostered the design of the 'G2X Correlator' for possible deployment in the next decade. That design is currently scoped to provide twice as many channels in high dispersion mode as the ALMA Correlator; more than an order of magnitude more in low dispersion mode. It provides support for the Atacama Compact Array (ACA), a set of a dozen 7m antennas designed to enhance the

submillimeter performance of ALMA (The ACA is currently beyond the scope of ALMA but a possible enhancement should Japan join the project). Furthermore, the G2X Correlator concept improved sensitivity through three-bit digitization in all modes rather than in the high dispersion modes only as provided by the ALMA Correlator. The concept envisions twice as many spectral windows (16) as the ALMA Correlator.

2.4. *Schedule*

The current ALMA Schedule is based upon the European construction phase beginning in mid-2002. The first antennas would then arrive in Chile sometime in late 2005. Commissioning observations would then begin late in 2006 and interim operations late in 2007. Under this schedule construction would be finished at the end of 2010 with full operations commencing in late 2011.

1. Alwyn Wootten & F. Schwab "Science with a Millimeter Array" NRAO Workshop Vol. No. 14, Charlottesville: National Radio Astronomy Observatory (NRAO), 1988, edited by A. Wootten and F. R. Schwab.

2. Alwyn Wootten "Science with the Atacama Large Millimeter Array" ASP Conf. Ser. 235, 2001, edited by A. Wootten.

3. S. Radford & R. Chamberlin "Atmospheric Transparency at 225 GHz over Chajnantor, Mauna Kea, and the South Pole" 2000, ALMA Memo 334.

4. S. Matsushita, H. Matsuo, J. R. Pardo & S. J. E. Radford Publ. Ast. Soc. Japan 51 (1999), 603 .

The Herschel Mission

Göran Pilbratt[1], Thijs de Graauw[2], Matt Griffin[3], Albrecht Poglitsch[4]

1   ESA ESTEC, Keplerlaan 1, NL-2201 AZ Noordwijk, The Netherlands
     (gpilbratt@astro.estec.esa.nl)

2   Space Research Organization Netherlands, 9700 AV Groningen, The Netherlands
     (thijsdg@sron.rug.nl)

3   Department of Physics and Astronomy, University of Wales, Cardiff,
     5, The Parade, Cardiff, CF24 3YB, UK (matt.griffin@astro.cf.ac.uk)

4   Max-Planck-Institut für extraterrestrische Physik, Postfach 1312, D-85741
     Garching, Germany (alpog@mpe.mpg.de)

Abstract: The Herschel Space Observatory (formerly known as FIRST) is the fourth cornerstone mission in the European Space Agency (ESA) science programme. It will perform imaging photometry and spectroscopy in the far infrared and submillimetre part of the spectrum, covering approximately the $60 - 670$ μm range. The key science objectives emphasize current questions connected to the formation of galaxies and stars, however, having unique capabilities in several ways, Herschel will be a facility available to the entire astronomical community. The three Herschel payload instruments and their capabilities are briefly described.

1. The Herschel Scientific Payload

Herschel will carry a 3.5 metre diameter passively cooled telescope. The science payload comprises two cameras/medium resolution spectrometers (PACS and SPIRE) and a very high resolution heterodyne spectrometer (HIFI). The focal plane units of the three instruments will be housed in a superfluid helium cryostat. Herschel will be placed in a transfer trajectory towards its operational orbit around the Earth-Sun L2 point by an Ariane V rocket (shared with the ESA cosmic background mapping mission Planck) in early 2007, and will have a minimum operational lifetime of 3 years. Roughly two thirds of the available observing time will be open to the general astronomical community through a standard competitive proposal procedure.

2. HIFI

The Heterodyne Instrument for the Far-Infrared, HIFI, has been optimised to address astrophysical investigations which require high spectral resolving powers and high sensitivity at far infrared and submillimetre wavelengths. HIFI combines the high spectral resolving power of the radio heterodyne technique with quantum noise limited detection from superconductor physics and state-of-the-art in microwave technology. The instrument will have continuous frequency coverage from 480 to 1250 GHz in five bands, and two further bands will provide coverage for 1410-1910 GHz. The resolving power will be up to 10^7 (300 - 0.03 km s^{-1}). Detection sensitivity will be within a factor 3 of the theoretical quantum noise limit, and both polarisations of the astronomical signal will be detected for maximum sensitivity. Calibration accuracy will be within 10%, with a goal of 3%

In order to cover the wide frequency range with high sensitivity, HIFI is designed to have 7 mixer bands and 14 LO sub-bands. The first five frequency-bands will each contain a pair of mixers using SIS junctions. Channels 6-Low and 6-High will contain two mixers based on the recently developed fast hot-electron bolometers (HEBs). HIFI will operate at one frequency at a time in both polarisations. For the Local Oscillator (LO), solid-state varactor/varistor frequency multipliers will be used. HIFI's instantaneous bandwidth of 4 GHz will be analysed in parallel by two types of spectrometers: acousto-optic spectrometers and autocorrelators.

3. PACS

The Photodetector Array Camera and Spectrometer (PACS) is designed to carry out imaging photometry and spectroscopy in the $60 - 210$ µm range. The integral field spectrometer uses an image slicer in conjunction with a cold grating. The spectrometer detectors are two arrays of Ge:Ga photoconductor (stressed and unstressed) with 16x25 pixels each. The spectrometer field of view is 50x50 arcseconds, resolved into 5x5 pixels, with an instantaneous spectral coverage of 1500 km s^{-1} and spectral resolution of 175 km s^{-1}. The photometer has two filled arrays of 0.3-K Si bolometers, one with 16x32 pixels and one with 32x64 pixels. Two bands are imaged simultaneously: $60 - 90$ or $90 - 130$ µm and $130 - 210$ µm, over a field of view of 1.75x3.5 arcminutes, with full sky sampling in each band. Both modes will allow spatially chopped observations by means of an internal chopper mirror with; this chopper will also be used to switch alternatively two calibration sources into the field of view.

4. SPIRE

SPIRE, the Spectral and Photometric Imaging Receiver, uses feedhorn-coupled NTD "spider-web" bolometers operating at 0.3 K. The instrument comprises a three-band imaging photometer covering the 250-500 μm range, and an imaging Fourier Transform Spectrometer (FTS) covering 200-670 μm. The photometer has a field of view of 4x8 arcminutes which is observed simultaneously at 250, 350 and 500 μm, with dichroic beam dividers separating the three spectral bands. Its angular resolution is determined by the telescope diffraction limit, with FWHM beam widths of approximately 17, 24 and 35 arcseconds at 250, 350 and 500 μm, respectively. An internal beam steering mirror can be used for spatial modulation of the telescope beam, and observations can also be made by scanning the telescope without chopping. The FTS has a field of view of 2.6 arcminutes and an adjustable spectral resolution of 0.04-2 cm^{-1} ($\lambda/\Delta\lambda$ = 20-1000 at 250 μm). It employs a dual-beam configuration with novel broad-band intensity beam dividers to provide high efficiency and separated output and input ports.

5. Further Information

Further information on Herschel and its instruments can be found on the Herschel Science Centre website *http://astro.estec.esa.nl/SA-general/Projects/First/first.html*, including links to the instrument web sites and to the proceedings of most recent symposium devoted to the mission: *The Promise of the Herschel Space Observatory*, 12-15 December 2000, Toledo, Spain (ESA SP-460).
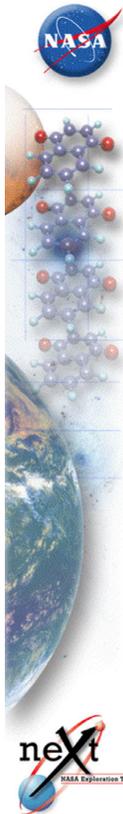
The Lunar L1 Gateway Concept:

Supporting Future Major Space Science Facilities
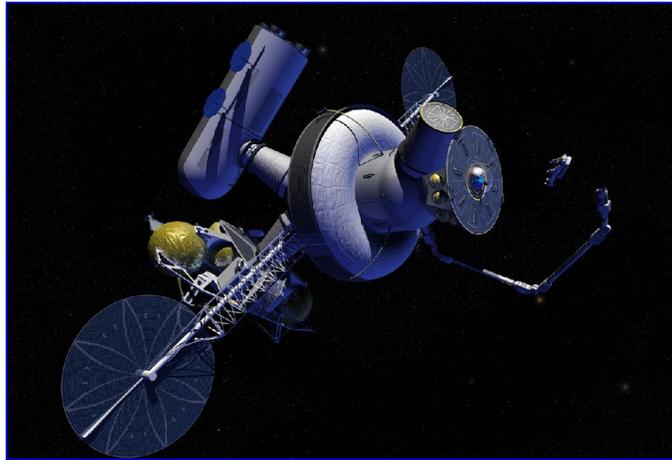
H. Thronson (NASA HQ), J. Geffre (JSC), S. Prusha (JPL),
L. Caroff, C. Weisbin (JPL), the JSC Advanced Design
Team, and the JPL Advanced Projects Design Team
March 2002
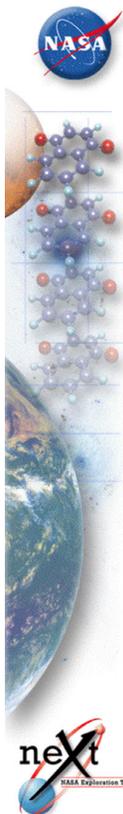
Concepts for Optimized Human/Robotic Support of Advanced Science Facilities

The Challenge :

Ambitious science facilities, such as post-NGST astronomical telescopes, will be extremely difficult to deploy, construct, rescue, service, and repair in space without sophisticated capabilities for manipulation. Such capabilities might include advanced robots, autonomous or remotely-operated systems, and/or humans on-site.

The Goals of This Study :

We report here on a series of ongoing studies to evaluate alternative architectures for future space science facilities and how robots, humans, and autonomous systems might be optimally used to support them.

This presentation outlines one scenario -- a "Gateway" at the Earth-Moon L1 point for supporting multiple options beyond Low Earth Orbit -- plus our process for evaluating human/robotic activities to construct telescopes.

Gateway Architecture

Evaluating the Earth-Moon L1 Point: Options for Exploration Beyond LEO

Crew departs from and returns to ISS

Earth L2

Moon

Mars

Crew Transfer Vehicle
- Transports crew between ISS and Gateway (4-6 day trip)
- Nominal return to ISS with contingency direct Earth return

L$_1$ Gateway
- "Gateway" to the Lunar surface
- Outpost for staging missions to Moon, Mars and science facility construction

Siting a human-occupied "Gateway" at the Earth-Moon L1 point has several advantages in the event that humans are important to support a major in-space science facility:

- After construction, such facilities may be transferred to Earth-Sun libration points (or beyond) with very modest Delta-Vs
- Humans may return to Earth relatively quickly in the event of emergency
- Long-term habitation at this site may be supported relatively easily from Earth
- Capabilities may be developed at this site for longer-term, deeper-space operations while still within short travel-time to Earth

---

LUNAR L1 GATEWAY MISSION PROFILE
[BASELINE CONCEPT]

LEO Operations

Launch Gateway on DELTA IV-H → Activate Critical Systems, Inflate & Checkout Gateway

Launch Shuttle with Gateway Outfitting Crew → Shuttle Rendezvous and Docking with Gateway → Outfit & Checkout Gateway

Launch SEP on DELTA IV-H → Autonomously Deploy SEP Solar Arrays

SEP Autonomously Dock with Gateway

Gateway and SEP spiral to LL1 (unmanned)

LL1 Operations

Crew Arrives at Gateway in CTV

Deliver Lunar Lander to Gateway (unmanned)

Gateway Logistics Resupply / Cargo Delivery (unmanned)

Up to 15 days*

Lunar Surface Mission

30 days

Telescope Mission

30 days

Science Mission

Crew Returns to Earth in CTV

*Reflects crew time spent in Gateway

Far-IR Telescope Concept Construction
[Baseline Concept]

Hardware Support

- Docking for crew transfer vehicle and telescope component delivery module
- SSRMS-class large manipulator
- Small, dexterous robot to aid inspections and assembly/maintenance tasks
- EVA Airlock and teleoperator control station
- Unpressurized partially enclosed work area
- Structure/platform to restrain the telescope during work
- EVA and robotic-compatible storage areas for tools and telescope components

Mission Support

- Complete assembly at Lunar L1: **2 weeks** for **2 teams** of EVA crew; **6-8 EVA sorties**
- For telescope maintenance missions, assume **1 team** of EVA crew for **2 weeks**
- **Total Mission Time at Gateway: 25 days**

Gossamer Struts — Fold Mirror — 10 m trough reflector — Sunshield
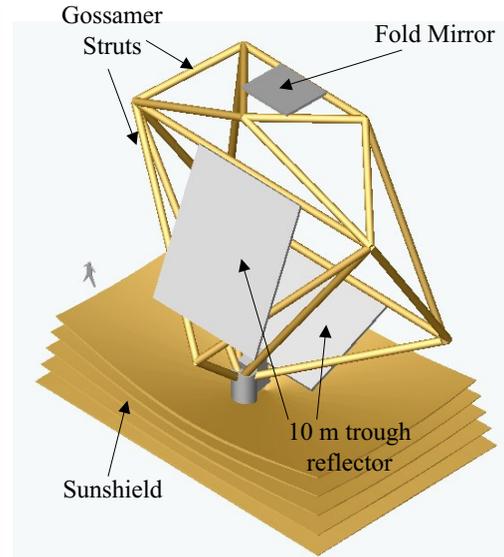
Gateway Concept Summary

- **Destination:** Lunar L1
- **Element Design Lifetime:** 15 yrs
- **Crew Size:** 4 persons
- **Mission Duration:** 10-30 days
- **Element Mass:**
  - **Launch:** 22,827 kg
  - **Outfitting:** 588 kg
  - **Post-outfitting:** 23,415 kg
- **Element Volume:**
  - **Launch:** 145 m^3
  - **Inflated:** 275 m^3
  - **(TransHab: ~340 m^3 for 7 persons)**
- **Power provided:**
  - **Photovoltaic Array:** 12 kW Nominal
  - **Energy Storage:** Li-ion Batteries
- **Support Missions:**
  - **Outfitting at LEO:** One mission/architecture
  - **Human Consumables:** Two missions/year
  - **Life Support resupply:** One mission/two years

Radiators — Docking Ports (3) — Cupola — RMS — EVA Work Platform/ Telescope Assembly Site
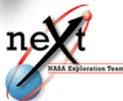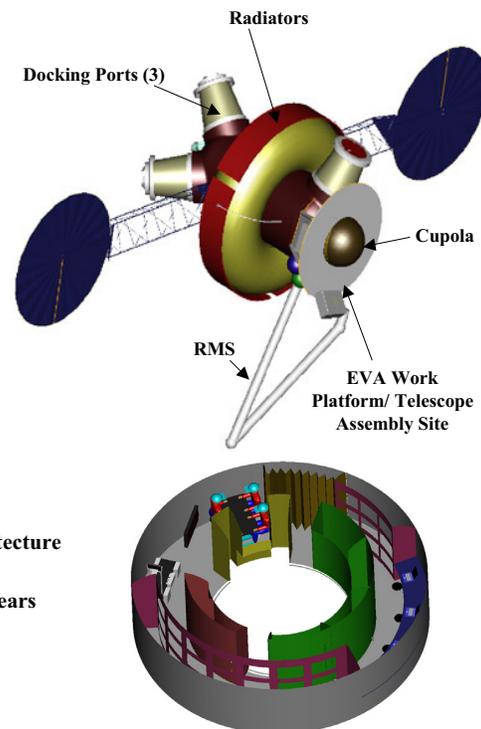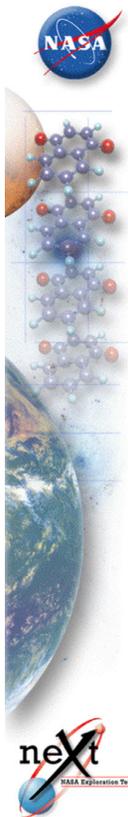
Gateway Configurations

Launch Configuration
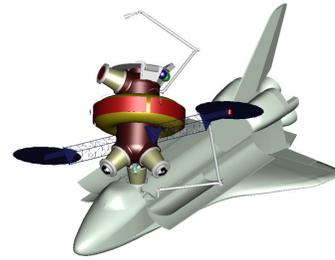
Gateway Outfitting in LEO
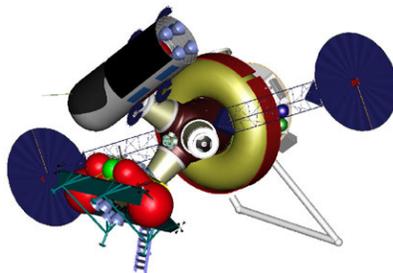
Lunar Surface Expedition
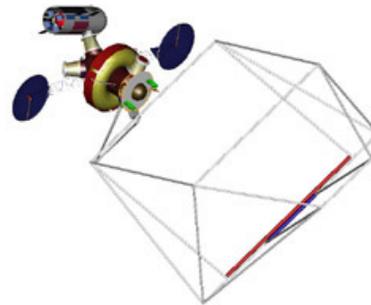
Telescope Assembly Mission

In-Depth Quantitative Analysis to Assess Human-Robot Optimization in Future Space Operations

➢ Relative strengths of humans and robots in performing a wide variety of tasks is well-established CONCEPTUALLY
  ▪ Humans are unequaled in unstructured, unpredictable, innovative scenarios
  ▪ Robots are best at high-risk access, many repetitive tasks

➢ There is much EXPERIENCE to validate these general notions
  ▪ "Rescue" of HST and CGRO, Armstrong's lunar terminal descent maneuver, multiple examples on ISS
  ▪ Robots have gone to "worse-than-hell" places (Venus, Jupiter) not currently accessible to humans

➢ Opinions and hunches about the value of humans/robots in space SIGNIFICANTLY EXCEED in-depth study and formal assessment
  ▪ Need standardized METRICS to quantify performance
  ▪ Need rigorously defined criteria to EVALUATE relative performance

Tools Available for Space Telescope Support

Humans		Robots	
1. EVA Astronaut Pairs of astronauts work in conjunction with robotic agents to assemble space telescopes. Two pairs of two EVA crew assumed on alternating EVAs.		**1. Robonaut** Dexterous anthropomorphic robot to complement human assembly agents. Provides fine motor skills, telerobotically controlled.	
2. RMS Operator RMS controlled from vehicle interior by IVA crewmember. Also controls RMS cameras and Mini-AERCam.		**2. Remote Manipulator System (RMS)** Shuttle/Gateway-based robotic arm for worksite support and payload manipulation	
3. Robonaut Operator Dexterous robot controlled via telepresence equipment. Operator may be IVA crewmember or Earth-based operator.		**3. Assembly Table** Notional concept for aiding telescope assembly. Robotic features may include worksite tilt, rotation, and elevation capabilities.	
4. Mission Control Provides mission support, guidance, and additional problem solving capability. May be used for telerobotic control in conjunction with IVA crewmember control.		**4. Mini-AERCam** Free-flying camera for close-proximity inspection. Controlled by IVA crewmember. Utilizes inert Xenon propulsion system to minimize contamination.	

Performance Case Study Process: Evaluating Options for Humans and/or Robots

"Primitives" [aka, tasks]: Go to Site, Transport Load, Mate Objects . . .

Decompose Scenario: Facility Assembly

Parameters: EVA duration, transport distance, load mass & size, etc.

Quantify Primitive Parameters

Scores per primitive for each H/R system option: data base, thought experiments; models; etc.

Determine "Aptitude"

Compute Composite Scores

EXAMPLE "PRIMITIVE" (TASK): 4.7-AGENT MOVES TO TASK SITE (SHUTTLE-BASED SCENARIO)

Human-Robot (H/R) System Options Modeled:
- Astronauts-Only EVA
- Robot / Manipulator-Only EVA
- Astronauts Ride Transport Vehicle
- Robot-Assisted EVA
- Embedded Robotic Automation

Sum of all the task scores: e. g., probability of success for this H/R option

Antarctica as a Stepping stone to Space

J.W.V. Storey, M.G. Burton and M.C.B. Ashley

School of Physics
University of New South Wales
Sydney NSW 2052
Australia
Email: j.storey@unsw.edu.au

The high Antarctic plateau offers a unique observing environment that is in many ways intermediate between earth and space. In the infrared, for example, the sky brightness is typically 10 to 100 times darker than at "temperate" observatories. The sub-mm transmission is superior to that measured anywhere else on earth. However, perhaps the most important feature of Antarctica is the extraordinary stability of the upper atmosphere, which dramatically reduces scintillation and offers large gains in astrometric precision. In this presentation we will discuss the site-testing results to date, the science that can be best be done, and the technology demonstrators that can best be tested in Antarctica.

1. Introduction

The Antarctic plateau is the coldest and driest place on earth. The highest points are at elevations above 4,000 metres. The atmosphere above the plateau is extraordinarily stable, as there is no jet stream, very little wind, and – at the South Pole – no diurnal variation. There is nowhere else on earth that approximates a space environment better than does Antarctica.

2. What is known about site conditions

For many people, mention of Antarctica conjures up images of icebergs, blizzards and leopard seals. However, none of these things is present on the high Antarctic Plateau, where astronomical site-testing work has concentrated. Dome C, for example, enjoys lower average wind speeds than most – if not all – US cities. At South Pole, the wind speed averages around 6 m/s, and is remarkably constant for days at a time in both speed and direction.

Aurora, while spectacular, do not affect infrared or sub-mm observations.

Temperatures, of course, are extremely low. At South Pole the temperature can drop below –75C, while sites higher on the plateau can experience still lower temperatures.

2.1 Infrared sky brightness

Across the infrared spectrum, modern infrared array detectors are capable of background-limited performance (BLIP). The sensitivity of cameras and spectrometers

at infrared wavelengths is therefore generally determined by the background flux, most of which comes from the atmosphere and the telescope mirrors themselves. On the Antarctic plateau the very low ambient temperatures result in low mirror temperatures, greatly reducing their emission – especially in the near-infrared. Additionally, the sky is both cold and dry, and both these factors result in a lower sky background flux.

Measurements of the near-infrared sky brightness by Ashley et al (1996) and Nguyen et al (1996) showed reductions in background flux of up to a factor of 100 relative to established, temperate-zone observatories such as Mauna Kea. A more comprehensive data set was taken by Phillips et al (1999). These data not only confirmed the earlier results, but also showed that the reductions in sky-brightness were substantial out to at least 5.5 microns.

The region around three microns is one of the two darkest places in the interstellar spectrum (the other is around 300 microns). At 2.4 microns, we find that on occasions the sky brightness can come within a factor of a few of the zodiacal light, as measured by rocket experiments.

In the mid-infrared, Smith & Harper (1998) showed reductions of over one order of magnitude in the sky brightness; a result that has since been confirmed by the more detailed measurements of Chamberlain et al (2000).

Finally, these gains have been demonstrated to be achievable in real astronomical observations, for example by the deep images obtained with the SPIREX telescope.

2.2 Submillimetre transparency and sky stability

Over a decade ago, Townes & Melnick (1990) analysed water vapour measurements from Vostok and predicted that far-infrared/sub-mm conditions the Antarctic Plateau would be extraordinarily good. The first measurements by Chamberlin, Lane & Stark (1997) confirmed that there was extremely good transmission at 492 GHz. These results have been extended to 860 GHz by Chamberlin (2001). More recently, Lay & Halverson (2001) have observed exceptionally low fluctuations in the microwave sky brightness.

2.3 Microthermal turbulence

Marks et al (1996), Marks et al (1999) and Marks (2002) were able to show, from a combination of tower and balloon-borne microthermal measurements, that the atmospheric turbulence above the Antarctic plateau is confined to a very thin boundary layer that extends to just a couple of hundred metres above the ice. More recent acoustic radar (SODAR) measurements (Travouillon et al, 2002) have detailed the extent and distribution of this boundary layer turbulence throughout the entire year.

2.4 Continuous observing
From the South Pole, every astronomical source outside the solar system moves around the sky at constant elevation and is, of course, continuously visible. This confers immense advantages to the study of time-varying phenomena, ranging from helioseismology to exo-planet searches.

For CMBR studies the constant elevation of any patch of sky allows integration over days or even months without need for correction of elevation-dependent ground-pickup or atmospheric emission.

At sites away from the South Pole, such as Dome C, sources suffer only a small diurnal elevation change and many remain circumpolar. The sun, moon and planets are variously observable at particular times of the year, although always at relatively low elevation.

The flip-side of this is that not all the sky is visible, and indeed the area of sky that can be seen is significantly smaller than at temperate locations. Partly compensating for this is that by sheer good fortune many of the most significant astronomical objects are well south of the equator – important examples include the Galactic Centre and the Magellanic Clouds,

3. The synergy between Antarctica and Space.
Antarctic observatories can play an important role in the development of space missions. This role is made up of the following main contributions: astronomical measurements before, during, and after the mission, and technology demonstrators.

3.1 Preliminary measurements
The unique observing conditions in Antarctica make it particularly well suited to carrying out preliminary surveys that can identify potential targets and weed out unsuitable sources.

3.2 Complementary measurements
In some cases it may be possible to perform some aspects of a proposed space mission from the ground, with little or no loss of capability. This enables the space craft to be designed to perform only those measurements that must be done from space. The result can be substantial cost savings, or a re-direction of effort into more mission-critical areas.

3.3 Follow-up measurements
Any successful mission creates more questions than it answers. Earth-based telescopes can follow up these questions immediately, while the next relevant space mission may be years away. This is a powerful argument for having capable Antarctic observatories in place and operating well before the launch of a space craft.

3.4 Technology demonstrators

Between now and the next-generation far-infrared/sub-mm space missions lies the need to develop and validate new technologies. As discussed above, there are many qualities of the high plateau Antarctic sites that render them uniquely suitable for this task. This is particularly so for interferometer experiments, where Antarctica can provide an attractive combination of a cold, stable environment, unlimited space, and a uniquely transparent and stable atmosphere.

4. Final thoughts

At first sight it may seem that the difficulties of building an observatory in Antarctica are a major obstacle. However, this is not the case. It is no harder than, but rather different to, building an observatory at a temperate site. The temperature range encountered throughout the year is about 60C, comparable to the range experienced in some inland US cities. The observatory must simply be designed for a "mean" temperature of around –50C, rather than +5C, and this is a simple matter of following good engineering practice.

While the difficulties of building a telescope in Antarctica may seem obvious, there are considerable logistical advantages. For example, the South Pole is currently the only observatory site in the world that is within a few hundred metres of a heavy-lift airport. (Concordia Station, now under construction at Dome C, will be the second.) Wind speeds on the Plateau are very low, and peak speeds for which structures need to be designed are also correspondingly low. Because the temperature never rises above zero, there is no liquid water and hence no re-freezing problems.

Digging a trench through snow is a lot easier than digging it through rock. Snow is a superb thermal insulator so, a few metres below the surface, the temperature remains remarkably constant year round. In such a trench could be placed interferometer delay lines and other critical optical components, obviating the need for elaborate temperature control systems (Lloyd et al 2002). This, combined with the absence of a high altitude jet stream, makes Antarctica a particularly attractive location for interferometer test-beds that will be needed for SIM, TPF and Darwin.

5. Acknowledgments

6. References

M.C.B. Ashley, M.G. Burton, J.W.V. Storey, J.P. Lloyd, J.W. Briggs and D.A. Harper, South Pole observations of the near-infrared sky brightness, *PASP,* **108**, (1996), 721

M.A. Chamberlain, M.C.B. Ashley, M.G. Burton, A. Phillips, J.W.V. Storey and D.A. Harper, Mid-Infrared Observing Conditions at the South Pole, *Astrophys. J.,* **535,** (2000) 501

R.A. Chamberlin, A.P. Lane and A.A. Stark, A.A. 1997, The 492 GHz atmospheric opacity at the Geographic South Pole, *Astrophys. J.,* **476,** (1997), 428

R.A. Chamberlin, South Pole submillimeter sky opacity and correlations with radiosonde observations, *JGR,* **106,** (2001) 20,101

O.P. Lay and N.W. Halverson, The impact of atmospheric fluctuations on dgree-scale imaging of the cosmic microwave background, *Astrophys. J.,* **543,** (2000) 787

J.P. Lloyd, B.R. Oppenheimer and J.R. Graham, The Potential of Differential Astrometric Interferometry from the High Antarctic Plateau, *PASA,* **18,** (2002), electronic

R.D. Marks, J. Vernin, M. Azouit, J.W. Briggs, M.G. Burton, M.C.B. Ashley, and J.F. Manigault,, Antarctic site Testing - microthermal measurements of surface-layer seeing at the South Pole, *A&A Suppl,* **118,** (1996), 385

R.D. Marks, J. Vernin, M. Azouit, J.F. Manigault and C. Clevelin, C., Measurement of optical seeing on the high antarctic plateau, *Ast. Ap. Suppl. Ser.,* **134**, (1999), 161

R.D. Marks, Astronomical seeing from the summits of the Antarctic plateau, *A&A,* (2002), in press

H.T. Nguyen, B.J. Rauscher, S.A. Severson, M. Hereld, D,A. Harper, R.F. Loewenstein, F. Mrozek, and R.J. Pernic, The South Pole near-infrared sky brightness, *PASP,* **108**, (1996), 718

M.A. Phillips, M.G. Burton, M.C.B. Ashley, J.W.V. Storey, J.P.Lloyd, D.A. Harper and J. Bally, The near-infrared sky emission at the South Pole in winter, *Astrophys J.,* **527,** (1999), 1009

C.H. Smith and D.A. Harper, Mid-infrared sky brightness testing at the South Pole, *PASP,* **110,** (1998), 747

C.H. Townes & G. Melnick, Atmospheric transmission in the far-infrared at the South Pole and astronomical applications, *PASP,* **102,** (1990), 357

T. Travouilon et al, (2002) in preparation.

SOFIA - Current Performance and Future Instrument Upgrades

Göran Sandell [1], Jürgen Wolf [2], and Jacqueline Davidson [1]

[1] USRA[1], NASA Ames Research Center, MS 144-2, Moffett Field, CA 94035
email: gsandell@mail.arc.nasa.gov
[2] DLR, NASA Ames Research Center, MS 144-2, Moffett Field, CA 94035

Abstract

The joint U.S. and German SOFIA project to develop and operate a 2.5-meter infrared airborne telescope in a Boeing 747-SP is now well into development. First science flights will begin in late 2004. Once fully operational the observatory will do 960 observing hours/year, with an expected lifetime of over 20 years. The advantages and drawbacks of an airborne observatory relative to space missions are briefly discussed, and the sensitivity, characteristics and science instrument complement are presented.

SOFIA will take advantage of new technology development, which allows the observatory to upgrade or replace instruments when new technology, for example more sensitive detectors or larger arrays become available. Some examples of possible future instruments are discussed.

Introduction

The development of the SOFIA observatory is now nearing its completion. For a recent summary of the observatory, its characteristics and first suite of instruments, see the Proceedings of SPIE 4014, Airborne Telescope Systems (Melugin & Röser, 2000), in which there is more than 40 papers describing different aspects of the SOFIA observatory. All the instrument flyers can also be viewed on the web (http://sofia.arc.nasa.gov/Science/instruments/sci_instruments.html).

In this short paper we want to emphasize the advantages of an airborne observatory and its unique potentials for implementing new technology due to the long mission lifetime, and the simplicity of changing and upgrading instruments.

The SOFIA Observatory

SOFIA is expected to have a lifetime of at least 20 years, much longer lifetime than a space mission. The Boeing 747-SP aircraft, which was chosen for SOFIA, is big enough to host the 2.5 m telescope and can carry relatively large and bulky instruments (weight limit 600 kg). SOFIA will typically fly one instrument at a time, but it is easy to change, repair, modify and add new instruments. One can even make changes and additions to the telescope, e.g. add tip-tilt system, change guide cameras or the secondary mirror. SOFIA can therefore more readily take advantage of technology improvements, like new, more sensitive detectors, bigger arrays, better filters or grisms etc.

Although SOFIA will fly at an altitude of 12.5 km with a pre-cooled telescope, the atmosphere and telescope still adds a considerable background, and SOFIA is therefore not as sensitive as for ex. SIRTF. On the other hand SOFIA can readily observe bright objects, which may saturate the low background instruments on spaceborn telescopes.

---

[1]Universities Space Research Organisation (USRA), is the prime US contractor for the SOFIA project.

Potential Future Instruments

SOFIA has a suite of nine instruments, which will be available at first light or shortly thereafter. Three of these are facility instruments, i.e. maintained and supported by USRA staff, five are principal investigator instruments, which are supported by the development teams, and one of them (HIPO - an optical CCD camera) is a specialty instrument.

Since SOFIA has no polarimeter, one of the first new instruments is likely to be a facility imaging polarimeter. Neither does SOFIA have a high resolution FIR spectrometer, although such an instrument was planned to be in the first instrument suite. One can also expect that detector technology in the mid- and far-IR regime will continue to produce more sensitive detectors and larger arrays. One can therefore expect that SOFIA will eventually have a FIR camera utilizing the whole FOV (8 arcmin). The rapid development in heterodyne mixer and local oscillator technology is also expected to continue and we therefore expect to get "broadband" heterodyne receivers that operate below 100 μm, and even imaging heterodyne systems. After 2010 SOFIA may also have a new secondary, which could give a 1-degree FOV in the cavity focus. One may therefore envision an IR imager that would take advantage of this large FOV.

References

[1] Melugin, R.K., Röser, H.-P., 2000, Airborne Telescope Systems, SPIE, Vol. 4014

COSMOG - Cosmology Oriented Sub-mm Modeling Of Galactic foregrounds

A. Kashlinsky[1] and D. Leisawitz[2]

[1]SSAI, Code 685, Goddard Space Flight Center, Greenbelt, MD 20771

[2] Code 630, Goddard Space Flight Center, Greenbelt, MD 20771

With upcoming missions in mid- and far-IR there is a need for software packages to reliably simulate the planned observations. This would help in both planning the observation and scanning strategy and in developing the concepts of the far-off missions. As this workshop demonstrated, many of the new missions are to be in the far-IR range of the electromagnetic spectrum and at the same time will map the sky with a sub-arcsec angular resolution.

We present here a computer package for simulating foreground maps for the planned sub-mm and far-IR missions, such as SPECS (Leisawitz et al 2002). The package allows to study confusion limits and simulate cosmological observations for a specified sky location interactively and in real time. Most of the emission at wavelengths long-ward of $\sim 50\mu$m is dominated by Galactic cirrus and Zodiacal dust emission. Stellar emission at these wavelengths is weak and is for now neglected. Cosmological sources (distant and not-so-distant) galaxies for specified cosmologies will be added later.

Briefly, the algorithm goes through the following steps:

• A template map at 100 μm and IRAS angular resolution ($\sim 6'$) uses the data of the Galactic dust emission from Schlegel, Finkbeiner & Davis (1998, hereafter SFD). The maps are based on the SFD detailed reanalysis of COBE/DIRBE and IRAS datasets.

The templates serve to normalize the to-be-observed region of the sky to IRAS data for the cirrus emission at 100 μm . IRAS data are pixelized with $2.37' \times 2.37'$ pixels and the beam FWHM is $6.1'$. The extra- and inter-polation will be to the required angular scales and wavelengths.

• The user can then specify the following parameters for the output maps:

1. the required angular resolution (an interferometer baseline or a filled-aperture diameter);

2. the required pixelization;

3. the requested sky location or Galactic (b, l) coordinates;

4. the field of view;

5. the wavelength of the observation; and

6. an option to add the zodiacal light component for the desired solar elongation.

• The extrapolation to the requested wavelength is done using the template Galaxy spectrum from Fixsen et al (1998). The Galaxy spectrum is essentially a dust modified black body of $T_{dust} \sim 20K$ and the emissivity index $\alpha \simeq -1.7$.

• The zodiacal light spectrum and intensity is taken from COBE DIRBE data (Hauser et al 1998). It is evaluated at the specified Solar elongation angle and wavelength using inter/extra-polation of the the DIRBE zodiacal emission model (Kelsall et al 1998). The angular structure of the zodiacal light emission is not modeled as it is expected to be very small.

• The smaller scale structure of the Galactic cirrus is then generated using Fast Fourier Transforms (FFT) normalized to the coarser IRAS/DIRBE maps.

First the power spectrum of the IRAS sky in the vicinity of the chosen sky location is evaluated. Cirrus is extrapolated to smaller angular scales assuming Gaussian random process and using the spatial spectral index specified by the user. The spectral index indicated by the current observations is $n \sim -2 - 3$ (Gautier et al 1998). The angular structure of the simulated field is normalized to the IRAS power spectrum and large-scale harmonics from the IRAS sky are kept on the relevant angular scales modified by the coupling matrix resulting from transition from coarser IRAS resolution to the user requested pixels.

• The program computes the confusion noise parameters with and without the zodiacal component in the given direction and produces maps with the specified angular resolution and pixelization.

• For a 512×512 pixel field of view the maps are generated in only $\sim 1 - 2$ secs CPU time on

a 300 MHz computer.

Fig. 1 shows the histogram of the 512×512 pixels patch in the Galactic plane $(b_{\text{Gal}} = 1°, l_{\text{Gal}} = 1°)$. The size of the pixel is $\simeq 0.3°$. No zodiacal component has been added and the slope of the cirrus power spectrum was chosen to be $n = -2$. The emission has been "observed" by a SPIRIT-like instrument (interferometer baseline of 30 meters) at $\lambda = 100\mu$m.
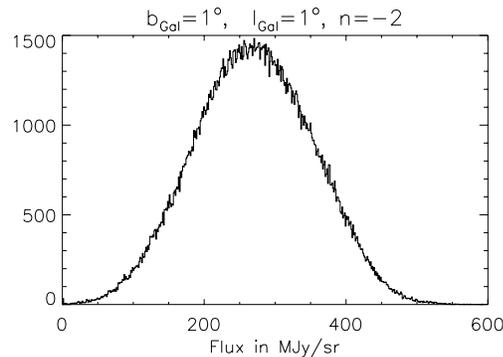
Fig. 1.—

Figure 2 shows the maps of putative observations by SPIRIT instruments of 3 patches at different Galactic locations assuming different slopes of the small scale cirrus power spectrum. The middle panels have Zodiacal component added at Solar elongation $e_\odot = 119°$, the maximal in that direction.

Figure 3 illustrates the dependence on the wavelength of observation for $n = -2.5$ and Galactic patch of 512×512 $0.3''$ square pixels at $(b_{\text{Gal}}, l_{\text{Gal}}) = (-45°, 150°)$. The interferometer resolution is $\simeq 1.2\text{baseline}/\lambda$ and the longer wavelengths will have super-arcmin resolution even with SPIRIT.

Finally Figure 4 shows the dependence on the baseline of the interferometer for observations at $\lambda = 1$ mm and assuming $n = -2.5$. Clearly the higher resolution of SPECS would be better for observations in the presence of Galactic confusion.

REFERENCES

Hauser, M. et al 1998,Ap.J.,508,25
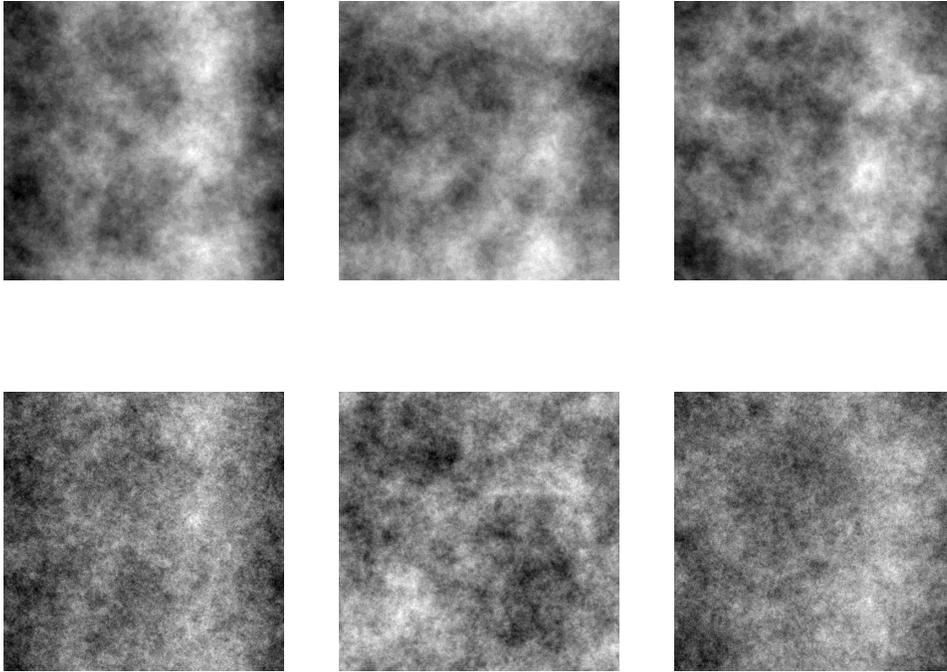
Fig. 2.— Top panels correspond to $n = -3$ and lower panels to $n = -2$. The panels from left to right correspond to $(b_{\mathrm{Gal}}, l_{\mathrm{Gal}}) = (1°, 1°), (1°, 100°), (80°, 200°)$.

Gautier, T.N. et al. 1992,AJ,103,1313

Kelsall, T. et al. 1998, Ap.J., 508,44

Leisawitz, D. et al. 2002, astro-ph/0202085

Schlegel, D. J., Finkbeiner, D. P. & Davis, M. 1998, Ap.J.,500,525 (SFD)
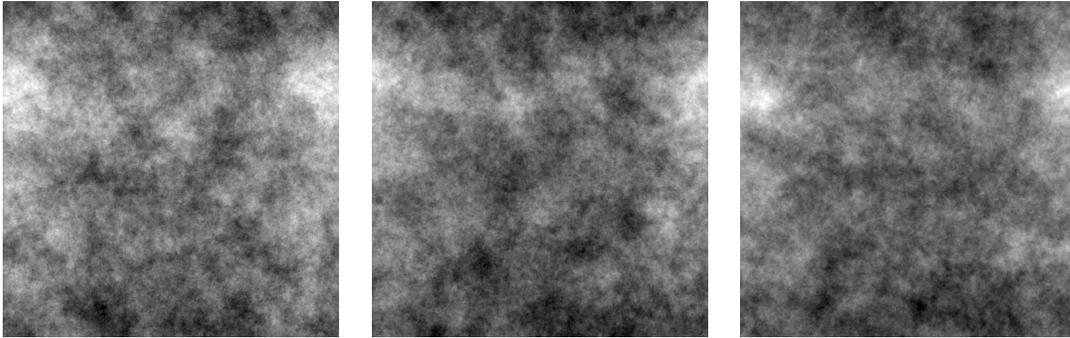
Fig. 3.— Same patch with $n = -2.5$ but observed with SPIRIT at 100, 500 and 1000 μm from left to right.
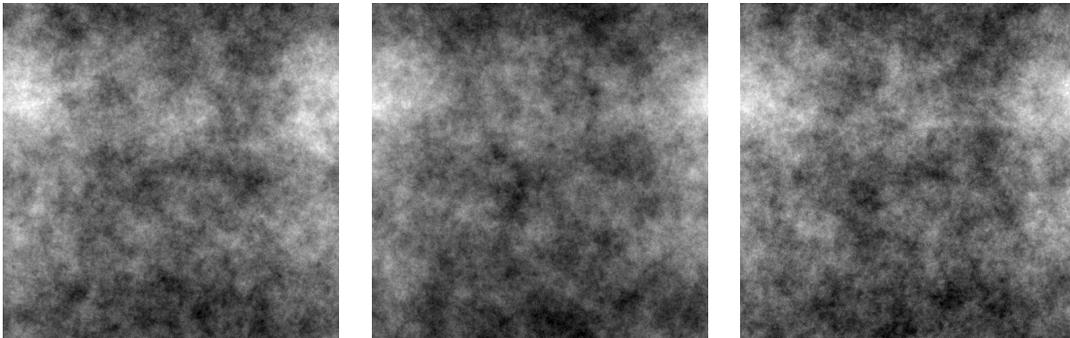
Fig. 4.— Same patch "observed" with space interferometer with baseline = 100, 300, 1000 meters from left to right.

APPENDIX

Below are shown examples of the IDL procedure interface and results of the runs:

```
IDL

>.r cosmog Background data files must be loaded ONLY in the first IDL run

Were the background data files file loaded?  Yes/No - y

Were the background data files file loaded?  Yes/No - y

choose Galactic latitude of your area:  bgal = -85

choose Galactic longitude of your area:  lgal = 10

enter wavelength in micron:  lambda = 100
```

Would you like to add zodiacal component to Galactic cirrus? Yes/No - n

Default map shows a square patch FOV of N=4 on the side - i.e. 4x4 IRAS pixels (50.55 arcmin2)

Would you like to change it? Type: Yes/No - y

Choose the new N (must be power of 2): 16

Choose to use either filled aperture telescope or interferometer

Is it interferometer? Yes/No - ? y

SPIRIT is the default baseline (30 m) or 3.44*lambda(mm) arcsec resolution

Would you like to change it? Type: Yes/No - n

Your resolution is: 0.343775 arcsec

Default is Npix=512 pixels per FOV: i.e. pixel = (4.44375 arcsec)2

Would you like to change it? Type: Yes/No - y

Choose the Npix from [...64,128,256,512...] (CPU time goes like Npix ln Npix) 128

Your new pixel = (17.7750 arcsec)2

choose cirrus spectral index (-3 < ns < -2) ns = -2.

The maximal, minimal, mean and median flux values are =

Fmax = 0.061137522 MJy/sr

Fmin = 0.042085866 MJy/sr

Fmedian = 0.0515721 MJy/sr

Fmean = 0.051615465 MJy/sr

Fzodi = 0.000000 MJy/sr

mean Fcirrus = 0.92955385 MJy/sr

```
sigmaF = 0.0030275102 MJy/sr

Specify window number for histogram:  nw = 3

The default values for flux in the map are Fmax, Fmin

Would you like to change the flux limit in the map?  Yes/No - nb

Default map is always shown as (512,512) array.

Specify window number for 512x512 map:  nw = 4

Would you like to show the map in log?  Yes/No - n
```

MAXIMIZING SCIENCE CAPABILITY FOR FAR-INFRARED SPACE MISSIONS

Dominic Benford, Dave Leisawitz, Harvey Moseley, Johannes Staguhn[†], George Voellmer

NASA / Goddard Space Flight Center   †SSAI

ABSTRACT

The far-infrared and submillimeter region (20μm–800μm) has perhaps the greatest potential of all wavelengths for advancement in astronomy. When viewed in terms of the cosmic backgrounds, the far-IR is extremely important: half of the total luminosity in the Universe is emitted at rest wavelengths ~80–100 μm. At the highest known galaxy redshifts (z~6) this energy is redshifted to ~600 μm. Existing and planned missions have a broad range of capabilities, defined in terms of their spectral coverage, spectral resolution, angular resolution, mapping speed, and sensitivity. In this 5-dimensional parameter space, the far-IR is substantially be-hind most other wavelength bands. The opportunity for future missions with great discovery potential is evident. Such missions will be well-suited to answering fundamental questions about the history of energy release in the Universe, the formation and evolution of galaxies, and formation of stellar and protoplanetary systems. We discuss the parameter space that can be filled by a few well-chosen space missions, specifically a submillimeter all-sky survey and a far-IR to submillimeter observatory. Ultimately, a long baseline submillimeter interferometer is necessary to provide sensitivity and angular resolution.

INTRODUCTION

Your experience – if you are like most FIR/Submm astronomers – will fail you.

The majority of astronomers who observe in the FIR/Submm are familiar with the difficulties involved in using ground based (CSO, JCMT, HHT, etc.) telescopes and airborne (KAO, balloon) platforms. Much of the difficulty involved in making sensitive measurements is purely due to the weather and the warm telescope you're looking through. Even in space (Herschel), the telescope emission alone can be a tremendous degrading influence on the sensitivity of observations. Contrary to what this experience tells you, the FIR/Submm is the darkest region of the spectrum, as shown in Figure 1.
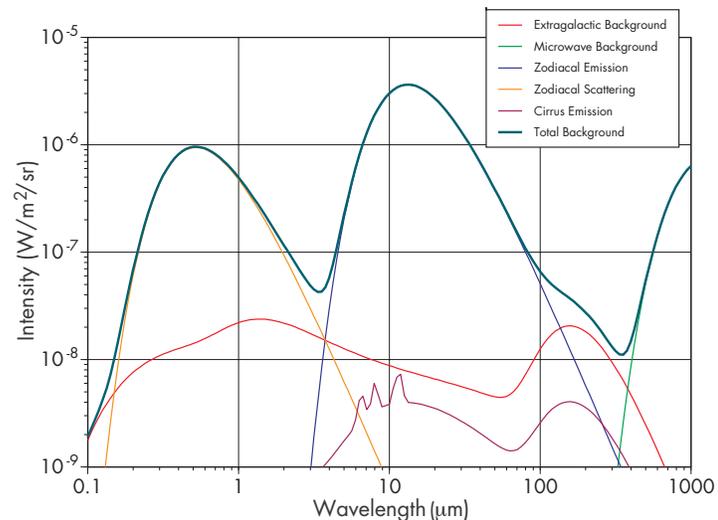
Figure 1. Extragalactic background light components; the darkest regions of the spectrum are at ~3µm and ~300µm.

Furthermore, the light from galaxies - shown in Figure 2 is a spectrum of the Milky Way - happens to be peaked at very nearly the darkest two regions of the UV to radio spectrum.

The ~3μm and ~300μm infrared windows on the Universe - particularly the 100-500μm band - are vastly underexploited as compared to other spectral regions.

Part of the reason for this is that any telescope must be cooled until its thermal emission is less than that of the background; this can require a 4K telescope for the longest wavelengths, as shown in Figure 3.
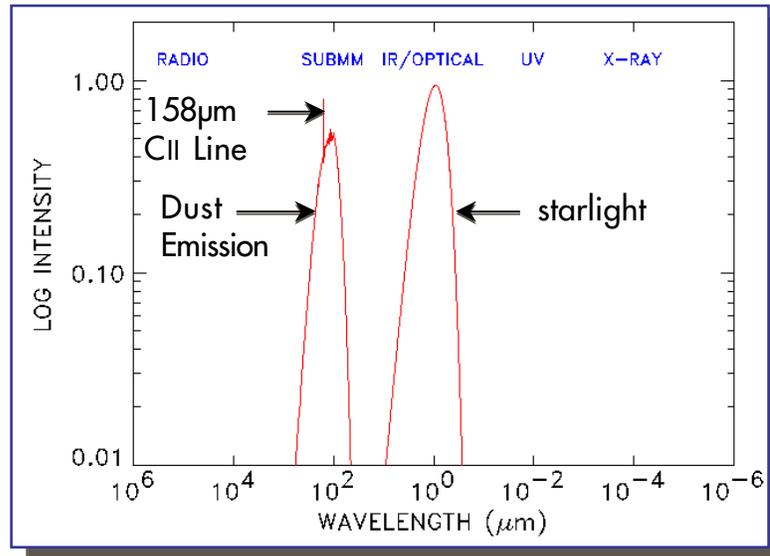
Figure 2. Spectrum of the Milky Way from radio to X-ray wavelengths.

In order to image to a given sensitivity limit, telescopes need both angular resolution and collecting area. At radio wavelengths, widely separated, small telescopes are needed to achieve both; in the optical, a single mirror suffices. The dividing line is near 100μm, as shown in Figure 4. In this region, both filled aperture and interferometers are useful; in fact, the optimal telescope might well be a partially-filled single aperture telescope.
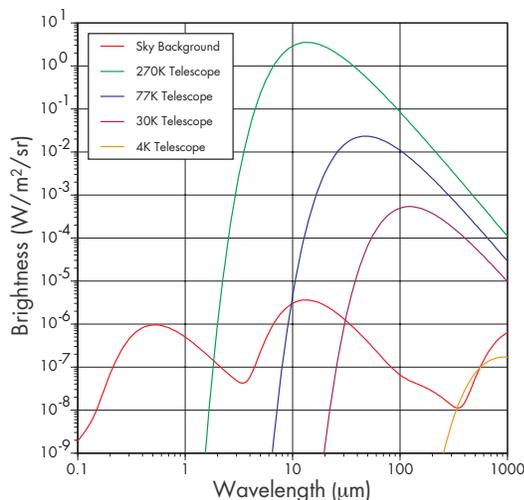
Figure 3. Brightness of the sky compared with that of a 5% emissive telescope.
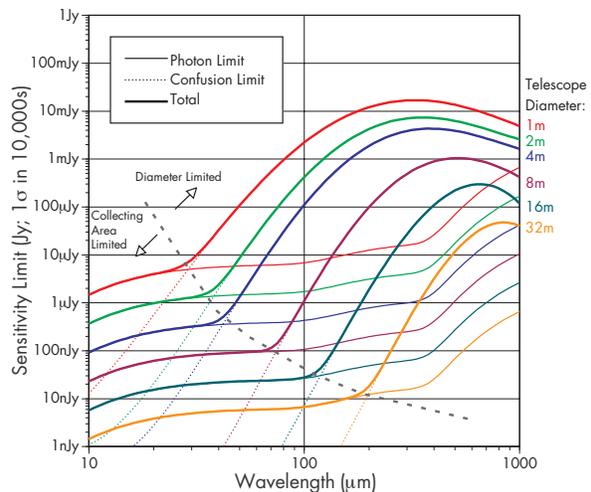
Figure 4. Sensitivity of telescopes as a function of diameter; longer wavelengths need angular resolution whereas shorter wavelengths require collecting area.

POSSIBLE FUTURE FAR-INFRARED MISSIONS:

We consider the potential role played by four mission concepts for far-IR space-based telescopes:
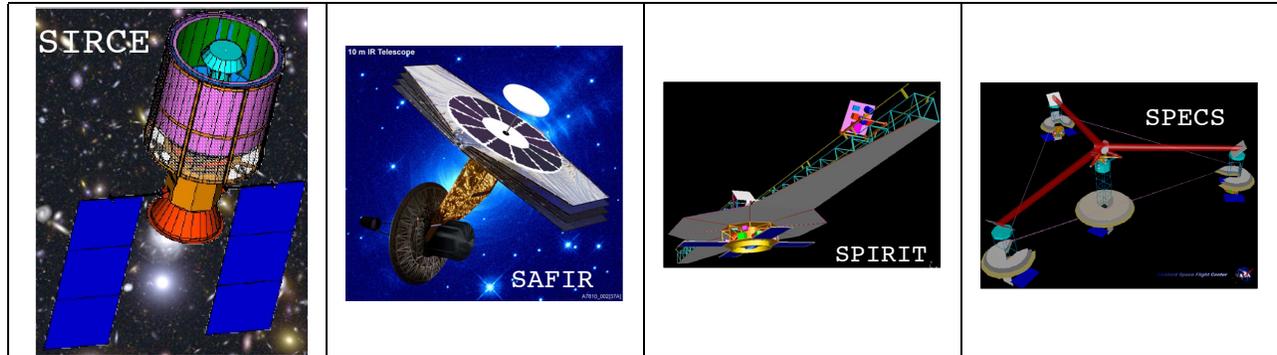
SIRCE: The Survey of Infrared Cosmic Evolution is a 2m-class cryogenic telescope with rapid broadband imaging capability optimized for an all-sky far-IR survey; it can see ultraluminous galaxies to z~5.

SAFIR: Single Aperture Far IR telescope, recommended by the Decadal Review, is envisioned as an 8m-class observatory for the far-IR. It will provide unprecedented sensitivity at wavelengths longward of $30\mu m$ to complement NGST.

SPIRIT: Space IR Interferometric Telescope is a scientific & technical pathfinder for far-IR interferometric imaging.

SPECS: The Submillimeter Probe of the Evolution of Cosmic Structure is a 3-element, long baseline interferometer capable of achieving sensitivity and resolution comparable to HST in the optical, NGST in the near-IR, & ALMA in the radio.

OBSERVATORY PARAMETER SPACE

Five key parameters for observatories are: spectral coverage, spectral resolution, angular resolution, mapping speed, and sensitivity. Different science investigations require more performance in some parameters than in others, so it is useful to compare missions in this parameter space to see what capabilities exist and which are neglected. Here, we shall plot all parameters versus wavelength, which illustrates spectral coverage. Four planned missions - NGST, SIRTF, Herschel, and ALMA - shall be used as demonstrative of the existing capability. (See Thronson et al. 1995 for similar discussions)

SENSITIVITY:

This is often considered to be the most important criterion for new observatories. If we look at the existing missions (Figure 5; in black), we see that a large gap exists in the $30\text{-}300\mu m$ range. Note that warm telescopes (e.g., SOFIA, JCMT, CSO) are typically at or off the top of this scale. The four possible future missions fill in this gap to yield comparable sensitivity across the visible to radio spectrum. SAFIR should be flown within a decade in order to fill in the crucial gap between NGST and ALMA.
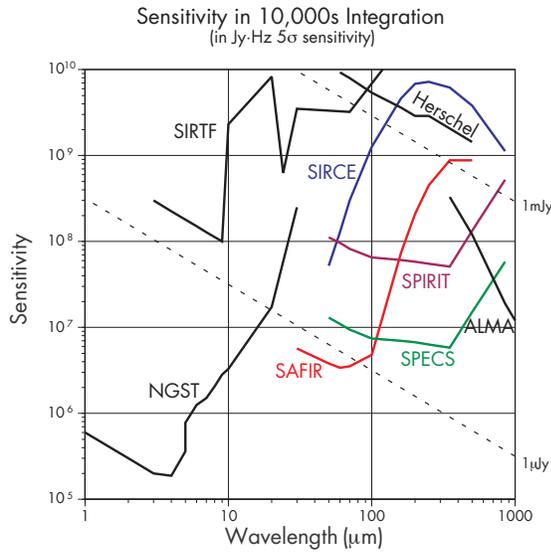
Figure 5. Sensitivity in Jy·Hz (5σ) in a 10,000s integration. Flux limits are shown as diagonal lines.
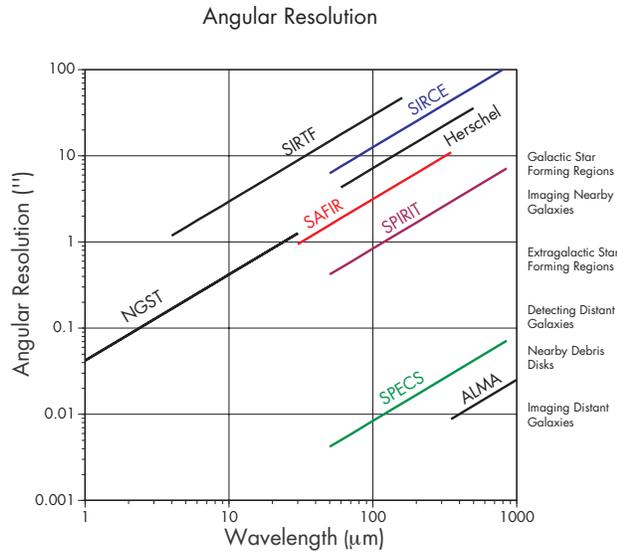
Figure 6. Angular resolution for a variety of single aperture and interferometric telescopes.

ANGULAR RESOLUTION:

Many science investigations require high – often sub-arcsecond – angular resolutions (Figure 6). NGST will provide this capability at λ<28μm, while interferometers such as ALMA and the SMA will provide it at submillimeter wavelengths where the atmosphere permits observations. In the 30-300μm regime, a long-baseline interferometer like SPECS will ultimately be required. The Decadal Review indicates such a mission should be started at the end of the decade.

MAPPING SPEED:

This parameter combines sensitivity and field of view to represent how quickly an area can be imaged. Although the angular resolution of SPECS and ALMA is unmatched, their small field of view means that these observatories will never conduct blind searches for objects (Figure 7). An all-sky survey such as SIRCE would be an ideal scientific precursor for ALMA, SAFIR, and SPECS.
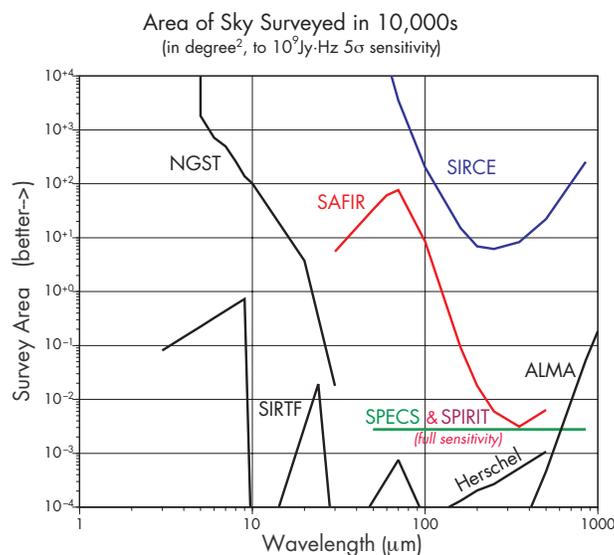
Figure 7. Mapping speed (survey area to a fixed sensitivity of 10^9 Jy·Hz, 5σ, in 10,000s). This comparison is necessarily somewhat qualitative.

SPECTRAL RESOLUTION:

If we momentarily disregard Herschel (which has a capable instrument complement, but is not extremely sensitive due to its warm telescope), it is clear that both continuum (R<10) and high resolution (R>1000) instruments do not exist for 30-300μm. Any of the four possible missions can help to fill this gap. The design of the instrument package in each mission will determine the spectral resolution and wavelength coverage. SAFIR is a likely candidate for very high resolution (R~10^6) spectroscopy.
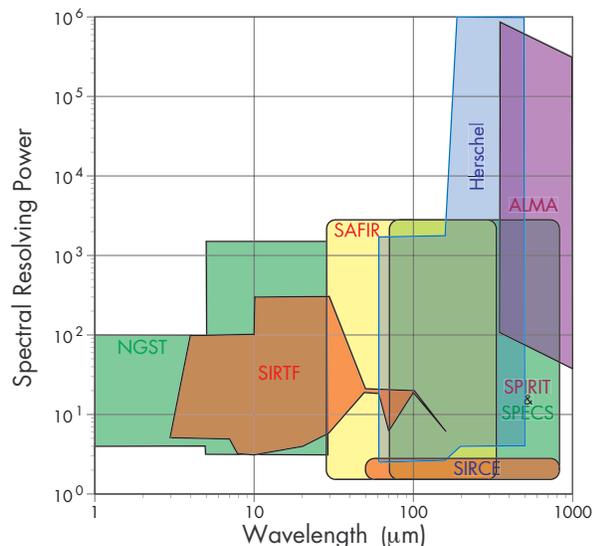
Figure 8. Spectral resolution for present and possible future missions.

CONCLUSION

To quote John Mather (these proceedings), "we must do everything simple and cheap before we can do anything complicated and costly . In this sense, we must plan on a logical mission line that ranges from simple to complicated, cheap to costly. At the same time, we should seek to gain scientific knowledge and technical experience so that each successive mission is better and more cheaply executed.

The parametric view of missions presented in this paper is not supposed to replace the detailed scientific cases for each of these missions, but rather to bolster them. There is so much yet unknown about the universe as seen with far-infrared eyes that it seems impossible to predict accurately what will be discovered by a far-future mission such as SPECS. Therefore, it makes sense to design missions that will maximize the amount of new capability present in the far-infrared as these new missions are built.

As a summary of the findings, it seems a logical mission progression is evident: in the coming decade, SIRCE and SAFIR should be flown, followed closely by SPIRIT as a predecessor to SPECS.

REFERENCES

Mather, J.C. 2002, "Complementarity of NGST, ALMA, and far IR space observatories", these proceedings

Thronson, H.A., Rapp, D., Bailey, B., & Hawarden, T.G. 1995, "Ecological Niches in Infrared and Sub-Millimeter Space Astronomy: Expected Sensitivity as a Function of Observatory Parameters", *PASP*, 107, p.1099

Session V

Instrumentation Issues

WaFIRS, a Waveguide Far-IR Spectrometer:
Enabling Space-Borne Spectroscopy of High-z Galaxies in the Far-IR and Submm.

C.M. Bradford[1], J.J. Bock[1,2], M. Dragovan[2],
L. Earle[3], J. Glenn[3], B. Naylor[1], H. Nguyen[1,2], J. Zmuidzinas[1]

[1]Mail Code 320-47, California Institute of Technology, Pasadena, CA,
[2]Jet Propulsion Laboratory, Pasadena, CA,
[3]CASA, University of Colorado, Boulder, CO

Author Contact: bradford@caltech.edu

ABSTRACT

The discovery of galaxies beyond z~1 which emit the bulk of their luminosity at long wavelengths has demonstrated the need for high-sensitivity, broadband spectroscopy in the far-IR/submm/mm bands. Because many of these sources are not detectable in the optical, long-wavelength spectroscopy is key to measuring their redshifts and ISM conditions. The continuum source list will increase in the next decade with new ground-based instruments (SCUBA2, Bolocam, MAMBO) and the surveys of HSO and SIRTF. Yet the planned spectroscopic capabilities lag behind, primarily due to the difficulty in scaling existing IR spectrograph designs to longer wavelengths.

To overcome these limitations, we are developing WaFIRS, a novel concept for long-wavelength spectroscopy which utilizes a parallel-plate waveguide and a curved diffraction grating. WaFIRS provides the large (~60%) instantaneous bandwidth and high throughput of a conventional grating system, but offers a dramatic reduction in volume and mass. WaFIRS requires no space overheads for extra optical elements beyond the diffraction grating itself, and is two-dimensional because the propagation is confined between two parallel plates. Thus several modules could be stacked to multiplex either spatially or in different frequency bands. The size and mass savings provide opportunities for spectroscopy from space-borne observatories which would be impractical with conventional spectrographs. With background-limited detectors and a cooled 3.5 telescope, the line sensitivity would be better than that of ALMA, with instantaneous broad-band coverage. We have built and tested a WaFIRS prototype for 1-1.6 mm, and are currently constructing Z-Spec, a 100 mK model to be used as a ground-based $\lambda/\Delta\lambda\sim350$ submillimeter galaxy redshift machine.

1. Scientific Motivation

The advent of large-format bolometer arrays for wavelengths around 1 mm (SCUBA, MAMBO) had revealed a new class of galaxies which are likely at medium to high redshift. These sources are cosmologically significant – their counts reproduce much of the diffuse far-IR / submillimeter background radiation, representing the energy generated by all galaxies over the history of the universe (Blain et al.1999, Barger et al, 1999). These submillimeter galaxies are luminous systems similar to the nearby IR galaxies discovered with IRAS. Of the nearly 200 submillimeter galaxies discovered thus far, only a small fraction have confirmed spectroscopic redshifts and well-determined properties at other wavelengths. This is because the sources are very dusty with high extinction at short wavelengths, the optical and UV energy is almost entirely reprocessed and reradiated between $\lambda = 50$ μm and 1 mm, making the optical counterparts too faint to be detectable. While there are spectral features that could be used in the millimeter / submillimeter, the instantaneous bandwidth of heterodyne millimeter-wave receivers is currently a small fraction of unity, so searching for lines in sources with unknown redshifts is impractical.

The long-wavelength continuum source list will only increase in the next decade with new ground-based instruments (SCUBA2, Bolocam) and the confusion-limited surveys of HSO and SIRTF. The recently discovered submillimeter galaxies, and their soon-to-be-discovered far-IR cousins demonstrate the need for broad-band spectroscopy in the far-IR / submillimeter / millimeter bands. Long wavelength spectroscopy with large instantaneous bandwidth is the key to measuring these sources' redshifts, which constrain their luminosities, sizes, and masses. Moreover, the wide variety of spectral features in the mid- and far-IR provide information on the conditions in the interstellar medium and contrain the luminosity source(s). The types of spectral features include:

Fine Structure Lines.   Species include Ne^+, S^{++}, Si^+, C^+, C^0, O^0, O^{++}, with luminosities from 10^{-4} to 3 x 10^{-3} of the total bolometric luminosity. These lines also measure the gas conditions and UV field properties in regions where stellar or AGN luminosity is input into the ISM. Mid- and far-IR fine structure lines have been used to study the starburst conditions in nearby galaxies (Stacey et al., 1991, Lord et al, 1994, Colbert et al. 1999, Carral et al. 1994, Malhotra et al, 2001) and recently, in as an AGN / starburst discriminator in ULIGs (Genzel et al, 1998).

PAH Bands. These features are very prominent in starburst systems, a well-studied redshift template with luminosities typically 1-4 % of L_{bol}. (see Helou et al., 2001, Tran et al., 2001.)

Molecular Rotational Transitions. Millimeter and submillimeter CO rotation is the dominant coolant of molecular gas and a probe of its temperature and density. Mid- and high-J lines trace warm gas associated with UV or shock heating of dense molecular gas (Jaffe et al, 1985, Harris et al, 1993 Ward et al., 2002, Bradford et al., 2001). Though not energetically important, other abundant species such as OH and

CH, constrain molecular gas column densities and abundances through absorption transitions (Bradford et al, 1999, Smith et al, 2001, Fischer et al 2001). These lines also have potential as a redshift probe for extremely obscured sources (those weak in fine structure lines) like Arp 220.

Table 1 (below) examines the various ISM probes available in the far-IR and submillimeter. The redshift range is that which would be observable with a spectrometer operating from λ = 20 μm to 1 mm.

TABLE 1. Far-IR SPECTROSCOPIC PROBES

SPECIES	WAVELENGTH	DIAGNOSTIC UTILITY	REDSHIFTS
		IONIZED GAS	
O IV	54.9	Primarily AGN.	0 - 17
S IV	10.5	Probes of the gas density and	0.9 - 100
O III	51.2, 88.4	UV field hardness in star-	0 - 18
S III	18.7, 34.8	formation H II regions.	0 - 50
N III	57.3	Provides effective temperature	0 - 16
Ne II	12.8	of hottest stars.	0 - 75
N II	122, 205	Diffuse interstellar H II regions.	0 - 7
		NEUTRAL ATOMIC GAS	
C II	158	Density and temperature probes	0 - 5
Si II	34.81	of photodissociated neutral-gas	0 - 30
O I	63, 145	interface between H II regions	0 - 15
C I	370, 610	and molecular clouds.	0 - 1.7
		MOLECULAR GAS	
H2 rotation	28.1 and shortward	Arises in dense, 100 - 1000 K molecular gas - often shock heated. Coolant of first gravitational collapse.	0 - 100
CO rotation	2600 and shortward	Primary coolant of molecular gas - probes pressure and temperature. Isotopes provide column densities, total gas mass	0 - 3
		DUST	
PAH	7.7, 11.3	Indicates star formation	0.8 - 100

2. Sensitivity from Space

Unfortunately, though there are a wealth of diagnostics for dusty galaxies in the far-IR and submillimeter, the wavelength range is not generally accessible from the ground.   The atmospheric windows shortward of λ~700 μm are accessible only occasionally from mountaintop sites. Between 30 and 200 μm, the atmosphere is completely opaque from all

terrestrial sites. While there are space missions planned to observe these wavelengths in the continuum (SIRTF, HSO), there is very limited spectroscopic capability planned for these wavelengths from space. The sensitivity attainable from a modest (diameter ~ 3.5 m) cool (T < 15 K) space telescope is dramatically better than what is currently planned between SIRTF and ALMA (see Figure 1). The effective line survey speed, proportional to the inverse of the sensitivity squared divided by the instantaneous bandwidth shows a more
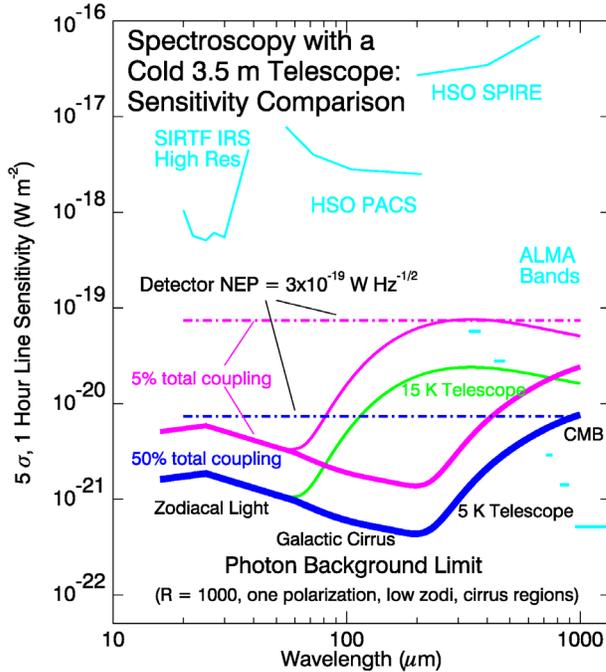
Figure 1. Spectral sensitivity on a cold 3.5 m space telescope. The 50% and 5% total couplings are conservative bounds to what might be achieved with a real spectrometer. Background noise is calculated from the fluctuation in the Zodical and cirrus backgrounds, toward patches of low intensity. If historical trends continue, background-limited performance at these frequencies will be possible, and gains of 2-4 orders of magnitude in sensitivity could be achieved.
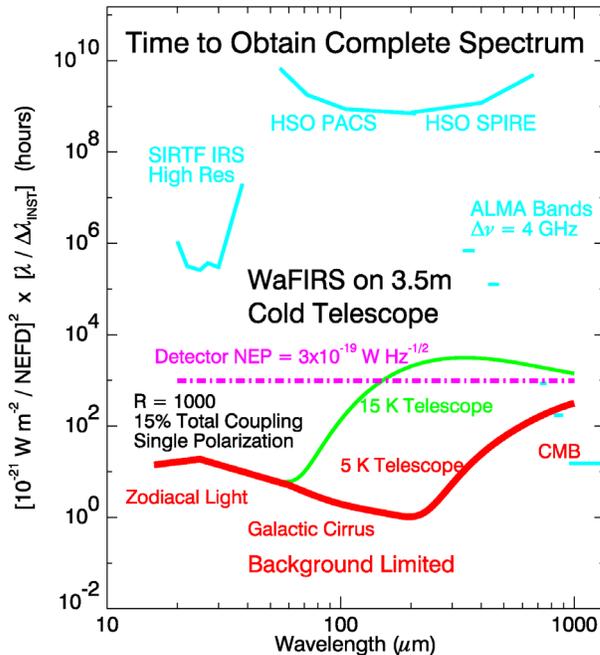
Figure 2. Sensitivity calculations similar to those of Figure 1, but including the effect of instrumental bandwidth. Under reasonable assumptions (15% efficiency, one polarization only), a broad-band spectrometer on a cold 3.5 meter telescope will have line survey speeds 6 or more orders of magnitude faster than the spectrometers of HSO, making far-IR line surveys practical. Such an instrument would be faster than ALMA for λ<600 μm.

dramatic gain of several orders of magnitude (see Figure 2). These large potential gains are possible in part because the raw sensitivity improves substantially by cooling the telescope to 15 K or lower (HSO is expected to operate at about 60 K). Another key aspect for observing sources with unknown redshifts is that the spectrometers on HSO are not optimized for line surveys. PACS is an imaging spectrometer with only a 1% instantaneous bandwidth, and SPIRE is a Fourier transform instrument with noise from the full band on the detectors at once. For an 8 meter class cold telescope such as SAFaIR, the sensitivity advantages are improved an additional factor of more than 5 beyond what is plotted in Figures 1 & 2.

3. Technical Background: Why a Waveguide Spectrometer

While there are a variety of options for a far-IR and submillimeter spectrometer, some are better suited to the science goals outlined above – observing high-redshift dusty galaxies with high sensitivity. Sources will be taken from preceeding continuum surveys, and will

Far-IR / Submm Line Surveys: Spectrometer Technology

AIRES on SQFIA
$R = 1.2 \times 10^6 \, \mu m / \lambda$
Total BW = 0.0008

PACS on HSO
R=1700
Total BW = 0.009

IRS on SIRTF
R=600 over 19-37 μm
Total BW = 0.5

Ebert-Fastie Grating
Total BW = 0.4

WaFIRS
Total BW = 0.4

Cryogenic Optical Volume in Units of λ^3

Resolving Power x Total Bandwidth
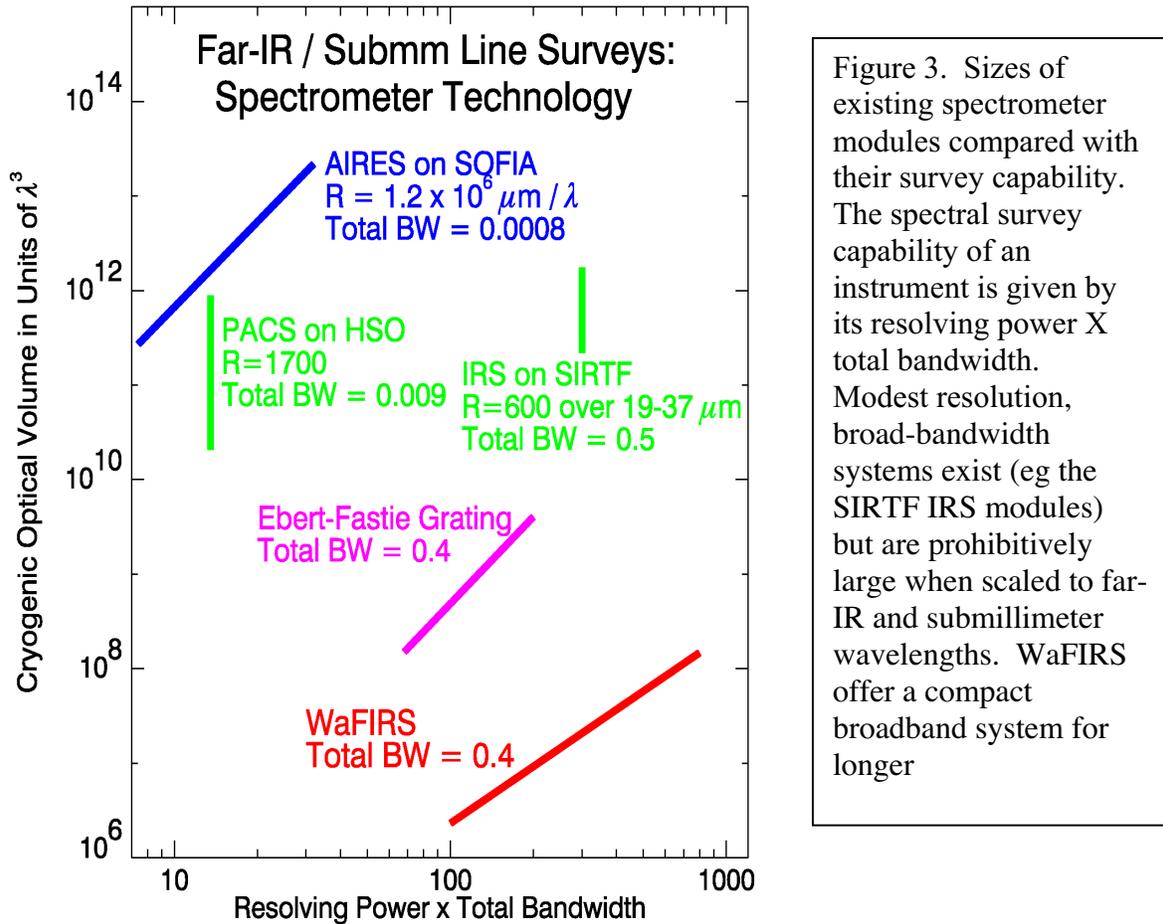
Figure 3. Sizes of existing spectrometer modules compared with their survey capability. The spectral survey capability of an instrument is given by its resolving power X total bandwidth. Modest resolution, broad-bandwidth systems exist (eg the SIRTF IRS modules) but are prohibitively large when scaled to far-IR and submillimeter wavelengths. WaFIRS offer a compact broadband system for longer

be spatially unresolved, so that imaging is not particularly important. In most cases, the redshifts will not be known in advance and the diagnostic lines are distributed over a broad spectral range, so a large instantaneous spectral bandwidth is critical. Given that in the far-

IR and submillimeter, the total number of detectors is typically a constraint, it therefore desirable to have the detectors arrayed spectrally rather than spatially, and an imaging monochromator such as a Fabry-Perot is not the instrument of choice. For ultimate sensitivity, a Fourier transform spectrometer (FTS) is not ideal because it places the entire spectral bandwidth and its associated photon noise onto a single detector. An FTS is appropriate only when using a detector which is not background-limited at the spectrometer resolution.

The obvious choice for background-limited point-source spectroscopy is a diffraction grating. Gratings have been used in astronomy for decades, and recently in the Infrared Space Observatory (ISO) as cryogenic, space-borne infrared spectrometers. When operated in first order, a grating naturally provides an octave of instantaneous bandwidth, and the resolution can be increased by increasing the grating size, roughly $d \sim \lambda \times R/2$. At far-IR and submillimeter wavelengths, this size quickly becomes prohibitively large, especially since real instruments are typically larger than the fundamental limit because they include collimating and imaging mirrors as well as order-sorting elements. For example, each of the spectrometer modules on SIRTF, measures about 40 x 15 x 20 cm, with a maximum $\lambda \times R$ product of 2 cm (R = 600 at 37 μm). To scale such an instrument up for a wavelength of 200 μm would result in a long dimension of over 2 meters, prohibitive for a space mission. Another example is the PACs spectrometer for HSO, an image slicing spectrometer which provides R=1500 out to 200 μm. The size of the cryogenic enclosure is quite large, roughly 80 cm x 80 cm x 30 cm, and because the instrument is designed for imaging spectroscopy; it only provides 16 spectral resolution elements, or 1% instantaneous bandwidth. The sizes of existing spectrometers are shown in Figure 3 in units of λ^3, plotted against the total number of spectral resolution elements.
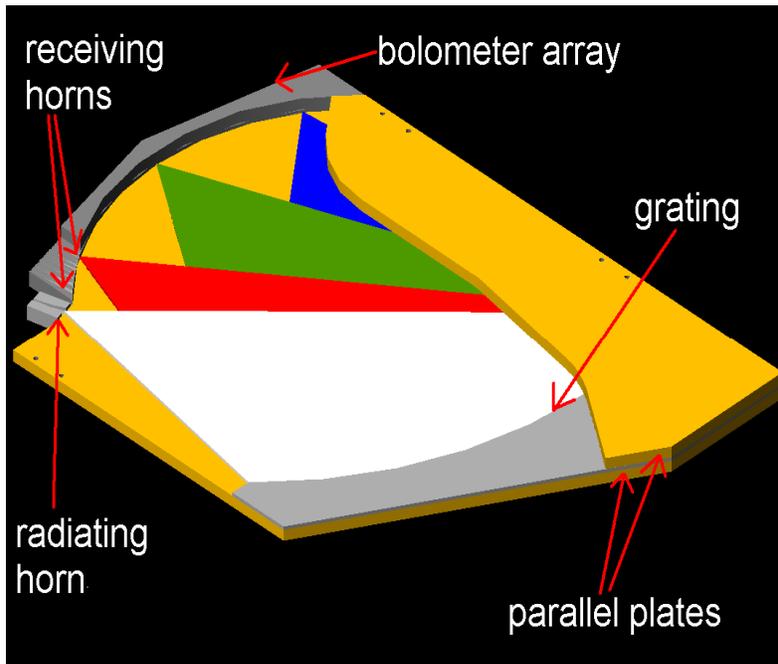
Figure 4. WaFIRS Concept: Rowland grating in parallel plate waveguide.
The light enters the parallel plate medium from the radiating horn and illuminates the curved diffraction grating. Each facet of the grating is individually positioned. Bolometers are positioned behind receiving horns on the focal curve.

To provide spectroscopic follow-up capability for the continuum surveys and to provide the foundation for further spectroscopic study, we are developing a new technology for a compact, broad-band spectrometer for the far-IR and submillimeter. WaFIRS, the Waveguide Far-IR Spectrometer, consists of a curved diffraction grating, entrance feed horn and detector feed horns all inside a parallel-plate propagation medium (see Figure 4). The spectrometer need be only a few wavelengths thick due to the two-dimensional geometry. Several spectrometers could easily be stacked to provide multiple wavelength bands, multiple spatial pixels, or both. The curved diffraction grating is the most space-efficient grating configuration possible because it both disperses and focuses the light. The grating can be nearly as large as the largest spectrometer dimension, with very little overhead, thus providing the maximum resolving power for a given cryogenic volume. Furthermore, the grating is used in first order which provides up to an octave of instantaneous bandwidth for any given module. The dramatic reduction in volume relative to conventional spectrometers is illustrated in Figure 3. The design has no moving parts and, once assembled, is completely light tight. While the concept has not been applied in the far-IR or submillimeter before, similar systems have been produced for near-IR and optical applications. The compact, lightweight geometry and robust construction make WaFIRS extremely well-suited for airborne, ballon and space-based spectroscopy.

4. **WaFIRS Technical Discussion:**

WaFIRS is conceptually similar to the slit spectrometers with curved gratings used by Rowland, Wadsworth, Eagle and other shortly after Rowland made the first curved grating around the beginning of the 20th century (see Born & Wolf, 1999). Though WaFIRS uses the same basic layout, it is based on propagation of a single electromagnetic mode in a two-dimensional medium bounded by parallel, conducting plates. The propagation mode is analagous to the TE_10 mode in rectangular waveguide -- the electric field is normal to the direction of propagation, with a half-wave vertical profile which vanishes at the top and bottom. [The dispersion relation is that of waveguide, namely as the frequency decreases toward cutoff, the wavelength increases infinitely.] Light is injected into a WaFIRS module with a horn which provides a suitable illumination pattern on the grating. The grating is in first order, and diffracts the light to a circular focal curve which extends over nearly 90 degrees of arc, on which the feed-horn coupled bolometers are arrayed. Figure 4 shows a sketch of the WaFIRS concept.

We have designed and built a prototype for wavelengths of 1-1.6 mm. The spectral resolving power is between 180-250 and the overall size only 56 cm x 42 cm x 2.5 cm. The key to the design is the placement of each facet individually such that for two frequencies, the change in propagation phase from the input to the output is exactly 2π between two adjacent facets, providing perfect (stigmatic) performance at these two frequencies. In our λ=1-1.6 mm prototype, there are 400 facets, and the resulting grating curve has a length of 51 cm. To evaluate the spectrometer designs, we perform diffraction calculations which account for the amplitude and phase produced by the input horn at each facet, then sum the contributions from all the facets at each output location. The model therefore includes

diffraction and geometric optics, at a scalar approximation. The illumination of the grating is important -- a larger pattern from a smaller input horn produces higher spectral resolution, but lower efficiency, due to power which is lost beyond the edges of the grating. For the first prototype, we have chosen a 3.5 mm input horn which illuminates the grating with a power pattern of FWHM = 110 (180) facets at 1.0 mm (1.6 mm). The spectral resolution ($\nu / \Delta\nu_{FWHM}$) that results is 180 (250) at 1.0 (1.6 mm). Because the facet positions are individually calculated, the geometric aberrations are completely negligible, and the system is strongly diffraction limited.

4.1 Spectrometer Efficiency

Estimates of the spectrometer efficiency must include losses from: 1) Waveguide propagation with finite-conductivity plates, 2) diffraction efficiency into the proper order (i.e. blaze efficiency), and 3) illumination losses. Figure 5 plots our calculations of these contributions, and their product. The waveguide propagation loss is given by standard expressions (see Pozar, 1998), the 3 % loss for total propagation from input to detector can be achieved by polishing and gold-plating the parallel plates. In our prototype operated
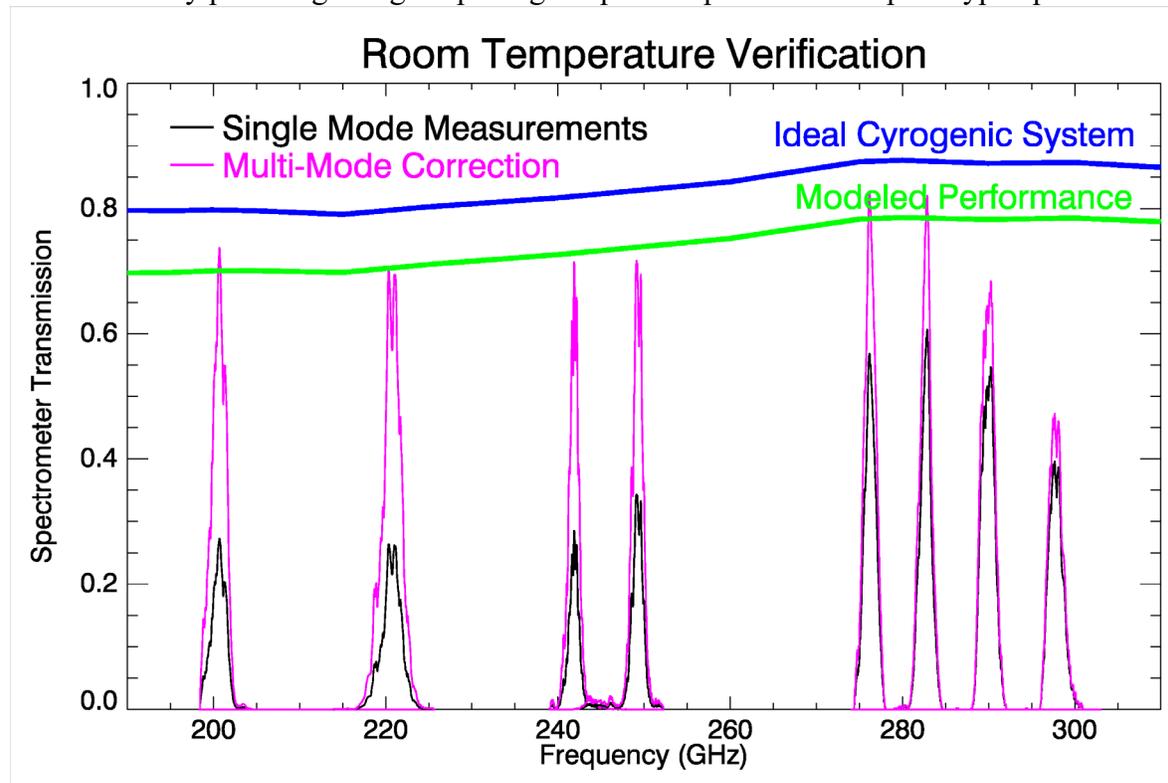
Figure 5. Testing of the first WaFIRS prototype. Measurements were made with a detector behind a single-mode feed, which does not couple all the power in the profile. Accounting for this coupling inefficiency produces the corrected curves which are close to performance predicted by the waveguide propagation loss, spillover loss. The high-frequency degradation is due to a blaze inefficiency, which will be corrected in future designs.

with only nickel coated plates and an aluminum grating, the propagation efficiency is somewhat lower. The blaze efficiency is calculated with commercially available software, applicable since with the waveguide propagation mode, the radiation is effectively in TE grating mode. These same calculations are the basis of our choice of a 29° blaze angle. Spillover losses are discussed above, they range from 89% to 94%. The net efficiency of a gold-plated cryogenic system is expected to be higher than 80% across our band.

4.2 Prototype Testing

We have measured the performance of our prototype using a backward wave oscillator (BWO) as a sweeping millimeter wave source (Figure 5,6, below). The power was measured with a diode detector in single-mode waveguide. At each frequency, a very small feed horn was used to measure the size of the intrinsic profile, then the total coupling as measured with a larger (but still single mode) feed is deconvolved using this profile. The results are quite close to our predictions over much of the band. We are investigating the reason for the loss in performance at the highest frequencies, it is likely a blaze efficiency.

Center Freq. = 218.8 GHz
Full Width = 1.2 GHz
Resolution = 178.

Center Freq. = 249.3 GHz
Full Width = 1.2 GHz
Resolution = 211.

Center Freq. = 290.0 GHz
Full Width = 1.1 GHz
Resolution = 260.

Figure 6 a,b,c .
Measured resolving power of the first WaFIRS prototype. To within the measurement uncertainties, profile widths are equal to those calculated in the design model. The second generation spectrometer will have resolving power between 250 and 400.

5. Scalability: Ground Based Observations and WaFIRS Modules for the Far-IR

In addition to our λ=1 mm prototype, we are constructing a cyrogenic version, Z-Spec for observations from ground-based submillimeter / millimeter observatories. Z-Spec will demonstrate the capability of WaFIRS to provide broadband spectroscopy with background-limited sensitivity. Z-Spec is discussed in more detail by J. Glenn et al in these proceedings (Glenn et al (2001)). We have produced designs with the same size and shape for a variety of shorter wavelengths, including a system which provides R=2000 at λ=100 μm. As wavelength is shortened, the number of facets and spectral resolution elements increases while the facet size and the physical size of a resolution element decrease. The grating remains in first order and the total fractional instantaneous bandwidth is constant. This extension to shorter wavelengths is possible because the design produces a stigmatic geometry for the grating, and geometric aberrations are small. The table below shows three examples of spectrometer designs:

Design Parameter	1st Prototype	Z-Spec Module	Far-IR Module
Frequency Range	195-310 GHz	195-310 GHz	1.9-3.0 THz
Stigmatic Frequencies	273, 204 GHz	296, 199 GHz	2.8, 2.0 THz
Number of Detectors	------------	160	~500
Number of Facets	400	500	4000
Resolving Power	180-250	250-400	1000-2000
Plate Spacing	2.5 mm	2.5 mm	0.6 mm
Spacing Tolerance	0.08 mm	0.04 mm	0.005 mm
Longest Dimension	50 cm	61 cm	55 cm
Illumination Efficiency	.87-.91	.78-.85	.9

6. References

Barger et al., 1999 AJ 117, 2656.
Blain et al., 1999. astro-ph /9906141
Blain et al., 2002, astro-ph /0202228
Born and Wolf, 1999. '*Principles of Optics*,' Cambridge U. Press. pp. 458-462
Bradford, C.M., PhD Thesis, Cornell University, 2001
Bradford, C.M., et al., 1999, in '*The Universe as Seen by ISO*,' Cox & Kessler, eds., ESA SP-427.
Genzel, R. et al., 1998, ApJ 498, 579.
Harris et al., 1991, ApJ 382, L75
Helou G., Malhotra, S., Hollenbach, D.J., Dale, D., Contursi, A.., 2001 ApJ 548, L73.
Jaffe, D.J. et al., 1985, ApJ 290, L59
Malhotra et al., 2001, ApJ, 561, 766.
Pozar, D.M. '*Microwave Engineering*.' J Wiley & Sons. Chap. 3.
Stacey, G.J. et al. 1991, ApJ 373, 423.
Tran, Q.D. et al., 2001, ApJ 552, 527.
Ward, J., Zmuidzinas, J., Harris A.I. Issac, K. 2002 ApJ, in press.

Detector technology for SCUBA-2, the new-generation submillimetre imager for the JCMT

W.D. Duncan, W.S. Holland, M.D. Audley, B.D. Kelly, D. Gostick,
P.T. Peacocke, M.J. MacIntosh, T. Hodson
(UK ATC)

K.D. Irwin, G. Hilton
(NIST)

A. Walton, A. Gundlach, W. Parkes, C. Dunare
(Univ. of Edinburgh)

P.A.R. Ade
(Cardiff)

E.I. Robson
(JAC)

SCUBA-2 Specifications

- **Simultaneous observing** at 450 and 850 μm
- **Field of view:** 8×8 arcmin (16 times the area of SCUBA)
- **DC-Coupled Arrays:** 80×80 square pixels; 1/2Fλ for 850; Fλ for 450 μm
- **Photon noise limited:** 2×10^{-17} W Hz$^{-1/2}$ (850 μm); 10^{-16} W Hz$^{-1/2}$ (450 μm)
- **F number:** 2.67
- **Time constant:** 1 ms
- **Cryostat:** Pulse-tube cooler & dilution refrigerator (cryogen-free)
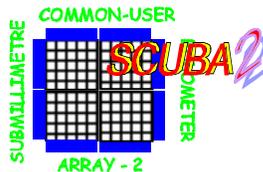
New Detector Technologies

Transition Edge Sensors

The SCUBA-2 bolometers use superconducting transition-edge sensors (TES; Irwin 1995) which are voltage-biased so that they operate with strong, negative *electrothermal feedback* (Lee et. al. 1995). For example, if the temperature drops slightly, so does the resistance. This causes the current to increase, and the increased Joule heating returns the detector to the original temperature.

Advantages of electrothermal feedback :

- automatically biases the detectors and regulates their temperature
- reduces the Johnson noise
- linearizes and speeds up the response
- increases the dynamic range
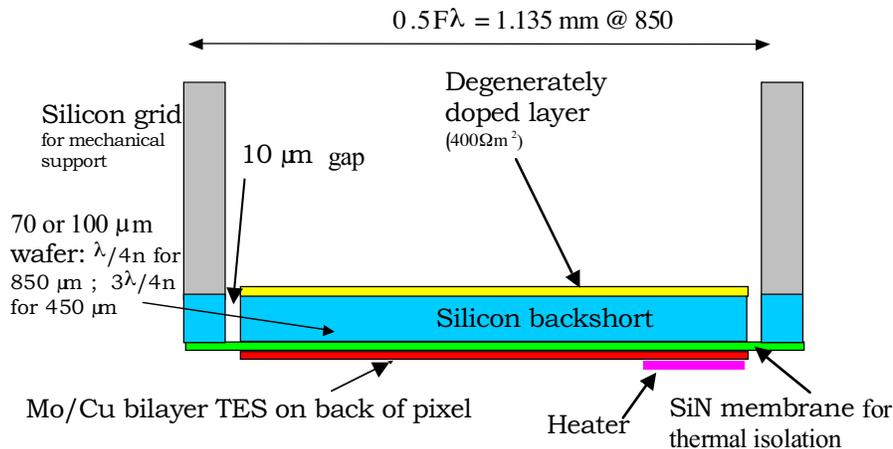
Cross-section of a single pixel The TES is on the underside of a backshort with an implanted resistive layer giving an impedance match to incoming radiation. The multiplexer wafer (not shown) is bump-bonded to the underside of the detector wafer. The heater is used to compensate for changes in the sky background intensity.

Pixel modelling

HFSS modeling indicates absorption of 80% of the incident light with this structure.

Monolithic Arrays

To fill an 8×8′ field requires 6,400 pixels for each array (c.f. 37/91 for SCUBA)! Because fabricating the pixels individually is impractical, the arrays are built from sub-arrays of 40×40 square pixels. Each sub-array is fabricated from a single wafer that has been polished to a multiple of a quarter wavelength in silicon. A silicon grid is diffusion-bonded to the detector wafer for mechanical support. The detector wafer is bump-bonded to another wafer that carries the SQUID multiplexers.

Modelled thermal parameters for pixels with $\gamma=3$ and $\alpha=50$							
	Tc (mK)	C (pJ/K)	τ (msecs)	G (pW/K)	P (pW)	P_{bias} (pW)	NEP $(10^{17}W/\sqrt{Hz})$
850 μm	120	8.8	2	1587	30	7.5	1.6
450 μm	120	8.8	1	7183	200	4	4.9
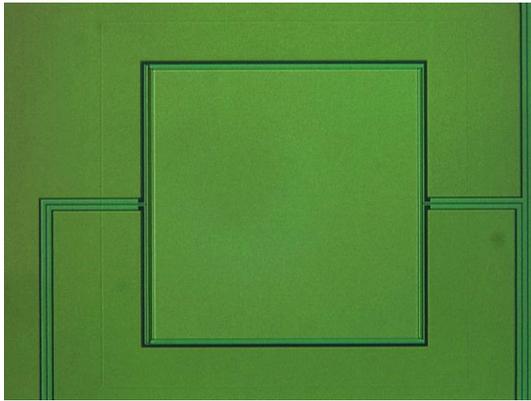
SCUBA 2 test pixel
1mm by 1mm MoCu TES on a Si Nitride
membrane. Pixel heater is at left side

SCT1-2 (A type) Die 4,8
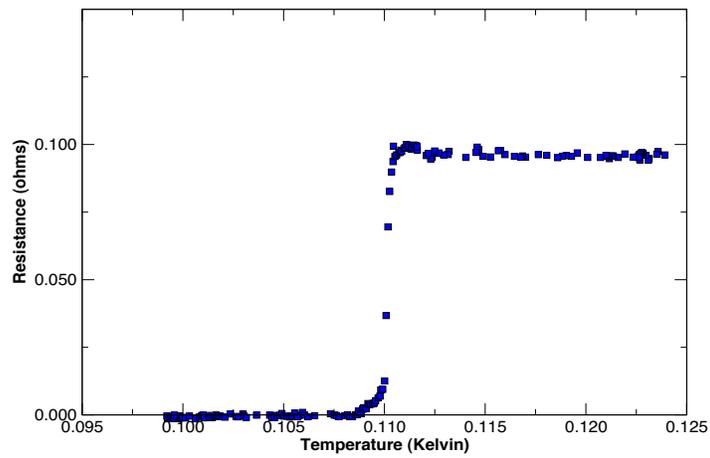
Thermal conductance of test pixel
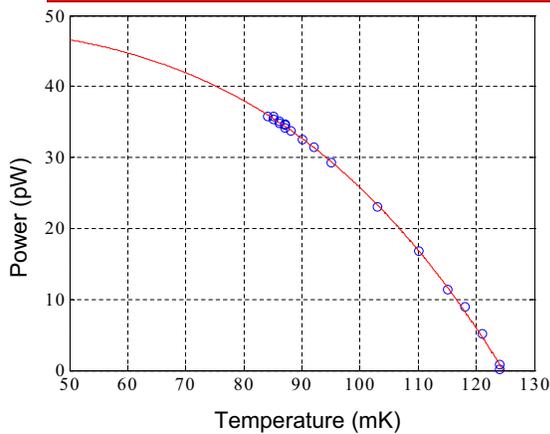
$P=K(T_c^n-T_{bath}^n)$

$K=5.87 \times 10^{-8}$

$T_c=124.7$

$n = 3.4$

$G=dP/dT$

$G=1.33$ nW / K

Multiplexed SQUID Readout

The TES detectors are read out by SQUID amplifiers. The first stage SQUIDs are on a separate wafer which is bump-bonded to the detector wafer. Running signal lines to all the detectors is not feasible so the signals are multiplexed. The SQUIDs in each column of a sub-array are wired in series, as are their feedback coils. The address lines of separate columns are wired in parallel and each row of SQUIDs is switched on in turn. Each column has a common output from which a SQUID array amplifier reads out a time-multiplexed signal. The SQUID array amplifiers are mounted close to the detectors at the 1 K stage.
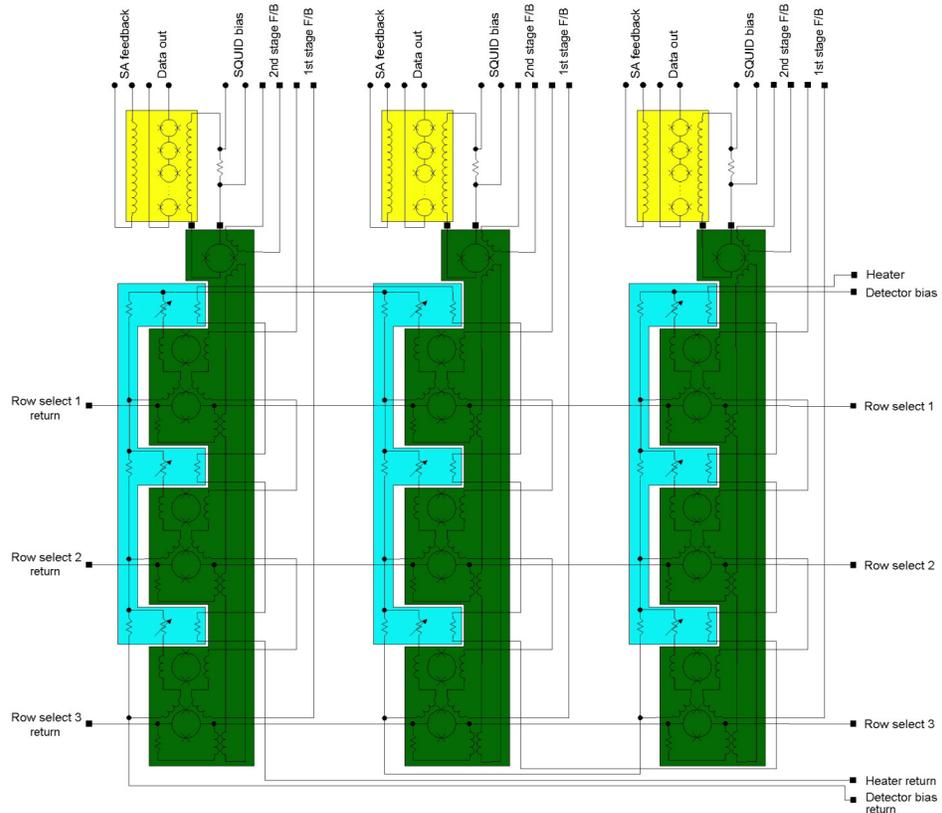
SQUID multiplexing scheme
The columns are addressed in parallel, which further reduces the
number of lines needed.
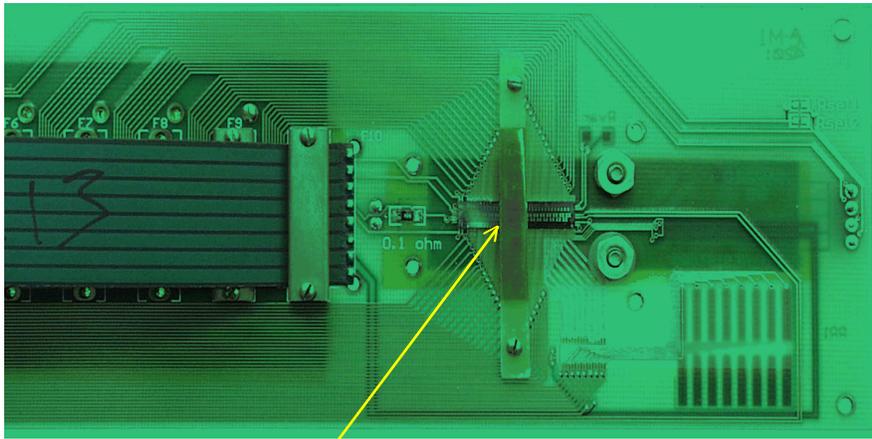
SCUBA 2 1 by 32 SQUID Mux chip in test probe
Clocks at up to 3 MHz
Across is 8 channel de-muxing of chopped signal

Dark SQUID experiment

Demultiplexed data

SQUID 1

Dark SQUID

Averaged DS

SQUID 1 - DS
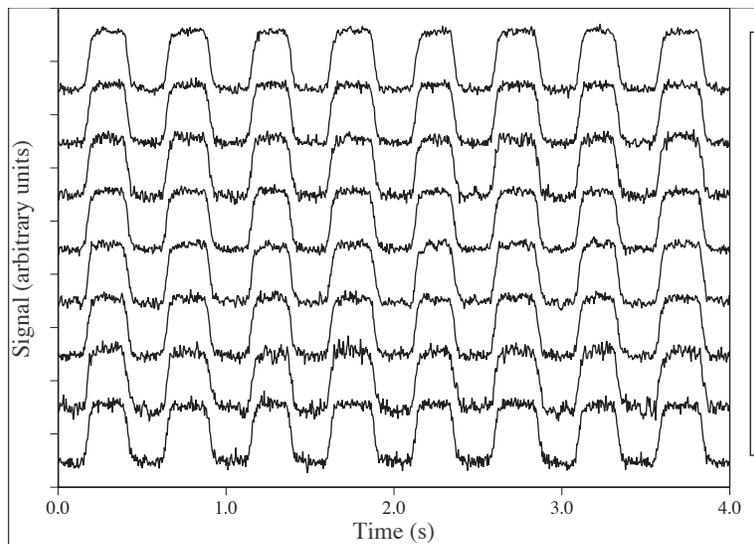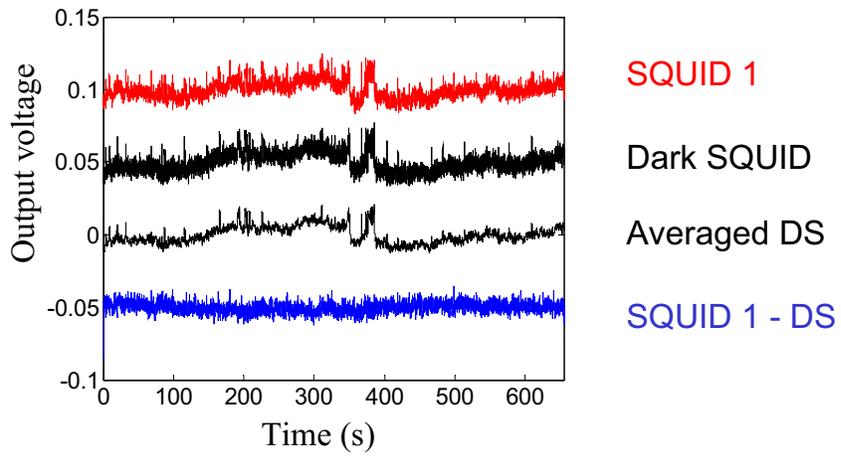
SCUBA 2 instrument - 4K and 1K optics and array modules

The wide field of the camera means that the instrument is large - the 4K box is close to 2m tall. The instrument will use 4K pulse tube coolers for the dilution refrigerator and to cool the optics

Integral Field Spectroscopy in the Far Infrared

Albrecht Poglitsch, Norbert Geis, Leslie Looney, and Walfried Raab

Max-Planck-Institut f. extraterrestrische Physik

Postfach 1312, 85741 Garching, Germany

e-mail: alpog@mpe.mpg.de

Abstract: We are presently building far-infrared integral field spectrometers for Herschel (PACS) and SOFIA (FIFI LS). These have become possible by (1) the development of relatively large, low noise photoconductor arrays and their associated cryogenic readout electronics and (2) a novel design for a reflective image slicer which – compared to equivalent designs in the visible or near-infrared – leads to a very compact unit capable of handling the large etendue (AΩ) required in a diffraction limited FIR system. In conjunction with a moderate size (30 cm x 8 cm) grating this concept allows for moderate resolution (R ∼ 2000), imaging spectroscopy perfectly suited for extragalactic astronomy.

If we project the recent progress in detector development (photoconductors and bolometers) in terms of number of pixels and sensitivity into the near future then this instrument concept could be readily extended in two ways which could contribute significantly to the scientific success of SAFIR: A line spectrometer with an instantaneous spectral coverage of a modest number of resolutions elements but a large field of view would ideally complement the photometric imaging capabilities of the mission for objects that can be spatially resolved. Applications could be spectral line maps of galaxies which would allow detailed diagnostics of phenomena like sequential star formation, galaxy interaction, and nuclear activity. Complementary to this, a small field of view in combination with very broad spectral coverage would be ideally suited for pointing-uncertainty tolerant spectroscopy, including spectroscopic redshift determination, of galaxies in the early universe, with the goal to study their formation and evolution (star formation and AGN formation).

1   Introduction

Integral field spectroscopy is becoming a well-established observing technique in the visible and near-infrared[1-5]. Instantaneous establishment of a 3-D data cube with two spatial dimensions and one spectral dimension can be achieved either with an imaging Fourier-transform spectrometer or with a grating spectrometer in connection with an optical "image slicer" which re-arranges a 2-dimensional field of view along the 1-dimensional entrance slit of the grating spectrometer. The latter method wins in terms of sensitivity if the dominant noise contribution comes from photon noise rather than from detector/read noise. While in the optical/NIR various schemes of image slicing (fibers, lenslet arrays, mirror slicers) are feasible, optical material properties in the FIR limit the

implementation of efficient slicers to purely reflective schemes. Flat mirror solutions are readily available in the optical/NIR, but because of the enormous increase in beam product ($A\Omega\sim\lambda^2$) at FIR wavelengths they cannot be simply scaled.

The advantage of integral field spectroscopy in observing efficiency has led us to developing a compact image slicer / grating spectrometer concept which is now being implemented in two FIR instruments for SOFIA and Herschel and which could be extended to larger fields of view or larger instantaneous spectral coverage for future missions like SAFIR.

2 Spectrometer Concept

Three objectives have been the main drivers for the design of our integral field spectrometer concept: For an all-cryogenic implementation in an airborne or space environment it had to be compact; it had to provide spectral resolution matched to the line width of extragalactic objects for optimum sensitivity under background-noise limited conditions; it had to make optimum use of the available detector pixels (which at present are still precious in the FIR).

For the last point, a trade-off can be made readily – at least for faint objects – as they usually allow observation of only a few lines within the covered wavelength band. Therefore, if the radial velocity of the object is roughly known, spectral multiplexing may be limited to a small number of resolution elements, enough to provide good baseline definition and to cover the velocity range within an object, and spatial multiplexing can be implemented to the extent that the size of the detector array allows, while the spectral lines under investigation can be observed sequentially. In most cases this will come at little extra expenses in terms of total observing time which will usually be dominated by the integration time on the weakest line to be detected. Additional aspects, like slit curvature of any (long-slit) grating spectrometer, or staggered slit patterns as formed by simple reflective slicer arrangements, should also be considered for optimization.

Compactness and spectral resolution have to be traded against each other, to some extent, but for wavelengths below 200 μm a resolving power of 1500…2000 can be reached with a grating of ~300 mm length. The width of the grating can be considerable less if an anamorphic collimator is used. For a relatively small instantaneous spectral coverage, the most compact implementation is a (pseudo) Littrow arrangement.

For the slicer unit, we have developed a simple configuration of 3 sets of mirrors which slice a 2-dimensional field, separate the pupils for each slice, and form a continuous (non-staggered) slit with pupil recombination, with reasonably fast optics (f/20…f/10), thus in a fairly compact volume.

3 Spectrometer Implementation

We are presently building two instruments which involve integral field spectroscopy in the FIR: the Far Infrared Field Imaging Line Spectrometer (FIFI LS) for SOFIA[6] and the Photodetector Camera & Spectrometer (PACS) for Herschel[7]. Both work in the

50…200μm range and employ 16 x 25 pixel arrays of photoconductive Ge:Ga detectors[8,9] (stressed and unstressed).
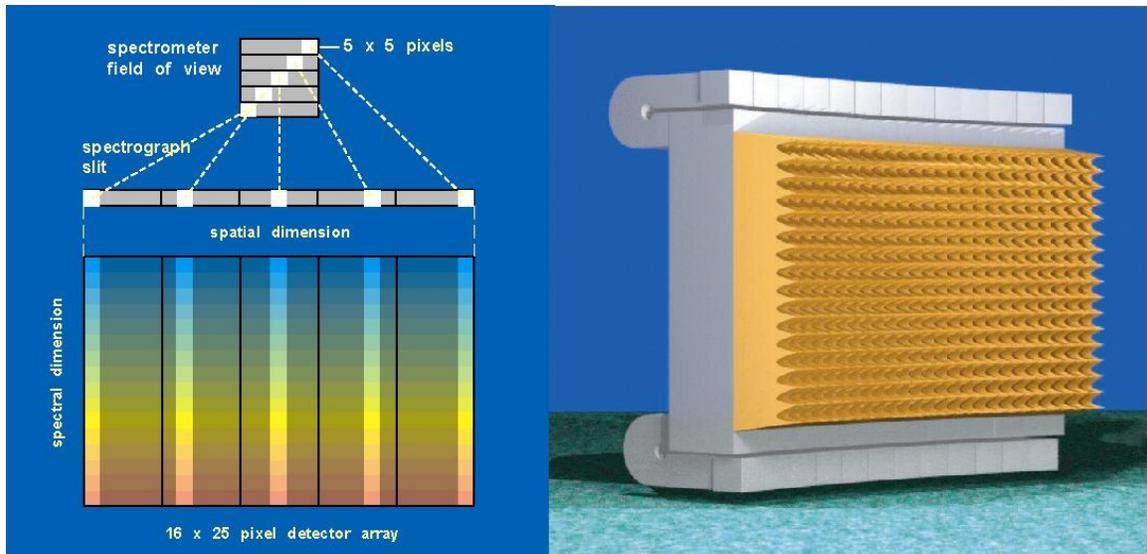
Figure 1: Integral field line spectrometer concept – a 2-dimensional field of view (5x5 pixels) is split into 5 slices which are re-arranged along one continuous line to form the entrance slit for a grating spectrograph which disperses the light from each spatial pixel into a spectrum (left) which is imaged onto the Ge:Ga photoconductor array (right). Each pixel of the array is contained in an integrating cavity. The light cone array in the front provides area-filling light collection into the detector cavities.
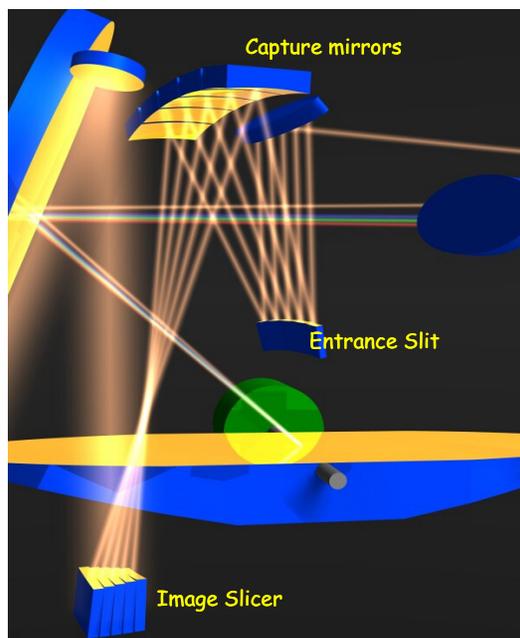
Figure 2: Image slicer unit. The telescope focus is re-imaged onto the image slicer which can be thought of as a twisted stack of mirrors. Each mirror in the stack is a field mirror which creates an image of the telescope aperture on one of the five capture mirrors such that the light from the individual slices is completely separated. The capture mirrors re-image the slices onto the set of field mirrors which form the entrance slit to the grating spectrograph. These field mirrors recombine the five pupil images into one which is then imaged onto the grating.

Both spectrometers employ a 5x5 pixel image slicer which is shown in Figure 2. The slicer consists of three groups of five mirrors, with optical power on all surfaces. To obtain a slit image along one continuous line, all mirrors have a tip- and a tilt angle which differs for each of the respective five elements in a group. This leads to a different image rotation for each slice in the slit image plane which can be controlled such that the slit curvature of the grating spectrometer can be compensated. This is, of course, strictly true only for one Littrow angle, i.e., for one center wavelength, but it minimizes the pixel overhead in the spectral dimension in any case.

For (near) diffraction-limited imaging, the slicer must cut into the diffraction pattern of the telescope, which introduces diffraction itself; however, the effect of this becomes primarily visible in images of the telescope aperture (i.e., in pupil planes) while at the same time the system is able to transfer a resolution *better* than that of the telescope in the cross-slice dimension, i.e., the slices are transferred with steep edges in the image plane.
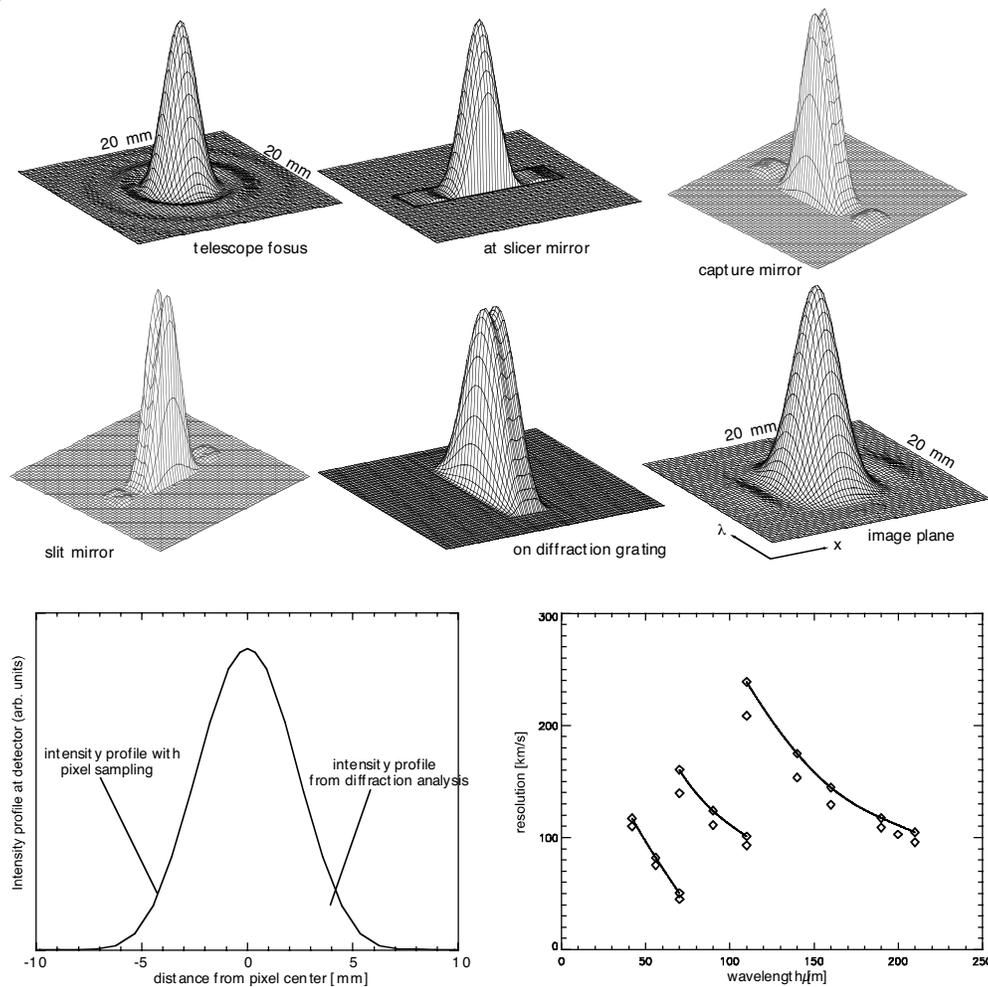
Figure 3: Physical optics analysis of the slicer/spectrometer optical train and derived spectral resolution including pixel sampling

Therefore, the full spectral and spatial resolution of a narrow-slit spectrometer can be maintained without the normally unavoidable loss of light, as long as all pupil-equivalent elements are dimensioned properly to not vignette the side lobes generated by the image slicer. We have verified this quantitatively by a physical optics analysis of the entire optical train of the spectrometer. Examples are shown in Fig.3.

3.1   The Far-Infrared Field-Imaging Line Spectrometer (FIFI LS) for SOFIA

FIFI LS is one of the first-light instruments for the Stratospheric Observatory for Infrared Astronomy (SOFIA). It accommodates two parallel spectrometers for the wavelength bands 42 - 110 μm & 110 - 210 μm together in one helium cryostat and allows simultaneous observation in both bands. Each spectrometer performs integral field spectroscopy with 5x5 spatial channels x 16 spectral pixels with pixels scales of 6" (short wavelength) and 12" (long wavelength). The resolving power ranges from 1000 to 5000, with an instantaneous velocity coverage of ~ 1500 km/s. The estimated sensitivity under the specified SOFIA telescope background conditions will be 5×10^{-17} W/m^2 @ 63 μm and 2×10^{-17} W/m^2 @ 158 μm (point source / 5σ / 1h integration time). The instrument will be available with the beginning of the observatory operation in the second half of 2004.

anamorphic collimator
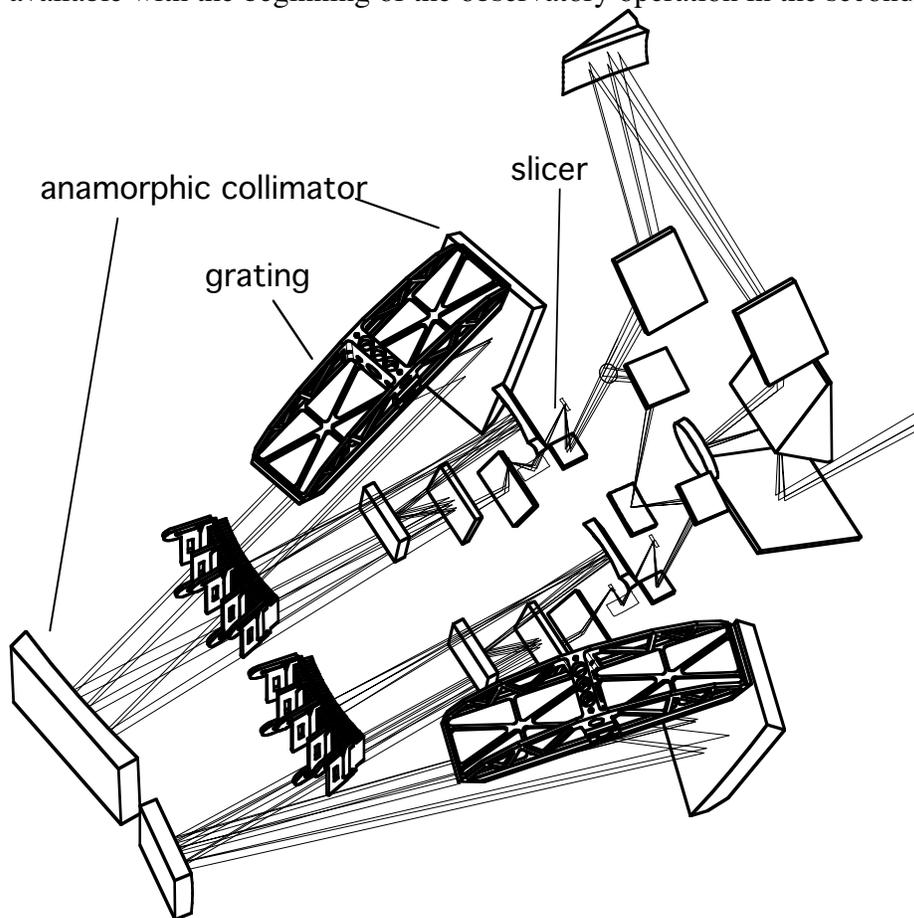
grating

slicer

Figure 4: FIFI LS optics. After the shared entrance optics a dichroic beam splitter distributes the light into the two independent spectrometer channels with their image slicer units, collimator optics, gratings, and Ge:Ga photoconductor arrays.

3.2 The Photodetector Array Camera & Spectrometer (PACS) for Herschel

PACS is one of the three focal plane instruments for ESA's cornerstone mission Herschel. PACS performs both, photometric and spectroscopic imaging. The photometer part of PACS employs filled arrays of micromachined silicon bolometers, operated at 0.3 K, with 64 x 32 and 32 x 16 pixels to simultaneously image two bands, 60-85 or 85-130 μm and 130-210 μm, over a field of view of ~ 1.75' x 3.5', with full beam sampling in each band. Under the telescope background conditions of Herschel we expect a point source detection limit of ~3 mJy (5σ / 1h), which is also approximately the confusion limit from the extragalactic FIR background. The spectroscopy part provides one 5 x 5 pixels integral field spectrometer of the type described above with a fixed pixel scale of ~9" and a wavelength range from ~57 μm to ~210 μm. The resolving power ranges from 1000 to 3000, with an instantaneous velocity coverage of ~ 1500 km/s. The estimated background-limited point source detection limit ranges from ~8×10^{-18} W/m^2 @ 60 μm to ~2.5×10^{-18} W/m^2 @ 180 μm (5σ / 1h). This will allow observations of the brightest FIR lines from galaxies out to redshifts of ~2.
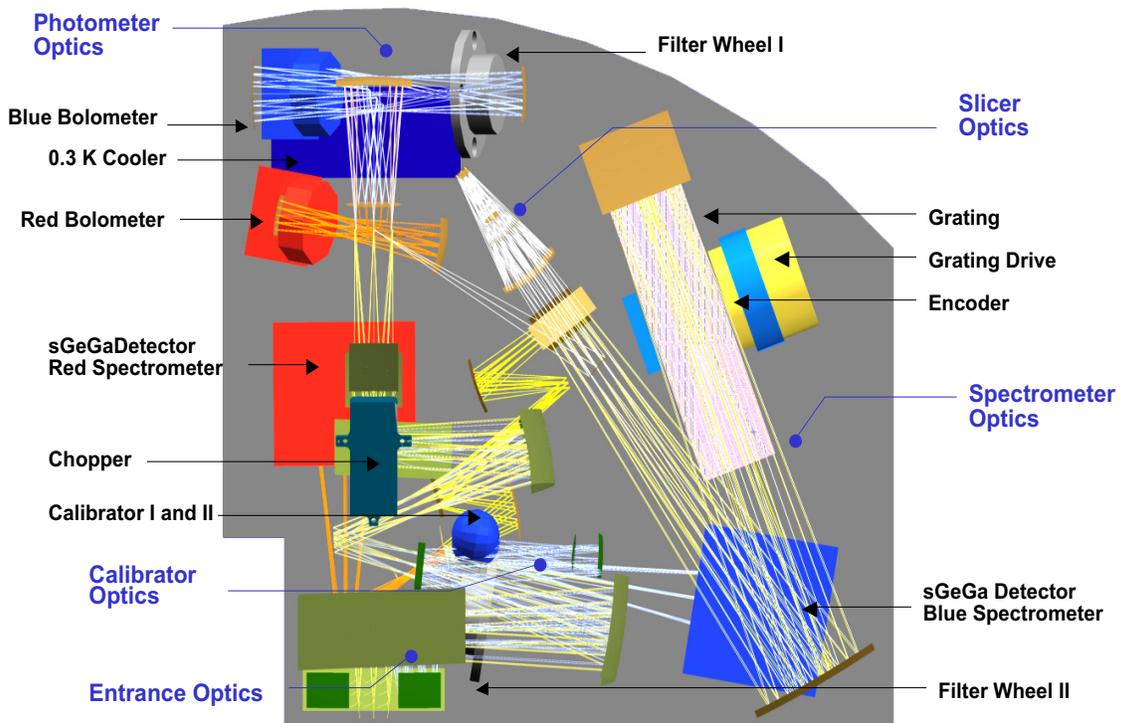
Figure 5: Layout of the PACS focal plane unit. Fields for photometry and spectroscopy are split off in an intermediate focus. The photometric bands are defined by filters and distributed to the two bolometer arrays by a dichroic beam splitter. The grating is used in 1st to 3rd order to cover the full wavelength range. Orders are selected by filters and distributed to the stressed or unstressed 16 x 25 pixels photoconductor arrays by a dichroic beam splitter.

The launch of Herschel is scheduled for early 2007.

4   Implications for future missions

With two integral field spectrometers for space or airborne applications under way, we can try to extrapolate what will be possible in the near future. It should be mentioned that the present developments are not limited by the optics, but by the available size of the detector arrays. FIR/submm detectors have made great progress over the last years, both in the field of photoconductors and – even more so – with new concepts of bolometers. We expect to see ultra-low noise, large format arrays of bolometric detectors in the next years, and new FIR photoconductors, like monolithic GaAs BIB arrays, are also under development. Since larger FIR/submm detector arrays are often assembled from one- or two-dimensional subarrays, and are often not buttable on all four sides, butted arrays with a substantially greater number of pixels in one dimension and a more modest number of pixels in the other may be the natural next step. Therefore, it seems realistic to think of two possible expansions of our integral field spectrometer design which both could be implemented in missions like SAFIR.

4.1   Imaging line spectrometer

An integral field spectrometer with a small number of spectral channels, as in our present design, but with a substantially larger number of spatial pixels, would enable spectral line mapping in resolved/extended sources with the full (and very attractive) spatial resolution and sensitivity offered by a large aperture, cryogenic space telescope.
Photometric mapping of dust enshrouded objects (star bursts, AGNs) would be ideally complemented by the full diagnostic power of imaging spectroscopy with the full advantage of the same low extinction as the photometric images.
The respective spectrometer would look very much like the present design, just with a larger image slicer, and with a detector array properly expanded in the "spatial" dimension.

4.2   Faint object spectrograph

As we are starting to resolve the FIR/submm extragalactic background, it is becoming clear that dramatic phases of star formation have occurred in massive galaxies already in the early universe. Still, the question remains how much of the detected luminosity comes from starburst activity, and how much from AGNs. Besides X-ray observations, FIR/submm spectroscopy provides the ideal tool to observe lines which suffer very little extinction, and which allow a discrimination of excitation mechanisms. Since the identification of the submm sources known to date with visible/NIR galaxies, and therefore also their redshift determination, is notoriously difficult, partly because of position uncertainties between different observatories, partly because these sources are not prominent at shorter wavelengths, an efficient spectroscopic observation of these

(faint) sources is best done with an instrument which provides a broad instantaneous spectral coverage and the best possible sensitivity and spatial resolution, but with a certain pointing tolerance. All these conditions would be ideally fulfilled by a wide-band spectrometer with a small integral field unit. Compared to our present development, this would simply mean a substantial extension of the detector array in the "spectral" dimension.

5   References

1. Le Coarer E., Bensammar S., Comte G., Gach, J.L., Georgelin, Y., 1995, A&A Suppl. Ser. 111, 359

2. Bacon R., Adam G., Baranne A., Courtès G., Dubet D., Dubois J.P., Emsellem E., Ferruit P., Georgelin Y., Monnet G., Pécontal E., Rousset A., Sayède F., 1995, A&A Suppl. Ser. 113, 347

3. Weitzel L., Krabbe A., Kroker H., Thatte N., Tacconi-Garman L.E., Cameron M., Genzel R., 1996, A&A Suppl. Ser. 119, 531

4. Kenworthy M.A., Parry I.R., Taylor K., 2001, PASP 113, 215

5. Bacon R., Copin Y., Monnet G., Miller B.W., Allington-Smith J.R., Bureau M., Marcella Carollo C., Davies R.L., Emsellem E., Kuntschner H., Peletier R.F., Verolme E.K., de Zeeuw P.T., 2001, MNRAS 326, 23

6. Looney L.W., Geis N., Genzel R., Park, W.K., Poglitsch A., Raab W., Rosenthal, D., Urban, A., Henning T., 2000, SPIE 4014, 14

7. Poglitsch A., Waelkens C., Geis N., 2001, ESASP 460, 29

8. Rosenthal D., Beeman J.W., Geis N., Looney L.W., Poglitsch A., Park W.K., Raab W., Urban A., 2000, SPIE 4014,156

9. Kraft S., Merken P., Creten Y., Putzeys, J., Van Hoof, C.A., Katterloher R.O., Rosenthal, D., Rumitz, M., Groezinger, U., Hofferbert, R., Beeman, J.W., 2001, SPIE 4540, 374

10. Richards P.L., 2002, (this volume)

11. Katterloher R.O., Barl, L., Jakob, G., Konuma, M., Haller, E.E., Frenzl O., Hermans, L., 2000, SPIE 4013, 100

The Millimeter-wave Bolometric Interferometer

S. Ali[1], P. A. R. Ade[2], J. J. Bock [3], G. Novak[4], L. Piccirillo[2], P. Timbie[1], G. S. Tucker[5]

[1]Dept. of Physics, University of Wisconsin—Madison, Madison, WI 53706

[2]Dept. of Physics and Astronomy, University of Wales—Cardiff, Cardiff CF243YB, Wales, UK

[3]Jet Propulsion Laboratory, Pasadena, CA 91109

[4]Department of Physics and Astronomy, Northwestern University, Evanston, IL 60208

[5]Dept. of Physics, Brown University, Providence, RI 02912

The Millimeter-wave Bolometric Interferometer (MBI) is a proposed ground-based instrument designed for a wide range of cosmological and astrophysical observations including studies of the polarization of the cosmic microwave background (CMB). MBI combines the advantages of two well-developed technologies — interferometers and bolometric detectors. Interferometers have many advantages over filled-aperture telescopes and are particularly suitable for high resolution imaging. Cooled bolometers are the highest sensitivity detectors at millimeter and sub-millimeter wavelengths. The combination of these two technologies results in an instrument with both high sensitivity and high angular resolution.

1. INTRODUCTION

Astrophysical observations are almost always limited by wavelength, angular resolution and/or sensitivity constraints. The Millimeter-wave Bolometric Interferometer (MBI) combines the high angular resolution of interferometers with the sensitivity of bolometers. The result is a proposed new instrument suitable for a wide range of cosmological and astrophysical observations. MBI will have a large baseline ($\sim$ 6 m) and will probe sub-arcminute angular scales. A prototype of this instrument, MBI-B, is a small baseline ($\sim$ 0.14 m) version of MBI designed to search for B-mode polarization in the CMB at degree angular scales.

With the large baseline MBI it should be possible to address the following scientific goals:

- Characterize the polarization of the CMB. MBI will measure simultaneously the temperature and polarization anisotropy of the CMB at angular scales from $30'$ to $40''$.
- Characterize clustering in the far-infrared background (FIRB).
- Measure and map magnetic fields near the center of the galaxy by measuring the Faraday rotation of synchrotron radiation.
- Measure the wavelength dependent properties of dust polarization.
- Image the Sunyaev-Zeldovich (SZ) effect in clusters of galaxies.
- Spectrally separate the kinetic and thermal SZ effects.
- Search for CMB and SZ point source foregrounds.

MBI will serve as a testbed for a possible future space-based interferometer mission to measure the polarization of the CMB. The angular resolution of space telescopes is normally limited by the aperture size which can fit in a rocket fairing. An interferometer can overcome this size limitation, by using an optical bench which extends after launch.

MBI-B will be constructed first to demonstrate the MBI concept. It's goals are to:

- Search for primordial B-mode polarization in the cosmic microwave background (CMB).
- Search for intracluster magnetic fields via Faraday rotation.
- Map magnetic fields near the core of our galaxy.

The main characteristics of MBI-B and MBI are presented in Table 1. The field-of-view (FOV) of each interferometer element is denoted by θ and the size of the synthesized beam size by ϕ.

TABLE 1: MBI Characteristics and Sensitivity.

	MBI-B 20×20 array		MBI-B		MBI	
λ	Sensitivity	Days to reach	θ	ϕ	θ	ϕ
(mm)	(mK$\sqrt{s}$)	3 μK/sky pixel	(degrees)	(degrees)	(arcmin)	(arcmin)
3	0.6 (1.1)	0.5 (1.4)	7	1.4	10	2.0
2	0.7 (1.3)	0.6 (2.2)	7	1.4	7.2	1.4
1	2.9 (5.4)	11. (37.)	7	1.4	3.6	0.7

Calculations assume equivalent temperature of atmosphere, optics and CMB is 40 K.
Each interferometer views 25 pixels on the sky simultaneously.
The sensitivity is the sensitivity for each sky pixel to fluctuations in a 2.7 K blackbody.
Sensitivity numbers and integration times are for background-noise-limited and,
    in parentheses, (detector-noise-limited) cases.

2. SCIENCE

2.1 CMB Polarization: Both the temperature anisotropy and spectral information of the CMB strongly constrain cosmological models, but there is another component — the *polarization* of the CMB — which has not yet been detected. In most models the amount of this polarization is $\leq 10\%$ of the temperature anisotropy (Bond & Efstathiou 1987). The fundamental nature of the prediction is such that the mere detection of polarization of roughly the correct amplitude would have great significance. Because the size of the expected polarization signal is so small, it is necessary to use the highest sensitivity detection scheme possible while simultaneously controlling and minimizing potential systematic effects.

Polarization on the sky may alternatively be separated into "electric" (E) and "magnetic" (B) components (Hu & White, 1997; Seljak & Zaldarriaga 1998; Kamionkowski et al. 1997). E and B multipole patterns are, respectively, gradient and curl decompositions of the polarization pattern, much like Q and U, but more directly related to the source generation mechanism. E type polarization is generated by scalar or tensor perturbations at the time of last scattering, while B type polarization is only produced by tensor perturbations and cannot be produced by scalar perturbations due to parity conservation. As a result a search for B mode polarization is a direct search for tensor modes in the early universe. The signal for B type polarization peaks up at degree angular scales.

For a noise-limited experiment the best observation strategy for detecting B-mode polarization is to spend more observation time on a restricted part of the sky rather than to survey a large part of the sky (Jaffe et al. 2000). A beam size smaller than $\sim 1°$ does not significantly improve the sensitivity to detecting B-mode polarization. At a wavelength of 2 mm MBI-B has a FOV of $\sim 7°$ and an effective resolution of $1.2°$. This choice of FOV and resolution represents a trade-off to most effectively probe the B-mode polarization signal. Figure 1 illustrates the trade-off between survey size and angular resolution and shows the choice of instrument parameters for MBI-B is nearly optimal.

The higher angular resolution of MBI is chosen to measure the CMB polarization power spectrum at smaller angular scales where the signal from acoustic oscillations is expected to appear. The expected polarization power spectrum at small scales is closely related to the temperature power spectrum at these scales.

2.2 Galactic Studies: MBI-B and MBI can also be used, for example, to measure the wavelength dependent polarization properties of dust. Observations of the thermal component of the dust radiation have revealed that the wavelength-dependence of the polarization is surprisingly strong. The observations are few, and are specifically limited to molecular cloud envelopes, but they give a consistent picture of a degree of polarization that falls with wavelength in the far-infrared,
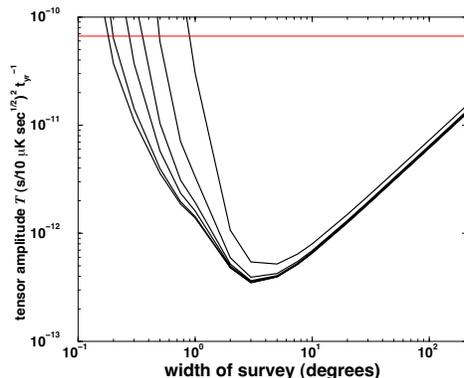
Figure 1. The smallest tensor amplitude that could be detected at 3σ with an experiment with a detector sensitivity of 10 $\mu K\sqrt{s}$ that observed for one year and maps a square region of sky of given width. The result scales with the square of the detector sensitivity and inversely with the duration of the experiment. The curves are (from top to bottom) the FWHM beamwidths of 1.0, 0.5, 0.3, 0.2, 0.1 and 0.05 degrees. The horizontal line shows the upper limit to the tensor amplitude from COBE. MBI-B will improve the COBE measurement by more than a factor of 10. (Reproduced from Jaffe et al. 2000.)

and then begins to rise in the submillimeter, near 350 μm wavelength. These results have been explained using a model in which the efficiency of grain alignment is correlated with exposure to radiation from both embedded stars and external sources (Hildebrand et al. 1999). The importance of understanding the physics of grain alignment lies principally in the fact that studies of polarized dust emission provide one of the few methods for mapping magnetic fields, especially in the dense regions where stars form. Without the physical understanding it is difficult to determine, for any given line-of-sight, which specific regions are being sampled, within an extended, heterogeneous and complex molecular cloud.

3. THE ADDING INTERFEROMETER

In a simple 2-element radio interferometer, signals from two telescopes aimed at the same point in the sky are multiplied (correlated) so that the sky temperature is sampled with an interference pattern with a single spatial frequency. In an interferometer that uses incoherent detectors, such as an optical interferometer, the electric field wavefronts from two telescopes are added and then squared in a detector — an "adding" interferometer as opposed to a "multiplying" interferometer (Rohlfs 1996). See Figure 2. The adding interferometer recovers the same visibility as a multiplying interferometer.

4. WHY USE A BOLOMETRIC INTERFEROMETER?

While inteferometers and bolometers are each highly developed, to our knowledge, no instrument which combines these two powerful technologies has ever been built. There are a number of advantages achieved in combining the two technologies, which we now describe.

For measurements which require high angular resolution, large single dishes are often impractical for a number of reasons including mass, deformations due to gravity and cost. Interferometers effectively enable high angular resolution by reproducing the resolution performance of a large dish; the trade-off is a reduction in collection area if the interferometer area is not filled. The limit of a filled interferometer is a single large dish. Thus for equivalent angular resolution, an interferometer can be substantially simpler and less costly than a single large aperture.

4.1 Better angular resolution for equivalent size. For a monolithic dish of diameter equal to the length of a two-element interferometer baseline, the interferometer has angular resolution
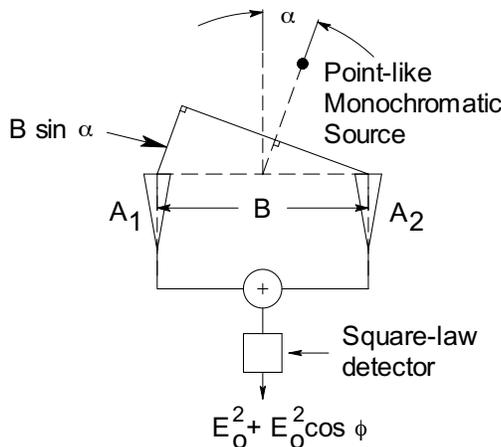
Figure 2. Adding interferometer. At antenna A_2 the electric field is E_0, and at A_1 it is $E_0 e^{i\phi}$, where $\phi = kB\sin\alpha$ and $k = 2\pi/\lambda$. B is the length of the baseline, and α is the angle of the source with respect to the symmetry axis of the baseline, as shown. (For simplicity consider only one wavelength, λ, and ignore time dependent factors.) In a multiplying interferometer the in-phase output of the correlator is proportional to $E_0^2 \cos\phi$. For the adding interferometer, the output is proportional to $E_0^2 + E_0^2 \cos\phi$. Modulation of the length of the baseline allows phase-sensitive detection to recover both the in-phase and quadrature phase interference terms and reduces susceptibility to low-frequency drifts ($1/f$ noise) in the bolometer and readout electronics.

roughly twice as good as that of the monolithic dish. The reason for this difference in angular resolution is the following: For acceptable sidelobe performance, the edge illumination of the monolithic dish must be tapered considerably, which reduces the effective aperture diameter. The effective edge taper of the interferometer is much smaller, so the effective diameter is nearly equal to the length of the baseline. The sidelobe rejection of an interferometer can also be significantly better than that from a single dish.

4.2 No chopping and scanning. Single dishes with either coherent or incoherent detectors typically use some form of "chopping" which is achieved either by nutating a secondary mirror or by steering the entire primary at a rate faster than the $1/f$ noise in the atmosphere and detectors. Similar approaches are used with arrays of detectors. Since an interferometer does not require this rapid chopping, the time constants of the bolometers used can be relatively long.

Interferometers provide direct 2D imaging and do not require scanning strategies; in addition individual maps may be mosaiced. Since beam steering is not required in an interferometer, for ground-based observations the signal from an interferometer is also significantly less affected by the atmosphere (Church 1995, Lay & Halverson 2001).

4.3 Multiplex advantage. For n apertures there are $n(n-1)/2$ baselines. The n beams can be combined in $n(n-1)/2$ pairs. However, the n beams can also be combined into a single beam, which results in a multiplexing advantage. There are two advantages to multiplexing. First, since the photon signal-to-noise ratio is increased, lower sensitivity detectors can be used. Second, by having all beams traverse symmetric paths, unmeasured path changes in the optics due to temperature variations, for example, are minimized. In order to accomplish multiplexing, each pair of telescope beams must be modulated at a unique frequency so the signal from telescope pairs can be recovered. With an array of N_D detectors, there is also an increase in observation speed since there are now effectively many interferometers operating simultaneously.

5. INSTRUMENT

The basic optical designs of the MBI-B and MBI are shown in Figure 3. They are each "Fizeau" or "image-plane" interferometers (as opposed to "Michelson" interferometers such as the proposed SPECS instrument described in Mather et al. (2000), Leisawitz et al. (2002) and Zhang et al. (2001)). The optical design has a number of desirable features. The interferometer elements

do not move with respect to one another which reduces systematic effects. Potential systematic effects are reduced through multiple levels of switching. The sky flats are translated sinusoidally by several wavelengths at a low frequency (< 1 Hz); the effective switching frequency is higher than that of the small mechanical motion. The entire interferometer can be rotated about the optical axis at a rate of $\sim 5°$/s.

MBI will build on MBI-B by providing a long baseline and thus much higher angular resolution. The concept for MBI is shown in Figure 4. The platform is similar to MBI-B in that it has altitude-azimuth-theta axes, but the baseline is much larger (~ 6 m).
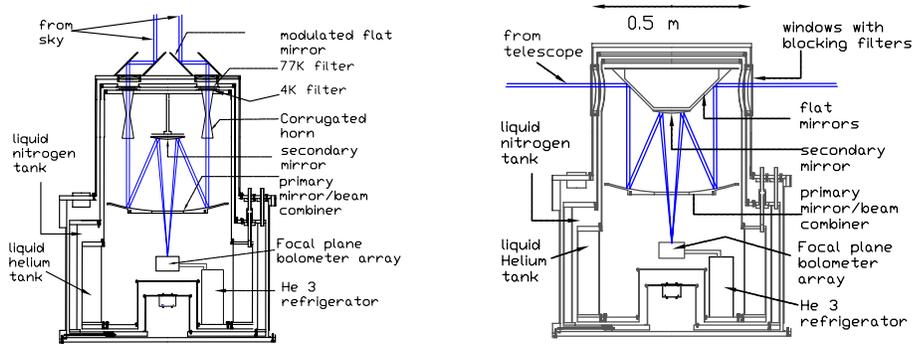
Figure 3. The MBI-B (left) and MBI (right) optical design. In the MBI-B design light from the sky is first reflected from two flat mirrors at $45°$. One of each pair of flat mirrors is modulated by a few wavelengths at a unique low frequency. Light enters the cryostat from the top, passing through blocking filters. The interferometer beams are defined with back-to-back corrugated feed horns; for clarity only two horns are shown. The beams are combined with a primary and secondary mirror in a Cassegrain telescope configuration. The interference pattern is imaged by a bolometer array in the focal plane. The bolometer array is cooled to below 0.3 K with a ^{3}He refrigerator. The MBI design is similar to the MBI-B design except that light enters the side of the cryostat and strikes a flat mirror at $45°$. The light comes from Cassegrain telescopes located up to several meters away. With only minor modifications the MBI-B receiver can be used for the MBI.

5.1 Bandwidth. While the sensitivity of a receiver to broadband signals increases as the square root of the bandwidth, for interferometers, the bandwidth restricts the angular range, θ, over which fringes are detected (Thompson, Moran & Swenson 1998; Boker & Allen 1999). If we assume the path lengths for a source at the center of the field of view (FOV) are equal, then the path length difference for a source at an angle θ from the center along the baseline axis is θB, where B is the baseline distance. If this path length difference is small compared to the coherence length of the light $\lambda^2/\Delta\lambda$, then the fringe contrast is not affected. Thus the FOV is determined by $\theta_{\text{FOV}} \leq (\lambda/\Delta\lambda)(\lambda/B)$. This equation indicates that for angles of the order of the product of the spectral resolution times the angular resolution, the fringe smearing is important. This relation imposes restrictions on the ratio between the maximum baseline achievable by the interferometer and the spectral bandwidth of the receiver. A choice of 20% spectral bandwidth will set the maximum baseline to about 5 times the diameter of each single telescope.

5.2 Sensitivity. Here we make an estimate for the expected sensitivity of MBI-B. These calculations are the basis for the projected sensitivity figures in Table 1.

MBI-B is an example of a Fizeau, or image-plane interferometer. Two main methods have been studied for combining the beams from the n different apertures in such an instrument (Prasad & Kulkarni 1989, Roddier & Ridgway 1999, Nakajima & Matsuhara 2001, Nakajima 2001). In one scheme, the $^{n}C_2$ interferometer, the radiation entering each aperture is divided $n-1$ ways and combined pairwise with the power from each of the other apertures to form an interference pattern on an array of detectors; there is one detector array for each of the $n(n-1)/2$ baselines. In
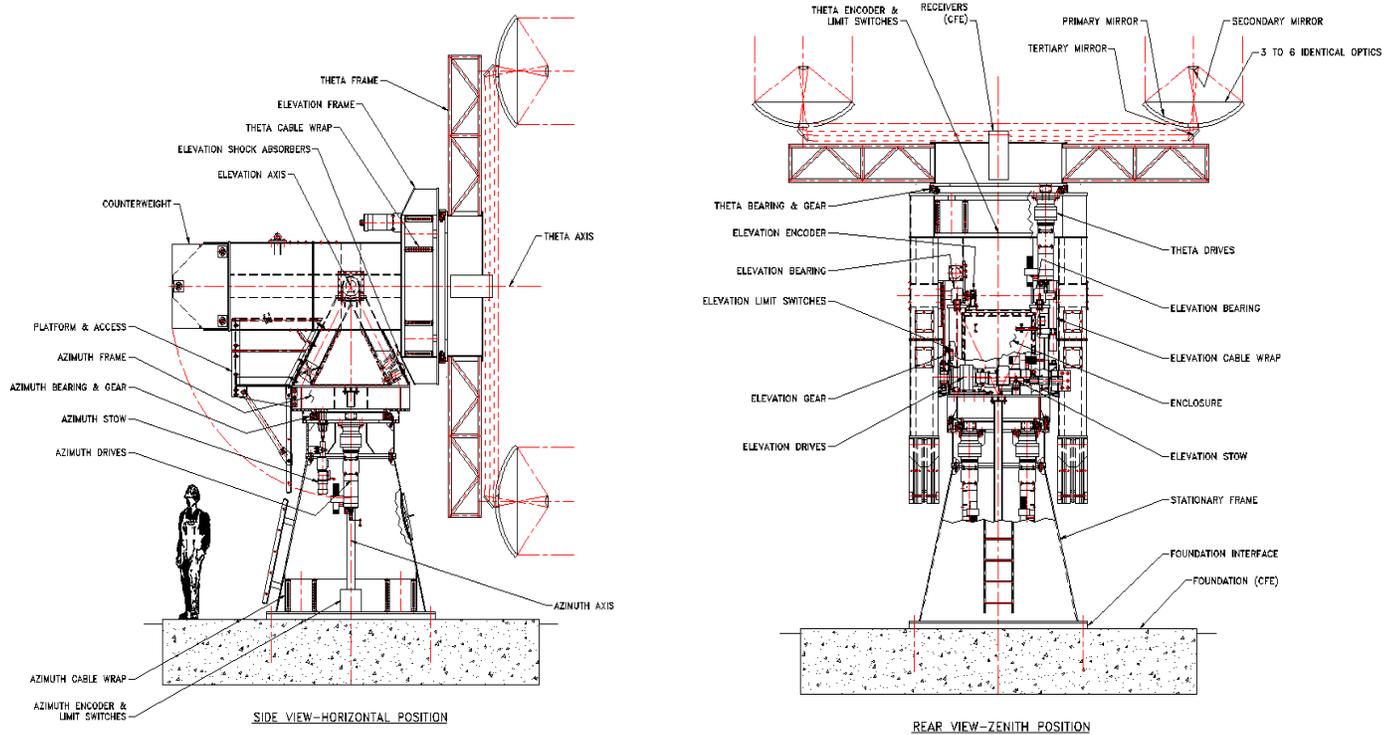
Figure 4. Concept for MBI (side and rear views). MBI will have eight 1.2 m telescopes with a maximum baseline of 6 m; for clarity only two telescopes are shown here. The telescopes are attached to a circular platform which can rotate about the optical axis — denoted here as the "theta" axis. The cryogenic beam combiner and receiver is located at the center of the circular platform. The platform is fully steerable in altitude and azimuth.

the second method, the ^nC_n interferometer, all n beams are combined simultaneously on a single 2-dimensional array of detectors. In the case where the sensitivity of the detectors is limited only by the photon-noise from the background radiation all forms of beam combination produce approximately the same sensitivity (Prasad & Kulkarni 1989). We have chosen the latter (^nC_n) approach because it is extremely simple to combine the n beams by bringing them to a focus in the focal plane of a telescope. No beam splitters are required. We locate the beam-combining telescope in a cryostat and cool the bolometer array in its focal plane to sub-Kelvin temperature. The array of bolometers can be considered as a multipixel correlator. A total of $n_c = n(n-1)/2$ fringe patterns or interferograms (one for each baseline) are superposed on this array in a "criss-cross" pattern. Each interferogram contains $2(\theta/\phi)$ fringe cycles, where θ is the primary antenna beam and ϕ is the resolution of the image corresponding to the maximum baseline. A total of $4(\theta/\phi)$ detector pixels are required along one dimension of the array for Nyquist sampling. Hence, a two-dimensional detector array must have a minimum of $N_D = 16(\theta/\phi)^2$ elements. Additional detectors do not provide any improvements in sensitivity.

For MBI-B full sampling of the interference plane requires $N_D = 400$. The detectors are single-mode polarization-sensitive bolometers (Turner et al. 2001), which will initially operate at a wavelength of 2 mm with a 20% bandwidth and optical efficiency $\eta = 0.3$. The ultimate sensitivity per detector pixel is achieved when the intrinsic bolometer noise is equal to or less than the statistical fluctuations of the background noise, the so called BLIP condition. We assume the optical loading on the bolometers is equal to the sum of the power loading from the optics, atmosphere and the CMB. At White Mountain, CA (altitude = 3800 m), for observations at a wavelength of 2 mm and at an observing angle 30° from the zenith, the effective temperature of the atmosphere is about 40 K (Grossman 1989). To calculate the Noise Equivalent Temperature

(NET), we divide the BLIP by $dB(T)/dT$ calculated for a blackbody at 2.735 K to arrive at an estimated estimated NET of $\sim 548 \ \mu K \sqrt{s}$, equivalent to a system noise temperature of about 67 K.

The expression for sensitivity for a direct detector interferometer depends on whether the detectors are limited by detector noise or by background photon noise. In all cases we assume that the interference pattern in the focal plane is fully sampled with N_D detectors according to the prescription above.

In the detector-noise-limited case, we obtain $\delta T = (T_s/\sqrt{\Delta \nu \Delta \tau 2n_c})(\theta/\phi)^2$ where $\Delta \nu$ is the optical bandwidth of the detectors, $\Delta \tau$ is the integration time, and the other quantities are as defined above. This expression is identical to that for a radio interferometer (Partridge 1995). δT is the noise in each pixel in the recovered image of the sky; the interferometer observes $(\theta/\phi)^2$ such sky pixels simultaneously. The expression is the same as that for a filled aperture, total power receiver with one detector and one pixel (of angular size ϕ) on the sky except for the factor of $(\theta/\phi)^2/\sqrt{2n_c}$. This factor accounts for the reduction in sensitivity arising because the interferometer array is not a completely filled aperture. For an interferometer with baseline length B and n apertures each of diameter D, $(\theta/\phi)^2/\sqrt{2n_c} \sim (B/D)^2/n$, which is the inverse of the aperture filling factor, q.

In the case of an interferometer in which the detectors are background-noise-limited the sensitivity scales as one over the square root of the aperture filling factor. As the filling factor is reduced, the signal reaching the detector array decreases linearly with q, but the photon noise on each detector decreases by the square root of q; the signal to noise is degraded as $\sqrt{q}$. Hence, in this case $\delta T = (T_s/\sqrt{\Delta \nu \Delta \tau n})(\theta/\phi)$ (Roddier & Ridgway 1999).

For MBI-B we have $n = 8$, so $n_c = 28$ at a single wavelength (2 mm). To measure these baselines fully, MBI-B requires $N_D = 20 \times 20 = 400$ bolometers in the array; it will observe a total of $(\theta/\phi)^2 = 25$ pixels on the sky simultaneously. In the detector-noise-limited case with $N_D = 400$ the noise per sky pixel in the synthesized image is then $\delta T = 1.3 \ mK \sqrt{s}$. The time required to integrate this image down to 3 μK noise per sky pixel is $(\delta T/3 \, \mu K)^2 = 2.2$ days.

According to Jaffe et al. (2000), a measurement at this sensitivity level with MBI-B's angular resolution and field-of-view can constrain the amplitude of the tensor modes (gravitational waves) created during inflation. The quantity of interest is the tensor-to-scalar mode ratio; with the parameters described here MBI-B will improve on the T/S ratio established by COBE by over a factor of 10.

This test of the MBI-B concept will observe in only one wavelength band at a time. The band is defined by an inductive-capacitive mesh filter placed in front of the array. In a future version of MBI-B, multiple bands could be observed simultaneously by using dichroic beamsplitters or frequency-selective-bolometers (Kowitt et al. 1996).

6. SUMMARY

We have outlined a concept for an adding interferometer that uses bolometer arrays for precision measurements in astrophysics and observational cosmology. We plan to construct a short baseline interferometer (MBI-B) to evaluate the design and use it to search for the extremely faint B-modes that are anticipated in the CMB at degree angular scales. Later, this same beam combiner/detector system could be coupled to an array of reflectors with $\sim$6 m baselines to explore smaller angular scales. Both of these instruments could serve as prototypes for more ambitious space-based interferometers.

7. REFERENCES

Boker T & Allen, J. A., "Imaging and Nulling with the Space Interferometer Mission", ApJSS, 125, 123 (1999)

Bond, J. R. & Efstathiou, G. "Microwave Anisotropy Constraints on Isocurvature Baryon Models," MNRAS, 226, 655 (1987)

Church, S.E., "Predicting residual levels of atmospheric sky noise in ground-based observations of the Cosmic Background Radiation," MNRAS, 272, 551 (1995)

Grossman, E. Airhead Software Inc., Boulder CO (1989)

Hildebrand, R. H. et al. "The Far-Infrared Polarization Spectrum: First Results and Analysis," ApJ, 516, 834 (1999)

Hu, W. & White, M. "A CMB Polarization Primer," New Astronomy, 2, 323 (1997)

Jaffe, A. H., Kamionkowski, M. & Wang, L. "A Polarization Pursuer's Guide," PRD, 61, 083501 (2000)

Kamionkowsi, M., Kosowsky, A. & Stebbins, A. "A Probe of Primordial Gravity Waves and Vorticity," PRL, 78, 2058 (1997)

Kowitt, M. S., Fixsen, D. J., Goldin, A. & Meyer, S. S. "Frequency Selective Bolometers," Appl. Opt., 35, 5630 (1996)

Lay, O. P. & Halverson, N. W. "The Impact of Atmospheric Fluctuations on Degree-Scale Imaging of the Cosmic Microwave Background," ApJ, 543, 787 (2001)

Leisawitz, D. et al. "Probing the Invisible Universe: The Case for Far-IR/Submillimeter Interferometry," Proceedings of the Second Workshop on New Concepts for Far-Infrared and Submillimeter Space Astronomy, University of Maryland, March 7–8, 2002

Mather, J. C., Moseley, S. H., Leisawitz, D., Zhang, X., Dwek, E. Harwit, M., Mundy, L. G., Mushotzky, R. F., Neufield, D., Spergel, D. & Wright, E. L. "The Submillimeter Probe of the Evolution of Cosmic Structure," submitted to the Review of Scientific Instruments (2000)

Nakajima, T. "Sensitivity of a Ground-based Infrared Interferometer for Aperture Synthesis Imaging," P. A. S. P., 113, 1289 (2001)

Nakajima, T. & Matsuhara, H. "Sensitivity of an imaging space based interferometer," Applied Optics, 40, 514 (2001)

Partridge, R. B. "3K: The Cosmic Microwave Background Radiation," New York: Cambridge University Press (1995)

Prasad, S. & Kulkarni, S. R. "Noise in Optical Synthesis Images I. Ideal Michelson Interferometer," J. Opt. Soc. Am. A, 6(11), 1702 (1989)

Roddier, F. & Ridgway, S. T. "Filling Factor and Signal-to-Noise Ratios in Optical Interferometric Arrays," P. A. S. P., 111, 990 (1999)

Rohlfs, K. & Wilson, T. L., Tools of Radio Astronomy, 2nd ed., Springer, New York (1996)

Seljak, U. & Zaldarriaga, M. "Polarization of the Microwave Background: Statistical and Physical Properties," in "Fundamental Parameters in Cosmology," proceedings of the XXXIIIrd Recontres de Moriond (1998)

Thompson, A. R., Moran, J. M. & Swenson, G. W. Jr. "Interferometry and Synthesis in Radio Astronomy," Krieger Publishing Company, Malabar, Florida (1998)

Turner, A. D., Bock, J. J., Beeman, J. W., Glenn, J., Hargrave, P. C., Hristov, V. V., Nguyen, H. T., Rahman, F., Sethuraman, S., and Woodcraft, A., "Silicon Nitride Micromesh Bolometer Array for Submillimeter Astrophysics," Appl. Opt. 40(28), 4921 (2001)

Zhang, X., Feinberg, L., Leisawitz, D., Leviton, D. B., Martino, A. J. & Mather, J. C. "The Wide-Field Imaging Testbed," IEEE Aerospace Conference Proceedings, 3, 1453 (2001)

BLAST – The Balloon-borne Large Aperture Submillimeter Telescope

Mark Devlin[1]*, Peter Ade[2], Jamie Bock[3], Simon Dicker[1], Matt Griffin[2], Josh Gundersen[5], Mark Halpern[6], Peter Hargrave[2], David Hughes[7], Jeff Klein[1], Philip Mauskopf[2], Barth Netterfield[8], Luca Olmi[9], Douglas Scott[6], Greg Tucker[10]

[1] University of Pennsylvania, Philadelphia, PA 19104, USA
[2] Cardiff University, Cardiff, UK
[3] Jet Propulsion Laboratory, Pasadena, CA, USA
[5] University of Miami, Coral Gables, FL, USA
[6] University of British Columbia, Vancouver, BC, CA
[7] INAOE, Puebla, MX
[8] University of Toronto, Toronto, ON, Canada
[9] University of Puerto Rico, Rio Piedras, Puerto Rico
[10] Brown University, Providence, RI, USA

ABSTRACT

BLAST is the Balloon-borne Large-Aperture Sub-millimeter Telescope. It will fly from a Long Duration Balloon (LDB) platform from Antarctica. The telescope design incorporates a 2 m primary mirror with large-format bolometer arrays operating at 250, 350 and 500 microns.

By providing the first sensitive large-area (10 sq. deg.) sub-mm surveys at these wavelengths, BLAST will address some of the most important galactic and cosmological questions regarding the formation and evolution of stars, galaxies and clusters. Galactic and extragalactic BLAST surveys will: (i) identify large numbers of high-redshift galaxies; (ii) measure photometric redshifts, rest-frame FIR luminosities and star formation rates thereby constraining the evolutionary history of the galaxies that produce the FIR and sub-mm background; (iii) measure cold pre-stellar sources associated with the earliest stages of star and planet formation; (iv) make high-resolution maps of diffuse galactic emission over a wide range of galactic latitudes. In addition to achieving the above scientific goals, the exciting legacy of the BLAST LDB experiment will be a catalogue of 3000-5000 extragalactic sub-mm sources and a 100 sq. deg. sub-mm galactic plane survey. Multi-frequency follow-up observations from SIRTF, ASTRO-F, and Herschel, together with spectroscopic observations and sub-arcsecond imaging from ALMA are essential to understand the physical nature of the BLAST sources.

THE BLAST INSTRUMENT AND SCIENCE GOALS

The primary advantage of BLAST over existing and planned sub-mm bolometer arrays such as SCUBA (Holland, Robson et al. 1999) on the JCMT, SHARC (Hunter, Benford et al. 1996) on the CSO (including their respective upgrades) is its greatly enhanced sensitivity at wavelengths <500 μm due to the dramatically increased atmospheric transmission at balloon altitudes. BLAST complements the Herschel satellite (formerly known as FIRST) by testing identical detectors and filters planned for the SPIRE instrument.

BLAST will be the first long duration balloon-borne telescope to take advantage of the bolometric focal-plane arrays being developed for Herschel. A LDB flight from Antarctica, providing the first surveys at 250, 350 and 500 μm, will significantly extend the wavelength range, sensitivity, and area of existing ground-based extragalactic and galactic surveys. The instrument parameters are given in Table 1. BLAST will conduct unique galactic and

---

* Sloan Fellow

extragalactic sub-mm surveys with high spatial resolution and sensitivity. Compared to the pioneering flights of PRONAOS (Lamarre, Giard et al. 1998), BLAST will have an advantage of >100 times the mapping speed. The scientific motivations for BLAST are similar to those of Herschel but are achievable within 2 to 4 years with a series of LDB flights. Additional information on BLAST can be found in Scott et al. (2001) and Hughes et al. (2001).

Using these unique BLAST surveys we expect to achieve the following science goals:

- Conduct a complementary series of wide (shallow) and narrow (confusion-limited) extragalactic 250—500 μm surveys, identifying the galaxy populations responsible for producing the far-IR and sub-mm backgrounds. BLAST will determine the amplitude of clustering of sub-mm galaxies on scales of 0.1—10 degrees.
- Measure the 250—500 μm spectral energy distributions (SEDs) and colors, from which one can derive rest-frame luminosities and star formation rates (SFRs) for sub-mm selected galaxies.
- Measure the sub-mm source-counts and place the strongest constraints to date on evolutionary models and the global star formation history of starburst galaxies at high-z (see Figure 2).
- Conduct Galactic surveys of the diffuse interstellar emission, molecular clouds, and identify dense, cold pre-stellar (Class 0) cores associated with the earliest stages of star formation.
- Observe solar system objects including the Kuiper-belt objects, planets, and large asteroids.
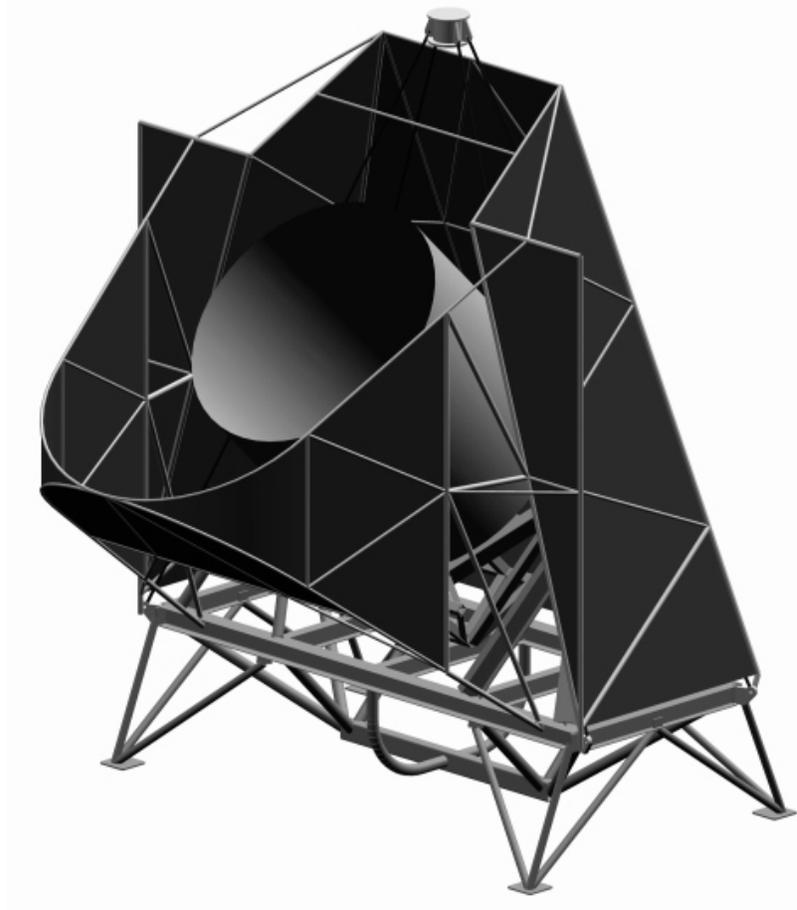
Figure 1 The BLAST Gondola

EXTRAGALACTIC SURVEYS

Observations at sub-mm wavelengths of starburst galaxies in the high-z universe have a particular advantage compared to observations in the optical and FIR because a strong negative k-correction enhances the observed sub-mm fluxes. By early 2002, SCUBA and the millimeter camera MAMBO (operating on the IRAM 30 m) will have com-

Table 1 BLAST Telescope Parameters

Telescope			
Temperature		200 K	
Throughput for each pixel		λ^2	
Bolometers			
Central Wavelengths (microns)	250	350	500
Number of Pixels	149	88	43
Beam FWHM (arcseconds)	30	41	59
Field of View for Each Array (arcmin)		6.5×14	
Overall Instrument Transmission		30%	
Filter Widths		$\sim 30\,\%$	
Observing Efficiency		90%	

Table 2 BLAST sensitivities. Sensitivities for SCUBA and SOFIA are given for comparison.

	250 μm	350 μm	500 μm
Background power [pW]	25.6	18.3	13.5
Background limited NEP [W Hz$^{-0.5}$ x 10^{-17}]	20	14	10
NEFD [mJy $\sqrt{\text{sec}}$]	236	241	239
ΔS (1σ, 1hr) (1 sq. deg.) [mJy]	38	36	36
ΔS (1σ, 6hr) (1 sq. deg.) [mJy]	15.5	14.7	14.6
Comparison with SCUBA (average NEFD) [mJy $\sqrt{\text{sec}}$]	-	1100	1000
Comparison with SOFIA (calculated NEFD) [mJy $\sqrt{\text{sec}}$]	550		

pleted their first series of extragalactic sub-mm and mm (850 μm — 1.3 mm) surveys. Although covering areas ranging from 0.002 to 0.2 deg^2 (Smail, Ivison et al. 1997; Hughes, Serjeant et al. 1998; Lilly, Eales et al. 1999; Carilli, Owen et al. 2001), these ground-based surveys are hundreds of times smaller than the proposed BLAST surveys (see Table 2).

The following results from the first SCUBA (850 μm) and MAMBO (1.25 mm) surveys have made a significant impact on several cosmological questions, while at the same time demonstrating the necessity for larger area and shorter-wavelength (250—500 μm) sub-mm observations:

- 30—50% of the sub-mm/FIR background detected by COBE has been resolved into individual sub-mm galaxies. The existing confusion-limited surveys are within a factor of a few in sensitivity of resolving the entire sub-mm/FIR background.
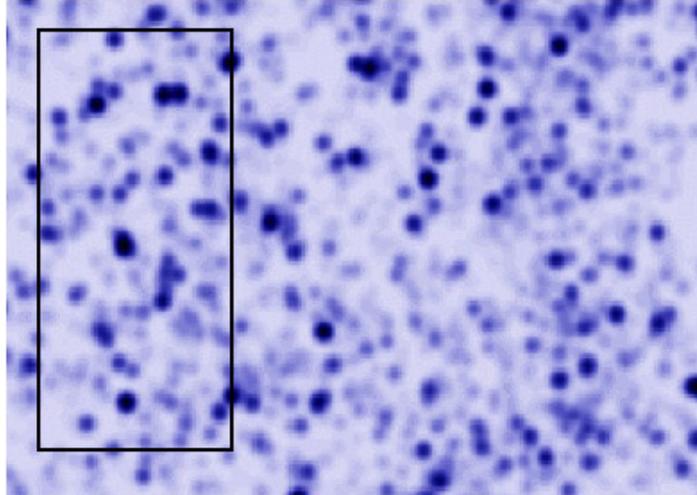
Figure 2   A simulated confusion-limited BLAST 250μm survey with a FWHM resolution of 30' and a 1σ sensitivity of 7 mJy. The assumptions for the evolution and the clustering of submm galaxies are described in Hughes & Gaztanaga 2000. The simulation area covers 0.1 deg^2 and is 1/10th the size of our smallest proposed survey. The rectangle represents the field of view of the 250 μm array.

- The majority of sub-mm sources appear to be associated with z>>1 Extremely Red Objects (EROs - IR galaxies that are extremely faint or undetected in the optical) and weak radio sources. There is still vigorous debate about the fraction of sources at z≥2.

- The sub-mm source-counts significantly exceed a no-evolution model and require strong evolution out to z~1, but place weak constraints at higher redshifts. The source-counts at bright sub-mm flux densities (N_{850} > 10 mJy) show a large scatter (about a factor of 5) between different surveys which may be the result of clustering on the scale of these surveys.

- The sub-mm surveys appear to find ~5 times the star formation rate that is observed in optical surveys at 2 < z < 4; how much of this discrepancy is due to the effects of dust obscuration and incompleteness in the optical is still being investigated (Adelberger and Steidel 2000).

BLAST/SCUBA OBSERVATIONS

A combined wide-area sub-mm survey is planned using SCUBA (at 850 and 450 μm) and BLAST (at 250, 350 and 500 μm) to (i) determine the star-formation history of massive, dust-enshrouded galaxies, (ii) identify whether these blank-field sub-mm galaxies are the progenitors of massive elliptical galaxies, and (iii) to provide photometric redshifts with sufficient accuracy to measure the evolution of the spatial clustering of the sub-mm sources. The planned SCUBA+BLAST survey will coordinate observations towards two "cirrus-free" fields covering a total of 0.5 sq. degrees that are accessible from Hawaii and a northern and southern hemisphere LDB flight.

The existing ground-based SCUBA surveys have identified more than 100 blank-field sub-mm sources. Unfortunately, due to the ambiguity in identifying a sub-mm galaxy with its optical counterpart, spectroscopic redshifts exist for only ~5 sub-mm sources. This "redshift deadlock" prevents a complete understanding of the nature and evolutionary history of sub-mm galaxies, since without redshifts, it not possible to measure the rest-frame luminosities, or the star-formation history for this population of optically-obscured galaxies.

The SCUBA+BLAST survey is designed to detect > 100-200 galaxies with accurate photometric sub-mm colors between 250 and 850 μm. Monte-Carlo simulations, that take into account the errors in the measurement and calibration, and the errors in the evolutionary models and spectral energy distributions of galaxies that fit the observed sub-mm number-counts, demonstrate that photometric redshifts with uncertainties of $_z = \pm 0.4$ can be obtained for galaxies over the redshift range $0 < z < 6$ (Hughes et al. 2001). Future millimeter-wavelength spectroscopic follow-up observations with the 100-m GBT and 50-m LMT will calibrate the accuracy this method. The cumulative redshift distribution for the population of galaxies detected in the SCUBA+BLAST survey, with FIR luminosities > 3 G 10^{12} L_{solar}, will provide the necessary information for a robust measurement of the SFR density between $0 < z < 6$ with an accuracy of ±20%. Furthermore, the combined SCUBA+BLAST survey will provide photometric redshifts with sufficient accuracy to remove any projection effects in the measured clustering signal, which itself is also a powerful test of whether the sub-mm galaxies are strongly-biased and trace high-mass peaks in the initial density fluctuations.

SCHEDULE

The BLAST instrument will have its first North American flight in the spring of 2003 from Ft. Sumner, New Mexico. The primary purpose of the first flight is to test the integrated system and qualify for the LDB flight from Antarctica in December of 2003. The test flight will incorporate only the 500 micron array. The 250 and 350 micron arrays will be installed in the summer of 2003 in preparation for the LDB flight.

ACKNOWLEDGMENTS

REFERENCES

Adelberger, K. and C. C. Steidel (2000). "Multiwavelength Observations of Dusty Star Formation at Low and High Redshift." ApJ **544**: 218.

Carilli, C. L., F. Owen, M. Yun, F. Bertoldi, A. Bertarini, K. M. Menten, E. Kreysa and R. Zylka (2001). Deep Millimetre Surveys - Implications for Galaxy Formation and Evolution, World Scientific.

Holland, W. S., E. I. Robson, W. K. Gear, C. R. Cunningham, J. F. Lighfoot, T. Jenness, R. J. Ivison, J. A. Stevens, P. A. R. Ade, M. J. Griffin, W. D. Duncan, J. A. Murphy and D. A. Naylor (1999). "SCUBA: a common-user submillimetre camera operating on the James Clerk Maxwell Telescope'." MNRAS **303**: 659.

Hughes, D., S. Serjeant, J. Dunlop, M. Rowan-Robinson, A. Blain, R. G. Mann, R. Ivison, J. Peacock, A. Efstathiou, W. Gear, S. Oliver, A. Lawrence, M. Longair, P. Goldschmidt and T. Jenness (1998). "High-redshift star formation in the Hubble Deep Field revealed by a submillimetre wavelength survey." Nature **394**: 241.

Hughes, D.H., Aretxaga, I., Chapin, E.L., Gaztanaga, E., Dunlop, J.S., Devlin, M.J., Halpern, M., Gundersen, J., Klein, J., Netterfield, C.B., Olmi, L., Scott, D., Tucker, G., (2000), "Breaking the Redshift Deadlock - I: Constraining the star formation history of galaxies with sub-millimetre photometric redshifts," , astro-ph/0111547 submitted MNRAS.

Hunter, T. R., D. J. Benford and E. Serabyn (1996). "Optical Design of the Submillimeter High Angular Resolution Camera (SHARC)." PASP **108**: 104.

Lamarre, J. M., M. Giard, Pointecouteau, B. E., J.P., , G. Serra, F. Pajot, F. X. Desert, I. Ristorcelli, J. P. Torre, S. Church, N. Coron, J. L. Puget and J. J. Bock (1998). "First Measurement of the Submillimeter Sunyaev-Zeldovich Effect." ApJ **507**: L5.

Lilly, S., S. Eales, W. K. Gear, F. Hammer, O. Le Fevre, D. Crampton, R. Bond and L. Dunne (1999). "The Canada-United Kingdom Deep Submillimeter Survey. II. First Identifications, Redshifts, and Implications for Galaxy Evolution." ApJ **518**: 641.

Scott, D., Ade, P., Bock, J., deBernardis, P., Devlin, M., Griffin, M., Gundersen, J., Halpern, M., Hughes, D., Klein, J., Masi, S., Mauskopf, P., Netterfield, B., Olmi, L., Page, L., Tucker, G., (2000), "The Balloon-Borne

Large Aperture Sub-Millimeter Telescope," Proc. Symposium 'The Promise of the Herschel Space Observatory' 12-15 December 200, Toledo, Spain.

Smail, I., Ivison and A. W. Blain (1997). "A Deep Sub-millimeter Survey of Lensing Clusters: A New Window on Galaxy Formation and Evolution." MNRAS **490**(L5).

Session VI

Detector Technology

FUTURE FAR INFRARED/SUBMILLIMETER DIRECT DETECTORS

P.L. Richards
Department of Physics
University of California
Berkeley, CA 94720

Abstract

The new mission concepts developed at this workshop will require large format arrays (1024 or more) of direct detectors at far infrared and millimeter wavelengths. The present status of such detectors will be reviewed and new technologies will be described which promise to meet anticipated mission requirements.

Semiconducting Photon Detectors

Our wavelength range can be divided into the 40-200µm band where Ge:Ga photoconductors are traditionally used and the 200-2000µm band where bolometers are used. Well developed existing technologies for the 4-200µm band include the SIRTF/MIPS Ge:Ga photoconductive arrays developed at Arizona. These have 1024 unstressed detectors at 70µm and 40 stressed detectors at 160µm and are ready for flight in early 2003. They operate at a few K, have detective quantum efficiencies of 15-20% and use MOSFET charge-integrating amplifiers and multiplexers. The performance of the SIRTF arrays is generally very satisfactory, but could be improved in several ways. Photoconductors require a relatively large detector thickness, giving significant sensitivity to cosmic rays. The quantum efficiencies could be better and there are both fast and slow responses which complicate observing strategies and data analysis. The major problem is the complicated stress mechanism required to produce 160µm response. This limits array size and reliability. Similar detectors are being developed for spectroscopy on Hershel/PACS.

Several projects are underway to produce improved photon detectors. Germanium blocked impurity band (BIB) detectors are under development at Berkeley and Rochester. If the performance of Si BIB's can be reproduced in Ge, they would have reduced cosmic ray cross section, higher quantum efficiency and long wavelength response without stressing. An alternative approach being explored in Berkeley is GaAs photoconductors for longer wavelengths than Ge.

In order to avoid stressed detectors, the CEA in France is implementing arrays of 2048 bolometers at 70µm and 512 bolometers at 170µm for photometry on the Hershel/PACS instrument. These ~300mK arrays use MOSFET amplifiers and multiplexers with very high impedance implanted Si thermistors to overcome the high amplifier voltage noise. The CEA group has developed a generally useful two-layer structure for filled arrays in which the thermistor and absorber grids on one wafer are bump-bonded to the backshorts and wiring on another. Their thermistor/amplifier technology is thought to be satisfactory for the NEP $= 10^{-16}$W/$\sqrt{\text{Hz}}$ required by PACS, but probably not for low background missions.

Well developed bolometer technologies exist for moderate sized arrays of bolometers for 200 - 2,000 μm which operate at temperatures from 100-300mK. They are fabricated with either Si or SiN membranes and use continuous metal film or grid absorbers, and either ion implanted Si or neutron transmutation doped Ge thermistors and JFET amplifiers at ~100K. An AC bias is used when low frequency stability is required. Bolometer architectures include the GSFC pop-up detectors or the CEA two layer structures for filled arrays and the JPL/Caltech spiderweb bolometers for horn-coupled arrays. The choice between these types of array depends on whether the factor that limits array size is focal plane area, or complexity and power dissipation associated with the number of detectors. Horn-coupled arrays produce a sparsely sampled pattern on the sky so that the beams must be dithered or scanned to fill in a map. Although the current generation of bolometers gives excellent performance in many applications, the JFET amplifiers severely limit the ability to produce arrays of more than a few hundred detectors. In addition to thermal and microphonics issues, the absence of an amplifier noise margin causes systems problems.

TES Bolometers

Because of these limitations, there is a need for revolutionary new long wavelength detectors. Fortunately, very promising new approaches are under active development. The voltage-biased superconducting bolometer with transition edge sensor (TES) and SQUID readout amplifier is a negative feedback thermal detector with many favorable operating properties. It can be made entirely by thin film deposition and optical lithography. The feedback reduces the response time, improves the linearity, and isolates the bolometer responsivity from infrared loading or heat sink temperature. There is also significant suppression of Johnson noise. The SQUID amplifiers operate at bolometer temperatures, dissipate very little power and have significant noise margin. These bolometers are being produced at Berkeley, Caltech/JPL, GSFC, NIST and elsewhere with appropriate architectures for 1024 element filled and horn coupled arrays. In addition, there is work on bolometers that are coupled to the optics with planar lithographed antennas and superconducting microstrip lines. In this implementation, the absorbing element is a resistive termination to the line, which can be very small. At low operating temperatures $\simeq$ 100mK, such terminations can be deposited directly on the dielectric substrate. The weakness of the coupling between electrons and phonons provides adequate thermal isolation without membranes or legs.

In antenna-coupled bolometers, the submillimeter wave signal is propagated in a superconducting microstrip transmission line. These lines can branch to form diplexers, and can incorporate high performance microstrip bandpass filters. The possibility exists of making a low resolution spectrometer with this technology. The Berkeley group is making crossed double-slot dipole antennas coupled to two bolometers. The difference between the bolometer signals is sensitive to the polarization of the signal illuminating the pixel. The Caltech/JPL group is studying an array of many slot antennas to give a narrow antenna pattern which couples easily to array optics with a wide frequency bandwidth that can be divided into several photometric bands.

Large format arrays of TES bolometers should have output multiplexing to avoid very large numbers of leads leaving the cryostat. Lines of ~50 detectors can be multiplexed before amplification using superconducting thin film technology. The NIST group has developed a time-domain multiplexer which uses a SQUID for each bolometer to switch the outputs sequentially through a single SQUID amplifier. Groups in Berkeley and Helsinki are developing frequency domain multiplexers which combine the signals from a row of bolometers, each of which is biased at a different frequency. The signals are then amplified by a single SQUID and recovered with ambient temperature lock-in amplifiers. The success of the TES bolometer technology ultimately depends on the success of one or both of these multiplexers.

Superconducting Photon Detectors

In a typical photon detector, the incident IR photon excites an electron into a (nearly) empty conduction band or hole into the valence band of a cold semiconductor. The current due to these carriers is then measured before they recombine. An analogous detector can be made in a superconductor. In this case, the electrons in a cold dark superconductor are bound into Cooper pairs. An incident photon can break a pair and create two single electrons, or quasiparticles. Given a readout scheme that is sensitive to the quasiparticles, but not the pairs, a superconducting photon detector can be made.

The inductance of a superconducting strip contains a contribution that is proportional to the density of quasiparticles. A detector is being developed at Caltech/JPL that places the inductor in a microwave resonant circuit that is AC biased off resonance, so that the transmitted signal depends on the quasiparticle density. A large number of such detectors can be read out (multiplexed) through a single broad band HEMT amplifier if each one is biased at a different frequency.

In a different development at Yale/Goddard, the quasiparticle density is measured with a superconductor/insulator/normal metal tunnel junction that transmits quasiparticles, but not pairs. The tunneling electrons go to the gate of a superconducting single-electron transistor (SET). The SET is the dual of the SQUID. It is so small that a single electron on the gate will measurably change the SET current. This device can thus operate as a submillimeter photon counter. Since photon rates at these frequencies are high even from dark regions of the sky, the Yale group is developing an RF SET with a high readout rate which is appropriate for multiplexed output.

Mission Requirements

The new missions under consideration at this workshop, such as SAFIR and SPECS, will require large format arrays of 1024 or more photon-noise limited detectors. Detectors for photometric bandwidths must have NEP ~ 10^{-18}W/$\sqrt{\text{Hz}}$. Detectors for dispersive spectrometers require 10^{-20} or 10^{-21}W/$\sqrt{\text{Hz}}$. Except possibly at 70μm, where the Ge:Ga photoconductors perform well, these requirements can be met with multiplexed voltage-biased TES bolometers. If such bolometers are used on any bands, then, for simplicity, they may well be used for all the far infrared submillimeter bands. Although tests of prototype detector arrays and multiplexers suggest strongly that this technology will meet mission requirements, there is very little experience with

assembling large TES bolometer systems for astronomical observations. Fortunately, such experience can be gained in time to benefit far-infrared/submillimeter missions. Development is now beginning on a large number of ground-based systems which will use TES bolometers in large format filled, or horn-coupled arrays with multiplexers at millimeter wavelengths. The SCUBA II development is well underway, but others will follow soon. Constellation X is proposing to fly large format arrays of TES X-ray calorimeters which have very similar cryogenic and readout requirements to those anticipated for TES bolometers on SAFIR at the end of the decade.

Polarization sensitive bolometers will be required to measure the polarization of the cosmic microwave background at submillimeter and millimeter wavelengths. Ground-based experiments with more than 10^3 antenna-coupled TES bolometers with multiplexed output are now being planned. A CMB polarization space mission is expected to occur which will certainly make use of these technologies.

With further development, it is possible that one or the other of the photon detectors will eventually displace the TES bolometers. In principle, they could operate at higher temperatures. The photon center could find a role in very high resolution dispersive spectroscopy.

The Role of Coherent Detection

J. Zmuidzinas[1]

[1]Division of Physics, Mathematics, and Astronomy
California Institute of Technology, 320–47
Pasadena, CA 91125
jonas@submm.caltech.edu

Abstract

Many interesting astronomical objects, such as galaxies, molecular clouds, PDRs, star–forming regions, protostars, evolved stars, planets, and comets, have rich submillimeter spectra. In order to avoid line blending, and to be able to resolve the line shape, it is often necessary to measure these spectra at high resolution. This paper discusses the relative advantages and limitations of coherent and direct detection for high resolution spectroscopy in the submillimeter and far–infrared. In principle, direct detection has a fundamental sensitivity advantage. In practice, it is difficult to realize this advantage given the sensitivities of existing detectors and reasonable constraints on the instrument volume. Thus, coherent detection can be expected to play an important role in submillimeter and far–infrared astrophysics well into the future.

Introduction

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. Indeed, coherent detection will play a key role in upcoming major submm/far-IR projects and missions, such as ALMA, SOFIA, and Herschel. However, coherent detection has one fundamental disadvantage, which is a limit to sensitivity that is imposed by quantum mechanics. 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, such as the large–aperture SAFIR mission envisioned for NASA. Will there be a role for coherent detection in future submillimeter space missions beyond Herschel ?

This paper examines the case for coherent detection in the submillimeter/far–IR, following the approach of an earlier paper. [1]. The case is straightforward: first, high spectral resolution is scientifically important; and second, although in principle direct detection does have a sensitivity advantage, it is difficult to realize this advantage in practice for the case of high resolution spectroscopy.

1. Submillimeter and Far–Infrared Spectroscopy

Spectroscopy in the submillimeter and far–IR, although still in early stages of development, has the potential to become a very powerful tool for astrophysics. Here one finds the ground–state transitions of numerous hydride species that are critical for understanding interstellar chemistry. One also has numerous diagnostics of warm gas, such as the high–J CO lines as well as the atomic and ionic fine structure lines of the lighter elements such as C, N, and O. At these long wavelengths, observations can penetrate through very large column densitites of dust, which are totally opaque in the infrared, optical, and UV, and only start becoming transparent again to X–ray photons.

So far, we have collectively obtained only a small glimpse of what submm/far–IR spectroscopy has to offer, through the pioneering work done on the Kuiper Airborne observatory and ground–based telescopes, followed by the substantial advances enabled by ISO. For example, the ISO far–IR spectra of nearby galaxies [2] show a wide variety of characteristics, and issue a challenge to all of us to decode their meaning. SOFIA and the Herschel Space Observatory will make enormous contributions, pushing beyond ISO to longer wavelengths, better sensitivity, better angular resolution, and higher spectral resolution.

In the future, it will be possible go beyond Herschel by many orders of magnitude in sensitivity, by using a colder telescope, optimized instrumentation, and better detectors. Although much attention has been focused on the *detection* of distant objects through submm/far–IR imaging, sensitive submm/far–IR *spectroscopic* observations of distant, high–redshift "submillimeter" galaxies will be a crucial next step toward understanding their redshift distribution, energy sources, and origin. Such distant–object studies will probably best be done with moderate resolution direct–detection spectrometers. However, developing an understanding of the basic physical phenomena involved will require the study of nearby objects at higher spectral resolution.

2. The Importance of High Spectral Resolution

There are many examples that can be given to illustrate the importance of high spectral resolution. Submillimeter line surveys, such as the recent 650 GHz CSO survey [3], have shown that the spectra of star–forming regions can be exceedingly rich, and that high spectral resolution (1 km s^{-1} or better) is necessary to deal with line confusion and line blending. The detection and subsequent analysis of the abundance, excitation, and physical origin of numerous chemical species would be impossible without high spectral resolution.
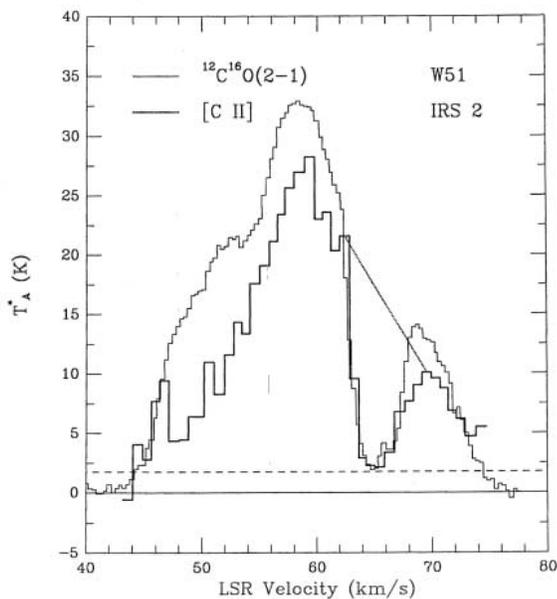
Figure 1: A high–resolution spectrum of the 158 μm [C II] line toward W51 IRS2 [4]. Note the sharp absorption feature, which in this case is most likely produced by a foreground cloud with a low–density PDR.

Water line observations from SWAS at 557 GHz provide another example. Here the line profiles are often observed to be "self–reversed", with deep, sharp absorption dips superposed on a broader emission profile [5]. This type of profile is expected to be common for the case of embedded sources, where the warmer gas located closer to the source is surrounded by cooler foreground material. The abundance of water vapor can vary greatly with location, and the large molecular dipole moment provides a strong coupling to the submm/far–

IR radiation field. These effects, along with large optical depths, can also be expected to yield complex line profiles. In such situations, comparison of optically thick and optically thin line profiles (e.g. $H_2^{16}O$ vs. $H_2^{18}O$) can lead to important insights.

Furthermore, complex profiles can occur even for species that are spatially widespread and have transition strengths and optical depths that are normally thought to be low. The 158 μm line of ionized carbon (C II) provides a good example. High resolution spectroscopy of galactic star–forming regions (performed on the Kuiper Airborne Observatory in the 1980's by Betz, Boreiko, and Zmuidzinas) often reveals complex line profiles, with deep, optically thick absorption dips (see Figure 1) due to foreground material. Ground–state transitions are particulary susceptible to this effect.

Finally, line profiles can often carry important information about the nature of the emission. A recent example is the submillimeter HCN laser in IRC+10216 [6], whose narrow line width is a characteristic signature of laser action.

3. The Connection Between Spatial and Spectral Resolution

Both spatial and spectral resolution are important. Although one often sees these plotted on orthogonal axes when describing instrument capabilities, in fact often there is a close connection:

- High spatial resolution often results in narrow line widths

- Spectral resolution can be used to gain spatial information

- Spectral resolution can help break through the broadband photometric spatial confusion limit

One example of the first point is the mapping of line emission from nearby galaxies with millimeter interferometers, where the synthesized beam covers a single molecular cloud, as compared to single–dish observations, where the beam picks up the entire rotation curve of the galaxy.

This behavior is not universal, however; in some cases the line profile becomes broader at higher spatial resolution, as occurs for galactic nuclei, or more generally when a localized gravitational potential plays an important

role in the dynamics. For some of these cases, one can use spectral information to gain spatial information. A good example is the analysis of the CO emission from the nuclear region of Arp 220 by Scoville and collaborators [7], [8], in which a self–consistent model of the mass distribution and dynamics was used to infer the structure at a scale smaller than the synthesized beam.

One does not necessarily need to have a dynamical model to extract spatial information from spectral data; a purely empirical approach is possible. In this case, one uses a second data set which has both high spatial and spectral resolution to perform the decoding from the spectral to spatial domain. There are well–defined algorithms for doing this. A recent example of this approach is the use of an IRAM CO(2-1) interferometer map along with a CSO CO(6-5) map to determine a line ratio map at somewhat higher spatial resolution than the CSO data [9]. In this context, it is intriguing to consider the use of ALMA maps in combination with high spectral resolution observations with SAFIR. For instance, what might this combination provide for the study of protoplanetary disks ?

4. Coherent Detection

Coherent and incoherent detection are easily distinguished. The fundamental difference is that coherent detection instruments respond to the complex amplitude of the field (amplitude and phase), whereas incoherent detection instruments respond to the intensity (power). Using the language of quantum mechanics, coherent instruments measure the value of a, the photon destruction operator, whereas incoherent instruments measure $a^{\dagger}a$, the photon number operator. The "quantum noise" limitation for coherent receivers arises from the fact that they measure both amplitude and phase (or "position" and "momentum") simultaneously, and these are not commuting obervables, and so are subject to an uncertainty principle [10]. 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.

In essence, coherent instruments amplify the field into the classical domain. The relevant quantity describing this translation from the quantum to the classical regime is the *photon number gain*, which is not necessarily synonymous with power gain. Figure 2 shows a block diagram of a typical heterodyne re-
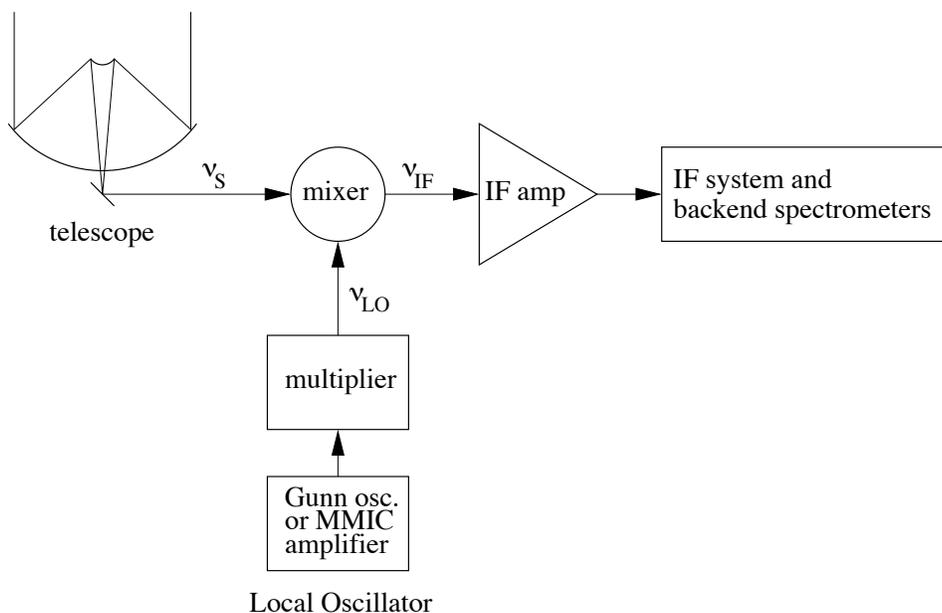
Figure 2: A block diagram of a submillimeter heterodyne receiver system. The signal from a telesope at a frequency ν_S is combined with a local oscillator at frequency ν_{LO} in a "mixer", which is a nonlinear device, to yield the difference or "intermediate" frequency $\nu_{IF} = |\nu_S - \nu_{LO}|$, which is then amplified and and spectrally analyzed.

ceiver. Even though the mixer may have power conversion loss, it most cases it has photon number gain, because the output frequency is much lower than the input frequency. Of course, the receiver system should always have overall photon number gain after the signal passes through the IF amplifier.

The physical origin of quantum noise can be illustrated for an ideal hypothetical amplifier, as shown in Figure 3. This device, consisting of an inverted population of atoms or molecules in a tube, clearly has photon number gain due to stimulated emission. It also has noise – quantum noise – which is due to spontaneous emission. One way of describing spontaneous emission is that it is emission "stimulated" by the zero–point quantum fluctuations of the electromagnetic field.

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
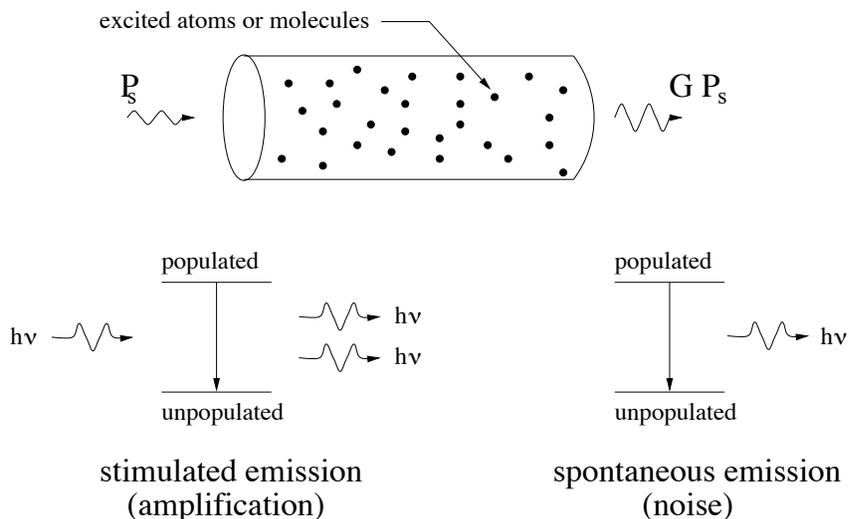
Figure 3: An illustration of quantum noise in a maser amplifier. This (fictitious) maser amplifier consists of a tube filled with a gas of molecules or atoms, which are pumped in a way that causes some transition with frequency ν to be inverted. A signal arriving at the input with power P_s is amplified 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 $P_s = 0$.

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, at radio or microwave frequencies, by the "backend". 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.

Another key point is that the *backend spectrometer does not need cryogenic cooling* because its noise is basically irrelevant, since the signal has been amplified. In contrast, the spectroscopic components of direct–detection instruments must be cold ($k_B T << h\nu$) to reduce the thermal background noise.

5. Direct Detection Spectroscopy

A wide variety of instruments have been used for direct detection spectroscopy: gratings, Fabry–Perots, Fourier Transform Spectrometers (FTS), etc. The choice depends on various requirements, including the wavelength range, the resolution and sensitivity required, the detectors available, the background level, size, cost, and so on. This plethora of approaches obscures the fact that *there is only one optimal 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.

In order to obtain maximum sensitivity, spectrometers should 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 wavelenth, 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 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. The conclusion is that for best sensitivity, only grating spectrometers (or their equivalent) should be used.

An interesting corollary is that correlation spectroscopy, widely used in heterodyne backends, is a fundamentally inferior technique when used for direct detection. In contrast, it is well known that there is no sensitivity penalty for using correlators as backends. What is the difference ? Heterodyne backends measure classical signals, with very high photon occupation numbers, whereas direct–detection instruments on cold telescopes operate in the low occupation number regime.

In this context we note that the SPIRE instrument on Herschel will use an FTS for submillimeter spectroscopy. While there were numerous practical reasons for this choice, including limitations on the available detector sensitivity, the choice of an FTS means that SPIRE cannot achieve the maximum possible

sensitivity for spectroscopy. *A grating spectrometer, along with a much colder telescope, can offer spectacular sensitivities for spectroscopy*, many orders of magnitude better than SPIRE/Herschel. This is a very interesting possibility for the future, and is discussed in more detail in the presentations and papers by J. Bock, C. M. Bradford, and J. Glenn at this workshop.

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 degradatation. While there are no fundamental limitations, there are obviously enormous practical problems with this approach.

One is forced to look at ways of reducing the volume. Using guided–wave propagation helps; this is the key idea behind the waveguide grating spectrometer concept described by Bradford et al. at this workshop. Other ideas have been suggested, such as using the slower propagation in high–index materials such as silicon or germanium. However, none of these approaches is very likely get to $R = 10^6$.

The remaining possibility is to fold 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.

6. Comparison of coherent and direct detection

There are two ways to perform this comparison. One can look at demonstrated sensitivities of existing instruments, or careful design studies, including the limitations of existing detectors. In many ways this is the best approach, since it includes all of the "real life" factors – filter inefficiencies, excess noise, etc. This is the approach that should be used before starting construction of a major instrument. On the other hand, it is not helpful for projecting into the future.

The opposite extreme is to assume that fundamental limits will be approached closely. This is likely to be a very accurate projection for the future, but on an uncertain time scale. This is (almost) the approach that I will take, as shown in Figure 4. The one arbitrary "real–life" adjustment I have made
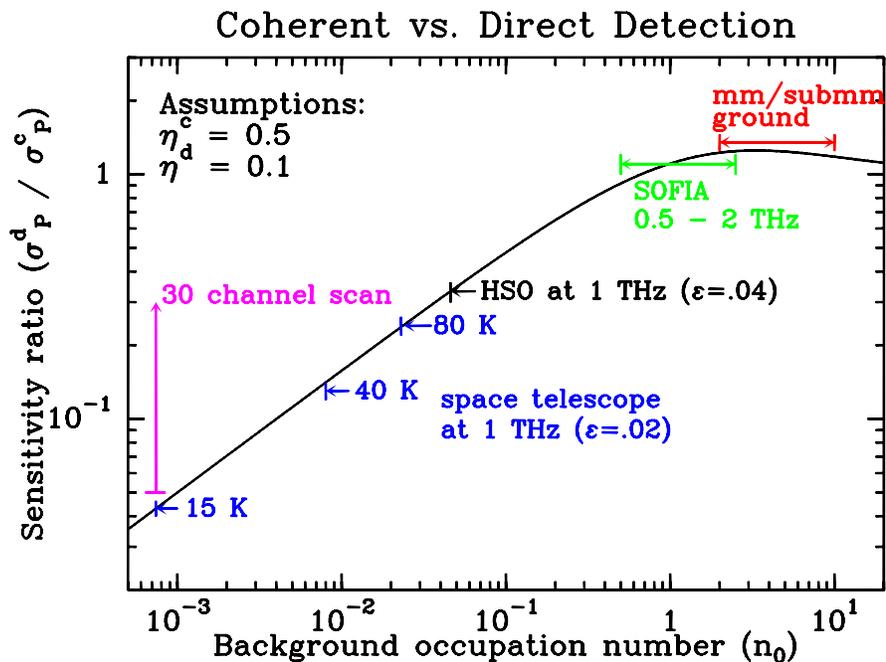
Figure 4: An idealized comparison of the relative sensitivities of coherent and direct detection. The vertical axis is the sensitivity ratio; a ratio less than unity favors direct detection. The vertical arrow shows the sensitivity penalty associated with a 30–channel sequential spectral scan. Sensitivity equations may be found in [1].

is to give the coherent instrument a better overall efficiency, since in reality it would have a much simpler optical system.

The sensitivity of ideal direct detection depends entirely on the background. This can be quantified by specifying the mean photon occupation number n. Ideal quantum–limited coherent detection corresponds to $n = 1$. For $n \sim 1$, there is no advantage to direct detection – which is the case for ground–based and airborne mm/submm/far-IR instruments. On the other hand, for $n << 1$, direct detection has an advantage which scales as $\sqrt{n}$, because coherent detection is quantum–limited with an equivalent background of $n = 1$.

For high spectral resolution, say $R \sim 10^6$, direct detection instruments
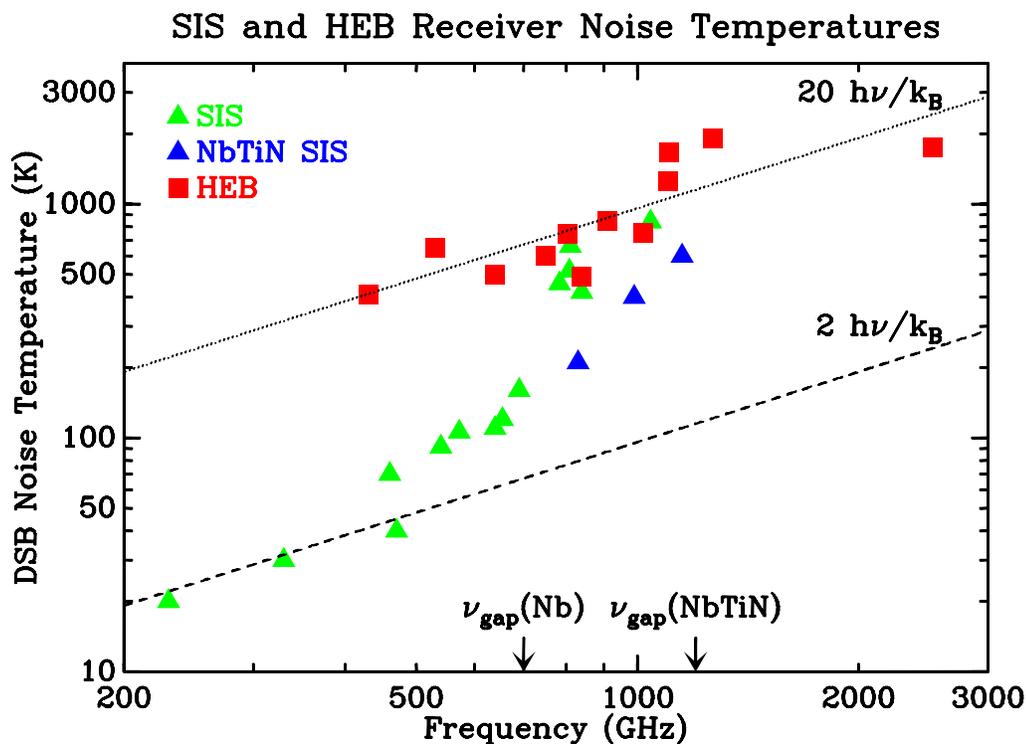
Figure 5: A selection of measured sensitivities for SIS and HEB receivers.

would use Fabry–Perots (as discussed in section 5) 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. 5 for the case $M = 30$. The overall result is that direct detection is still more sensitive, but not by an overwhelming factor. However, there are several other 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. 5)

- 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. $10^{-18}\,\mathrm{W\,Hz^{-1/2}}$)

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.

7. Status of submm/far–IR receiver technology

Figure 5 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.

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.

Future developments can be expected in several areas. First, and most importantly, the push toward quantum–limited sensitivities must be continued. It has been demonstrated that SIS mixers are capable of reaching the quantum limit at millimeter wavelengths. However, SIS mixers become increasingly difficult to produce at higher frequencies, and will not operate above 1.5–1.6 THz with current materials. For higher frequencies, HEB mixers are used. 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. 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. Finally, there is still substantial room for development and improvement of local oscillators.

Acknowledgements

References

[1] J. Zmuidzinas, "Progress in coherent detection methods," in *The Physics and Chemistry of the Interstellar Medium* (V. Ossenkopf et al., ed.), pp. 423–430, U. Cologne, GCA–Verlag Herdecke, 1999.

[2] J. Fischer, "Galaxies: The long wavelength view," in *Proceedings of the Conference "ISO beyond the Peaks", Villafranca del Castillo, Spain, 2–4 February 2000 (ESA SP–456, 2000).*

[3] P. Schilke, D. J. Benford, T. R. Hunter, D. C. Lis, and T. G. Phillips, "A line survey of Orion–KL from 607 to 725 GHz," *Ap. J. (Suppl.)*, vol. 132, pp. 281–364, February 2001.

[4] J. Zmuidzinas, *Heterodyne Spectroscopy of Neutral and Ionized Carbon in the Interstellar Medium.* PhD thesis, University of California, Berkeley, 1987.

[5] M. L. N. Ashby *et al.*, "An analysis of water line profiles in star formation regions observed with the Submillimeter Wave Astronomy Satellite," *Ap. J. (Lett.)*, vol. 539, pp. L115–L118, August 2000.

[6] P. Schilke, D. M. Mehringer, and K. M. Menten, "A submillimeter HCN laser in IRC+10216," *Ap. J. (Lett.)*, vol. 528, pp. L37–L40, January 2000.

[7] N. Z. Scoville, M. S. Yun, and P. M. Bryant, "Arcsecond Imaging of CO Emission in the Nucleus of Arp 220," *Ap. J.*, vol. 484, pp. 702+, July 1997.

[8] K. Sakamoto, N. Z. Scoville, M. S. Yun, M. Crosas, R. Genzel, and L. J. Tacconi, "Counterrotating Nuclear Disks in Arp 220," *Ap. J.*, vol. 514, pp. 68–76, Mar. 1999.

[9] J. S. Ward, J. Zmuidzinas, A. I. Harris, and K. G. Isaak, "A ^{12}CO $J = 6 - 5$ Map of M82: The significance of warm molecular gas," *Ap. J.*, 2001. (submitted).

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

Ideal Integrating Bolometer
A. Kogut[1], M. DiPirro[2], and S.H. Moseley[1]

We describe a new "ideal integrator" bolometer as a prototype for a new generation of sensitive, flexible far-IR detectors suitable for use in large arrays. The combination of a non-dissipative sensor coupled with a fast heat switch provides breakthrough capabilities in both sensitivity and operation. The bolometer temperature varies linearly with the integrated infrared power incident on the detector, and may be sampled intermittently without loss of information between samples. The sample speed and consequent dynamic range depend only on the heat switch reset cycle and can be selected in software. Between samples, the device acts as an ideal integrator with noise significantly lower than resistive bolometers. Since there is no loss of information between samples, the device is well-suited for large arrays. A single SQUID readout could process an entire column of detectors, greatly reducing the complexity, power requirements, and cost of readout electronics for large pixel arrays.

1. Introduction

A bolometer consists of a thermometer with heat capacity C mounted on an absorbing substrate, weakly coupled to a bath at temperature T_0 via a thermal conductance G (Figure 1). Resistive bolometers (including both composite and transition-edge bolometers) use the temperature dependence of a resistive element to measure the infrared signal incident on the absorber. In steady-state operation, the thermometer heats up until the dissipated power matches the power conducted through the thermal link to the bath,

$$\Delta T_{dc} = \frac{P_{elec} + P_{IR}}{G} , \qquad (1)$$

where P_{elec} and P_{IR} represent the electrical and infrared power, respectively. Larger temperature excursions — a larger signal — can be achieved by reducing the conductivity G of the bolometer to the bath. To mitigate low-frequency (1/f) noise, though, the infrared signal must be modulated at some frequency ω. The synchronous temperature change then becomes

$$\Delta T_{ac} = \frac{\Delta P_{IR}}{G\sqrt{1 + \omega^2 \tau^2}} , \qquad (2)$$

where $\tau = C/G$ is the time constant of the device. Equation 2 is equivalent to the dc response convolved with a low-pass filter and illustrates a fundamental limitation of resistive devices: the conductance G can not be lowered arbitrarily to increase the dc response without forcing the time constant τ to unacceptably large values. For fixed

[1] Code 685, Goddard Space Flight Center, Greenbelt, MD 20771. E-mail Alan.Kogut@gsfc.nasa.gov
[2] Code 552, Goddard Space Flight Center, Greenbelt, MD 20771

chopping frequency ω, the sensitivity reaches a plateau at $G \approx C\omega$ beyond which further reductions in G bring little improvement.
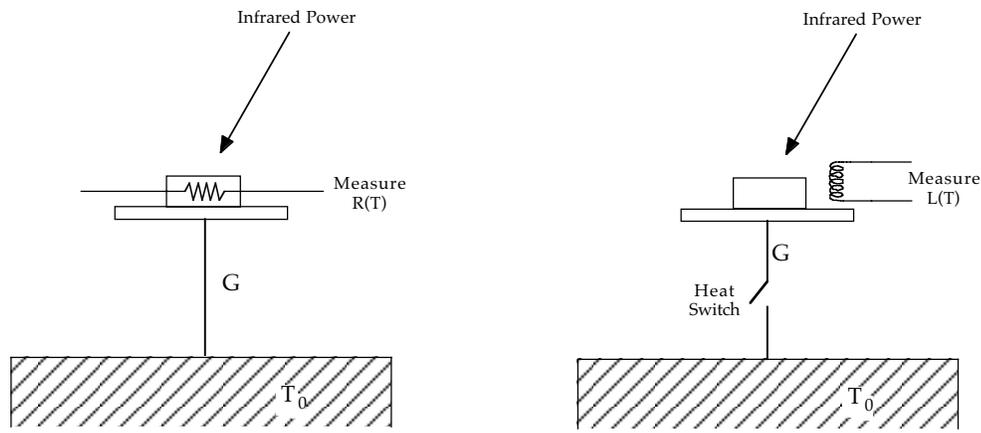
We have developed a new bolometer concept using a non-dissipative thermometer coupled with a heat switch to break the link between thermal conductance and performance. Penetration-depth thermometers use inductive coupling to measure the temperature of a thin superconducting film without dissipating heat into the system. Temperature changes in the bolometer thus depend *solely* on the integrated infrared power incident on the absorber. The bolometer time constant can then be made arbitrarily long, allowing the device to function as an ideal (lossless) integrator. A heat switch periodically increases the conductance G by a factor of 10^4 or more during short "reset" intervals. This forces the bolometer to the bath temperature on time scales short compared to the intrinsic time constant, reducing effects of 1/f noise and thermal drifts on time scales longer than the period between successive resets.

2. Integrating Bolometer

Figure 1 shows a schematic of the ideal integrating bolometer. It consists of a non-dissipative thermometer on an absorbing substrate, weakly coupled to a thermal bath through a heat switch with conductance ratio

$$\gamma = G_{on}/G_{off} . \tag{3}$$

The combination of a non-dissipative sensor with a resettable heat switch breaks the link between conductance and sensitivity that limits resistive bolometers. In the "on" state the conductance is high and the thermometer quickly cools to the bath temperature. In the "off" state the conductance is minimal and the thermometer heats as

$$\frac{dT}{dt} = \frac{P_{IR} - G(T - T_0)}{C} \approx \frac{P_{IR}}{C} \tag{4}$$

Fig 1: Schematic of resistive bolometer (left) and integrating bolometer (right). The integrating bolometer uses a heat switch and non-dissipative thermometer to provide greatly increased sensitivity and lossless integration.

Provided $G\Delta T \ll P_{IR}$, the bolometer responds as an ideal integrator, with temperature change linearly dependent on the integrated infrared power incident on the absorber.

Penetration-depth thermometers are an attractive technology for the lossless thermometer. They use the temperature dependence of the (partial) screening of an external magnetic field by a thin superconducting film to monitor the temperature of the film without dissipative losses or electrical contacts. The Meissner effect in a superconductor generates surface screening currents to block an external magnetic field; the simple case of a bulk superconductor in a parallel magnetic field B_0 leads to an internal field

$$B(x) = B_0 \exp(-x/\lambda) \tag{5}$$

where λ is the characteristic penetration depth for the field. Near the superconducting transition, the penetration depth has a strong temperature dependence

$$\lambda(T) = \lambda_0 \left[1 - \left(\frac{T}{T_c} \right)^4 \right]^{-1/2} \tag{6}$$

where T_c is the transition temperature and $\lambda_0 \approx 5\text{-}20$ nm is the penetration depth at T=0. For film thickness $d \sim \lambda_0$, the screening is incomplete and the attenuated magnetic field tracks the thermal dependence of the penetration depth.

Use of a non-dissipative thermometer allows the bolometer to operate as an ideal integrator, with temperature change linearly dependent on the integrated infrared power. The bolometer stays in the linear regime provided that $G\Delta T \ll P_{IR}$ and the integration time Δt is short compared to the time constant $\tau = C / G$. Both of these require small conductance, and hence a long time constant. We recover device speed and avoid unwanted 1/f contributions by using a heat switch to link the absorbing substrate to the bath. The switch effectively produces two time constants: a long time constant in the "off" state while integrating the sky signal, and a short time constant in the "on" state to reset the device to the bath temperature.

The simplest choice for a heat switch uses a thin metal strip toggled between its superconducting and normal states. At temperatures well below the superconducting transition, all conduction electrons are paired and the superconductor has small thermal conductivity. Applying a magnetic field drives the superconductor to its normal state, greatly increasing the thermal conductivity. Conduction ratios $\gamma > 10^4$ can be achieved using aluminum near 100 mK (Mueller 1978).

A superconducting heat switch requires a magnetic field large enough to drive the switch to the normal state, without interfering with the thin film fields nearby. We achieve this by sandwiching a thin aluminum film between niobium layers deposited on one support leg of the bolometer. The top and bottom niobium layers form a continuous circuit, generating an appreciable field within the aluminum strip between them, but cancelling at distances large compared to the 0.4 μm separation of the sandwich. A current 5 mA or less through 10 μm wide niobium strips will generate the 10^{-2} Tesla critical field needed to force the aluminum to the normal state, falling below 10^{-9} Tesla at

the penetration depth pickup coil. Since the niobium legs are deposited on top of each other, the fringe field at the coil is primarily parallel to the chip surface and will not affect the perpendicular component sensed by the bolometer. Measurements of cross-talk on a test switch agree well with calculations, providing assurance that cross-talk between the heat switch and the pickup coil will not limit the performance of the bolometer. The niobium remains in the superconducting state throughout the entire reset cycle, adding negligibly to the total thermal conductance. The inductance of the switch is small (L < 1 nH), allowing rapid toggling.

The integrating bolometer has a relatively simple readout. A persistent current stored in a superconducting drive coil generates an external field through the thin-film thermometer. As the film temperature changes, the inductance of the coil-film system changes. To conserve flux through the superconducting coil, the coil current changes and can be measured with a SQUID. Figure 2 shows a typical sampling sequence. When the switch is opened, the incident radiation begins to heat the bolometer. A readout SQUID measures the current in the pickup coil, then returns a time Δt later to re-sample the current (using the intervening time to sample other pixels in the array). The difference in measured currents yields the integrated power incident on the detector in the time Δt between samples. After the second sample, we close the heat switch to reset the device. Note that the measured signal depends only on the difference in sampled currents between successive readouts — the technique does not require the bolometer to return to precisely the same temperature each cycle, and is insensitive to drifts on time scales long
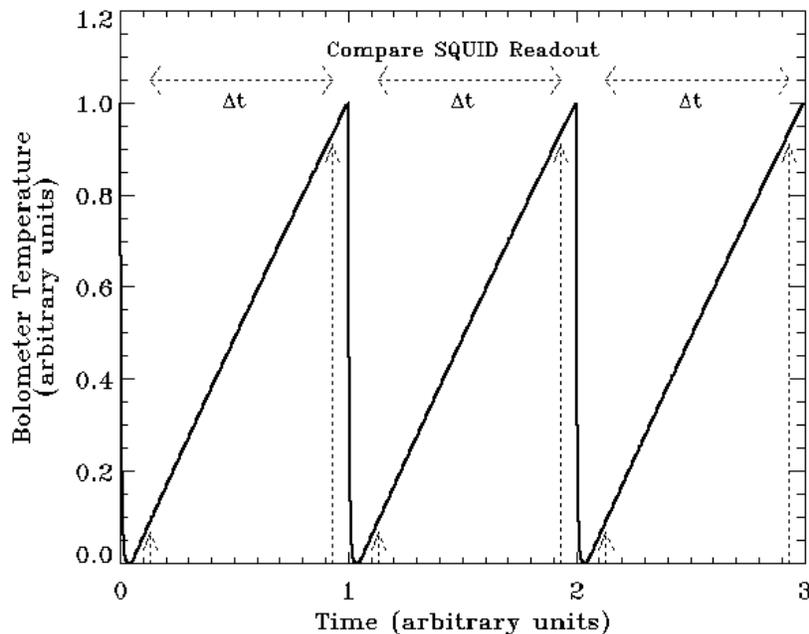
Figure 2: Three successive readout cycles for an integrating bolometer. A SQUID samples the temperature-dependent current after the heat switch opens, then again a time Δt later. The change in current depends on the integrated power absorbed by the detector. After the second sample, the heat switch closes to reset the device.

compared to the reset period Δt.

Figure 3 shows a schematic view of a prototype bolometer developed at GSFC. It consists of a 1 mm × 1 mm silicon nitride membrane 0.5 μm thick, suspended by four silicon nitride legs for low thermal conductance. A Mo-Au film 60 nm thick on the back side of the membrane provides the superconducting film, with a 25-turn niobium pickup coil on the front side. To ensure accurate alignment, the superconducting film and pickup coil are deposited on opposite sides of the same membrane. Niobium leads for the pickup coil occupy two of the four legs, while the niobium-aluminum heat switch ``sandwich'' occupies a third. A thin bismuth film on the membrane provides impedance matching for greater far-IR absorption. The niobium leads for the heat switch make electrical contact with the bismuth film; a final heater lead on the fourth leg allows a bias current to pass through the bismuth film, providing an additional source of heat for calibration and test purposes. The prototype bolometer is optimized for operation at 300 mK; greater sensitivity can be achieved by lowering the temperature to 50 mK.

3. Sensitivity

The sensitivity of a bolometer is described by the noise equivalent power (NEP), the incident infrared power required to produce the minimum detectable signal (see, e.g., Mather 1984 and references therein). The integrating bolometer has several advantages over resistive bolometers. Since it has no dissipative elements, the Johnson noise term vanishes. The heat switch cycle produces an effective response time Δt much smaller than the intrinsic time constant τ, resulting in reduced phonon noise. To see this, consider the spectral density of random thermal fluctuations ΔT within the absorber,
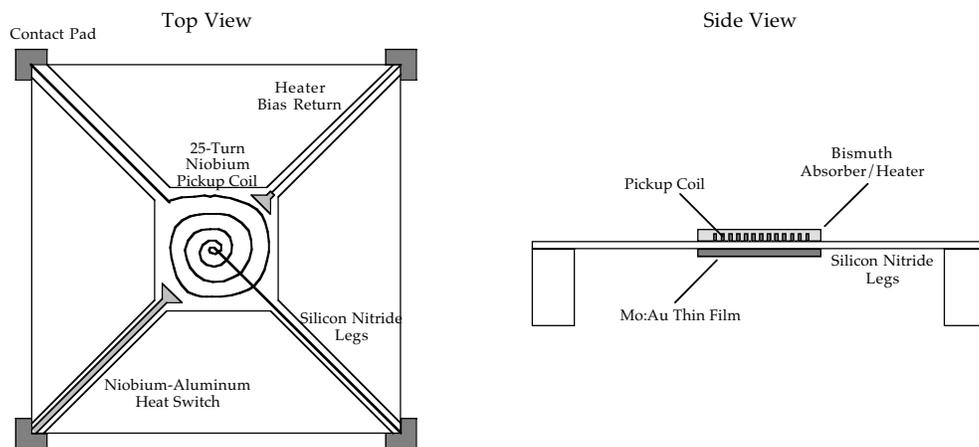
Fig 3: Design for prototype chip built at GSFC. The pickup coil and superconducting film are deposited on opposite sides of a silicon nitride membrane suspended by four legs. The niobium-aluminum heat switch occupies one leg, leads for the pickup coil are on two other legs, while the fourth leg provides a return for the heater current passed through the heat switch leads and bismuth absorber.

$$\tilde{T}(\omega)\tilde{T}*(\omega) = \frac{4\tau k T^2}{C(1+4\pi^2\tau^2\omega^2)},\tag{6}$$

where $\tilde{T}$ is the Fourier transform of the temperature and $*$ represents complex conjugation (Day et al. 1997). The temperature variance is simply the integral of spectral density over frequency,

$$\left\langle (\Delta T)^2 \right\rangle = \frac{4\tau k T^2}{C}\int_0^\infty \frac{d\omega}{1+4\pi^2\tau^2\omega^2}.\tag{7}$$

The heat-switched bolometer cannot respond to fluctuations on time scales longer than the reset period Δt, imposing a low-frequency cutoff on the integral. With this cutoff, the variance becomes

$$\left\langle (\Delta T)^2 \right\rangle = \frac{kT^2}{C}\frac{2}{\pi}\left(\frac{\pi}{2} - \arctan\frac{2\pi\tau}{\Delta t}\right) \approx 0.1\frac{\Delta t}{\tau}\frac{kT^2}{C},\tag{8}$$

or

$$\text{NEP}_{\text{phonon}} = \sqrt{0.1\frac{\Delta t}{\tau}4kT^2 G}.\tag{9}$$

A non-dissipative, heat-switched bolometer is more than an order of magnitude more sensitive than a resistive bolometer. This can be understood intuitively. An unswitched bolometer can have only one statistically independent sample per time interval τ. Activating the heat switch shortens the time constant by a factor $\gamma > 10^4$; hence, the
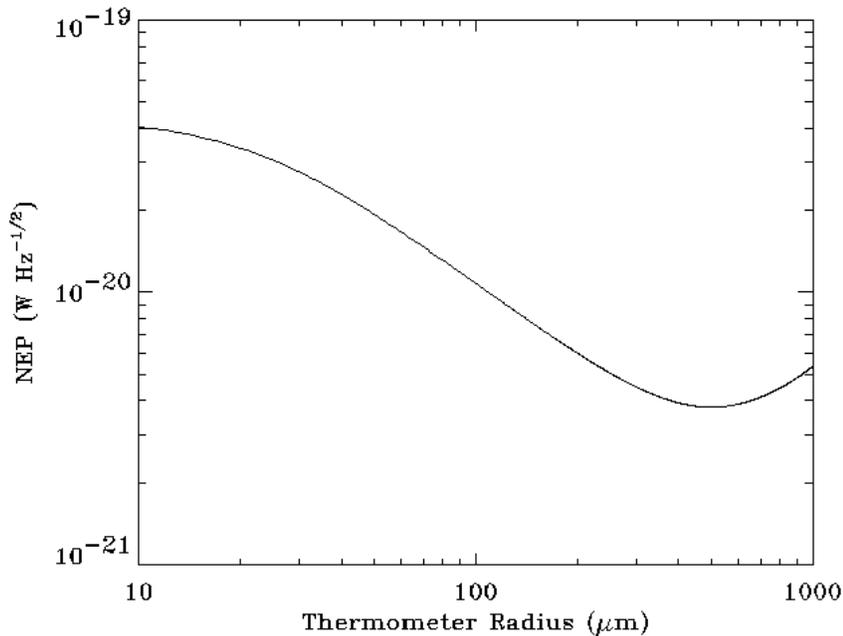
Fig 4: Predicted NEP vs sensor size for ideal integrating bolometer at 50 mK. Phonon noise dominates at small sizes, and SQUID noise for large sizes. A heat switch period $\Delta t = 0.1$ ms is shown.

sample taken after the heat switch is turned off again is statistically independent of previous samples. Using a heat switch allows us to acquire $N = \tau/\Delta t$ more samples in a fixed time interval. Since the variance in the mean of N samples scales as $1/N$, we thus reduce the phonon noise by a factor $\sqrt{\Delta t/\tau}$. An NEP well below 10^{-20} W Hz$^{-1/2}$ appears achievable.

Two-stage SQUIDs have typical flux noise below 10^{-6} Φ_0 Hz$^{-1/2}$ (100 times the quantum limit), with 1/f knee well below the 2 kHz sampling rate of the detector. The amplifier and phonon noise terms scale differently with physical device size. For fixed heat switch period, the phonon noise of an integrating bolometer scales as R^{-2} from the dependence of the cutoff term on the heat capacity C (assuming film thickness independent of radius). The flux resolution of the thermometer scales as R while the heat capacity scales as R^2; hence the amplifier noise $\sim C/\Phi$ scales as R. For a device operated at 50 mK, the NEP reaches a minimum below 10^{-20} W Hz$^{-1/2}$ for sensor size 1 mm to 100 μm (Fig 3), a range easily achieved by photolithographic techniques.

The integrating bolometer has several advantages over resistive devices. The heat switch allows the conductivity to be optimized for sensitivity without sacrificing device speed. The lossless thermometer allows the bolometer to operate as an ideal integrator between resets. The device speed, data rate, and dynamic range depend on the period between heat switch resets and can be changed in real time. *Signal information is derived solely by differencing "snapshot" current measurements, without requiring continuous monitoring throughout the integration period.* A single readout circuit can thus service multiple devices, reducing the complexity, cost, and power consumption of the readout electronics.

References

Day, P., et al. 1997, J. Low Temp. Phys., 107, 359
Mather, J.C. 1984, App. Opt., 23, 584
Mueller, M., et al. 1978, Rev. Sci. Inst., 49, 515

Progress towards high-sensitivity arrays of detectors of sub-mm radiation using superconducting tunnel junctions with integrated radio-frequency single-electron transistors.

T. R. Stevenson[1], W.-T. Hsieh[1,4], M. J. Li[1], D. E. Prober[3], K. W. Rhee[5],
R. J. Schoelkopf[3], C. M. Stahle[1], J. Teufel[3], and E. J. Wollack[2]

[1]NASA Goddard Space Flight Center, Detector Systems Branch, Code 553, Greenbelt, MD 20771
[2]NASA Goddard Space Flight Center, Infrared Astrophysics Branch, Code 685, Greenbelt, MD 20771
[3]Department of Applied Physics, Yale University, P.O. Box 208284, New Haven, CT 06520-8284
[4]Raytheon ITSS, Lanham, MD 20703
[5]Naval Research Laboratory, Electronics Sciences and Technology Division, Washington, DC 20375
(first author's e-mail address: thomass@pop500.gsfc.nasa.gov)

For high resolution imaging and spectroscopy in the FIR and submillimeter, space observatories will demand sensitive, fast, compact, low-power detector arrays with 10^4 pixels and sensitivity $< 10^{-20}$ W/Hz$^{0.5}$. Antenna-coupled superconducting tunnel junctions with integrated rf single-electron transistor readout amplifiers have the potential for achieving this high level of sensitivity, and can take advantage of an rf multiplexing technique. The device consists of an antenna to couple radiation into a small superconducting volume and cause quasiparticle excitations, and a single-electron transistor to measure current through junctions contacting the absorber. We describe optimization of device parameters, and results on fabrication techniques for producing devices with high yield for detector arrays. We also present modeling of expected saturation power levels, antenna coupling, and rf multiplexing schemes.

1. Introduction

To take advantage of very low background photon rates, space-based far infrared or submillimeter-wave interferometers will require large advances in detector sensitivity and speed. Integration of photon-counting detectors with low power readout electronics to make large-format arrays is desired. The Single Quasiparticle Photon Counter (SQPC) is a type of superconducting direct detector which has been proposed (Schoelkopf *et al.* 1999) to meet these requirements. This paper gives an overview of operation principles of the SQPC, the current state of development of this device, and its potential advantages.

2. Detector Concept and Fabrication

The SQPC is an antenna-coupled Superconducting Tunnel Junction (STJ) detector with integrated Radio Frequency Single-Electron Transistor (RF-SET) readout amplifier. (See Fig.1). STJs have been used for energy-resolving detection of single photons at visible to x-ray wavelengths (Peacock *et al.* 1996). In an STJ detector, a superconducting-insulating-superconducting tunnel junction is biased below its superconducting gap. At temperatures well below the superconducting transition temperature, very little current flows since most electrons are bound in Cooper pairs.
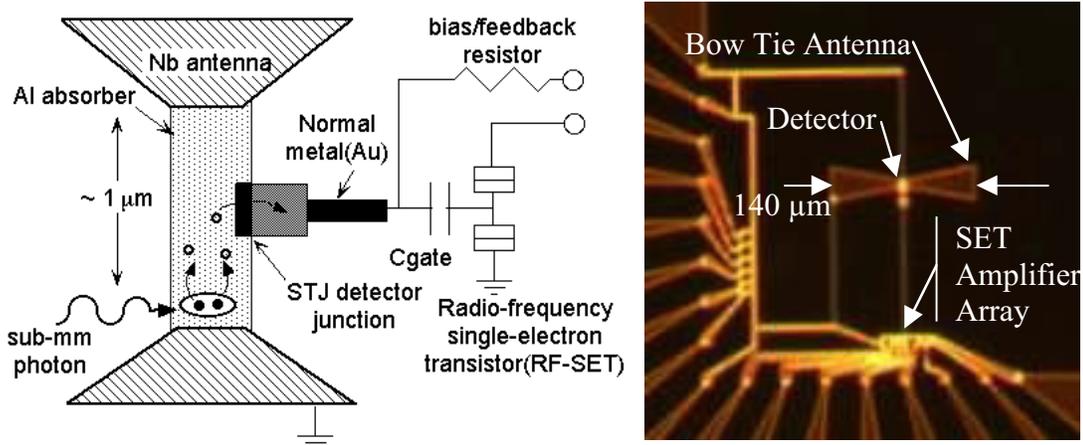
Fig. 1. Single Quasiparticle Photon Counter (SQPC): A niobium bow-tie antenna provides coupling to submillimeter radiation. Absorption of a submillimeter photon breaks Cooper pairs in the aluminum strip joining the halves of the bow-tie, and gives a current pulse through a tunnel junction connected to an RF-SET readout amplifier.

The number of thermally-generated unbound quasiparticles, and hence also the dark current and detector noise, decrease exponentially as the operation temperature is lowered. For aluminum junctions, operating temperatures of $\approx$ 250 mK or less are used.

When a photon is absorbed in one of the superconducting electrodes of the tunnel junction, or in a superconducting absorber film linked to a junction electrode, then much of the photon's energy goes into breaking Cooper pairs, and the quasiparticles released give a current pulse. The integrated charge in the pulse is a measure of the photon energy. At high count rates, overlapping pulses give a photocurrent proportional to the absorbed optical power.

The tunneling time which sets the detector speed can be quite fast ($\approx$ 1 μs) since the response is a non-equilibrium effect, and not limited by long thermal phonon relaxation times at low temperatures.

2.1 Use of an STJ detector for low energy photons

Figure 1 shows how an STJ is adapted for detection of millimeter or submillimeter-wave photons in an SQPC. The optimal volume of the STJ detection electrode scales in proportion to the maximum photon energy (see Sec. 4). For submillimeter detection, dimensions of the absorber and junction need to be submicron. Efficient coupling of long-wavelength radiation to the small absorber is provided by an antenna structure. Sensitive and high-bandwidth readout of photocurrents through the small, high-impedance tunnel junction is provided by an integrated RF-SET, as described below.

In our devices, in Fig. 2, we make the STJ using aluminum films for the electrodes and absorber, and we fabricate the antenna from niobium. For photon frequencies between the superconducting gap frequencies of aluminum (100 GHz) and niobium (700 GHz), the aluminum absorber strip appears to have its normal-state resistance, and can
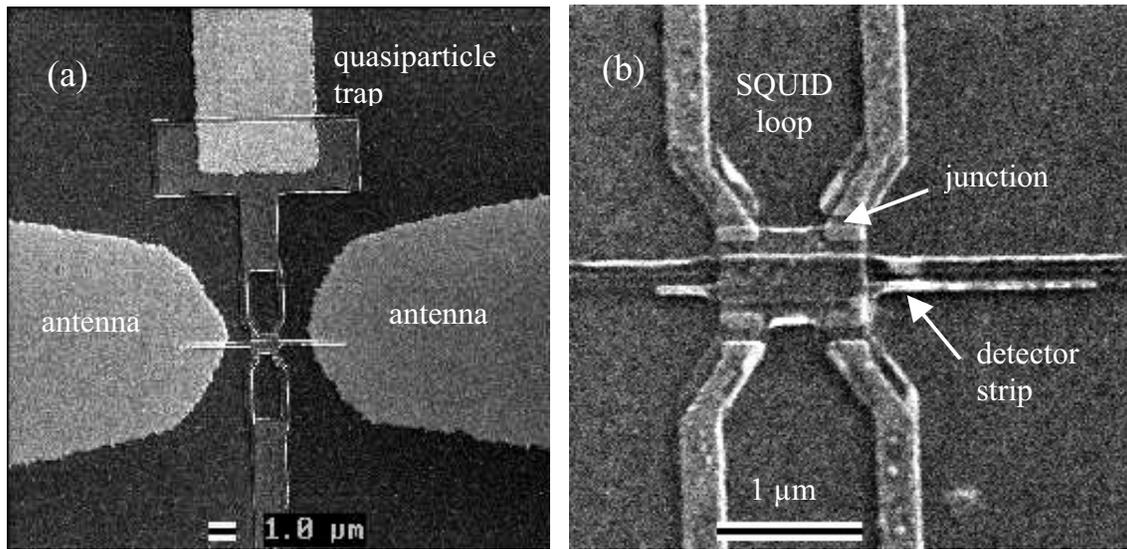
Fig. 2. (a) SQPC detector strip and tunnel junctions are located between two halves of a niobium bow-tie antenna for coupling to submillimeter radiation. A gold quasiparticle trap is included here in the wiring to just one of two dual detector SQUIDs. (b) Close-up view of detector strip and tunnel junctions made by double-angle deposition of aluminum through a resist mask patterned by electron beam lithography. Pairs of junctions form dc SQUIDs, and critical currents can be suppressed with an appropriately tuned external magnetic field.

present a well-matched impedance for absorbing energy at the center of the superconducting niobium bow-tie antenna.

Figures 1 and 2a show an additional feature: a normal metal (gold) section in the bias lead near the detector junction. This acts a quasiparticle trap which aids rapid diffusion of collected quasiparticles away from the junction after tunneling. This prevents the backtunneling effect (Wilson *et al.* 2001) which may otherwise slow the detector response time. Figure 2b shows that instead of one junction we actually use two junctions in parallel to form a dc Superconducting Quantum Interference Device (SQUID). This allows the critical current of the combined junctions to be suppressed nearly to zero, which is necessary for bias stability, and essential for achieving the lowest dark currents (see Sec. 3).

2.2 Integrated Readout Amplifier

At detector readout frequencies (<< 100 GHz), the aluminum is superconducting, and the tunnel junction has a subgap differential resistance of 100 MΩ or higher, and a capacitance of a femtofarad or less. A standard readout circuit for a high impedance photoconductor is the transimpedance amplifier (see Fig. 3). An ideal amplifier for implementing this readout circuit for an SQPC is the RF-SET (Stevenson *et al.* 2001). A Single Electron Transistor (SET) is a very high performance electrometer based on the
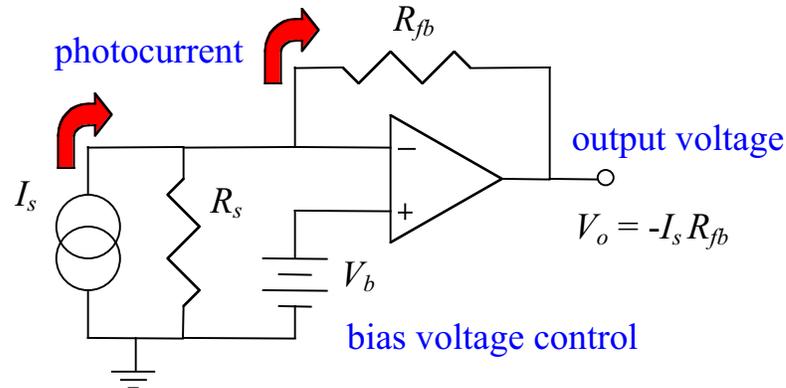
Fig 3. Measuring photocurrent with a transimpedance amplifier: feedback draws photocurrent out through a (cold) feedback resistor, keeping constant bias voltage across the detector.

Coulomb blockade effect (Fulton & Dolan 1987) with sub-femtofarad input capacitance. An RF-SET integrates the SET with a LC circuit resonant at ≈ 1 GHz to impedance match the typical 50 kΩ SET output impedance to a 50 Ω High Electron Mobility Transistor (HEMT) (located at 4K). Signal bandwidths of 100 MHz can be obtained.

To make an RF-SET transimpedance amplifier, we feed the room temperature output voltage of the RF-SET amplifier system back to the input gate of the SET via a cryogenic, high-value (100 MΩ) resistor integrated at the detector bias point (see schematic in Fig. 1). Although we have not yet done so, we intend to use electron beam lithography to fabricate a feedback resistor on-chip with physical dimensions small enough to not limit the readout bandwidth with stray capacitance.

2.3 Fabrication Process

We use optical lithography and thin-film processing techniques to fabricate substrates for SQPCs and SETs. The substrates include the antenna, inductors and capacitors for the rf circuits, and device contacts.

We use electron-beam lithography to fabricate both SETs and the SQPC detector in one process. Figure 4 shows one of our transistors, and a sketch of the self-aligned process used to form the small tunnel junctions required both for SET and detector. We use a standard SET fabrication process using a resist bilayer (Dolan 1977). The bottom resist layer is more sensitive to electron-beam exposure than the top, high-resolution layer. Consequently, development of the resists results in undercuts which can be made to merge and form free-standing bridges of the top resist. Evaporating aluminum films onto the substrate at two angles allows junctions to be formed under the resist bridge, as shown in Fig. 4a. A room temperature thermal oxidation step between deposition of the two aluminum layers forms an Al_2O_3 tunnel barrier in the junction area defined by the overlap of the layers.

We have refined this fabrication process with the goal of achieving the reproducibility needed for large-format arrays. By monitoring and controlling the sensitivity of the lower resist layer, we have recently attained large improvements in device yield. We now make chips with 20 or more functional SETs, and small (5 element) arrays of SETs
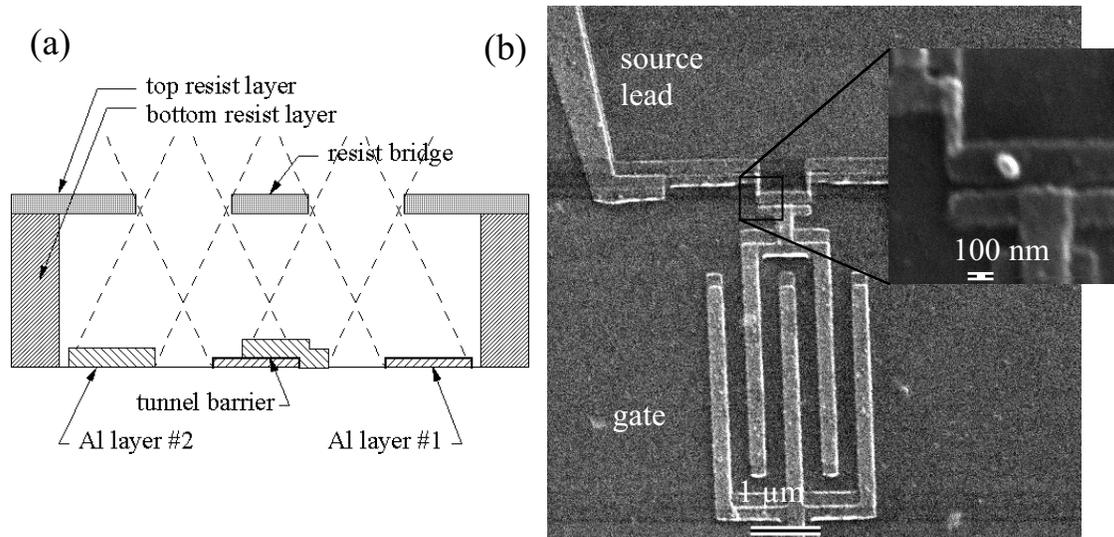
Fig. 4. (a) Double-angle deposition process used to form the self-aligned junctions for the SQPC and SET. (b) An SET with 0.5 fF input gate capacitor. Source and drain leads are connected to island via ultra-small tunnel junctions. Inset shows close up of a 60 nm x 60 nm junction.

with junction resistances clustered within 10% of a target value. We are continuing to study and make improvements in device yield, uniformity, and long term stability.

3. Optimization of Device Parameters and Experimental Results

Apart from efficiency of antenna coupling, the fundamental factors determining SQPC sensitivity are: (i) detector responsivity, (ii) shot noise on the dark current, (iii) noise of the RF-SET expressed as an equivalent voltage noise at its input, (iv) Johnson noise in the feedback resistor, and (v) impedance of the detector in parallel with the feedback resistor. We have investigated each of these issues, and predict Noise Equivalent Power (NEP) $\approx 1 \times 10^{-19}$ W/√Hz could be obtained with our existing SQPC prototypes based on the demonstrated levels of performance measured for each factor.

3.1 Detector Responsivity

The responsivity of the SQPC is ideally equal to e/Δ, where Δ is the energy gap of the absorber material and e is the electronic charge. For aluminum, $e/\Delta \approx 5000$ A/W. Efficient collection of photon-generated quasiparticles as a tunneling current depends on making the tunneling time short by confining the quasiparticles to a small absorber or trap volume, and on avoiding sources of quasiparticle recombination other than recombination with the thermal equilibrium concentration of quasiparticles. At low operating temperatures, thermal recombination rates are orders of magnitude slower than our 1-10 µs tunneling times. Using the dual detector SQUID structure shown in Fig. 2b, we have used one SQUID in a pair to electronically inject a quasiparticle current into the common absorber strip. We see a strong response in the subgap tunneling current of the

second SQUID. While we have yet to complete analysis, we believe the observed response indicates good confinement and collection in our devices.

3.2 Dark Currents

We have investigated dependence of dark current on device parameters. We found initial prototype detectors with ≈ 1 kΩ junctions deviated from BCS predictions for thermally-generated dark current below ≈ 250 mK. The temperature-independent dark current was proportional to the square of the (magnetic-field dependent) critical current of the SQUID, and was explainable as "rectification" of the Josephson oscillations occurring at a non-zero dc bias voltage (Holst 1994). Since the minimum critical current of a dc SQUID is limited to non-zero values by self-inductance, and by asymmetry between junctions, we could not fully suppress the critical current with a magnetic field as is done for UV/optical or x-ray STJ detectors.

Instead, we were motivated to try junctions with smaller areas and higher resistance-area products. For junctions with a Josephson energy $E_J = I_c \Phi_0 / 2\pi$ smaller than kT or the charging energy $E_c = e^2 / 2 C_J$, we observed extra suppression of the critical current by thermal or quantum fluctuations. The dark current at bias less than Δ/e was then greatly improved, and we measured BCS-like dark currents down to at least 250 mK with values of 2 pA or 0.5 pA for normal-state resistances of 10 kΩ or 40 kΩ respectively. Shot noise on 0.5 pA is 0.4 fA/$\sqrt{\text{Hz}}$, which corresponds to a limiting NEP $= 8 \times 10^{-20}$ W/$\sqrt{\text{Hz}}$.

3.3 RF-SET voltage noise

Nearly quantum-limited charge noise $\approx 10^{-6}$ e/$\sqrt{\text{Hz}}$ has been demonstrated for RF-SETs with small input gates (Aassime 2001). However, just as for dc SQUIDs, it is difficult to maintain quantum-limited sensitivity while providing strong coupling to an input signal. For SQPC readout, the figure of merit is the voltage noise, equal to the charge noise divided by the input gate coupling capacitance. As gate capacitance is increased, the voltage noise at first drops, but then levels off or increases as the charge noise of the RF-SET starts to degrade due to co-tunneling effects (Stevenson 2001). We have obtained record SET voltage noise of 30 nV/$\sqrt{\text{Hz}}$ with a 0.5 fF gate, and have been able to maintain a voltage noise close to this value during closed-loop transimpedance amplifier operation (Segall 2002).

3.4 Predicted Detector Sensitivity

Table I summarizes the noise budget for an SQPC detector using demonstrated parameters. We assume our 0.5 pA device is used with a 100 MΩ integrated feedback

Table I. Detector sensitivity for demonstrated parameters.

Noise source	Parameters	Effective current noise
Dark current	0.5 pA at 250 mK	0.4 fA/$\sqrt{\text{Hz}}$
RF-SET noise	30 nV/$\sqrt{\text{Hz}}$	0.3
Johnson noise in Rfb	100 MΩ at 250 mK	0.4
Total current noise		0.6
NEP		1.2×10^{-19} W/$\sqrt{\text{Hz}}$
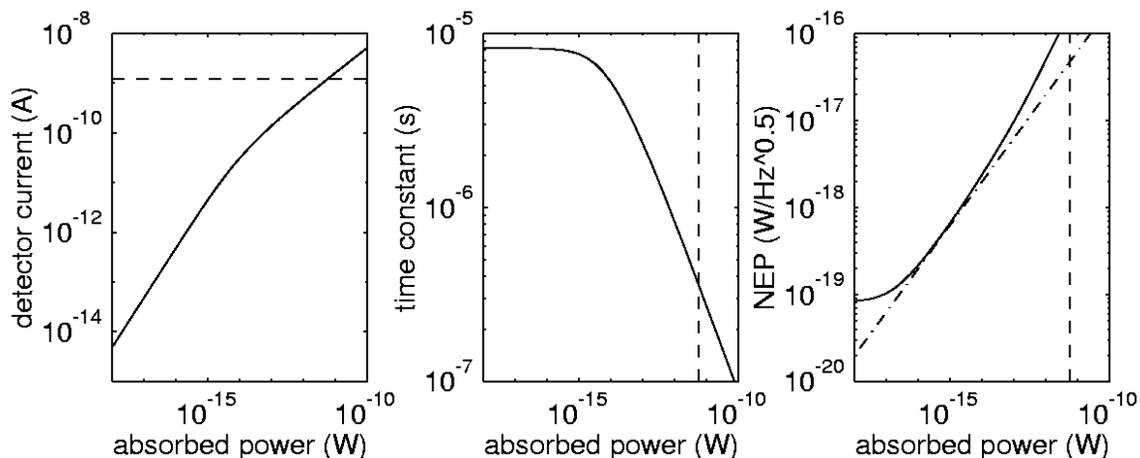
Fig. 5. Simulated saturation effects for 300 GHz photons absorbed in an SQPC with 40 kΩ normal resistance, 0.023 (μm)^3 absorber volume. Plots show power dependence of photocurrent, time constant, and NEP. Dash-dot line shows limiting NEP from just photon shot noise. Dashed lines estimate current or power levels at which device will be driven normal.

resistor at 250 mK, with higher subgap impedance for the biased detector. The expected NEP is 1.2×10^{-19} W/√Hz. Even better sensitivity is possible at lower temperature with a higher value feedback resistor.

4. Modeling of Detector Performance for Various Absorbed Power Levels

The maximum power level the SQPC can tolerate, and the expected background count rates in a space environment, will determine the maximum fractional bandwidth allowed for the incident radiation. As the absorbed power is increased, the peak or steady-state concentration of quasiparticles in an SQPC absorber strip will grow, and the rate of self-recombination of particles will increase. When the self-recombination rate per quasiparticle equals or exceeds the tunneling rate, then the efficiency of charge collection drops. We have performed Monte Carlo simulations of the effects on signal and noise in this case, and have found that the saturation of the detector is really quite gradual and weak.

Figure 5 shows preliminary results calculated for a device like our prototype with 0.5 pA dark current. The self-recombination and tunnel times become equal at an absorbed power of 3 fW. Above that power, the photocurrent grows as square-root of power, and the time constant of the response drops with the self-recombination time. The total NEP degrades very slowly, and remains close to the photon shot-noise limit until the power reaches 0.1 pW. With the expected background for space observations (Mather *et al.* 1998), this SQPC model allows background limited sensitivity for incident bandwidths of 0.1% to 100%. Only at powers above a pW will the device saturate "hard," when quasiparticle concentrations grow large enough to suppress the superconducting gap and drive the aluminum normal. We consider weak saturation one advantage of the SQPC.

5. Design of Antenna Coupling and Calibration

We have an experiment currently underway to attempt to measure for the first time the photoresponse of an SQPC to radiation. For this purpose, we have made a calibrated 200 GHz source with calculable coupling to the SQPC antenna.

5.1 STJ/Antenna Coupling

At this stage of sensor development, we employ a bow-tie antenna on a dielectric substrate to optically couple the device. This simple scheme has minimal impact on processing and will allow basic characterization of the device properties. In order to integrate the detector to the antenna, the impedance, field configuration, and interaction of the bias network topology with the radiating structure need to be considered.

We estimate the bow-tie antenna's input impedance by treating the structure as a radial transmission line in the quasi-static limit (Rutledge & Muha 1982). For the geometry used for the prototype devices, we estimate an input impedance of ~130 Ω. The rf impedance presented by a typical SQPC-SET sensor is 20~50 Ω. This impedance mismatch will be addressed in subsequent sensor iterations.

For an antenna on a substrate, the power radiated into the dielectric is greater than the power radiated into the vacuum by a factor of $e_r^{-1.5}$. For the silicon substrate in use, we note that ~97.5% of the power resides in the dielectric portion of the half-space. A quarter-wave-matching layer is employed between the dielectric and freespace to improve the coupling efficiency for the component of the radiation propagating away from the antenna into the dielectric. A substrate lens can also be employed to limit conversion of trapped rays into surface waves on an electrically thick substrate.

The beam efficiency, the ratio of the main beam-to-total beam solid angle, for a bow tie is relatively low. For this reason, we plan to transition to antenna structures with improved sidelobe control. This will enhance compatibility of the device development effort with needs of precision radiometric applications. Given the sensor's planar topology, processing requirements and RF impedance levels; dual slot and taper slotline configurations are presently under consideration to meet these demands.
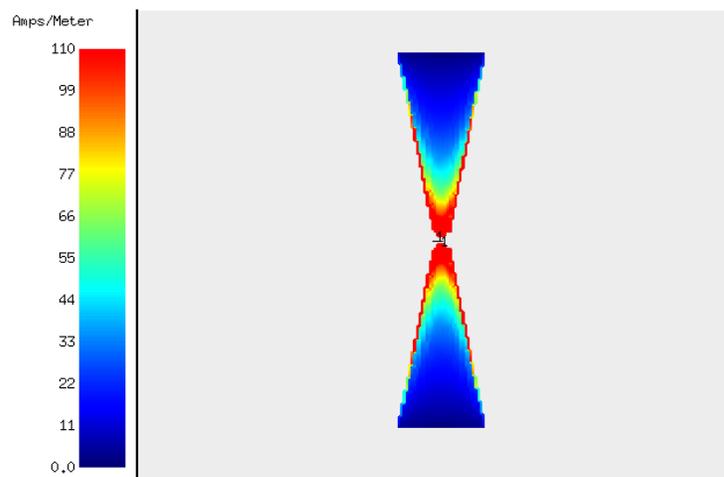
Fig. 6. Simulation of bow-tie antenna high-frequency current distribution.

5.2 200 GHz Calibration Source Development

A quasi-optical calibration scheme was chosen to allow maximum flexibility during sensor characterization. For the initial measurements, a quasi-optically coupled "reverse bolometer" will be used as a thermal source in the sensor's field of view to produce a calculable radiometric flux. The absorber physical temperature and emissivity as a function of frequency effectively control the source bandwidth.

The calibrator was fabricated as follows: The absorber was formed by evaporation of Pd/Au to realize ~400 Ω/   on an electrically thin silicon substrate at the anticipated operational temperature. This layer is held in a silicon frame at a spacing of ~410 microns, roughly a quarter wavelength, from a smooth oxygen free copper backshort. The emitter is supported on micro-machined conductive legs, which provide electrical connections for the thermometry and thermal isolation. Provisions are provided in the design to reconfigure the calibrator for use as a mount in WR05 waveguide for verification of the coupling efficiency via standard waveguide metrology techniques.

6. Multiplexing for Arrays

An advantage we see for the SQPC detector system is the multiplexing capability of RF-SETs. Multiplexing schemes will be crucial to the development of large-format arrays of SQPCs, or other low-temperature detectors.

RF-SETs have a natural wavelength division multiplexing capability, as shown in Fig. 7a. Each RF-SET is connected to one coaxial line by an rf tank circuit with a unique resonance frequency. A directional coupler allows rf carriers to be applied at each
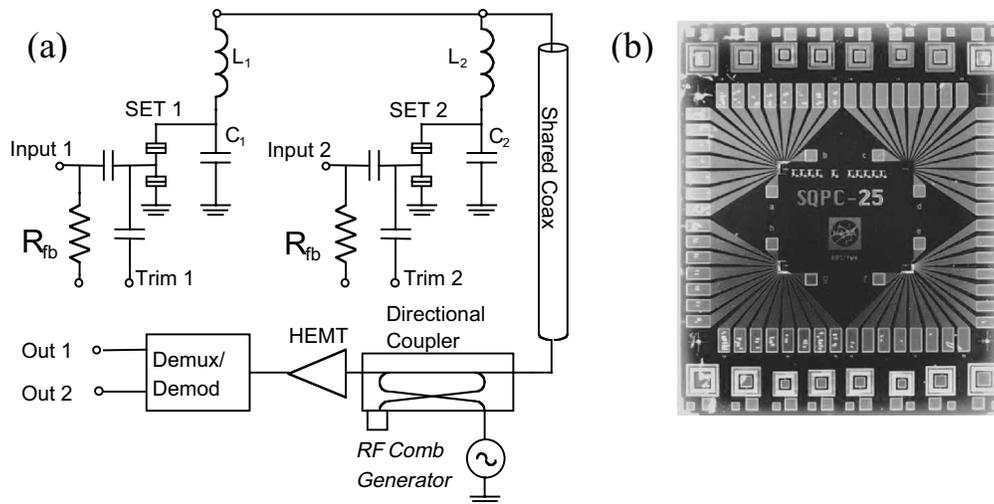
Fig. 7. (a) WDM multiplexing scheme. All inductors are connected to one coax running from subkelvin chip to components at 4K. Multiple carrier frequencies pass through directional coupler to excite tank circuits. Reflected power is directed to HEMT amplifier. Outputs are demultiplexed and demodulated at room temperature, and fedback to resistors for transimpedance operation. Detector bias can be changed using trim gates. (b) 8 mm x 10 mm chip with 16 lithographically defined tank circuit inductors.

resonance frequency and reflected powers to be monitored by a single HEMT following amplifier at 4 K. RF-SETs can be individually or simultaneously powered and read out.

We have made a two-channel rf multiplexing demonstration using discrete inductors for the rf tank circuits (Stevenson *et al.* 2001). The two input signals were successfully reconstructed with little cross-talk. Lithographic versions of the rf circuits (Fig. 7b) had measured parameters in agreement with electromagnetic modeling, with reduced cross capacitance and inductance. We have designed a 50-channel system with components based on measured parameters of our lithographic circuits (Stevenson *et al.* 2002).

The wavelength division multiplexing scheme still requires two wires for each pixel which is being read out simultaneously: one wire for connection to the transimpedance feedback resistor, and one for a weakly coupled SET trim gate. Combining wavelength division multiplexing with some form of time-division, or other, multiplexing will be required in to implement large arrays; however, the 50-fold reduction in output connections seems quite valuable.

7. Conclusion

We have refined a fabrication process for SQPC detectors and RF-SET amplifiers with the goal of achieving the reproducibility needed for large-format arrays, and have recently attained large increases in device yield. With proper device design, we have attained subpicoamp dark currents, and record RF-SET voltage noise levels, which allow sensitivities of 1×10^{-19} W/$\sqrt{\text{Hz}}$ at temperatures as high as 250 mK. There is potential for even better sensitivity at lower temperatures, and the expected detector saturation behavior and multiplexing capability of RF-SETs enhance the potential for application of the SQPC in space-based submillimeter-wave interferometers.

8. Acknowledgments

9. References

Aassime A, Gunnarsson D, Bladh K, Delsing P, and Schoelkopf R, 2001, Appl. Phys. Lett. 79, 4031.

Dolan GJ, 1977, Appl. Phys. Lett. 31, 337.

Fulton TA and Dolan GJ, 1987, Phys. Rev. Lett. 59, 109.

Holst T, Esteve D, Urbina C, and Devoret MH, 1994, Physica B 203, 397.

Mather JC, Moseley SJ, Leisawitz D, *et al.*, 1998, astro-ph/9812454.

Peacock A, Verhoeve P, Rando N, van Dordrecht A, Taylor BG, Erd C, Perryman MAC, Venn R, Howlett J, Goldie DJ, Lumley J, and Wallis M, 1996, Nature 381 135.

Rutledge DB and Muha MS, 1982, IEEE Trans. Antennas Propag., AP-30 535.

Schoelkopf RJ, Moseley SH, Stahle CM, Wahlgren P, and Delsing P, 1999, IEEE Trans. Appl. Supercond. 9 2935.

Segall K, Lehnert KW, Stevenson TR, and Schoelkopf RJ, 2002, in preparation.

Stevenson TR, Aassime A, Delsing P, Schoelkopf S, Segall K, and Stahle CM, 2001, IEEE Trans. Appl. Supercond.11, 692.

Stevenson TR, Pellerano, FA, Stahle CM, Aidala K, and Schoelkopf RJ, 2002, Appl. Phys. Lett., accepted.

Wilson CM, Frunzio L, Segall K, Li L, Prober DE, Schiminovich D, Mazin B, Martin C, and Vasquez R, 2001, Trans. Appl. Supercond. 11, 645.

Heterodyne Array Receiver development at KOSMA

S.Heyminck, U.U.Graf, E.A.Michael, S.Stanko, D.Rabanus, C.E.Honingh,
K.Jacobs, R.Schieder and J.Stutzki

KOSMA, I. Physikalisches Institut der Universität zu Köln,
Zülpicher Straße 77, 50937 Köln, Germany

Abstract

We present an overview of the array receiver development in the KOSMA receiver group. The central topic of the presentation is the 16 element 490/810 GHz array SMART (SubMillimeter Array Receiver for Two frequencies), which has been installed at the KOSMA telescope on Gornergrat near Zermatt/Switzerland in September of 2001. The opto–mechanical design of this receiver makes intensive use of our CNC machining capabilities, both to produce special optical components like imaging phase gratings for LO multiplexing, and also to manufacture relatively large receiver sub–assemblies as monolithic integrated optics blocks. We use our standard waveguide SIS mixers as detectors and the KOSMA Array–AOS as backends.

We give a short presentation of the receiver design and show results obtained with the instrument during its first astronomical observing season.

Based on the good experience with SMART, we are using a similar approach for the design of STAR (SOFIA Terahertz Array Receiver), a 16 element 1.9THz hot electron bolometer array, which we are currently developing.

Keywords: submillimeter array receiver, integrated optics, phase grating, Terahertz array receiver, SOFIA

1. INTRODUCTION

Developing an astronomical submillimeter heterodyne array receiver is quite different from building a single pixel receiver. But it is the only way of increasing the data rate of a single dish telescope since the mixer noise limit in the frequency range up to 1THz has reached nearly the quantum limit. Over the last couple of years several groups developed array receivers: examples are the CHAMP project of the MPIfR[1] or the Pole Star instrument on AST/RO.[2] The KOSMA

Further author information: E-mail: lastname@ph1.uni-koeln.de

submillimeter receiver group has been focused on this approach over the last 3 years.

Our first step in developing array receivers is the SMART instrument. It has been installed September 2001 on the KOSMA telescope and is operational since then. Its design includes new features like Fourier–gratings for the local oscillator supply, integrated optics, compact bias– and IF–electronic, and automatic mixer tuning. Therefore this instrument can be seen not only as a new astronomical receiver with high performance, but also as a proof of technology for our further receiver development. Many of these developments were triggered by the need to keep the instrument as compact as possible to match the size and weight limitation on our relatively small telescope.These constrains will be shared by our current development project, a 1.9THz 16 pixel heterodyne array receiver for the SOFIA telescope where the available space is similar to the situation at KOSMA.

2. GENERAL FEATURES OF SMART
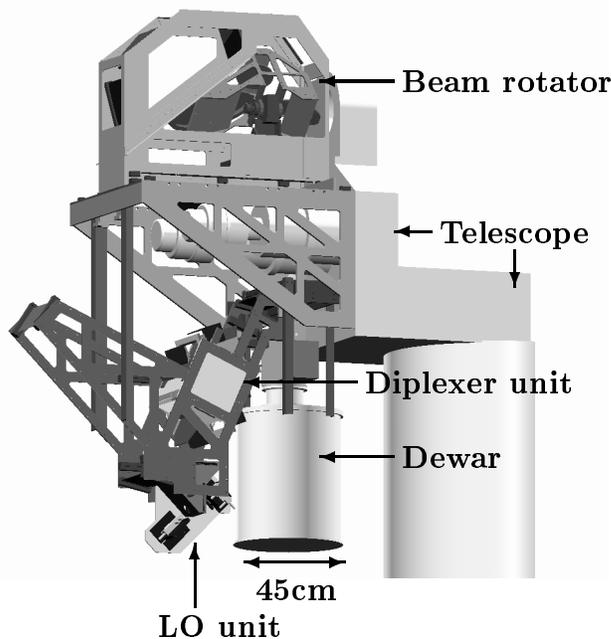
Figure 1. SMART, receiver geometry.[3]

SMART consists of polarization split subarrays of 2×4 mixers each. One of the mixer subarrays operates at 490GHz and the other one at 810GHz. We can thus observe eight positions, each in two frequencies simultaneously, which results in 16 heterodyne channels in total. In the moment only one of the two rows, with 4 pixels at each frequency, is equipped with mixers. The second half will be equipped in fall 2002, so that the receiver will be in full operation in winter 2002/2003.

The instrument is mounted at the left hand Nasmyth port of the KOSMA telescope.[4] The first optical element behind the elevation tube is a K–mirror type image rotator followed by two Gaussian telescopes. Between the two Gaussian telescopes the diplexer unit is located. It consists of two independent Martin–Puplett interferometers, one for each frequency band. The incoming signals are divided by a polarizer grid to pass the two interferometers and are recombined again by a second polarizer grid. The

focal plane of the telescope is reimaged by the first Gaussian telescope to the center of the interferometers, in order to minimize overlap losses.

The second Gaussian telescope reimages the diplexer focal plane to the mixer unit in the dewar. To keep the dewar window as small as possible and to minimize the heat load in the dewar, it is located between the two mirrors of the second Gaussian telescope at the location of an image of the telescope aperture, where the beam cross section is minimal.

The local oscillator (LO) power is multiplexed by a newly developed collimating Fourier grating, which is a phase grating with a smooth surface structure superimposed on a parabolic mirror to match the wave front curvature of the divergent LO–beam. This technique results in a very efficient and compact design of the LO unit. The LO–signals are combined with the sky signal in the diplexer unit.

The mixer assembly inside the dewar consists of a polarization grid to distribute the signal to the two subarrays. Each subarray is made up of a facetted mirror and two rows of four mixers, each. The individual parabolic mirror facet in front of each mixer is needed to collimate the divergent beam from the mixer before entering the common optics. This is required in order to pack the beams more closely in the array optics and on the sky. We achieve a beam spacing of $3.25\times$ waist size which corresponds to approx. $2.2\times$FWHM beams spacing on the astronomically source.

The facetted mirror with all of its reference and mounting surfaces was produced on our CNC–milling machine as an integrated optics unit in a single machining cycle to get the best possible accuracy. These two units, together with the mount for the polarizer grid and the first common mirror, are sandwiched between two precision machined plates. The machining accuracy allows to avoid any further alignment. A similar approach has been taken throughout all parts

Figure 2. Schematics of the cold optics inside the dewar with the two facet mirror assembles, the polarizer grid and the first common mirror. These elements are sandwiched between two plates (only one shown).

of the receiver optics. Making maximum use of our CNC machining capability, it has been possible to manufacture all assemblies with very high accuracy.

Due to this design the alignment effort for the whole array could be kept nearly the same as that of a single pixel receiver.

The main electronics system consists of three subsystems, a magnet supply to suppress the Josephson current in the mixer junction, a mixer bias box and an IF–processor. Each of these components is designed as a stand alone unit. All tree are connected by an address bus to lock them to the same receiver channel. They all can be remotely controlled by a personal computer. Additionally, two synthesizers for the LO–phase lock and a computer interface box are needed. The electronics rack with the components for the first 8 channels is shown in figure 3. A second rack with only the magnet supply, the bias box and the IF processor box will be installed together with the next 8 mixers.

Trough the interface box, the computer can monitor the temperature and the pressure inside the dewar, controls the image rotator, controls the phase lock loop circuits and sets the frequencies of the LO reference synthesizers. The computer runs under the LINUX operating system, which combines high stability and the possibility of remote operation. Since all electronics boxes are build as stand alone units, manual or partly manual operation of the system is also possible.

Figure 3. Photograph of the SMART electronics rack with the magnet supply (beneath the oscilloscope), bias box, IF processor, the two synthesizers, computer interface box and the control computer on the bottom.

To get the possibility of fast frequency change and easy tuning of the system we implemented a computer controlled auto–tune mode which sets the magnetic field and the optimum bias point for each of the mixers. The main problem was the setting of the magnetic field, since the field required is not always reproducible due to external and statistical effects. It is thus impossible to get the correct magnetic field by using a look up table. The tuning algorithm has to find the optimum magnetic field by measuring and interpreting the conversion curve or the IV–curve of each SIS mixing device.

Figure 4. Focal plane beam map measured simultaneously with 4 mixers at 490GHz (thin lines) and 4 mixers at 810GHz (thick lines). The contour levels are from 5% to 95% of the peak intensity of each beam.
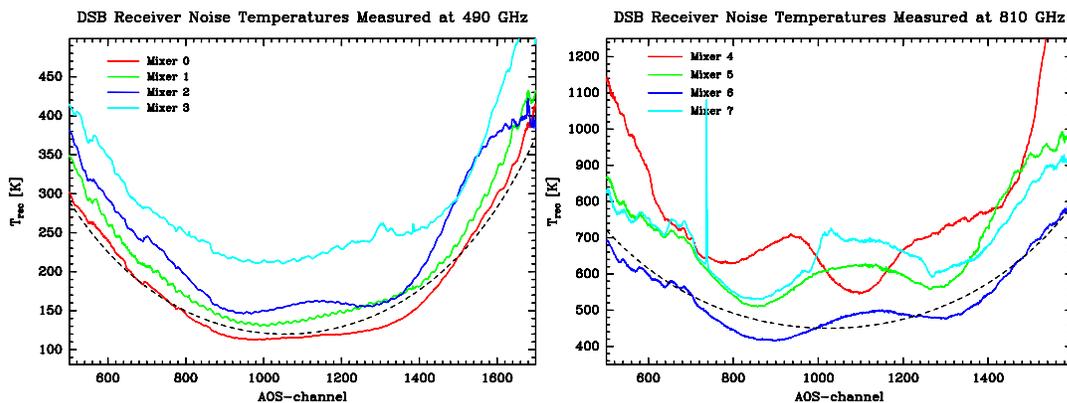
Figure 5. Measurement of the receiver noise temperature of the 490 GHz channels (left hand side) and the 810 GHz channels (right hand side). The IF center frequency in both measurements is 1.5 GHz (AOS channel 1100) with 1 GHz nominal bandwidth (approx. 1000 AOS channels). The dashed line indicates the noise temperature function caused by the diplexer transmission with a 120K(490GHz) / 450K(810GHz) receiver.

As mixers we use the in house developed and procured SIS (Superconductor Isolator Superconductor) mixers.[5] The backends are our in house developed 4 channel Array AOSs which combine high spectral resolution, high stability, compact design and low power consumption.

3. PERFORMANCE OF SMART

To measure the focal plane beam pattern (figure 4), we used a chopped cold load mounted on a XY–translation stage to scan the intensity distribution of the receiver. The measurements show that the optical setup performs as designed. The beams are round and they all have their nominal size to within 5%. The

490GHz beams and the 810GHz beams match well in position, and the spacing of the beams also agrees with the design value. To reach this result no individual mixer alignment was necessary, which demonstrates the power of the integrated optics approach in designing the instrument.

The noise temperature measurements (figure 5) where made at the telescope on the actual system used for astronomical observations, with two 4 channel array acusto optical spectrometers as backend. Since we have no sideband filter in the array all temperatures are double sideband temperatures. Over the central part of the IF–band we reach DSB noise temperatures of 150K-250K for the 490GHz and 500K-700K for the 810GHz frequency band.

The shape of the curves in figure 5 — in particular at 490GHz — is dominated by the transmission function of the Martin–Puplett–Diplexer. To reduce this effect in the 490GHz branch, we plan to increase the IF center frequency from 1.5GHz to 4GHz, which is possible with the current diplexer design. This would result in a flatter diplexer transmission curve over the 1GHz band and therefore in a better over all noise temperature. At 810GHz, we will stay at an IF center frequency of 1.5GHz because it enables simultaneous measurements of two important lines in opposite sidebands: the CO $J = 7 \to 6$ and the [CI] $^3P_2 \to {}^3P_1$ line.

4. FIRST ASTRONOMICAL RESULTS

The main astrophysical motivation for building this dual frequency receiver is to map the line intensity ratio of the [CI] $^3P_1 \to {}^3P_0$ and the [CI] $^3P_2 \to {}^3P_1$ emission with highest possible accuracy. Measuring both lines simultaneously eliminates a number of errors by the atmosphere and the calibration.[6]

As an example of one of the first measurements made with the new instrument in December 2001, we present 3 simultaneously observed maps of W3 IRS5. We mapped 176 positions with a total integration time of 160 seconds per position with half of the time spending on the on– and the other half on the off–position. This resulted in a total observing time of approximately 3 hours including all overheads from telescope movement and calibration. For the [CI] $^3P_1 \to {}^3P_0$ emission the atmospheric τ was between 0.7 and 1.1 and for the [CI] $^3P_2 \to {}^3P_1$ emission between 1.0 and 1.4. The baseline RMS of the received spectra at each position are between 0.1K and 0.4K for 490GHz and between 0.8K and 1.3K for the 810GHz band. The CO $J = 7 \to 6$ Line and the [CI] $^3P_2 \to {}^3P_1$ lines are measured simultaneously, each in one of the two sidebands of the 810GHz channels, while the CO $J = 7 \to 6$ line was detected in the 490GHz channels.
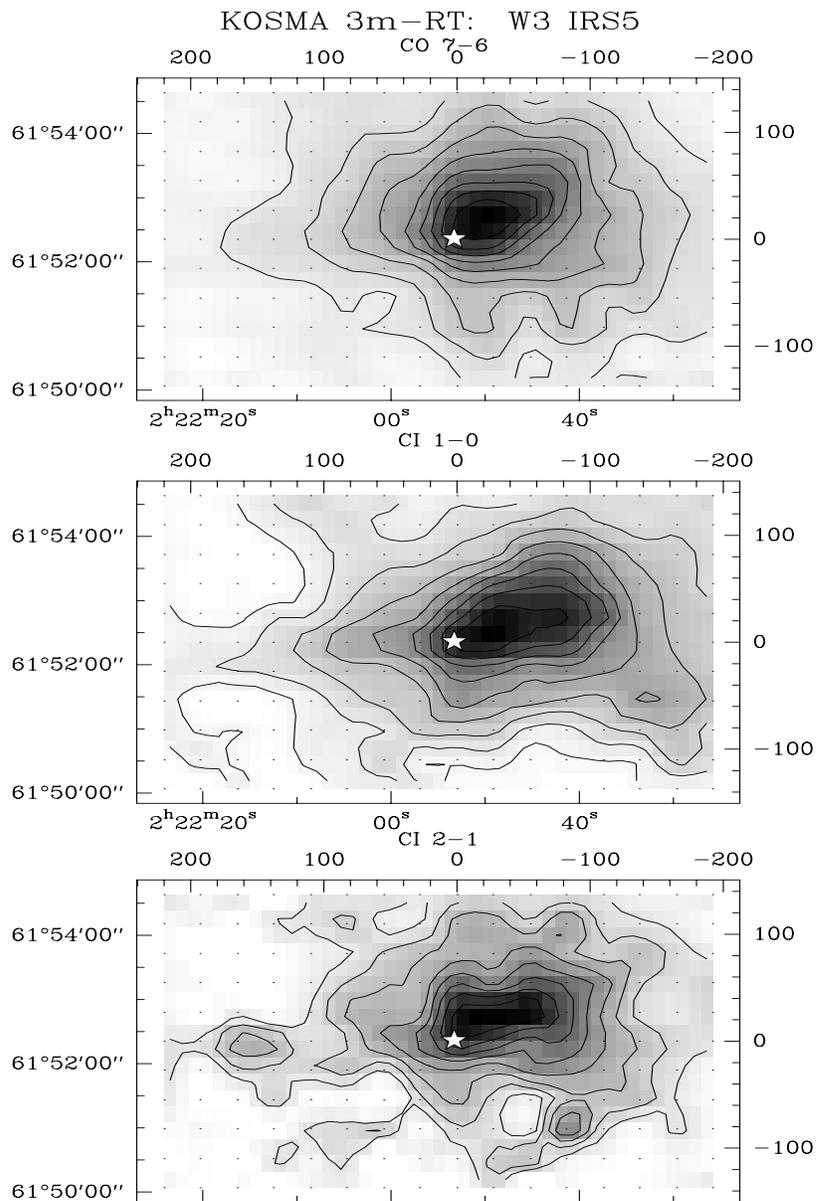
Figure 6. Example for a map done with SMART on W3 IR5S in December 2001. All three maps are received simultaneously.

5. FUTURE PROJECTS

The experience gained in designing, building and operating SMART is beneficial for the development of further array receivers. This holds especially for the concept of our electronics with its full redundancy and the fabrication of integrated optics with regard to STAR (SOFIA Terahertz Array Receiver). STAR will be a 4×4 pixel 1.9THz heterodyne array consisting of two interleaved 8 pixel

subarrays split by polarization.

Since Nb–SIS mixers will not work at 1.9THz, we are developing diffusion cooled hot electron bolometers (HEBs)[7] which we plan to use as detector device. One of the recent results in this development is a bandwidth measurement made with a 800GHz HEB device shown in figure 7, which indicates a usable bandwidth of approximately 4.7GHz.
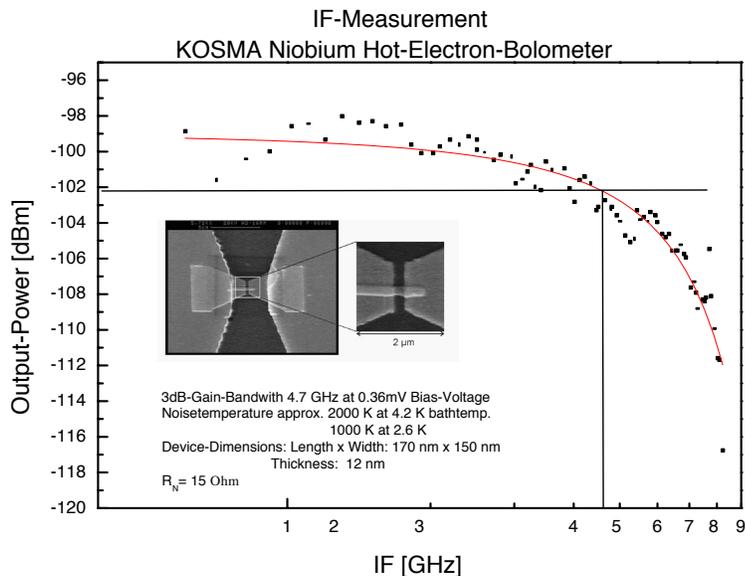
Figure 7. Bandwidth measurements of the KOSMA 800GHz Niobium HEB (by J. Stodolka).

The concept of the optical setup for the receiver is shown in figure 8. The SOFIA telescope can be defocused far enough that we do not need any additional imaging optics. This means we will have a K–mirror type image rotator directly followed by the diplexer unit consisting again of two independent Martin–Puplett interferometers. The first and only focusing element will be the individual parabolic mirror in front of each mixer, arranged again on a facetted mirror. This will give us the minimum number of optical components, and therefore the minimum optical losses, possible.

A schematic view of the facetted mirror and the mixer arrangement is shown in figure 9. In principle it is pretty much the same as in the SMART instrument used. In the final design we will again use one of these mixer arrangements per polarization. The design approach for the whole unit will be similar to the one we used in the SMART instrument (fig. 2).

For LO multiplexing we again use our collimating Fourier gratings.[8] First measurements for the beam pattern needed have been done at a frequency of 490GHz to test the design of the unit cell structure (fig. 10).

For the use as spectrometer backends, the development of small and stable wide band acusto optical spectrometers[9] is continuing. Here the newest design, the demonstration model for the HIFI instrument on Herschel in a new compact setup, is shown in figure 11.
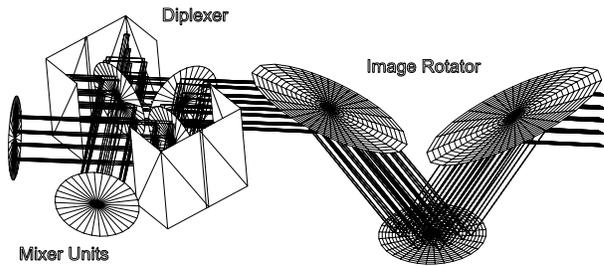
Figure 8. Concept for the optical arrangement of the STAR receiver. 4×4 pixels arranged in two interleaved subarrays to get closer beam spacing on the astronomical source.
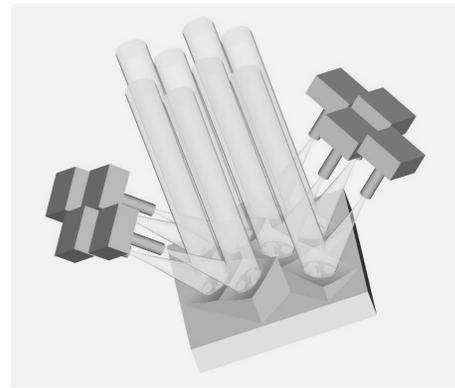
Figure 9. Mixer arrangement of one of the two STAR subarrays and the arrangement of the collimating mirrors in front of each mixer.
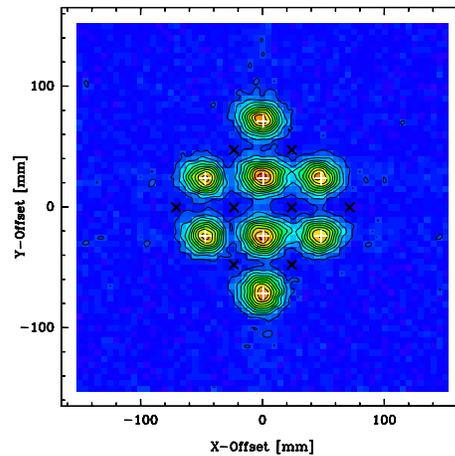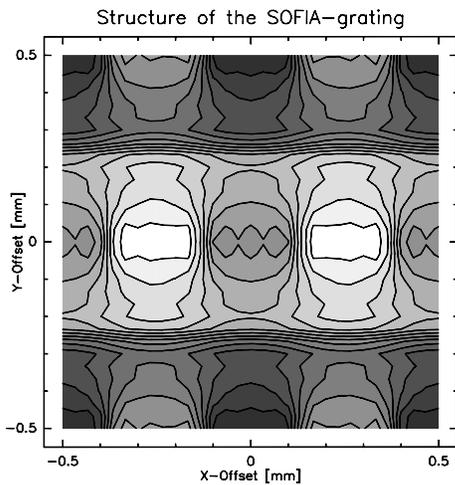
Figure 10. Surface topology of the unit cell (left hand side) producing the diffraction pattern shown in the right hand measurement. Contour levels are from -1.5 to 1.5 radians in steps of 0.25 radians. Measured intensity distribution (right hand side) in a 490GHz beam diffracted by a flat Fourier grating. The given pattern is one of the two subarrays, which will be operated at orthogonal polarizations to minimize the spacing between the pixels. The beam positions of the other subarray are marked by black crosses.

6. CONCLUSIONS

We built the first dual frequency array receiver and demonstrated its high performance in terms of sensitivity and usability. This shows the potential of our design concept which we will also use and extend for further receiver developments. For higher frequencies, the concept of integrated optics should work even better be-
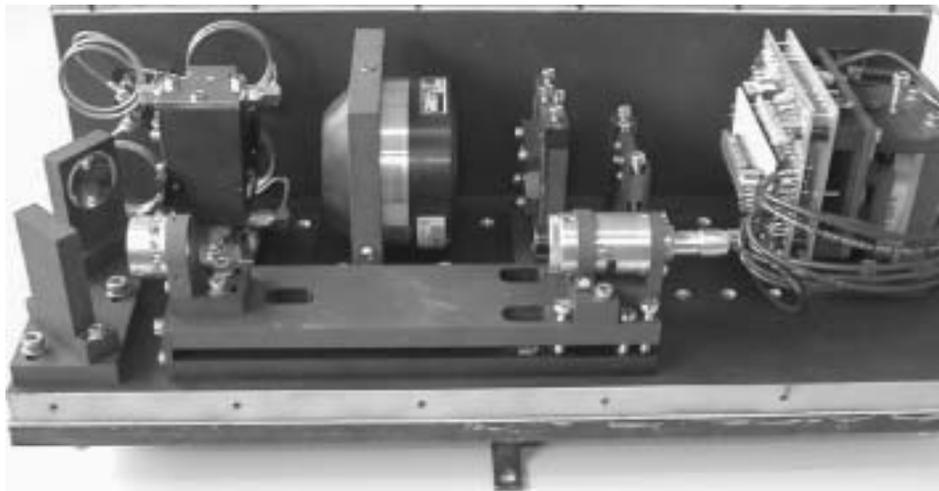
Figure 11. Photograph of the latest development stage of the array AOS optics. The design was done for the HIFI instrument on the Herschel satellite.

cause of the smaller dimensions of the whole optical arrangement. Within the limitations of our CNC milling machine this gives the possibility to produce even more optical elements in one machining cycle to yield the highest positioning and angular accuracy between the optically active surfaces. The surface RMS reachable with our CNC machining equipment is about $2\text{-}3\mu m$, which should be good enough for frequencies up to 3THz at least.

The electronics concept with its modular design and redundancy against computer failure showed high functionality in terms of stability, handling and performance. Also the automatic tuning system is a big advantage in terms of usability and tuning effort needed.

Acknowledgments

References

1. R.Güsten, G.Ediss, F.Gueth, K.Gundlach, H.Hauschildt, C.Kasemann, T.Klein, J.Kooi, A.Korn, I.Krämer, R.LeDuc, H.Mattes, K.Meyer, E.Perchtold, M.Pilz, R.Sachert, M.Scherschel, P.Schilke, G.Schneider, J.Schraml, D.Skaley, R.Stark, W.Wetzker, H.Wiedenhöver, S.Wongsowijoto, and F.Wyrowski, "CHAMP — the carbon heterodyne array of the MPIfR," in *Advanced Technology MMW, Radio, and Terahertz Telescopes*, T. G.Phillips, ed., Proceedings of SPIE Vol. 3357, 1998.

2. C.K.Walker, C.Groppi, A.Hungerford, C.Kulesa, D.Golish, C. d'Aubergny, K.Jakobs, U.U.Graf, C.Martin, and J.Kooi, "Pole star: An 810GHz Array Receiver for AST/RO," in *Proceedings of the 12th International Symposium on Space Teraherz Technology*, (San Diego, California), 2001.

3. U.U.Graf, S.Heyminck, E.A.Michael, and S.Stanko, "KOSMA's 490/810 GHz Array Receiver," in *Proceedings of the 12th International Symposium on Space Terahertz Technology*, (San Diego, California), 2001.

4. C.G.Degiacomi, R.Schieder, J.Stutzki, and G.Winnewisser, "The KOSMA 3 m submm telescope," *Optical Engineering* **34**(9), 1995.

5. S.Haas, C.E.Honingh, D.Hottgenroth, K.Jacobs, and J.Stutzki, "Low noise broadband tunerless waveguide SIS receivers for 440–500 GHz and 630–690 GHz," *Intl. J. IR mm Waves* **17**(3), 1996.

6. J.Stutzki, U.U.Graf, S.Haas, C.E.Honingh, D.Hottgenroth, K.Jacobs, R.Schieder, R.Simon, J.Staguhn, G.Winnewisser, R.N.Martin, W.L.Peters, and J.P.McMullin, "Atomic Carbon in M82: Physical conditions derived from simultaneous observations of the [CI] fine structure submillimeter wave transitions," *Phys. Rev.* **477**(L33), 1997.

7. J. Stodolka and K. Jacobs, "Fabrication and Receiver Measurements of a Diffusion-Cooled Hot-Electron-Bolometer at 800GHz," in *Proceedings of the Twelfth International Symposium on Space Terahertz Technology*, 2001.

8. U.U.Graf and S.Heyminck, "Fourier Gratings as Submillimeter beam splitters," *IEEE TransAP* **49**(4), April 2001.

9. R. Schieder, O. Siebertz, F. Schloeder, C. Gal, J. Stutzki, P. Hartogh, and V. Natale, "Wide-Band Spectrometer for HIFI - First," in *Proceedings of SPIE, UV, Optical, IR Space Telescopes and Instruments*, vol. 4013, SPIE, (Munich, Germany), 2000.

TES Detector Noise Limited Readout Using SQUID Multiplexers

J.G. Staguhn[1,3], D.J. Benford[1], J.A. Chervenak[1], , S.A. Khan[1,4],
S.H. Moseley[1], R.A. Shafer[1], S. Deiker[2], E.N.Grossman[2], G.C. Hilton[2],
K.D. Irwin[2], C. Reintsema

[1] NASA/Goddard Space Flight Center,Greenbelt, MD 20771, USA
[2] NIST, Boulder, CO 80305, USA
[3] SSAI, 5900 Princess Garden Parkway, Lanham, MD 20706, USA

Abstract. The availability of superconducting Transition Edge Sensors (TES) with large numbers of individual detector pixels requires multiplexers for efficient readout. The use of multiplexers reduces the number of wires needed between the cryogenic electronics and the room temperature electronics and cuts the number of required cryogenic amplifiers. We are using an 8 channel SQUID multiplexer to read out one-dimensional TES arrays which are used for submillimeter astronomical observations. We present results from test measurements which show that the low noise level of the SQUID multiplexers allows accurate measurements of the TES Johnson noise, and that in operation, the readout noise is dominated by the detector noise. Multiplexers for large number of channels require a large bandwidth for the multiplexed readout signal. We discuss the resulting implications for the noise performance of these multiplexers which will be used for the readout of two dimensional TES arrays in next generation instruments.

SCIENTIFIC MOTIVATION

Large arrays of fast, low temperature detectors are needed for a variety of disciplines, one of which is astronomy. The best suited wavelength range to study the early evolution of the universe and the assembly of the first prototype galaxies is the FIR/sub-mm regime. About half of the observable radiation energy in the universe is emitted in this wavelength range. The low photon energies combined with the requirement for large bandwidths and large arrays is still a challenge for detector development, since only bolometers meet these requirements. Semiconductor bolometers, which are the most commonly used devices at these wavelengths, cannot easily be multiplexed and are not well suited for high speed readout.

The Superconducting Transition Edge Sensor (TES) bolometer has been developed for use at wavelengths ranging from the sub-mm to X-rays. With its low impedance it is well matched for readout with SQUIDs. This scheme has a number of advantages, since SQUIDs can be operated at the detector temperature and are easily multiplexable. These properties, as well as the small time constants of TES devices, make it possible to develop large arrays of multiplexed TES detectors.

MULTIPLEXED READOUT OF DETECTOR ARRAYS

Large detector arrays require multiplexed readout

For practical reasons, such as limitation of (cryogenic) wiring and amplifiers, large detector arrays require multiplexed readout. The fundamental properties of a multiplexing system are in principle identical, independent of the scheme of choice, be it frequency- or time domain multiplexing: the required bandwidth for the readout increases linearly with the number of pixels. The noise bandwidth of the multiplexed signal is larger than that of the signals from each individual pixel and the noise requirements for the readout system are correspondingly higher, unless the de-multiplexed signal is appropriately filtered. For large arrays such filtering will be necessary and the methods to do so will differ, depending on the multiplexing scheme used.

Multiplexer design
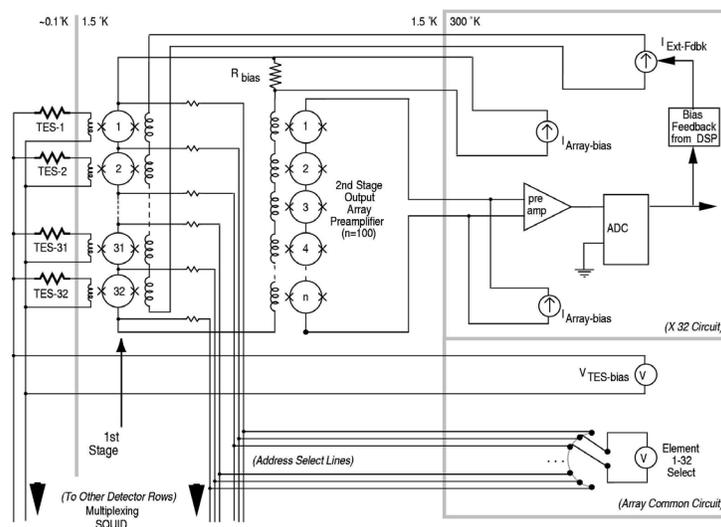
Figure 1. Schematics of the SQUID multiplexers

Figure 1 shows a schematic which illustrates the functionality of our NIST-designed SQUID multiplexers [1]. Each individual detector couples inductively into its own dedicated first stage SQUID. *n* first stage SQUIDs for readout of the same number of detector pixels are stacked in series with *n+1* electrical "address leads". By driving a bias current between one adjacent pair of address lines, only one SQUID will be operational while all other SQUIDs remain superconducting. Thus the output voltage over the entire series of first stage SQUIDs will be equal the voltage across the active SQUID. The TES is biased at all times and is low-pass filtered to a response time which is longer than the readout cycle of the entire first stage.

Noise performance measurements

The phonon dominated white current noise density of TES is approximately given by $i_n = \sqrt{8 k_b T / R_{TES}}$. The Johnson current noise density of a general resistor R is given by: $i_J = \sqrt{4 k_b T / R}$. In order to verify the detector noise dominated performance of the multiplexer it is sufficient to perform Johnson noise measurements of resistors with comparable values of resistance. We tested 8 channel multiplexers which were cut from wavers that are one revision earlier than the SQUID chips we used for the first astronomical observations with multiplexed Transition Edge Sensor bolometers [2]. For the comprehensive noise performance tests we replaced the TES pixels with calibrated resistors. Figure 2 shows a multiplexed measurement of the Johnson noise of resistors at a temperature of 480 mK and the corresponding theoretical predictions for the noise, including aliasing effects. The resistors were read out at a pixel frame rate of 10 kHz, yielding a bandwidth of 5 kHz which is sufficient for the required detector speed. The multiplexed signal was bandpass limited to 100 kHz. Correspondingly all signals up to 100 kHz are aliased into the 5 kHz band (an approximation of the actual filter roll-off above 100 kHz was incorporated in the calculations presented in this paper). The individual detector circuits contain low-pass filters. This allows the integration of the detector signal during one readout cycle. The individual low-pass filter device parameters and the model predictions for the noise are summarized in Table 1. All but one channel we measured are well modeled by a SQUID noise contribution of 4 $pA/\sqrt{Hz}$. This value is consistent with the predicted SQUID multiplexer properties (the mutual inductance in our SQUID multiplexers is ~ 500 pH). Only one channel (channel 5) has elevated SQUID noise, which is caused by a design error in the version of the multiplexers we used for the measurements. Figure 2 demonstrates the good agreement between theoretical predictions and measurements. If appropriately low-pass filtered, the Johnson noise of detectors operated at values higher than 100 mΩ becomes comparable or even less (depending on filter rolloffs) than the aliased SQUID noise of our current eight channel multiplexer. In current ground based operations however we do not operate our detectors above ~100 mΩ.

Conclusion

We have demonstrated the capability of our SQUID multiplexers to perform TES noise dominated readout. The mutual coupling between the SQUIDs and the detector circuits will be increased in the next generation of multiplexer chip designs leading to further improvements in the noise performance. Future large arrays with significant numbers of multiplexed pixels will nevertheless reach the limit at which the SQUID noise contribution is not negligible. In this case, however, fast sampling will make digital bandpass filtering possible. This will allow detector noise limited readout for a larger number of pixels, limited only by the sampling rate of the readout electronics.
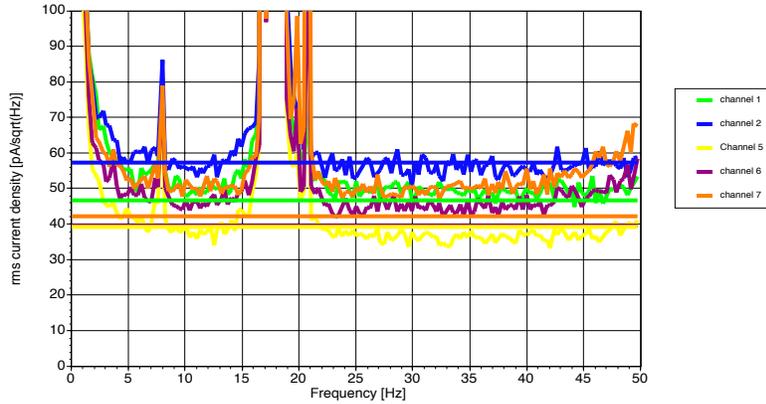
Figure 2. Multiplexed Johnson noise measurement with superimposed model predictions. The peak at 17 Hz corresponds to an acoustic resonance frequency of the wires in the dewar.

TABLE 1. Johnson noise density versus SQUID multiplexer noise density

T=480 mK	R [mΩ]	R/2πL [kHz]	Johns. Noise [pA/$\sqrt{Hz}$]	aliased J.N. [pA/$\sqrt{Hz}$]	SQUID Noise [pA/$\sqrt{Hz}$]	aliased SQUID n. [pA/$\sqrt{Hz}$]
Chan. 2	8	1.78	51.5	37.2	4.0	18.8
Chan. 1	18	3.55	36.4	37.9	4.0	18.8
Chan. 0	52	9.59	22.2	37.4	4.0	18.8
Chan. 7	106	19.18	15.7	36.3	4.0	18.8
Chan. 6	215	38.55	11.0	34.1	4.0	18.8.
Chan. 5	Inf.	inf	0.00	0.00	8.0	37.5

Note: Column 4 "aliased Johnson noise " shows the mean, low-pass filtered and aliased Johnson current noise density in the 5 kHz band of the multiplexed channel readout. This value is less less than the non-low-pass filtered Johnson noise of a resistor with the same resistance as channel 2. The low-pass filter device parameters are shown in column 2. The theoretical values for channel 0 are displayed in the table even though it did not operate properly during the measurement and is not displayed in Fig. 1.

REFERENCES

1. Chervenak, J.A., Irwin, K.D., Grossmann, E.N., Martinis, J.M., Reintsema C.D, and Huber, M.E., *Appl. Phys. Letters*, **74** (26), 4043 (1999).
2.   Staguhn, J.G. et al. "First Astronomical Use of Multiplexed Transition Edge Bolometers", these proceedings

First Astronomical Use of
Multiplexed Transition Edge Sensor Bolometers[*]

J.G. Staguhn[1,2], T.A. Ames[1], D. J. Benford[1], J.A. Chervenak[1],
E.N. Grossman[3], K.D. Irwin[3], S.A. Khan[1,4], B. Maffei[5],
S.H. Moseley[1],F. Pajot[6], T.G. Phillips[7], J.-C. Renault[8],
C.D. Reintsema[3], C. Rioux[6], R.A. Shafer[1], C. Vastel[6,9] & G.M. Voellmer[1]

1. NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
2. SSAI, Lanham, MD 20706, USA   *3. NIST, Boulder, CO 80305, USA*
4. Imperial College, London SW7 2BW, UK   *5. University of Wales, Cardiff, CF24 3YB, Wales*
6. IAS-CNRS, 91405 Orsay, France   *7. Caltech, Pasadena, CA 91125, USA*
8. IAP-CNRS, F-75014 Paris, France   *9. CESR, 31028 Toulouse, France*

Abstract. We present performance results based on the first astronomical use of multiplexed superconducting bolometers. The Fabry-Perot Interferometer Bolometer Research Experiment (FIBRE) is a broadband submillimeter spectrometer that achieved first light in June 2001 at the Caltech Submillimeter Observatory (CSO). FIBRE's detectors are superconducting transition edge sensor (TES) bolometers read out by a SQUID multiplexer. The Fabry-Perot uses a low resolution grating to order sort the incoming light. A linear bolometer array consisting of 16 elements detects this dispersed light, capturing 5 orders simultaneously from one position on the sky. With tuning of the Fabry-Perot over one free spectral range, a spectrum covering $\Delta\lambda/\lambda = 1/7$ at a resolution of $\delta\lambda/\lambda \approx 1/1200$ can be acquired. This spectral resolution is sufficient to resolve Doppler-broadened line emission from external galaxies. FIBRE operates in the 350 μm and 450 μm bands. These bands cover line emission from the important star formation tracers neutral carbon [CI] and carbon monoxide (CO). We have verified that the multiplexed bolometers are photon noise limited even with the low power present in moderate resolution spectrometry.

SCIENTIFIC MOTIVATION

Spectroscopy of distant galaxies in the far-infrared and submillimeter has lagged behind continuum studies at the same wavelengths (e.g., with SHARC at the CSO [1]) and spectroscopic studies at longer wavelengths (e.g. OVRO [2]). In large part, this is due to a relative lack of available instrumentation combining high sensitivity and large bandwidth. Observing an emission line from a typical galaxy with velocity-broadened linewidth of ~300km/s in the 350 μm and 450 μm atmospheric windows (850 GHz and 650 GHz, respectively) requires a spectrometer with a bandwidth of at least 0.5 μm (1 GHz). Additionally, detecting this line is easiest if the spectral resolution is approximately this width. Hence, a spectrometer with a spectral resolution of slightly more than 1000 is optimal for the detection of faint galaxies in the far-infrared and submillimeter.

---

[*] Contribution of the U.S. Government, not subject to copyright.

The Fabry-Perot Interferometer Bolometer Research Experiment, FIBRE, is an instrument designed to demonstrate a suite of advanced technologies suitable for sensitive detection of far-infrared light. This includes superconducting transition edge sensor (TES) bolometers, SQUID multiplexed amplifiers, and a cryogenic Fabry-Perot interferometer. These components are being developed for the SOFIA imaging Fabry-Perot spectrometer SAFIRE and for a complement of ground-based instruments.

INSTRUMENT DESIGN

Superconducting TES Bolometers and SQUID Amplifiers

The superconducting TES bolometer has been developed for use at wavelengths from the submillimeter to X-rays. It combines high speed with high sensitivity and can be read out by SQUID amplifiers, which are well suited to multiplexing. Unfortunately, the development of these detectors is too detailed to discuss at length here [3,4]. The FIBRE bolometers were used in the laboratory to demonstrate that multiplexed detection using the NIST-designed SQUID multiplexer of Chervenak et al. [4] was possible [5]. Further measurements, as presented by Staguhn et al. [6] in these proceedings, have verified that the noise performance of these detectors is limited by phonon and Johnson noise contributions as predicted by theory.

FIBRE features two 1×8 monolithic bolometer arrays consisting of 1 mm×1 mm absorbers with a 50 μm ×150 μm Mo/Cu bilayer TES. This is shown in Fig. 1.
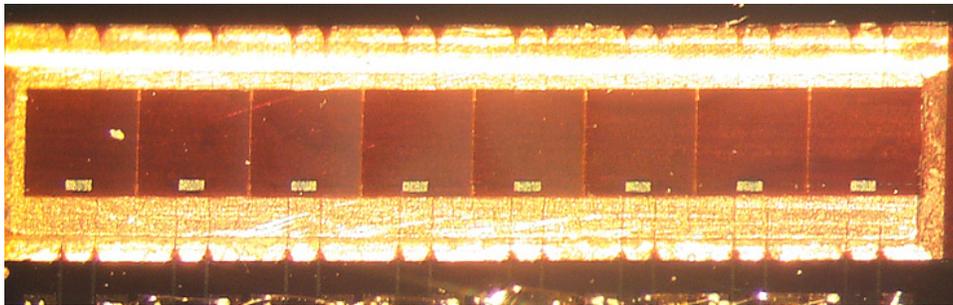
FIGURE 1. Photograph of a single 1×8 monolithic bolometer array. Each 1 mm×1 mm pixel is a 1 m thick silicon membrane supported by 4 legs approximately 5 m wide. The TES is the small pale rectangle at the bottom center of each detector.

Optical Design

The optical design uses a single Fabry-Perot etalon followed by an order-sorting grating [7]. The grating is blazed to operate in its first order, which is broad enough to transmit orders 40-45 of the Fabry-Perot for 350 μm operation and orders 32-35 for 450 μm operation. The grating disperses these Fabry-Perot orders along the array such that they are separated onto adjacent sets of pixels. In this manner, a spectrum consisting of several orders of the Fabry-Perot is collected simultaneously. By stepping the Fabry-Perot over one free spectral range, a complete spectrum is accumulated.
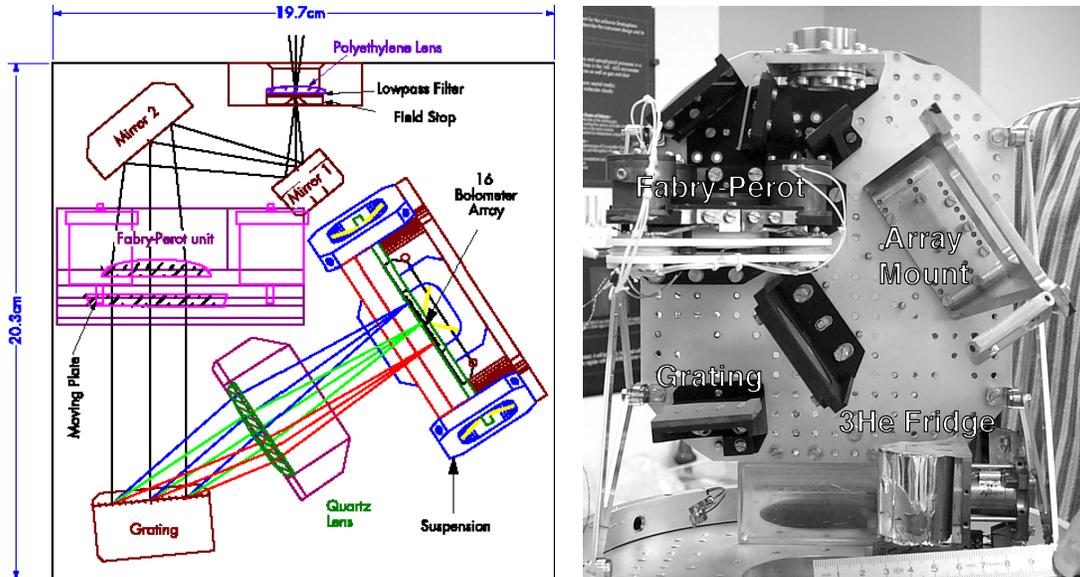
FIGURE 2. (Left) Diagram of spectrometer optics, showing three orders being dispersed onto the bolometer array. (Right) Optics being assembled, with the detector array and baffles yet to be added.

Commissioning Observations

FIBRE was delivered to the Caltech Submillimeter Observatory (CSO) on Mauna Kea, Hawai'i in May 2001. During six nights of poor weather, the instrument was operated and found to work very well. The bolometers were read out in multiplexed fashion (Fig. 3). Those bolometers that were illuminated at a given Fabry-Perot tuning were found to have ~20 times the noise of the dark bolometers. The expected photon noise contribution is approximately 10 times the intrinsic (phonon + Johnson) noise of the detectors, so the system noise is near the theoretical performance and the bolometers are background-limited with a net NEP of 3×10^{-17} W/$\sqrt{\text{Hz}}$. A spectrum (Fig. 4) was taken using a local oscillator source operating at 372 μm (807 GHz). The spectral resolving power was measured to be 1200, for a velocity resolution of 250 km/s, as predicted from the known performance of the Fabry-Perot.

The opacities at the zenith during the observing run were measured by skydips using both FIBRE and the CSO facility 350 μm taumeter. These measurements yielded zenith opacities of $\tau_{350\ \mu m} \sim 4$ during most of the run. No scientific data could be taken in such poor conditions, but in order to demonstrate a multiplexed detection using TES bolometers, we observed the limb of the Moon at 365 μm.
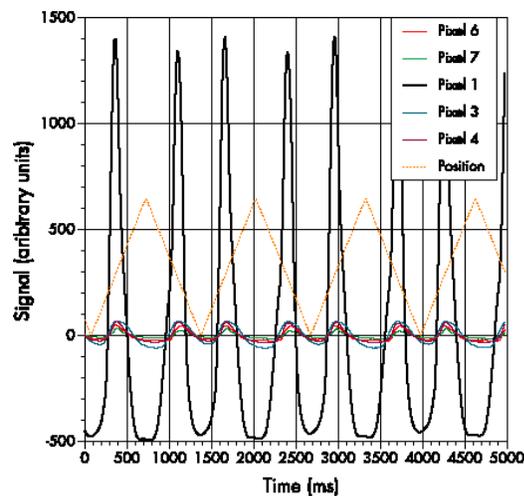
FIGURE 3. Multiplexed readout of the FIBRE bolometers while tuning the Fabry-Perot and observing a local oscillator source. The signal is seen strongly on one channel, weakly in others, demonstrating good optical performance for spectroscopy.

The secondary mirror was nutated to subtract the atmosphere, so we obtained a high signal-to-noise detection of the Moon emission despite a transmission of ~1% (Fig. 5).
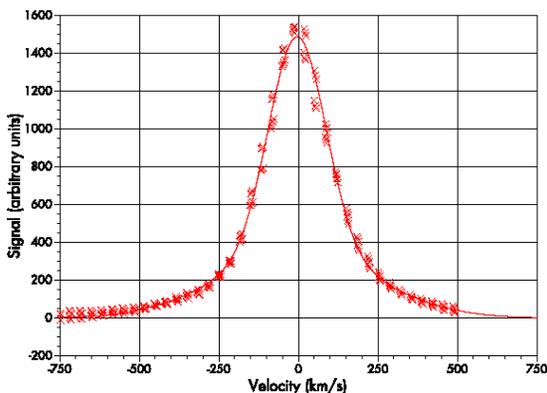
FIGURE 4. Calibration spectrum taken using a local oscillator operating at 372 μm (807 GHz), calibrated in velocity units. The resolution is 250 km/s , for a resolving power of 1200.
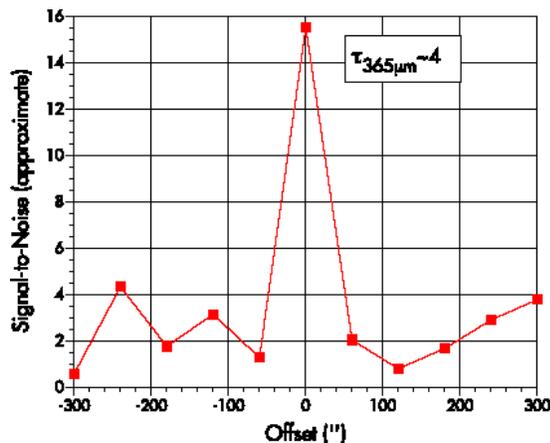
FIGURE 5. Detection of the limb of the Moon, taken while chopping so that signal is seen only when exactly on the limb. Each data point contains 3 seconds of on-source time.

CONCLUSION

FIBRE achieved first light at the CSO, detecting the Moon at 365 μm in bad weather with an atmospheric transmission of ~1%. The spectrometer was operating with a spectral resolving power of ~1200, and the signal amplitude and noise were consistent with expectations. The TES bolometer and SQUID multiplexer technology has been thus validated in an astronomical application. We anticipate future observations to study galaxies in the fine-structure line of CI and the CO rotational lines, and to continue to refine the multiplexed TES detectors in astronomical applications.

ACKNOWLEDGMENTS

REFERENCES

1. Benford, D.J., Cox, P., Omont, A., Phillips, T.G. & McMahon, R.G., *ApJ* **L518**, 65 (1999).
2. Blain, A.W., Frayer, D.T., Bock, J.J., & Scoville, N.Z., *MNRAS* **313** (3), 559-570 (2000).
3. Benford, D.J. et al., *ASP Conference Series* **217**, 134-139 (1999).
4. Chervenak, J.A, et al, *Appl. Phys. Letters* **74** (26), 4043 (1999).
5. Benford, D.J., et al., *Int. J. IR MM Waves* **21** (12), 1909-1916 (2000).
6. Staguhn, J.G., et al., these proceedings.
7. Maffei, B., et al., *Infrared Physics & Technology* **35**, 321 (1994).

Dual Transition Edge Sensor Bolometer for Enhanced Dynamic Range

J. A. Chervenak, D. J. Benford, S. H. Moseley
NASA, Goddard Space Flight Center, Greenbelt, MD 20771

K. D. Irwin
NIST, Boulder, CO 80303

Abstract

Broadband surveys at the millimeter and submillimeter wavelengths will require bolometers that can reach new limits of sensitivity and also operate under high background conditions. To address this need, we present results on a dual transition edge sensor (TES) device with two operating modes: one for low background, ultrasensitive detection and one for high background, enhanced dynamic range detection. The device consists of a detector element with two transition temperatures (T_c) of 0.25 and 0.51 K located on the same micromachined, thermally isolated membrane structure. It can be biased on either transition, and features phonon-limited noise performance at the lower T_c. We measure noise performance on the lower transition 7×10^{-18} W/rt(Hz) and the bias power on the upper transition of 12.5 pW, giving a factor of 10 enhancement of the dynamic range for the device. We discuss the biasable range of this type of device and present a design concept to optimize utility of the device.

Introduction

Space-based astrophysics at submillimeter and millimeter waves will rely on detectors with extremely low noise equivalent power (~10^{-18} W/rt(Hz)) such as that achievable with a transition edge sensor (TES) bolometer [1,2]. Many groups have demonstrated superconducting and semiconducting devices on thermally isolated structures with performance suitable for astrophysics. However, such devices have suffered from a relatively low bias power, which make them susceptible to saturation or degradation of sensitivity under high background conditions. The alternatives are to apply stringent limits on the throughput of the telescope at the focal plane or to devise a detector that maintains a suitable level of operation during high background observations. The ideal solution would not throw away photons and maintain photon noise limited detector performance.

We propose a high dynamic range device in which, on a single thermally isolated structure, detector and readout for low background limits are connected in series with a lower performance detector element, which operates under high background when the ultrasensitive detector is saturated. One technical path to enhancing dynamic range is modulation of the thermal conductance G, which controls the thermal transport between the detector element and the thermal bath. In many cases, the geometry of membrane thickness and aspect ratio of micromachined features set detector G. Concepts for intrinsic modulation of G are being pursued, for example, in the integrating bolometer work [3]. Here, we explore the modulation of dynamic G of the device with temperature. We present the design of a superconducting TES device exhibiting two distinct bias temperatures integrated into the same geometry and report on a successfully fabricated and tested device.

Superconducting TESs offer linear response over the entire biasable region of the transition, that is, the range of voltages over which the superconducting film is held with strong electrothermal feedback (ETF) in the narrow region between the superconducting and normal states of the detector element. Since the detector self-regulates at T_c, a voltage biased TES has the same bias power and hence G at any point on the transition. As show below, a device with two T_cs can be designed to have two Gs and exhibit two distinct regions of bias voltage which have different bias powers. When biased at the low T_c, the high T_c portion is superconducting

and contributes no noise or power, enabling operation under low background conditions at the best achievable sensitivity of the device. Under high background conditions, the device remains baisable with strong ETF, albeit with a large parasitic resistance from the normal state of the low T_c device, such that the degradation in sensitivity can be well calibrated.

Performance of a dual transition device

We fabricated a dual T_c device by depositing Al and Ag films through a shadowmask onto a moveable mount holding a silicon substrate. The shadowmask is a silicon wafer with etched holes spaced several millimeters apart. It is mounted as an aperture plate directly in front of a 1 mm^2 silicon membrane (1 μm thick) with micromachined, thermally isolated legs, which can be manipulated in vacuum to deposit a chosen feature through the mask and onto the Si. A small (50x100 μm^2) Al/Ag proximity bilayer is deposited through two apertures such that the Ag film completely covers the Al film, ensuring normal metal boundary conditions in the superconducting element. A larger aperture is used to cover a fraction of the first bilayer with thicker Ag so that, for this device design, the dual T_c is exhibited by two segments of the same Al film. Then AlO$_x$ is deposited (e-beam Al in 5x10^{-6} torr background pressure of O$_2$) on the legs to create contacts between the bilayer and frame of the chip. These leads are both resistive in the normal state (>10 Ω/sq) and high T_c (~2.5 K), ensuring that the leads make a negligible contribution to G at sub-Kelvin temperatures.

To obtain the current-voltage characteristic of the TES, the current through a shunt resistor in parallel with the TES is varied, generating a known voltage bias across the detector element. The current through the detector is then read out with a two-stage SQUID in which the first stage input coil is in series with the detector element in a superconducting loop. The resistance versus temperature is measured by noise thermometry at zero bias with the same SQUID amplifiers, which are impedance matched to have much lower input current noise than the Johnson noise of the detector element in its normal state. The low frequency current noise reflects the sum of the detector, shunt, and parasitic resistances in the loop, and the detector is presumed to be the only temperature dependent resistance over the sub-Kelvin temperature range.
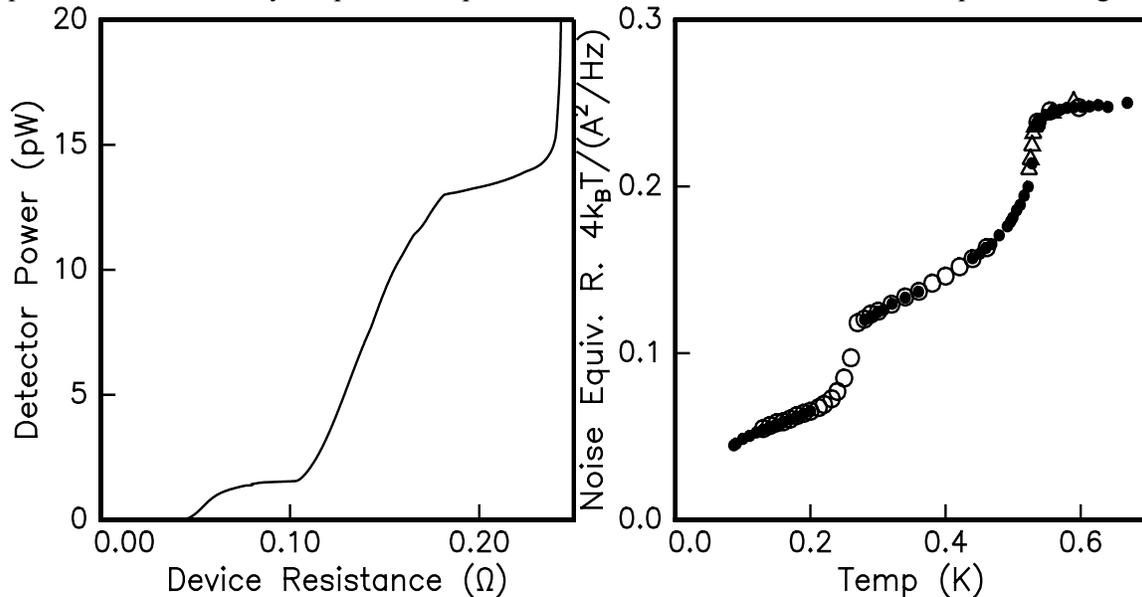
Figure 1(a) Power vs. resistance in a dual-T_c TES (b) Resistance vs. temperature of the same device

We show an I-V curve converted to detector bias power versus resistance in Fig. 1 (a). There are two regions of negative differential resistance in the IV that appear as flat power

regions of 1.2 and 12.5 pW, near 0.1 and 0.2 Ohms respectively, throughout which the detector exhibits strong electrothermal feedback. This factor of roughly ten in bias power constitutes the increase in dynamic range for this device, defined in terms of the total power throughput to the focal plane required to saturate the detector. In between these relatively flat regions is a sloping current-biased region in which the power in the detector keeps it on the upper transition but the parasitic resistance in the circuit prevents strong ETF because the detector is not voltage biased. The slight slope to the upper biasable region is also attributed to the parasitic, which perhaps varies as a function of bias voltage for this device. The base temperature for the thermal bath is regulated at 100 mK for the IV characteristic and the noise measurements presented in Fig. 2. The structure in the IV is corroborated by measurements of the resistance versus temperature curve for the device, shown in Fig 1(b). The RT curve exhibits two distinct sharp drops at 0.51 and 0.25 K and film resistance above each drop that corresponds to the resistance at the top of the flat power regions in the IV curve. It also shows a gradual resistance decrease in between the two sharp drops, likely a consequence of the method of fabrication, and a substantial parasitic resistance in the loop at the lowest temperatures.
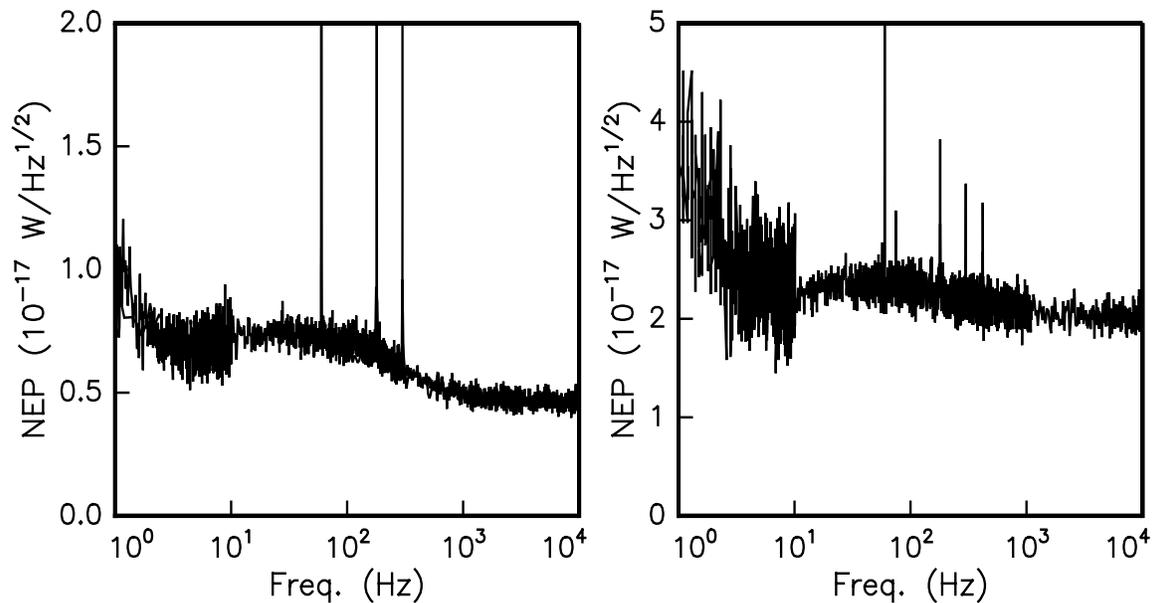
Figure 2(a) Noise measured in a dual-T_c device biased on the lower T_c (b) Noised measured at the high T_c.

We have taken noise measurements while the detector is biased at each T_c that are shown in Figs. 2(a) (low T_c) and 2(b) (high T_c). The current noise in the SQUID is measured and multiplied by the voltage at which the detector is biased. The low T_c device noise shows phonon-limited performance between 1-300 Hz and rolls off to the Johnson noise of the device at frequencies above the signal band. The measured value of 7×10^{-18} W/rt(Hz) is in good agreement with the theoretical limit $(2k_B T_c^2 G)^{1/2}$ that is predicted for strong ETF. The high T_c device noise shows excess noise in and out of the signal band but, measure at 2.3×10^{-17} W/rt(Hz), is still within a factor of two of the phonon-noise limit. The heat capacity of Si is comparable to that of the detector element at the high T_c and plays a role in this noise degradation [4].

Device design for enhanced dynamic range

We have shown that placing a well-thermalized superconductor of similar heat capacity in series with the detector element does not affect the performance of an ultrasensitive TES bolometer. However, it is apparent from Fig. 1 that the challenging part of the dual T_c detector is

designing a high T_c detector element that is biasable (i.e. at voltages where ETF is exhibited) over the largest possible range. A dual transition device can be implemented in several different ways including the series arrangement presented here in which a single element exhibits two T_cs, two separate sensors also connected in series, or independently biased detector elements on the same thermally isolated membrane. We first discuss the device with two detector elements wired in series.

The range of the biasable region on the high T_c device is determined by the size of the current-biased region in which the detector bias resistance is comparable to the sum of the normal state of the low T_c device (R_{lo}), shunt, and parasitic resistances. A device in which $R_{lo} \sim R_{hi}/4$ would provide a substantial biasable region for the higher T_c. Since a current-biased region of some extent is difficult to avoid, saturating the low T_c sensor with signal power may drive the device into a region that is difficult to calibrate quickly. Thus, to implement a two T_c device, a resistive heater element should be integrated onto the bolometer membrane with separate leads to supply quiescent power to the device prior to saturating with signal. The bolometer can then be calibrated to operate in a region of strong ETF at high T_c without having to adjust the bias voltage of the detector. Similarly, two separate detector elements connected in series will exhibit the same behavior. The advantage to this method is that a cleaner RT curve can be achieved, although the longer thermalization times may result in non-optimal noise behavior.

Note that the series impedance always limits the biasable range at high T_c. It also potentially suppresses the noise margin of the SQUID. The SQUID, in each case, must be optimized for the multiplexing limit for the low T_c current noise. Matching the detector noise at the top of the two transitions (for this example $T_c(low)/T_c(high) \sim 0.25$ when $R_{lo} = R_{hi}/4$) makes best use of the dynamic range of the SQUID.

Independent biasing appears to be the most straightforward method to two T_cs. However, it requires separate bias line wiring, which, unless connected in series on the frame of the chip can result in more complicated room temperature electronics. Also, disconnecting the detector elements necessitates thermalization of their electron temperatures through the membrane. Either effect can lead to complicated transients and unwanted time constants in the detector response. This arrangement does leave open the possibility that separate SQUIDs be used to optimize amplifier noise for both operation modes.

In conclusion, a dual-T_c transition edge sensor bolometer has been fabricated and tested. The device exhibits the theoretical phonon-noise limit biased at the low T_c and a factor of ten dynamic range enhancement when biased at the higher T_c.

References

[1] D. J. Benford, S. H. Moseley, and J. A. Chervenak, "Mission requirements for ultralow-background large-format bolometer arrays,,. Proc. Far-IR, Sub-mm, & mm Det. Tech. Workshop, Monterey, CA 2002, *in press*. D. J. Benford, et. al., "Surveying galaxy evolution in the far-infrared all-sky survey concept,,, This proceedings: Proc. Far-IR Missions Workshop, College Park, MD 2002, *in press*.
[2] K. D. Irwin, G. C. Hilton, D. A. Wollman, and John M. Martinis "X-ray detection using a superconducting trnsition edge sensor with electrothermal feedback,,. Appl. Phys. Lett. **69**, 1945 (1996).
[3] A. Kogut, M. Dipirro, J. Panek, and S. H. Moseley, Proc. Far-IR, Sub-mm, & mm Det. Tech. Workshop, Monterey, CA 2002, *in press*
[4] H. F. C. Hoevers, et. al. "Thermal fluctuation noise in a voltage biased superconducting transition edge thermometer,,. Appl. Phys. Lett. **77**, 4422 (2000).

Least–Squares Self–Calibration of Imaging Array Data

R. G. Arendt[1,2,3], S. H. Moseley[1], D. J. Fixsen[1,2]

[1] *Code 685, NASA/GSFC, Greenbelt, MD, 20771*
[2] *Science Systems & Applications, Inc.*
[3] *arendt@stars.gsfc.nasa.gov*

ABSTRACT

When arrays are used to collect multiple appropriately–dithered images of the same region of sky, the resulting data set can be calibrated using a least–squares minimization procedure that determines the optimal fit between the data and a model of that data. The model parameters include the desired sky intensities as well as instrument parameters such as pixel–to–pixel gains and offsets. The least–squares solution simultaneously provides the formal error estimates for the model parameters. With a suitable observing strategy, the need for separate calibration observations is reduced or eliminated.

We show examples of this calibration technique applied to HST NICMOS observations of the Hubble Deep Fields and simulated SIRTF IRAC observations.

1. Overview

Dithered images of the sky can contain sufficient information to determine both the sky intensities and the instrument properties without the use of separate calibration data. The best set of parameters describing the sky and the instrument can be obtained through a least–squares solution of the over–determined system of linear equations that model the data.

In this process, each point on the sky serves as a stable calibration source which is used to set the relative calibration of a subset of the detector pixels. The overlap between subsets for various points on the sky establishes the relative calibration of all detector pixels.

A much more complete description of the least–squares calibration can be found in Fixsen, Moseley, & Arendt (2000) and Arendt, Fixsen, & Moseley (2000). The notation used here follows that used in these references.

2. Data Models

The current version of the code calculates least–squares calibrations for data D^i that are described by any of the following three models:

$$D^i = S^\alpha G^p \tag{1}$$

$$D^i = S^\alpha G^p + F^p \tag{2}$$

$$D^i = S^\alpha G^p + F^p + F^q \tag{3}$$

where S^α is the sky intensity at position α, G^p is the detector gain at pixel p, F^p is the detector offset at pixel p, and F^q is a detector offset that can vary from frame to frame but

is constant over a) a whole frame, b) each quadrant of a frame (e.g. *HST* NICMOS), or c) alternating columns of a frame (e.g. *SIRTF* IRAC).

3. Algorithm Description

Our algorithm is a least–squares fit. As such, the solution for both the sky and detector parameters, X, can be computed as

$$X = (HWH)^{-1}HWD \tag{4}$$

where D is the data, W is the weight or inverse of the covariance matrix, and H is the mapping from the data to the detector pixels and sky positions (i.e the data model). Our algorithm takes advantage of the facts that 1) W is diagonal, 2) H is sparse and maps each datum to a single sky position, and 3) known singularities can be treated. The matrix HWH is divided into detector (A) and sky (C) portions as

$$HWH = \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \tag{5}$$

The mapping within H means that C is diagonal. Hence we express the portion of the inverse matrix related to the detector parameters as

$$(HWH)^{-1} = (A - BC^{-1}B^T)^{-1} = A^{-1/2}(I - TT^T)^{-1}A^{-1/2} = A^{-1/2}\left[\sum(TT^T)^n\right]A^{-1/2} \tag{6}$$

where $T \equiv A^{-1/2}BC^{-1/2}$ and T retains the sparse features of H.

4. Code Description

Direct evaluation of $(HWH)^{-1}$ would require inverting a $P \times P$ matrix where P is the number of parameters in the model. For a modest 256×256 array and a model including gains and offsets, $P > 2 \times 10^5$. By instead using the series expansion and making use of the sparse structure of H, our code avoids the direct inversion of such large matrices and never needs to handle arrays larger than twice the size of the data set which is being calibrated. Calculation of formal statistical uncertainties are a minor variation of the calculation of the solution itself. The code is written in IDL, with some calls to simple C functions. For efficiency in speed and memory, the structure of the code is related to the types of parameters in the data model.

5. Benefits and Features

A short list of some of the advantages of using a self–calibration procedure are:
1) Data can be calibrated without time lost to dedicated calibration observations.
2) Since the science data determine the calibration, the calibration is perfectly matched to the time and conditions of the science data collection.
3) The data model (H) clearly specifies any assumptions, avoiding hidden assumptions of ad hoc calibration plans.
4) Calibration can be derived in cases with few dithers if several fields are observed and

instrument parameters are stable or modeled.

6. Dithering Issues

A good dither pattern establishes direct calibrations for many pairs of pixels on all spatial scales regardless of the calibration method to be used. Any desired subpixel dithering may be superimposed on the main pattern. Subpixel dithering may be desirable for improving the resolution of data collected with detectors that undersample the instrument PSF. Figure 1 shows examples of one bad and two good dither patterns.
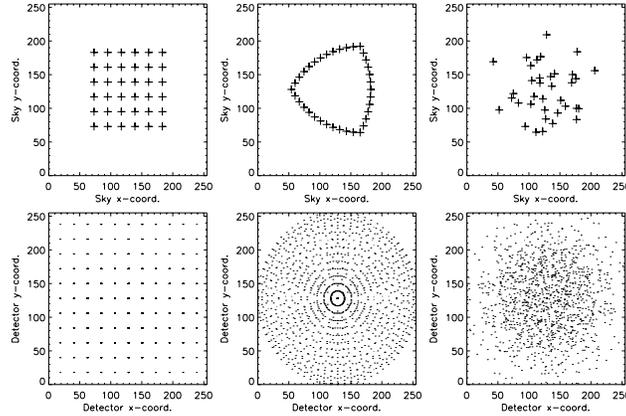
Fig. 1.— The top row shows three examples of 36–position dither patterns. The bottom row shows the corresponding patterns of all detector pixels that can be directly calibrated against the center pixel (for this example) of the array, by virtue of having observed a common position on the sky. The grid pattern at left is bad because it only provides calibrations between detector pixels at fixed intervals. The Reuleaux triangle and random patterns are good at providing calibrations on a wide range of scales.

7. Examples

The least–squares calibration code has been tested using *HST* NICMOS data for the HDF–N and HDF–S. These data were calibrated using the data model given by eq. (3), in which the F^q term contains the derived corrections for the "pedestal" effect that afflicts NICMOS data. The procedure contains no requirements on the spatial locations of the data being calibrated, only that the sky and detector parameters are fixed over the time of the observations. Therefore, we have processed HDF–N and HDF–S data as a single data set. This allows the HDF–N data which was collected with a relatively poor grid–like dither pattern to benefit from the more favorable dithering that was used for the HDF–S. Figures 2 and 3 show the resulting calibrated images for data taken with the F110W and F160W filters. For the HDF–N F160W observations, some of the HDF–N observations exhibit a different sky background level than the main field. Since these data would invalidate the assumption that the sky intensities S^α are constant, we processed these data as if the were at a third location (leftmost field in Fig. 3). In effect, the sky index α can be used to indicate both position and epoch of the observations in cases where the sky intensity varies.

Examination of the calibration of individual frames from the F110W data shows that

some frames are better calibrated than others (Fig. 4). The poorly calibrated frames turn out to be those observation for which the angle between the HDF and earth limb was smallest, and which appear to have elevated sky background levels. The results in Figure 4 show how the calibration of some of the data may be affected in cases where variable sky intensities are present, but ignored.

8. Dark Frames

Including dark frames as part of the data set to be calibrated can help prevent degeneracy between detector gain and offset parameters for low contrast targets. In this degeneracy a term x^p, which is fixed over time but may vary across the detector, may be subtracted from the gain and added to the offset after scaling by the mean sky intensity, without significantly changing the fit to the data:

$$D^i = G^p S^a + F^p \approx (G^p - x^p)S^\alpha + (F^p + x^p \langle S^\alpha \rangle) \tag{7}$$

If dark frames are unavailable, the degeneracy can be broken by observing a sky field or fields with significant contrast, such that $S^\alpha \neq \langle S^\alpha \rangle$, as in the examples below. Also, dark frames provide an absolute zero point reference for the sky and detector which can only be recovered from the sky observations with additional assumptions. About 1% of the HDF data in the previous examples were dark frames.

REFERENCES

Fixsen, D. J., Moseley, S. H., & Arendt, R. G. 2000, ApJS, 128, 651

Arendt, R. G., Fixsen, D. J., & Moseley, S. H. 2000, ApJ, 536, 500

Arendt, R. G., Fixsen, D. J., & Moseley, S. H. 2001, ADASS XI Conference Procedings, in press. (A long form of this is also available at http://cfa-www.harvard.edu/cfa/oir/Research/irac/arendt_hdf_preprint.pdf)
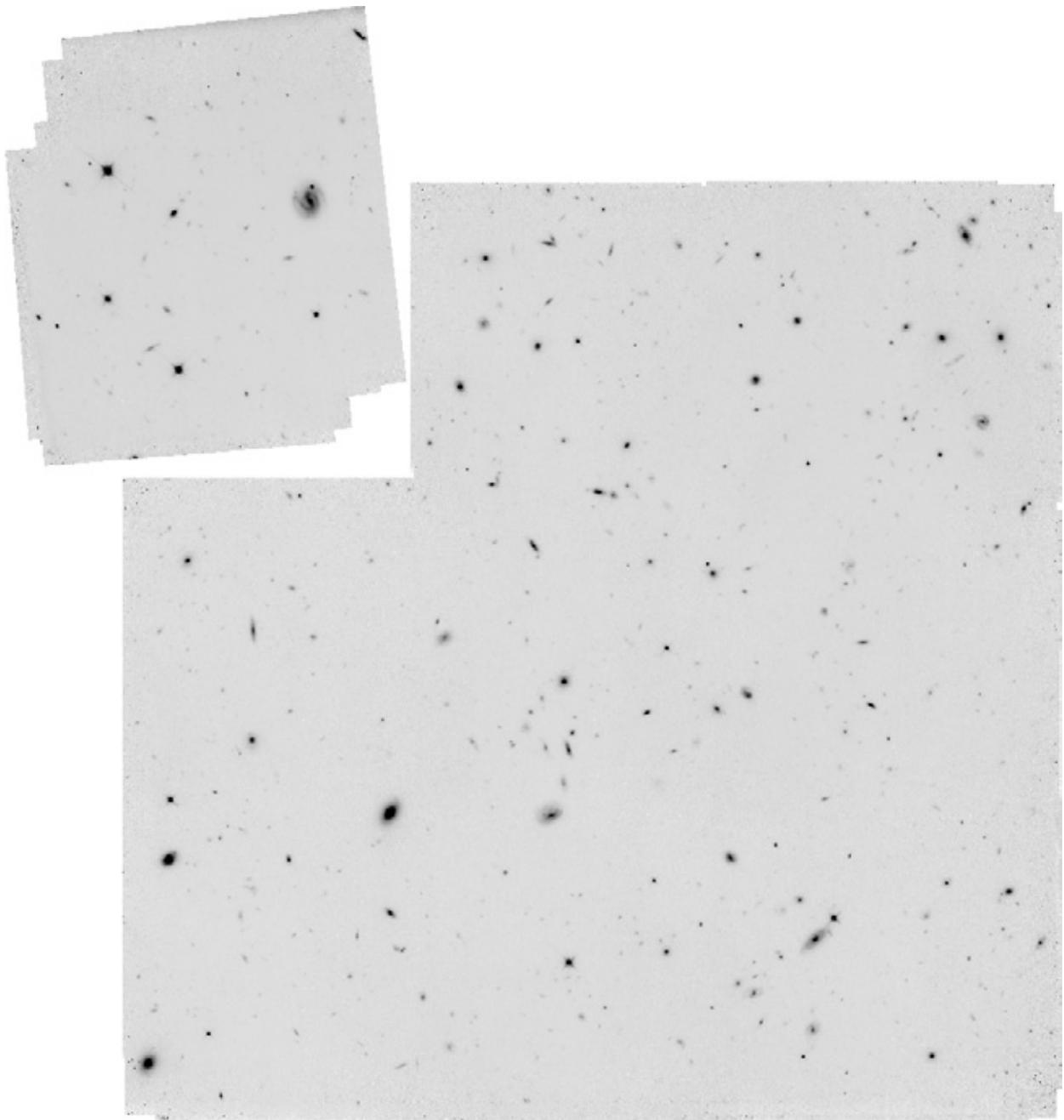
Fig. 2.— The least–squares self–calibrated image of the HDF–N (large area) and HDF–S (small area) in the NICMOS F110W filter. Displayed with square root intensity scaling.
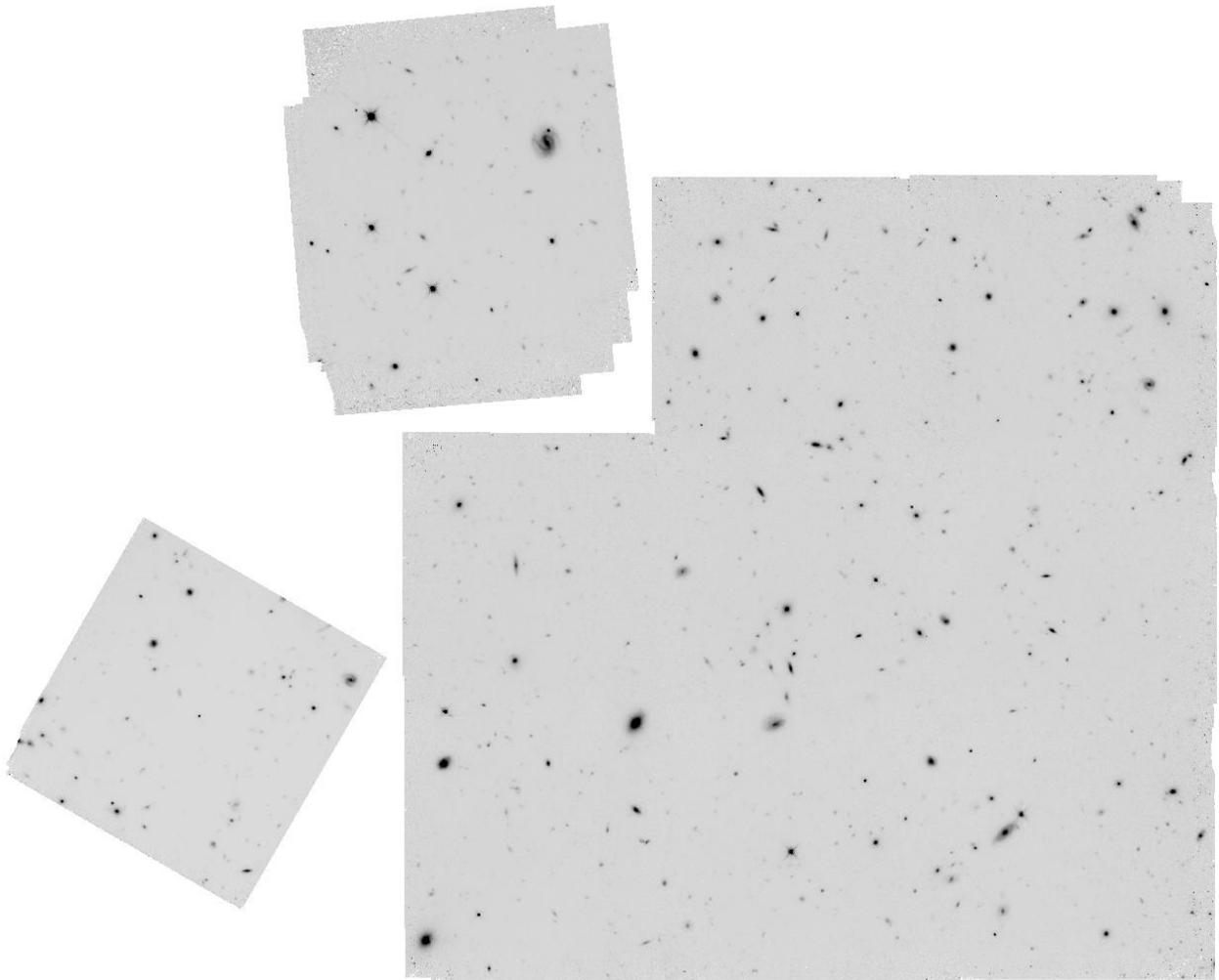
Fig. 3.— The least–squares self–calibrated image of the HDF–N (large area) and HDF–S (small area, upper left.) in the NICMOS F160W filter. The small region at left, is a portion of the HDF–N field that was observed at a different epoch and with a different background level, and was thus processed as if it was a third sky location. Displayed with square root intensity scaling.
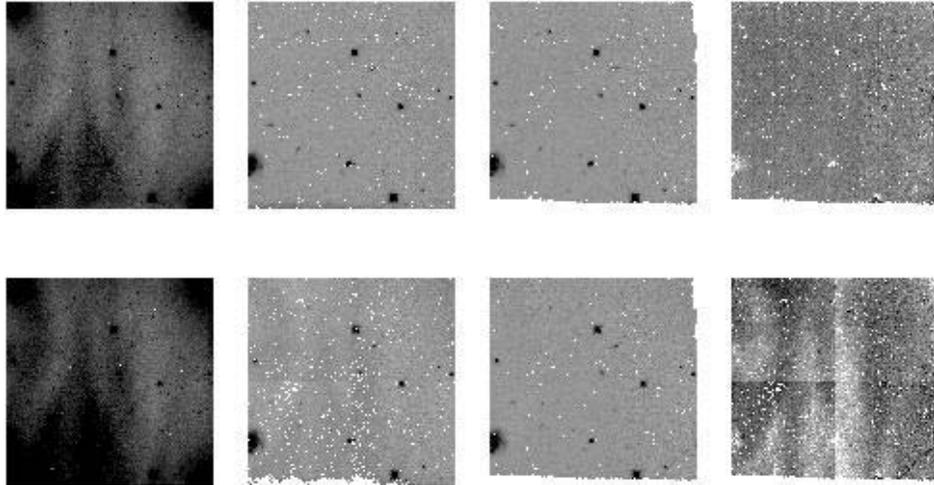
Fig. 4.— Left to right are: a single raw NICMOS frame, the same frame with calibrated via the least–squares self–calibration, the same frame as calibrated by processing at STSCI, and the difference between the two calibrated frames. The top row shows a frame for which the self–calibration solution works well. The second row shows a frame which had different sky background level than most of the data, and is consequently poorly calibrated by the same solution.
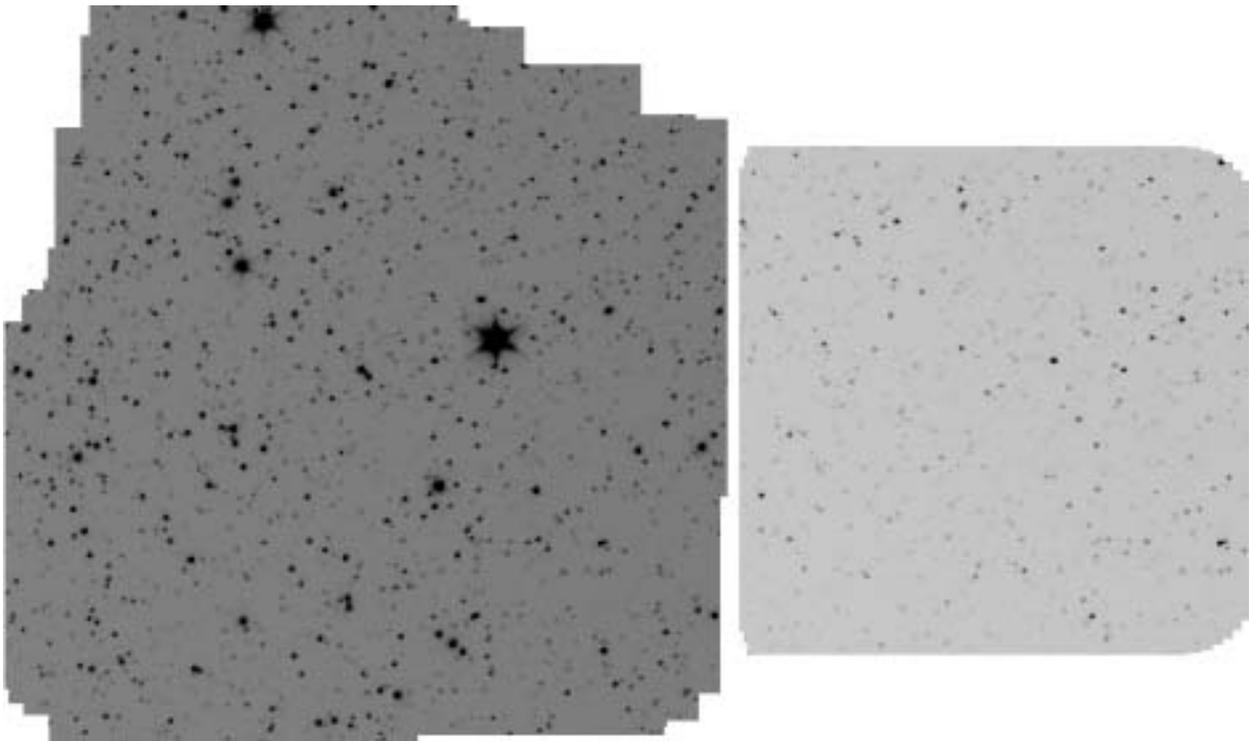
Fig. 5.— An example of simulateously processing fields with two different sky levels, to avoid degeneracies in the self–calibration solution without the use of dark frames. These are simulated IRAC data at low and high ecliptic latitudes.
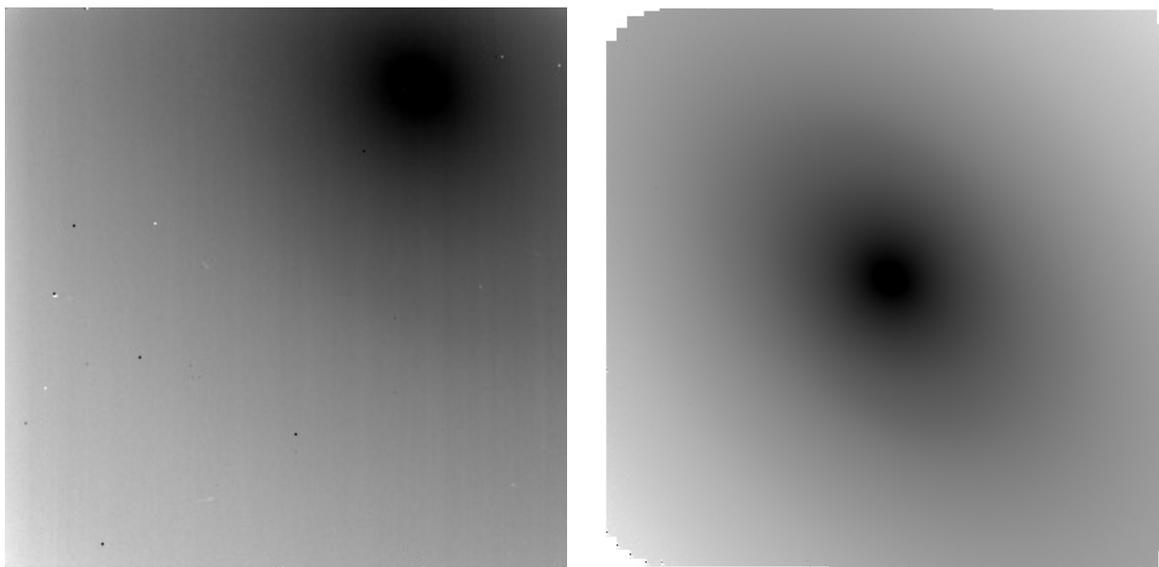
Fig. 6.— An example of using a single field with high contrast to avoid degeneracies in the self–calibration solution without the use of dark frames. This field is simulated IRAC observations of a model of the M 31 surface brightness. On the left is a single raw frame. Detector gain variations are small compared to the sky intensity variation, and may be hard to see in this reproduction. On the right is the calibrated final image. The full image is reduced in scale by a factor of ~ 2 compared to the single frame at left.

SESSION VII

OPTICS TECHNOLOGY

Wavefront Sensing and Control Technology for Submillimeter and Far-Infrared Space Telescopes

Dave Redding
NASA / JPL

Abstract

The NGST wavefront sensing and control system will be developed to TRL6 over the next few years, including testing in a cryogenic vacuum environment with traceable hardware. Doing this in the far-infrared and submillimeter is probably easier, as some aspects of the problem scale with wavelength, and the telescope is likely to have a more stable environment; however, detectors may present small complications. Since this is a new system approach, it warrants a new look. For instance, a large space telescope based on the DART membrane mirror design requires a new actuation approach. Other mirror and actuation technologies may prove useful as well.

Acknowledgements

Imaging and Wavefront Control

Mirror — Rays
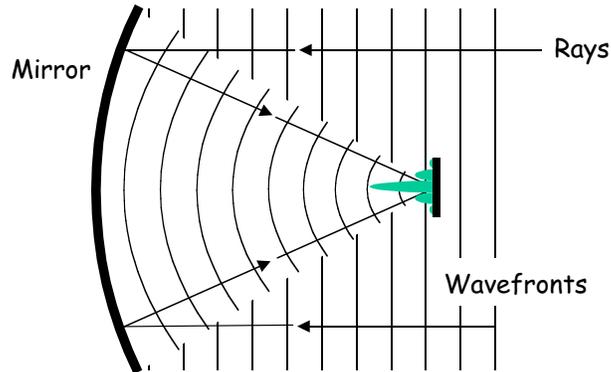
Wavefronts

- A perfect optical system converts incoming wavefronts to concentric spherical wavefronts converging to a point image on a detector
- Imperfections arise from fabrication error, temperature changes, alignment shifts, strain relief, long-term dimensional change
  - Traditionally minimized using massive structures
- Wavefront control uses moving and deforming elements to compensate imperfections after launch
  - Replaces massive structures with computers and actuators

Principle of the wavefront control approach for NGST

Fringes
where $\lambda = 2m\delta$

prism — F
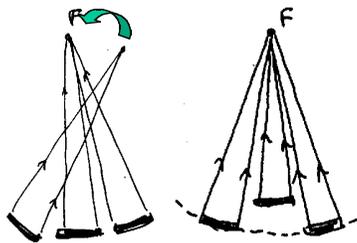
1. COALIGNMENT AND COFOCUSING

- Aligned to the accuracy of a single element telescope.
- Primary mirror piston ~ 5λ (10 microns) (limited by depth of focus of individual segment)
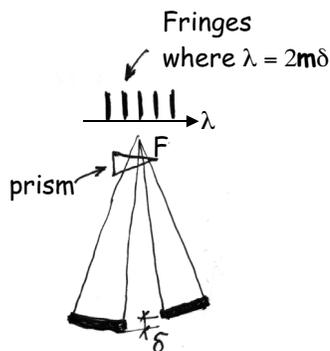
2. COARSE PHASING

Dispersed fringe sensing

WF error < λ
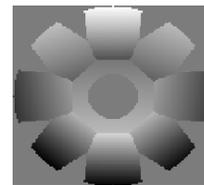
3. FINE PHASING

Phase retrieval
- WF measurement error < $\lambda/100$

- WF control error < $\lambda/20$
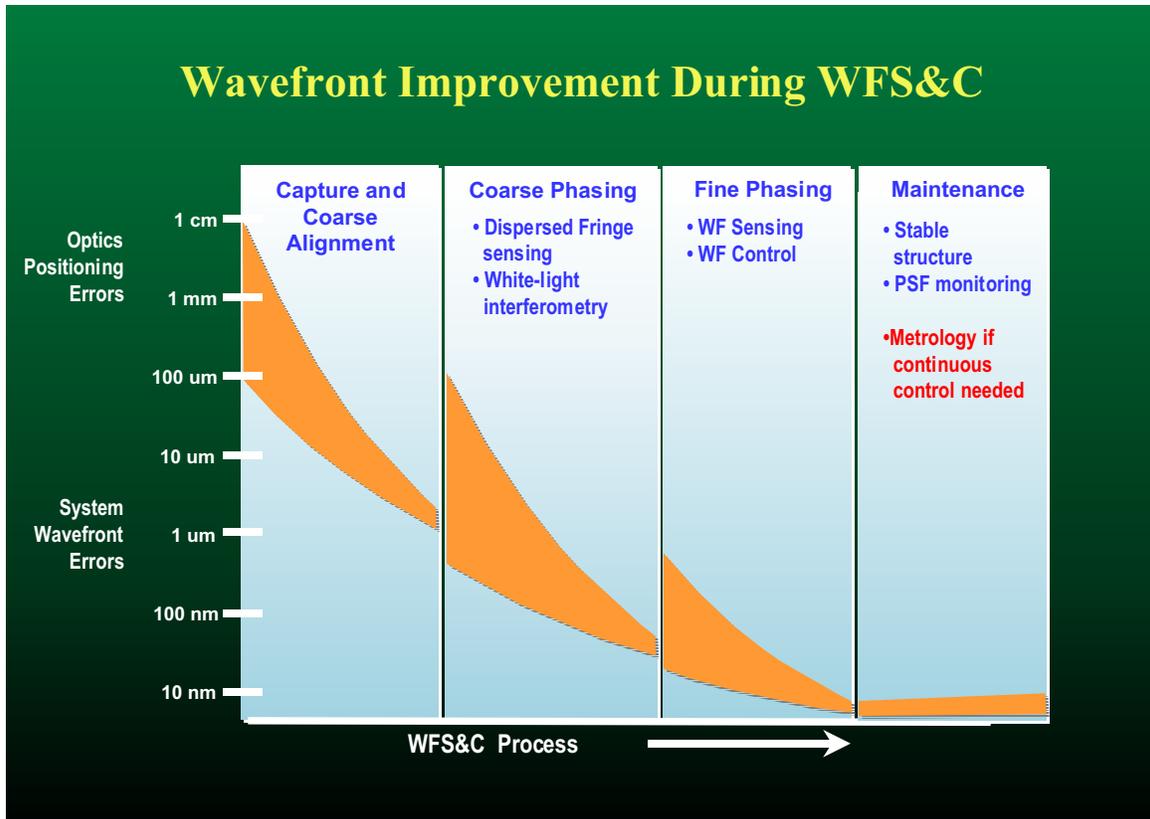
Coarse Phasing

WCT-2 Testbed Layout

Telescope Simulator

3–Segment PM

Segmented Mirror Cluster

OAP

Camera

Filter Wheel (Pupil Lens)

Stage CCD

Flip-in DFS

B/S

AO DM

FSM

DM

OAP2

Aberrator DM

Selector mirror

Filter Wheels Flip-in Optics

B/S

OAP1

White Light

5μm Pinhole

Laser - λ.633nm

Fold

Rotating Polarizer

Source Module

OAP

Segmented Aperture Compared with DM

AO/DM
Φ = 125 mm

WCT-2 Aperture
Φ = 38 mm

Dispersed Fringe Sensor (DFS)

δL

L

x

Dispersion

Phase

Fringe

x

λ Increases

Dark Band

DFS Fringe

- DFS Modeled Fringe: DCATT Model
- Wavelength range: λ = 624.8 - 658.8 nm
- Piston error: δL = 7 μm

WCT-2 Demo: DFS Analysis (Segs #2 & #3)
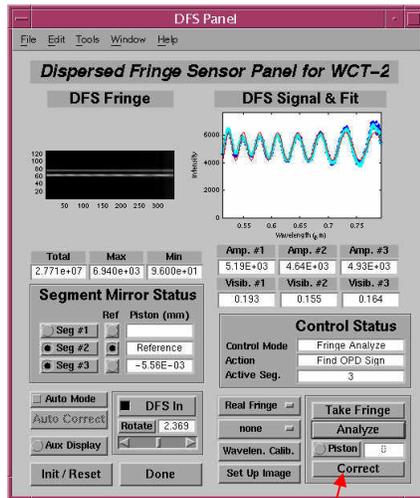
- Processed DFS image (Seg #1 tilted out)
- Processed DFS fringe from Seg #2 and #3 (dotted lines).
- DFS fitted curve (solid lines)
  - Fringe period determines piston magnitude
  - Relative phase between sidelobe traces determines the sign (up or down) of the piston
  - DFS analysis result:
    Relative Seg #3 Piston = -5.56 μm

Push here to implement piston correction

WCT-2 Demo: After Correction (Segs #2 & #3)
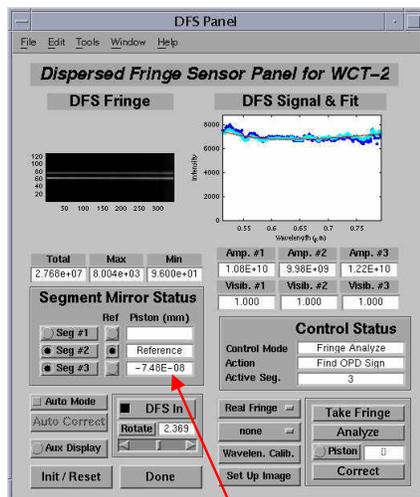
- Processed fringes after implementing correction show very little modulation
  - Modulation goes to 0 when segments are phased
  - Control has achieved sub-λ residual piston error

Detected piston reduced to near zero

White Light Interferometry (WLI) on WCT-2

- **WLI scans one segment in piston, using relative to another segment or segments, while observing in broad-band light**
- **Peak occurs when segments are phased, secondary peaks occur as** $\delta = 2n\lambda_{center}$

- **Plots show the WLI signal and its fit before and after the correction of Seg #1 relative to Seg #3 (reference), while Seg #2 is tilted away**
- **Piston error before correction was** -0.68 μm
- **After correction WLI detected** 0.0 μm **piston**
- **Independent check using IPO measured** 13 nm **residual piston**

- **WLI detection error is dominated by** 20 nm **scan step size**
- **Other error sources include image under-sampling, lab seeing, and jitter**

WLI PSF: Un-phased and Phased

WLI signal and Fit: Befroe Correction

Data
Fit

Normalized Intensity

Piston Scan (μm)

WLI signal and Fit: After Correction

Data
Fit

Normalized Intensity

Piston Scan (μm)
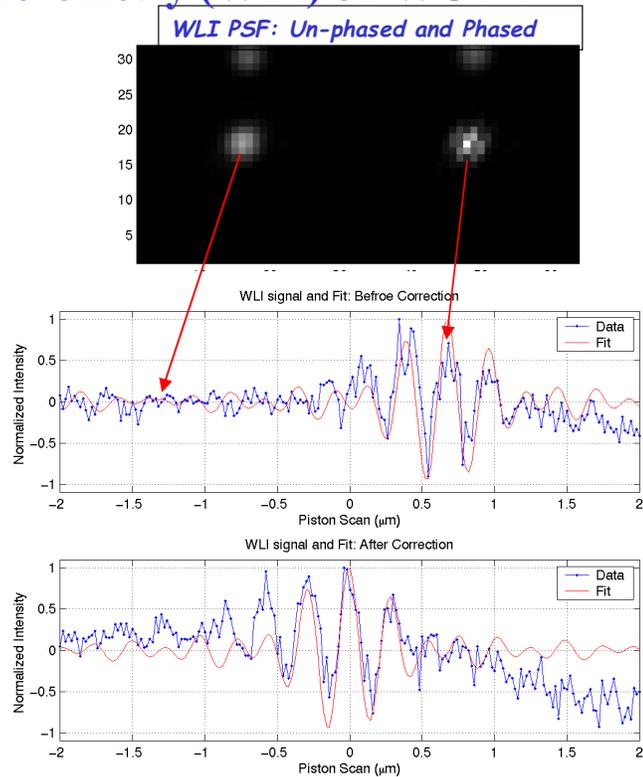
Coarse Alignment and Coarse Phasing Summary

- WCT-2 performance exceeds requirements and expectations
- Coarse alignment segment capture exceeds expected misalignments
  - Focusing algorithm -- camera stage test:
  - Before: > 10 mm (limited by camera FOV chosen)
  - After: < depth of focus (± 0.35 mm @ 633 nm)
- Segment mirror tilt errors:
  - Before: > 0.6 mrad (limited by segment actuator stroke)
  - After: < 4 μrad (limited by the jitter & seeing)
- Segment mirror piston errors:
  - Before: ~ 10 - 20 μm (limited by segment actuator stroke)
  - After: < 0.1 μm (DFS) (confirmed by PR and IPO)
          < 0.02 μm (WLI) (confirmed by PR and IPO)

Fine Phasing
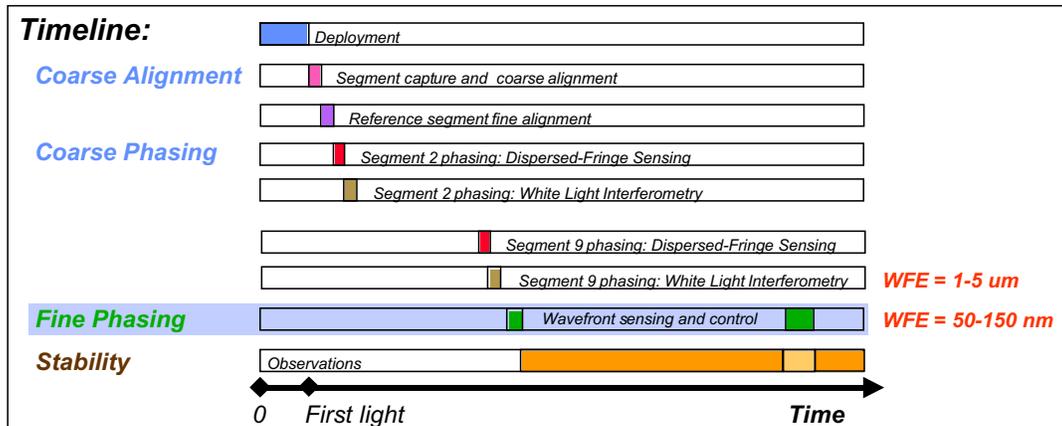
WF Sensing Using Images
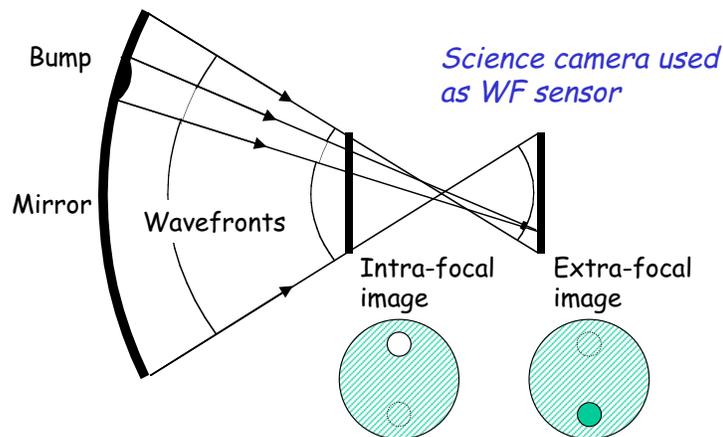
- A bump on the mirror surface shifts the focus of a patch of the beam
- This shows up as a bright spot on one side of focus and a dark spot on the other
- Computer processing of multiple defocussed images correlates the intensity variations in each, derives common WF phase map
- This phase map is then used to compute new control settings

Modified Gerchberg-Saxton Algorithm
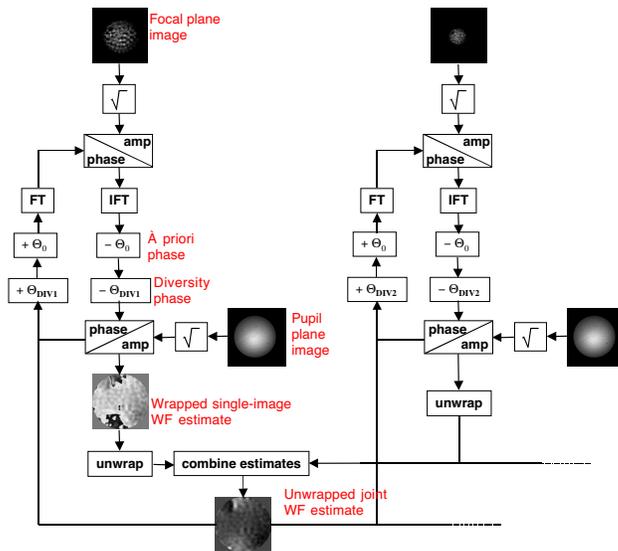
- Uses pupil image data to halve number of unknowns
- Uses defocussed images to improve visibility of aberrations
  - Reduces contrast between low, high-f effects
  - Reduces impact of jitter, other blurring
- Subtracts known phase (Θ_0, Θ_{DIV}) from the iteration to reduce iteration dynamic range

- Multiple images overdetermine solution to ensure uniqueness
  - Provides more data without introducing new unknowns
- Phase unwrapping allows estimation of WFE > λ
  - Joint unwrapping improves unwrapping robustness

Example from NGST WCT-1 Testbed
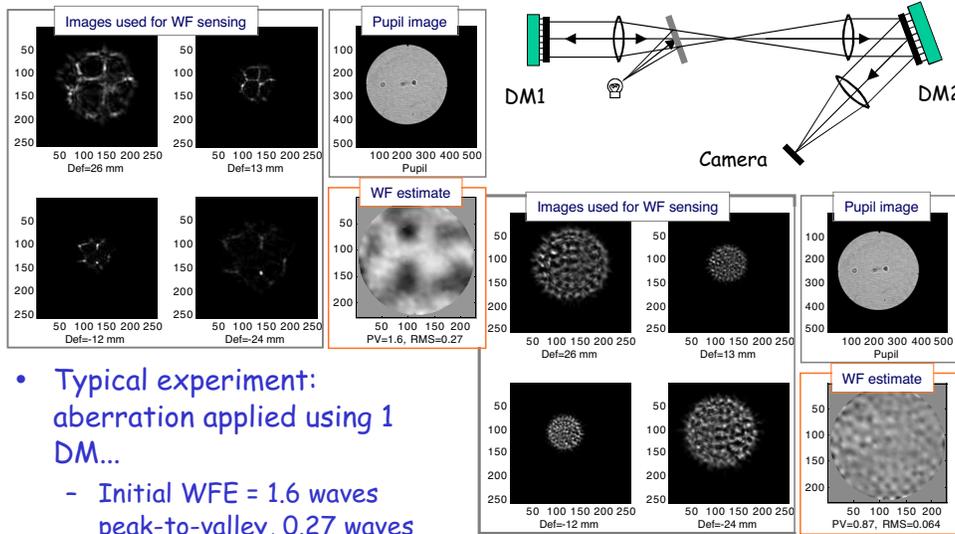
- Typical experiment: aberration applied using 1 DM...
  - Initial WFE = 1.6 waves peak-to-valley, 0.27 waves RMS

- ... and corrected using second DM
  - After control WFE = 0.87 waves PV, 0.064 waves RMS

Example from NGST WCT-2 Testbed
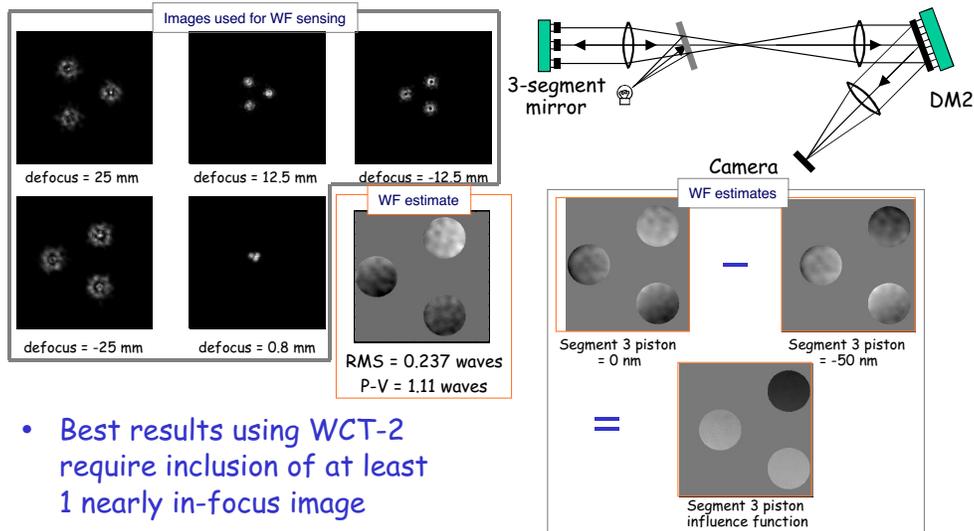
- Best results using WCT-2 require inclusion of at least 1 nearly in-focus image

- Segment piston influence function determined by "poking" actuators and subtracting WFs

High Dynamic Range Retrieval & Control Examples

- WCT-1 Testbed Results

- High aberration injected at testbed SMDM

- Phase retrieved & unwrapped at 632.8 nm

- Phase controlled at testbed AODM

SIRTF Brutus Test

Defocussed Images, Field Pt. 1
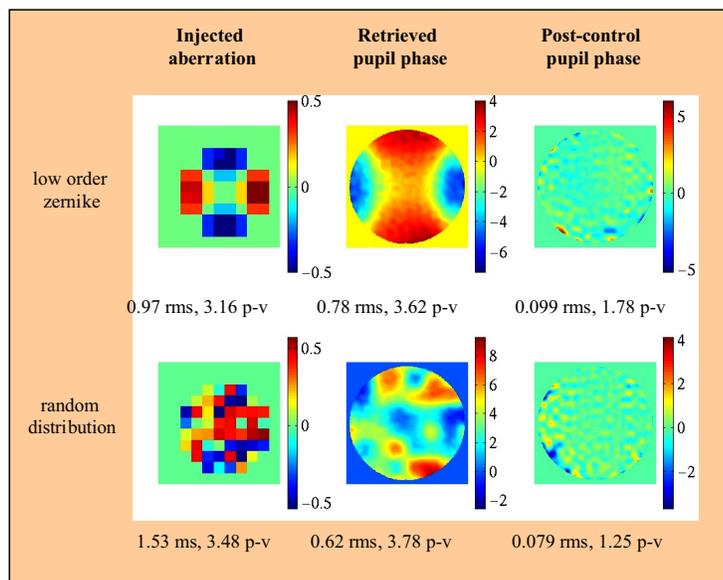
Telescope set up
for vibe test
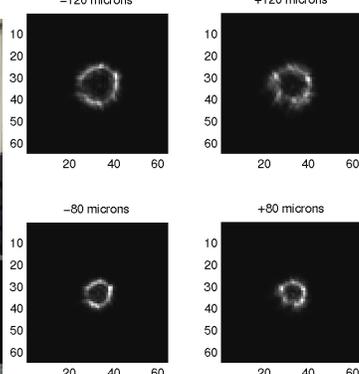
Wavefront Change with Field Angle

- SIRTF telescope end-to-end cryogenic optical test
  - Internal light source and autocollimating flat provide for in-chamber test capability
- Image-based WF sensing successfully measured WF at multiple field points
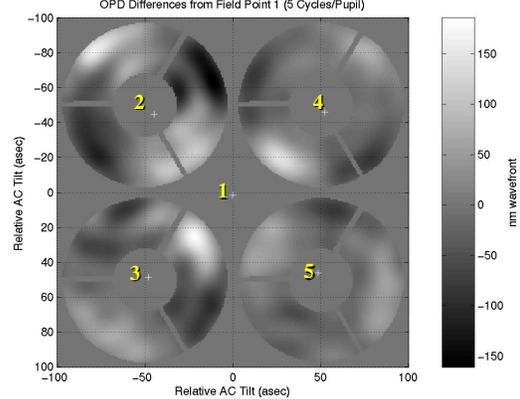
Mars Observer Camera Example

Image M1699: Actual and Simulated Images
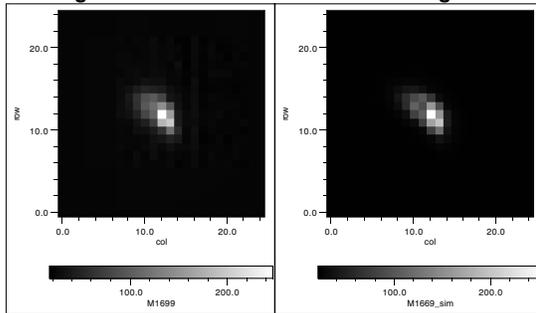
Actual image   *Simulated image*

– Image-specific parameter results
  » **Focus: 1 Watt hub heater yields 0.1814 um** » **Intensity: 2.961 (unitless)**
  » **Field: 0.4485 mrad** » **Background: 20.828 (DN)**

Image M1535: Actual and Simulated Images
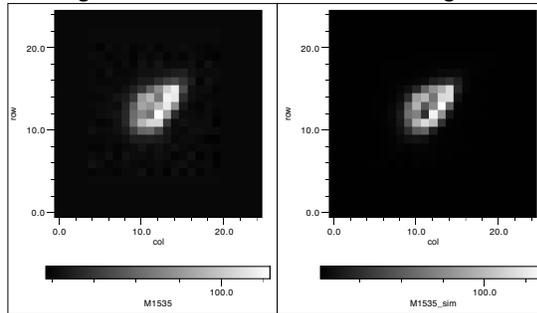
Actual image   *Simulated image*

– Image-specific parameter results
  » **Focus: 3 Watt rim heater yields -0.2908 um** » **Intensity: 2.8958 (unitless)**
  » **Field: 1.806 mrad** » **Background: 20.498 (DN)**

- Diagnostic data taken en route to Mars
- Illustrates prescription retrieval with relatively low resolution, low SNR data
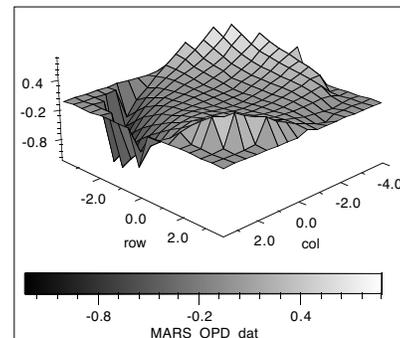
Image-Based WF Sensing Heritage
Includes Hubble Space Telescope
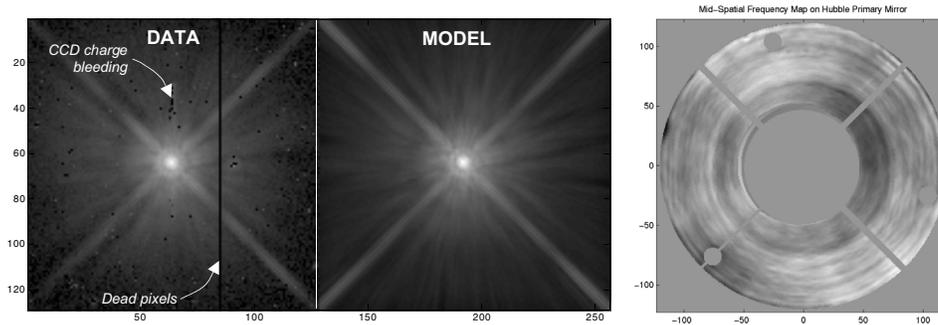
- "DATA" image is a composite of three exposures of star Feige 23, taken Oct 24, 1998 as a calibration image, displayed with a log10 stretch
  - Exposure times of 1, 4, 100 seconds provide high dynamic range
  - Taken with PC1 camera, F606W filter
- "MODEL" image is computed using HST model incorporating retrieved high-resolution mirror map
  - Map estimated using WF sensing operating on archival data
  - Model further optimized to match this image using prescription retrieval
  - Ref. J. Hutchings, D. Frenette, R. Hanisch, J. Mo, P. Dumont, D. Redding, S. Neff, "Imaging of z~2 QSO host galaxies with the HST," Astron. J., in press.

Phase Retrieval Camera (PRC)
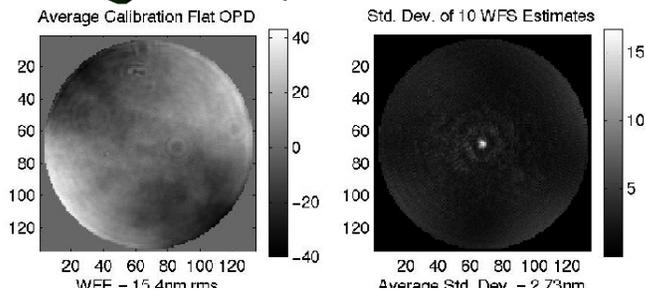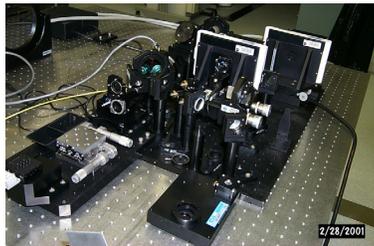
- An imaging camera for WF sensing
  - Provides a "portable WFC testbed" for use with NMSD, AMSD, other optics
  - Provides large optics WFC experience before NAR
- Calibrated performance (using internal calibration flat on flip stage)
  - 15.4 nm static WF aberration
  - 1/325 wave repeatability for low spatial frequency WF errors
  - Performance verified by comparison with Zygo interferometer

Cryogenic Deformable Mirror Technology
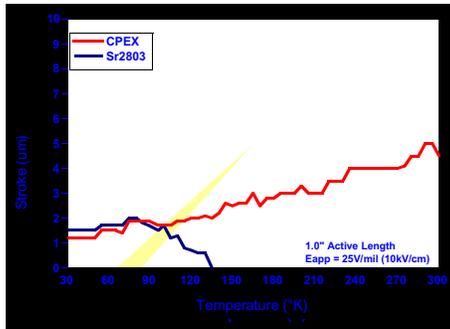
Cryoceramic materials for DM actuators meet stroke objectives over NGST operational temperature range

Modular actuator array with integrated electrical connections provides better performance, lower risk for cryo DM

Modular Cryo DM

349 Actuator Cryo DM
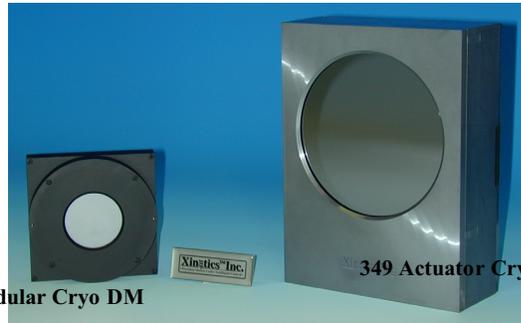
- Deformable Mirror (DM) enables correction of primary mirror aberrations with a small optic elsewhere in the beam train
- Xinetics Inc. successfully developed Cryogenic DM technology and demonstrator mirrors meeting NGST requirements under a SBIR III
  - Developed an electrostrictive electroceramic for 35-65°K
  - Developed a piezoelectric electroceramic for 35-375°K
  - Completed 2 349-channel cryogenic DMs
  - Developed modular cryo DM technology and completed demonstrator DM
  - Successfully thermal cycled DMs to 35°K and demonstrated DM actuation at 50°K
- NGST contractors chose not to baseline a DM; however, technology is now available if actuated primary mirror does not meet performance requirements

Maintenance: PSF Monitoring and/or Metrology

404
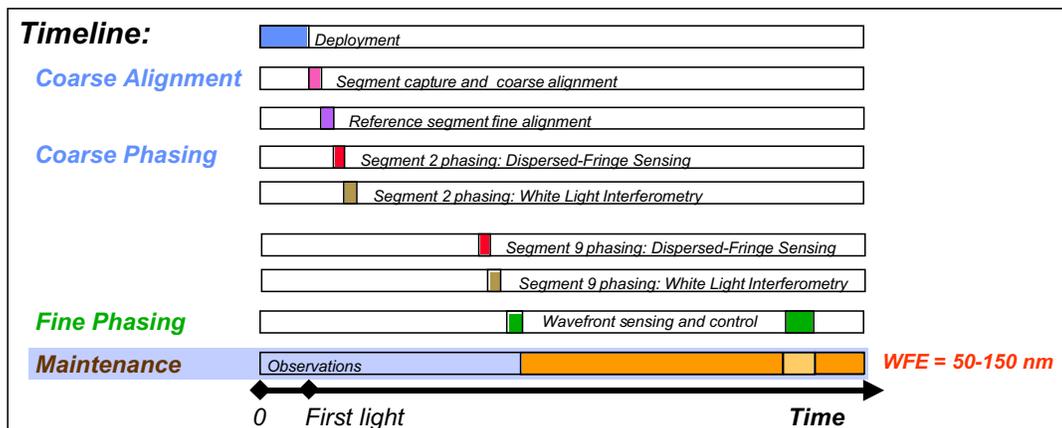
PSF Monitoring During Science
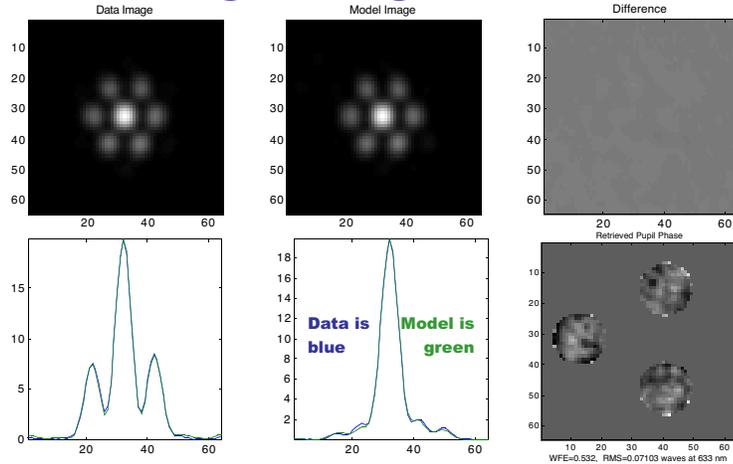
- "Infocus PSF Optimizer" (IPO) estimates WF from infocus imagery
- Experiments using WCT-2 show robust, accurate low spatial-frequency WF control
  - Measure and control tip-tilt-piston for 3 segments
  - Sensing range of $\lambda/2$, accuracy of $\lambda/100$ demonstrated
- More complex 9- and 36-segment NGST apertures being studied in simulation

Piston Accuracy with PSF Magnifier (λ=900 nm)
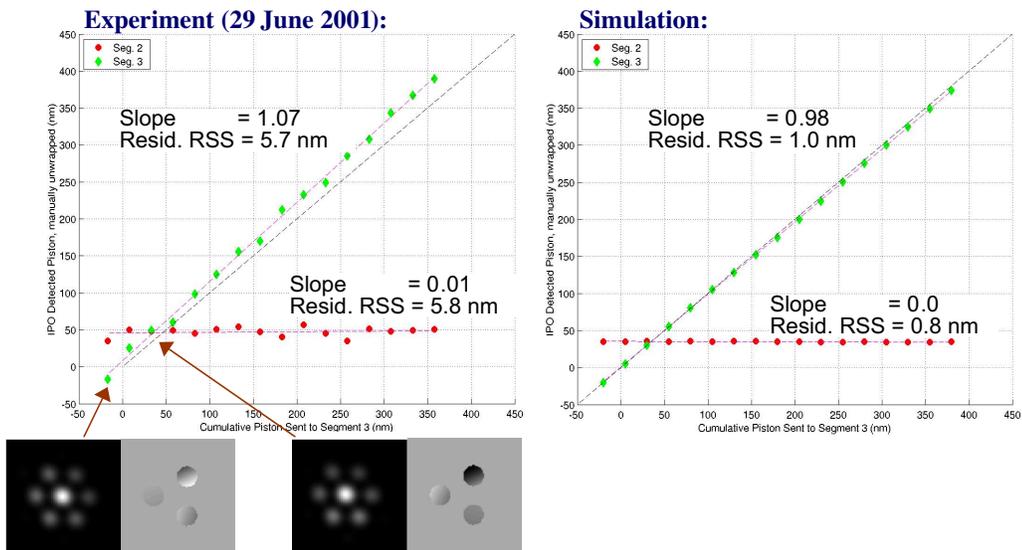
- Piston Seg. 3 in steps of +25 nm. Detected piston was manually unwrapped after 225 nm ($\lambda/4$).
- Residual errors show ~6 nm piston detection uncertainty (RSS), which is on the same order as the 5 nm PZT accuracy.

405

Control of Segmented PM using Metrology
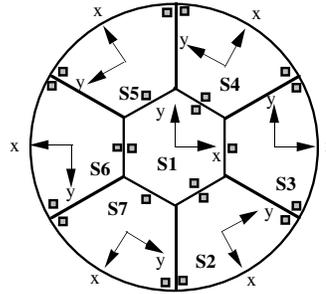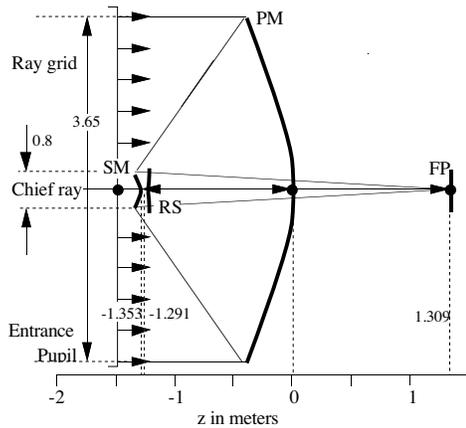
Does Sensor Configuration Observe Wavefront Errors?

Configuration A	Configuration B	Configuration C	Configuration D	Configuration E	Configuration F
Edge sensors only	Edge sensors only	Edge and gap sensors	Edge and gap sensors	Edge and gap sensors	Optical truss
NO	**NO**	**NO**	**Yes**	**Yes**	**Yes**
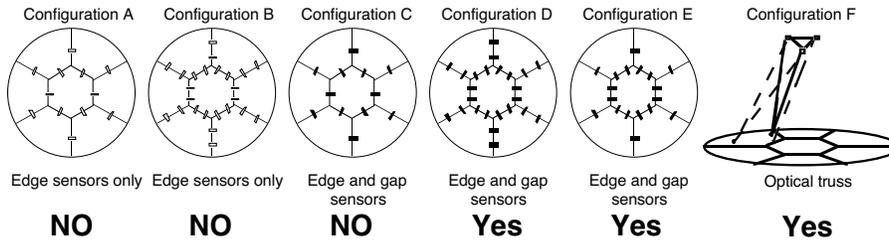
Laser Truss Application of SIM Metrology

Laser gauges form "optical truss," measure optical configuration from PM to detector

Instrument package

Multichannel beam launcher derived from SIM COPHI technology

PM retro-reflectors, e.g.
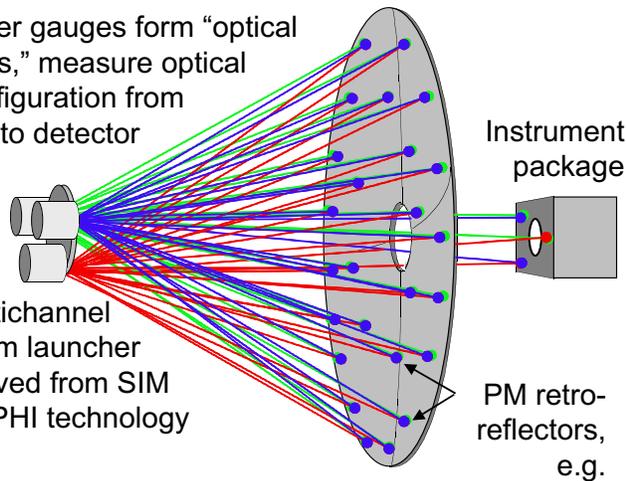
Laser truss can fully measure motions that affect optical quality

3 retroreflectors per segment for rigid-body motions, more for flex deformations

Absolute mode for large dynamic range, low accuracy

Relative mode for high accuracy

Optical fiber couples laser on warm side to cold beam launcher

Other configurations may be useful (backside of PM, for instance)

COPHI demonstrated performance in relative mode:

Sensing bandwidth	1Hz	100 Hz	10 kHz
Resolution	0.005 nm	0.02 nm	0.1 nm
Thermal Stability	Better than 1nm/K		

406

WFC Technology Heritage Matrix

Approximate Technology Readiness Levels (TRLs) for Cryo SubMM/FIR missions

Control Mode	NGST	PSR	SIM	SIRTF	Other
Capture	TRL4	TRL2			
Coarse Alignment	TRL4	TRL2			
Coarse Phasing					
White-Light Interferometry (WLI)	TRL4				TRL2?
Dispersed-Fringe Sensor (DFS)	TRL4				
Keck Phasing Camera					TRL2?
Fine Phasing					
Image-based WFS	TRL4			TRL6	
Dedicated WFS					TRL2?
Maintenance					
Edge sensing systems		TRL2			???
Laser truss metrology		TRL2	TRL4		
Cryogenic Components					
Segment actuators	TRL5				
DM actuators	TRL4				
Cryo edge/gap sensors		TRL2			???
DART membrane actuators					TRL2?

Conclusion

- NGST WFS&C system will be developed to TRL6 over the next few years
  - Tested in cryo-vac environment with traceable hardware

- SubMM problem is probably easier
  - Scales with wavelength
  - Likely to have a more stable environment
  - Detectors may present small complications

- New system approaches need a new look
  - DART membrane mirrors require a new actuation approach
  - Other mirror and actuation technologies may prove useful as well

Cryogenic Delay Line for Long-Baseline Interferometry in the Far-Infrared

P.R. Lawson, M.R. Swain, P.J. Dumont, J.D. Moore, and R.F. Smythe

Jet Propulsion Laboratory
MS 306-388
4800 Oak Grove Drive
Pasadena, CA 91109
lawson@huey.jpl.nasa.gov

Abstract: We discuss the design, current status, and ongoing development of a cryogenic delay line for long-baseline direct-detection interferometry in the far-infrared.

1   Introduction

Direct-detection interferometry in space at far-infrared wavelengths will ultimately allow high angular resolution studies of galaxy and star formation and overcome the confusion due to the extra-galactic background, opening up an entirely new domain of astrophysical research (cf. Rieke *et al.* 1999). In support of these science goals, we are now at JPL developing new technologies needed to make long-baseline far-infrared interferometry possible. The design of a direct-detection interferometer for the far-IR is extremely challenging. For it to be background-limited in space in the region of 40-400 microns, its optics and servo-mechanisms must operate at near liquid helium temperatures. Our current efforts are focused on the commissioning of a precision cryogenic delay line and integrated optics beam combiner. The preliminary tests of our prototype delay line, at a temperature of 120 K, have shown the design to be fundamentally sound. We are in the process of implementing the servo control system for the delay line and have completed the fabrication and assembly of a novel cryogenic mechanical amplifier for phase-measurement interferometry. The current status and ongoing development of our program are described.

2   Overview of the Cryogenic Delay Line

Our objective has been to design a precision delay line that does not use axles, is capable of accepting 10 cm optical beams, and which provides 50 cm of optical delay while operating in hard vacuum at temperatures as low as 4 Kelvin. We choose a design wavelength of 100 μm and adopted a cat's eye optical design to facilitate path compensation with a high-speed servo. Our requirement that the fringe visibility loss be less than 1.0% is achieved if the straightness of travel is 250 μm over the full stroke and the piston jitter is less than 1.2 μm rms at a wavelength of 100 μm. As a goal
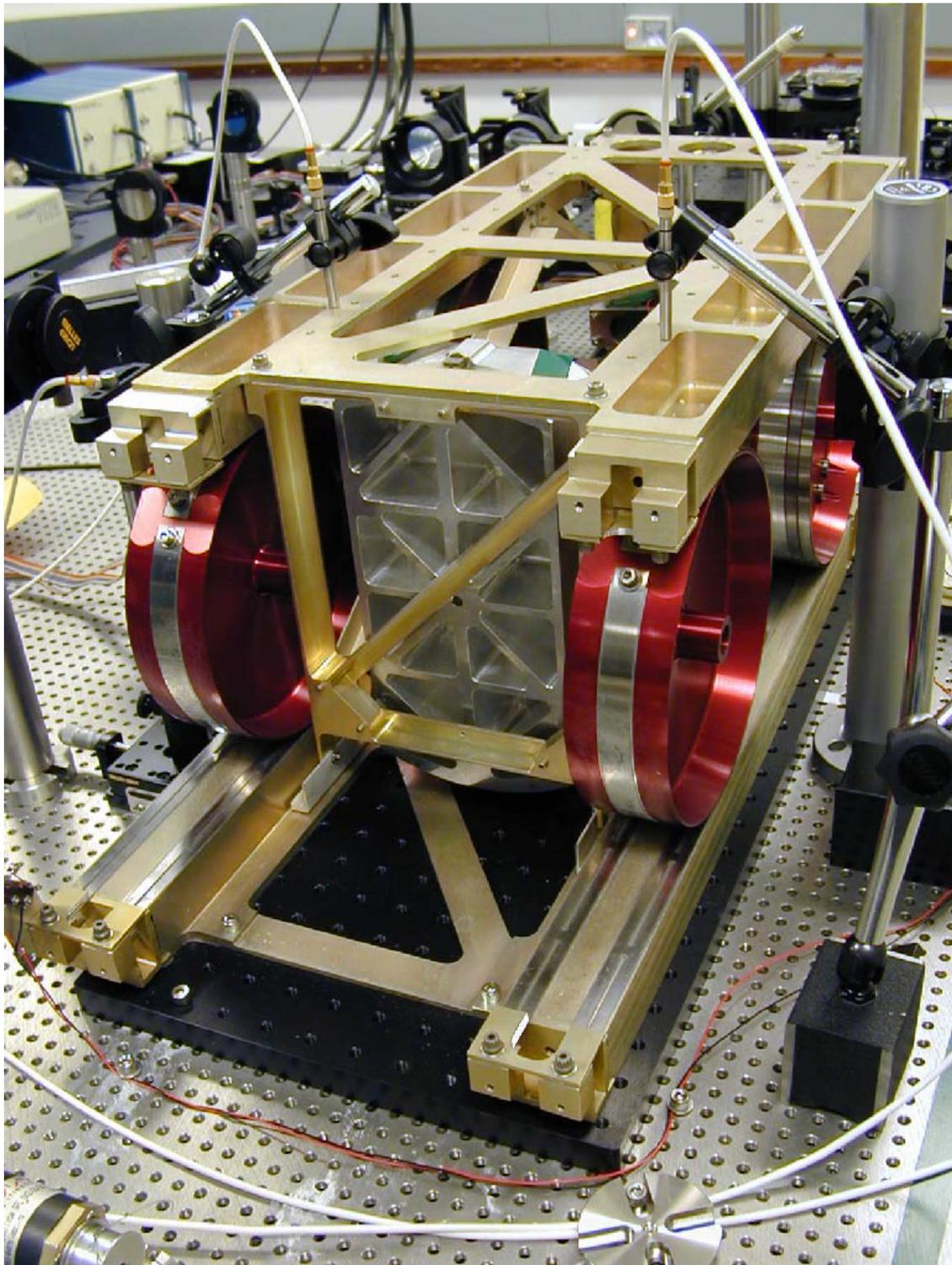
Figure 1: Cryogenic delay line with wheels constrained by straps. The upper platform supports the optics and rolls along the top of the wheels. The back of the lightweighted primary mirror is also visible.
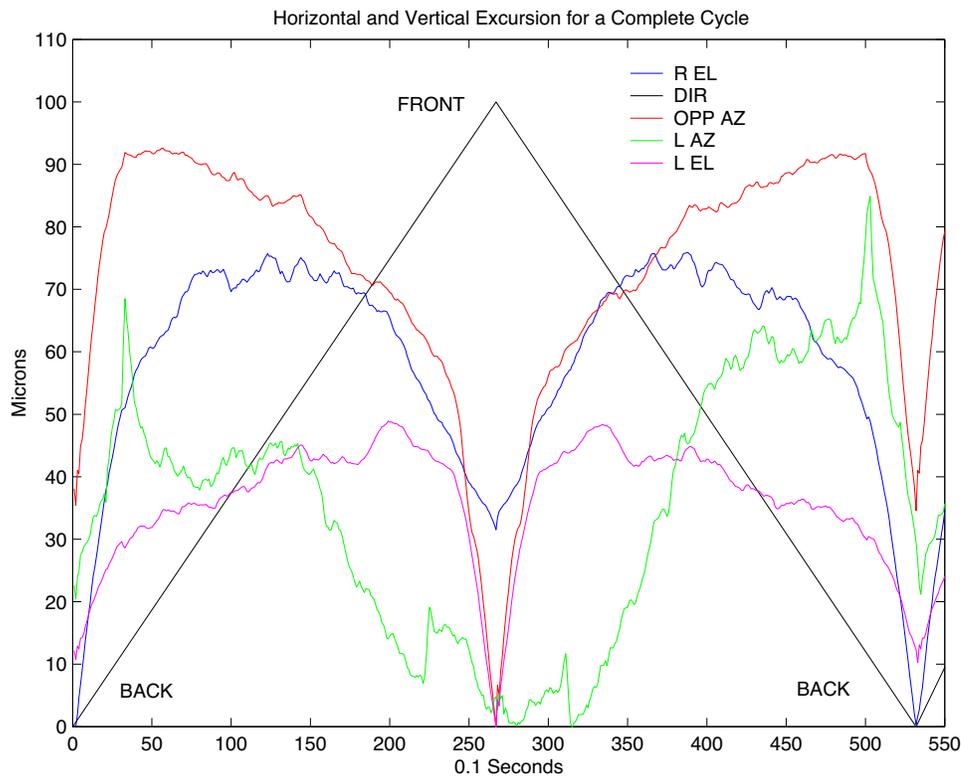
Figure 2: Performance curves for forward and backwards scans of the delay line. The amplitude of the motions are approximately 40 microns, with the carriage being lifted noticeably when brought fully front or fully back.

we would like to demonstrate the operation of a cryogenic delay line with pathlength stability at the 5–10 nm, in support of the Terrestrial Planet Finder program.

Our initial studies focused on a "double porch swing" design using flexure pivots and developed by Donald E. Jennings at the Goddard Space Flight Center. The limited angular travel of flexure pivots motivated us to explore other architectures, and ultimately led us to an entirely new design shown in the photograph of Fig. 1, and reported previously by Lawson *et al.* (2000) and Swain *et al.* (2002). The principle of operation is simple: an upper carriage carrying the optics sits on four wheels that roll to translate the delay line. There are no axles in this system; the wheels are constrained only by metal straps and a magnetic preload. The carriage has been tuned for a run-out of 25 μm over its full travel—a factor of 10 better than our requirement. Figure 2 shows plots of performance tests at room temperature, showing misalignments with an amplitude of 40 μm. Lab tests have shown that the alignment is maintained well within tolerances over 1000 cycles (16 hrs of cycling).
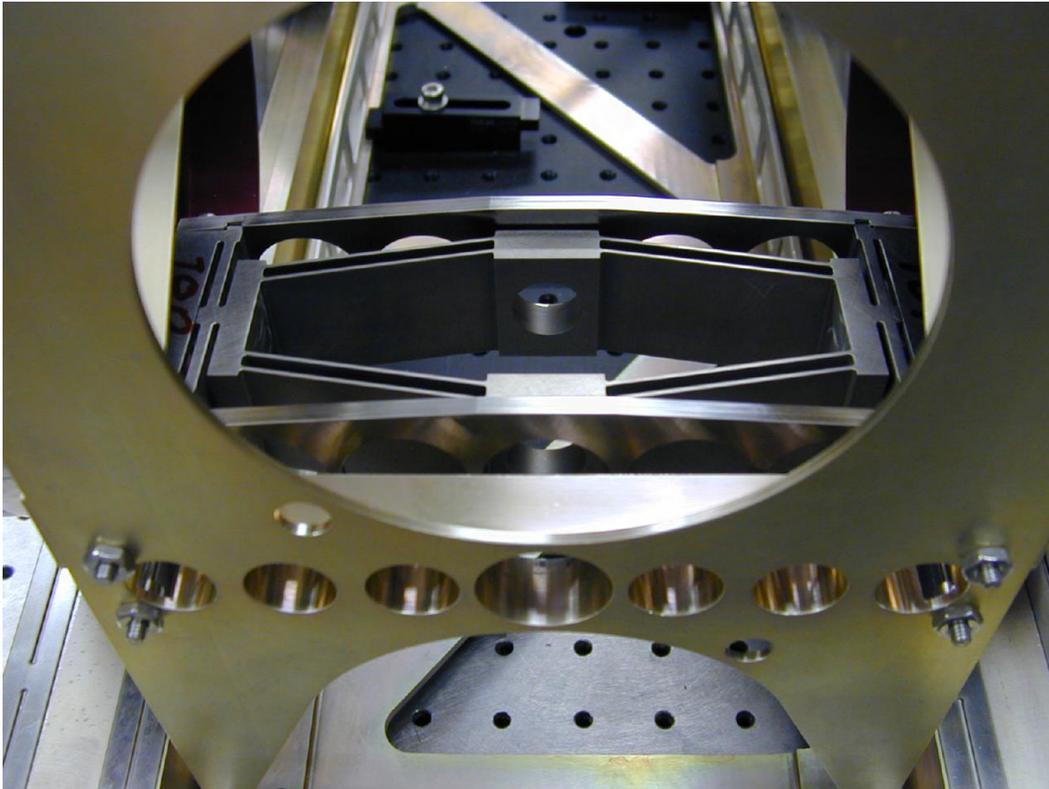
Figure 3: View of the front of the delay line looking into the mechanical amplifier for the piezo stack. When installed, the stack sits inside the parallelogram and mirrors, shown in Fig. 4, are located fore and aft.

3 Novel Mechanical Amplifier

The ability to measure fringe phase is key to many aspects of astronomical interferometry, and piezo transducers are ideal modulators for most methods of phase measurement, but primarily at visible wavelengths where the stroke lengths are relatively short. At infrared wavelengths and operating at cryogenic temperatures piezos are ill-suited as modulators, because their throw is reduced by as much as a factor of two, and even a wavelength or two of modulation is beyond their capability. The largest commercially available piezo stacks are about 5 inches in length and have a throw of about 180 microns at room temperature and only about 90 microns at 4 Kelvin.

To compensate for the loss of travel at cryogenic temperatures, the PZT is mounted in a mechanical amplifier, shown in Figs. 3 and 4, that supports one of the mirrors of the interferometer. The device was designed based on an original concept from Rob Calvet at JPL dating from 1993. The mechanical amplifier resembles an elongated parallelogram with pairs of parallel flexures along each side. The piezo transducer is
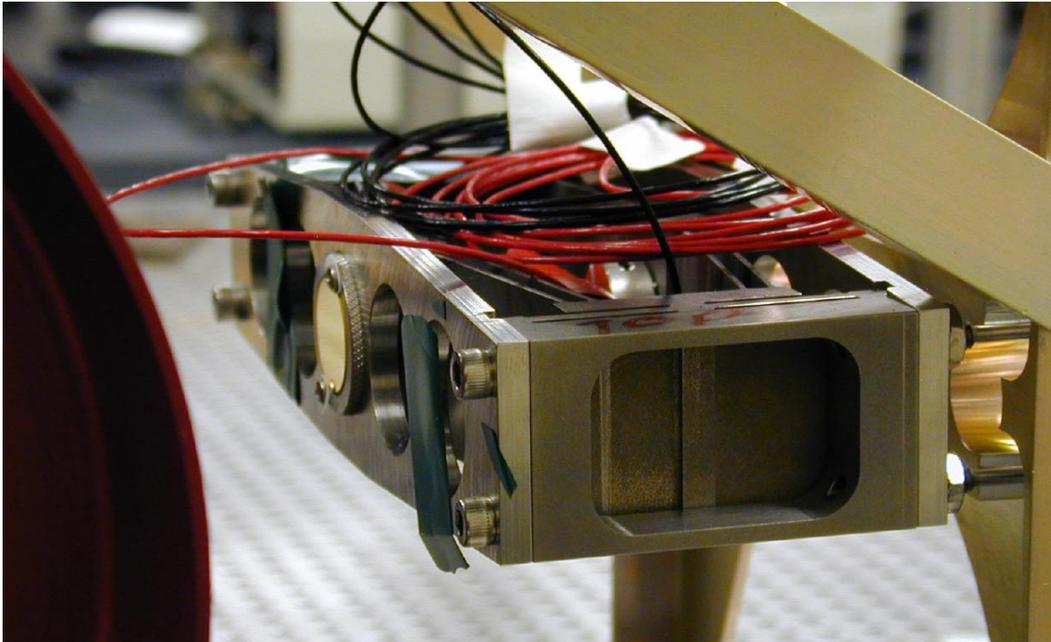
Figure 4: Side view of the mechanical amplifier with the cat's eye secondary mirror now installed.

compressed along the axis of the long diagonal of the parallelogram by support flexures at each end. The expansion of the piezo along the long diagonal causes the ends of the short diagonal to move towards each other with a motion amplified by a factor of 3 or 4. The parallel flexures are used to eliminate unwanted twisting and vibration modes so the small mirror will not tilt when translated by the amplifier. The support flexures that hold the piezo allow a symmetrical expansion of the piezo within the amplifier. The design is completely symmetric and balanced such that inertial forces are nulled. This provides mechanical stability that allows rapid (100 Hz) sampling without inducing vibrations. Optical interferometers normally obtain the mechanical stability and momentum compensation by using an additional piezo stack mounted back-to-back with the first piezo so that the second one has motions that are equal but opposite in direction. By mounting the stack symmetrically with the support flexures the stack expands equally about its center, does not induce vibrations, is more compact, and does not require momentum compensation. The device is made of titanium and was machined using a Wire EDM process so as to be as strong and lightweight as possible.

4   Preliminary Cryogenic Testing

In August 2001 we conducted preliminary tests of the delay line at liquid nitrogen temperatures with the aid of a 150 quart Coleman Marine Cooler. The temperature

Figure 5: Preparations for cryogenic tests at liquid nitrogen temperatures.

of the base of the delay line was held at 77 K, which by conduction and after two hours of cooling brought the top of the delay line to 120 K. The delay line was cycled back and forth by a stepper motor for about two hours, operated well, and moved smoothly throughout the testing. Encouraged by these results, we are now in the process of designing a custom-made dewar to be used for future tests. The dewar will be designed with space for a beam combiner, now also under development.

5   Current Status and Future Plans

Initial tests of the delay line have shown the design to be fundamentally sound and operating well within requirements. Further cryogenic tests will be undertaken when the new dewar is completed. Work is now also underway to implement the pathlength servo controls and to test the new mechanical amplifier. Plans are being made to integrate the delay line in a far-infrared interferometer testbed.
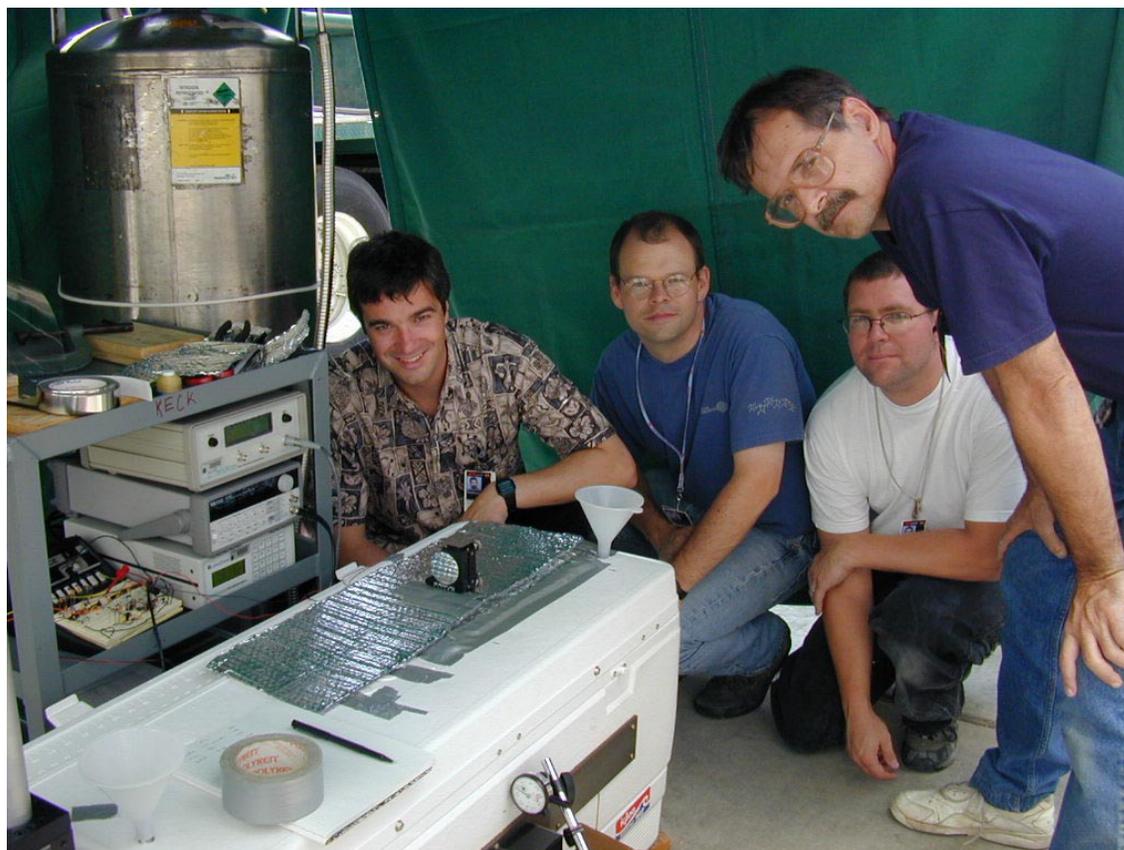
Figure 6: Mark Swain, Peter Lawson, Robert Smythe and Jim Moore.

References

Lawson, P.R., et al. 2000, Bull. Am. Astron. Soc. 32, 1426

Rieke, G.H., et al. 1999, http://mips.as.arizona.edu/MIPS/fircase3.pdf

M.R. Swain *et al.* 2002, Proc. 36th Liège International Astrophys. Colloq.

Acknowledgements

THE DART SYSTEM FOR FAR-IR/SUBMILLIMETER SPACE MISSIONS

MARK DRAGOVAN FOR THE DART GROUP[*]

1. Introduction. The current generation of telescopes, both ground and space based, can trace their design and fabrication methods back to the original telescopes of the 17th century[1]. There is no *a priori* reason that a space telescope should look anything like a ground based one. In the space environment the mechanical elements of the telescope are in free fall and hence do not feel the effects of gravity (other than tidal forces, which are too small to have any effect), so constraints imposed by gravity are nonexistent. Nor is there any reason that the traditional methods of fabrication, essentially the rubbing of two pieces of glass together with some abrasive grit in between, should be used to figure the optical surfaces used in a telescope. The overriding consideration is that the space telescope be large, lowmass, and diffraction limited over a reasonable field of view.

The DART[2] is a system of two cylindrical-parabolic reflectors. One reflector will produce a line focus; two reflectors, properly oriented, will produce a point focus. This system is ideally suited to using tensioned membranes for the reflective elements, and hence a lowmass telescope system. For FarIR/submillimeter missions the DART presents a compelling new telescope architecture that is scalable to large apertures, and with its large membrane area is well suited to passive cooling. Two other white papers are referencing the DART system as a possible telescope to enable each mission concept.

2. DART Optical Layout and Analysis. An intrinsic property of any surface is its Gaussian curvature. A surface with zero Gaussian curvature is either flat or has the shape of a trough, so that one of the principal curvatures is always zero. Such a surface can be formed by tensioning along only one axis. If the shape of the surface in the curved direction is a parabola, then a line focus results for an incident plane wave. To produce a point focus, a system of two trough-shaped reflectors properly oriented with respect to each other must be used. A perspective view of such a system is presented in figure 1.

In order for this system to focus and have a completely unobstructed aperture the focal lengths of the two individual reflectors must be unequal. The aberrations of the system are identical to those of an off-axis paraboloid with focal length f_1 in the direction which the first reflector focuses, and f_2 in the orthogonal direction, with the subscripts referring to the first or second reflector. For the specific system displayed in figure 1, the extent of the focal surface is ± 25 resolution elements[3] fully sampling the

---

[*] JPL: E.Cohen, J.Dooley, A.Chmielewski; Lockheed-Martin: J. Tolomeo, G. Cross, D. Tenerelli, P. Forney

[1] e.g. Gregory(1663), Newton(1668), and Cassegrain(1672)

[2] Dual (or Dragovan) Anamorphic Reflector Telescope; Astro-ph/0001241

[3] A resolution element is the *radius* of the Airy disk.
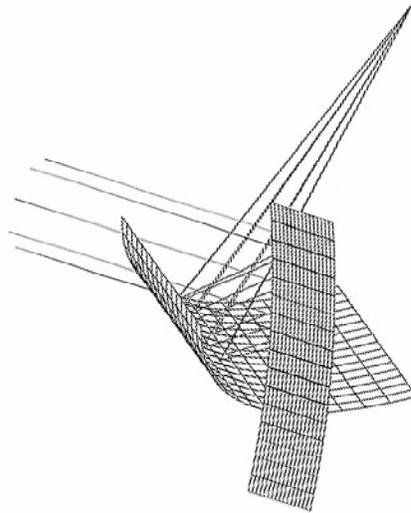
FIG. 1. The layout of a two mirror reflector system where the individual reflectors are parabolic cylinders. The orientation and curvatures of the individual reflectors are chosen so that a point focus results for an incident plane wave. The reflectors as illustrated are greatly oversized to emphasize the curvatures of each reflective element. It is clear by inspection that the system is completely unobstructed. What is less obvious is that the system has a reasonable field of view, as illustrated in figure 2.
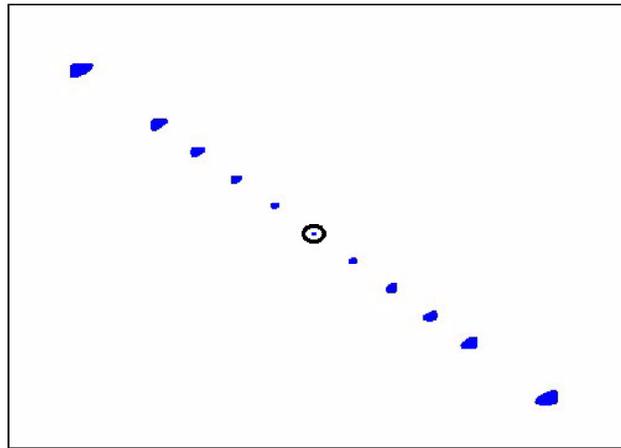
FIG. 2. *A geometrical optics ray tracing spot diagram for the DART system of figure 1. The Airy disk is illustrated by the ellipse at the center. The Strehl ratio is greater than 0.75 over the ± 25 resolution element field illustrated; thus most of the field is diffraction limited. The field of view scales as $f_1 \cdot f_2$.*

focal surface. The Airy disk is not circular, but has eccentricity

$$e = (1 - (f_1/f_2)^{-2})^{1/2}.$$

For the system shown in figure 1, $f_1/f_2 = 4/3$ or $e = 0.66$.

In figure 2 is displayed a geometrical ray trace of the $f4/3$ layout. The aberrations scale with the product of the individual focal lengths $f_1 \cdot f_2$ since the dominant aberration
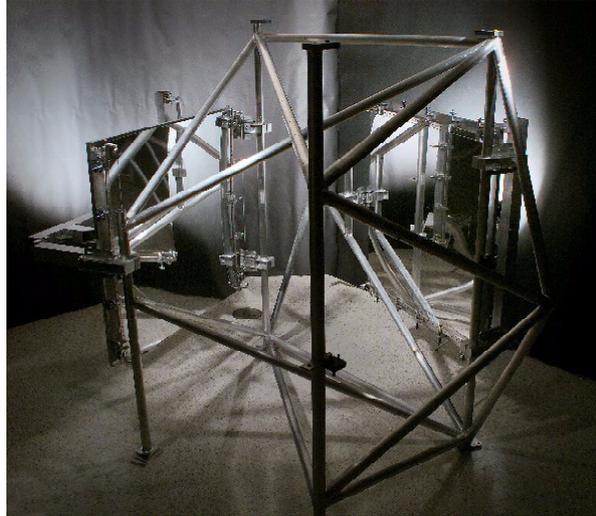
FIG. 3. *The two reflectors are mounted on a rigid truss structure. The reflector on the left has a six degree of freedom mount to allow for precision alignment of the system. A collimated beam enters from the left, hits the righthand mirror, continues on to the left most mirror, and exits the system on the right.*

is coma, similar to a traditional optical system where the comatic aberration scales as f^2.

The parabolic-cylindrical surfaces are formed by tensioning a reflective foil over a frame which has a parabolic contour along one axis and is rigid enough to support the tensioning. The alignment of the two reflectors is critical to the performance of the system. An arrangement of sixadjustable rigid struts connecting the two reflectors completely constrains all degrees of freedom while allowing the adjustment of the relative orientation of the two reflectors[4].

3. Physical Implementation of the DART System. A 1.2 m DART prototype/testbed was developed under the New Millennium Space Technology 6 study phase program. The working system (located at Lockheed-Martin in Sunnyvale, CA) is diffraction limited at 40 μm, and has a mass density of 7kg/m^2 for each individual reflector. Figures 3 and 4 illustrate the system, and demonstrate the imaging capability at $10 \mu m$. The constructed system was designed, built and tested in less than 8 months.

4. Discussion.

4.1. Shape Control. The shape of the reflective surface is determined by the tensioning of a membrane over a stiff boundary. The shape of the boundary is determined by the bending of a beam. By choosing the correct application of forces and moments at the edge of the beam a parabolic shape is obtained[5].

The membrane surface will have predominantly a cylindrical shape with a slight negative curvature due to the Poisson effect (the fact that a material shrinks a small

---

[4] Stewart, D. 1965 *Proc. Instn. Mech. Engrs.* **180**, 371.

[5] J. Tolomeo, ST6 Final Report, Aug. 2001.
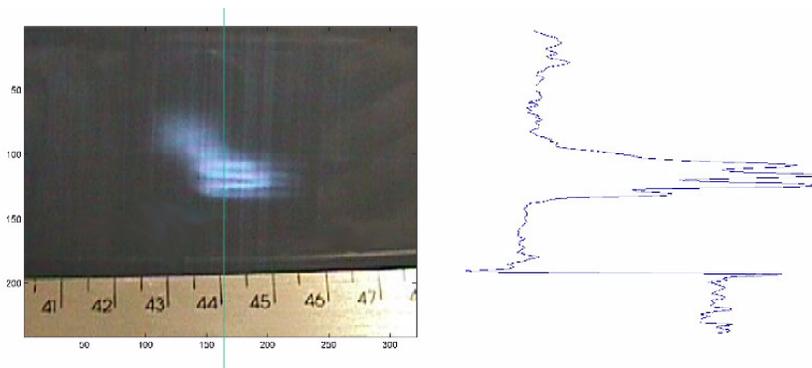
FIG. 4. *The image of a hot coil of wire as imaged through the DART system at 10μm. The intensity profile is displayed to the right of the image. The peaks are clearly evident. The ghost to the upper left is residual scattering from a slight misalignment of the collimator. The image is not sharp because the telescope is diffraction limited at 40μm.*

amount in the direction perpendicular to the applied force). Several methods are being investigated to minimize this effect; for the FarIR the magnitude of the effect is less than $\lambda/10$.

4.2. Scaling relations. Current technology millimetric telescopes have densities of order 10kg/m^2, a factor of $\sim 10^3$ between the mass of the reflecting layer and that of the support structure. For optical telescopes the situation is much worse where the current state-of-the art has density of order 150kg/m^2; the supporting substrate a factor of $\sim 10^6$ more massive than the reflecting layer.

The areal density of the reflecting layer is given by $\sigma_r = \rho_r t$ with t the thickness of the reflecting layer, and ρ_r the density. The thickness of the reflecting layer of a high electrical conductivity metallic film can be determined, to good approximation for a specific reflecting material, by considering the skin depth

$$\delta = 1/\sqrt{\pi c \mu \sigma_e / \lambda},$$

where σ_e is the conductivity of the reflecting surface, λ is the wavelength, and $\mu = 400\pi\text{nH/m}$. With a thickness of $t = 5\delta$, the surface is opaque and the wave is reflected with low loss[6]. For a very good conductor such as copper $\sigma_e = 5.7 \times 10^7 (\Omega\text{m})^{-1}$. In the case of optical light ($\lambda = 0.5\mu\text{m}$) the film only has to be $\sim 50\text{nm}$ thick to reflect the incident light; for millimeterwaves ($\lambda = 1000\mu\text{m}$) a $\sim 1\mu\text{m}$ thickness is required. This gives an areal density $\sigma_r \sim 8 \times 10^{-3}\text{kg/m}^2$ in distinct contrast to the areal density of the substrate material, which can be many orders of magnitude greater.

By examining existing telescopes one finds that the areal mass density of the supporting substrate (generally some form of glass) is $\sigma \propto d^\beta$, where d is the aperture diameter and $\beta \sim 0.5$. This is independent of the technology used, or the epoch when

---

[6] Goldsmith, P. *Quasioptical Systems*, IEEE Press, 1998.; Bock et al. *Applied Optics* **34**, 4812-16 (1995).
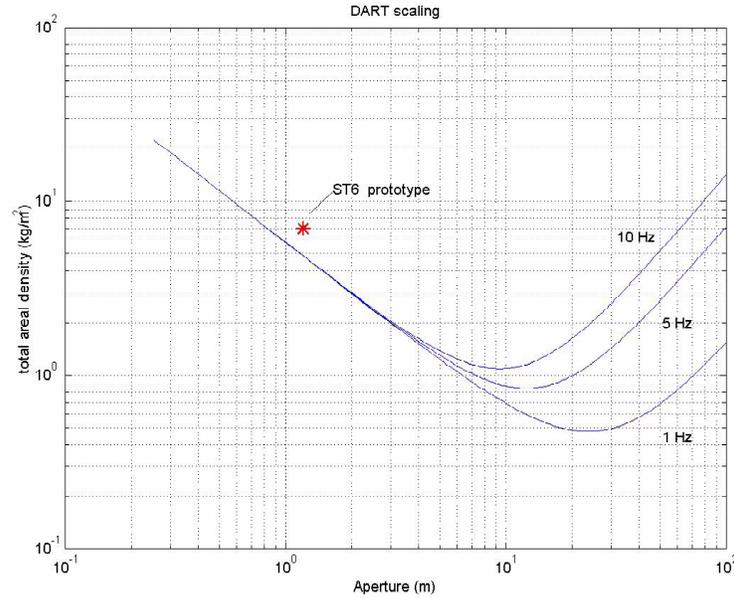
FIG. 5. *A scaling relationship for DART telescopes. The existing prototype has an areal mass density of 7kg/m², falling just above the theoretical line.*

the telescope was constructed. In comparison, the areal density of a membrane reflector system scales differently, and is straightforward to calculate. For the reflective membrane

$$\sigma_r(d) = \rho_r t_r(d).$$

For the supporting boundary

$$\sigma_b(d) = 4\rho_b h(d)\Delta d/d$$

where $h(d)$ is the functional dependence of the boundary thickness with diameter, and Δd is the width of the boundary. The total density is simply the sum

$$\sigma(d) = \sigma_r(d) + \sigma_b(d) = \rho_r t_r(d) + 4\rho_b h(d)\Delta d/d.$$

The areal density decreases with greater aperture; only if the ring has $h(d) = h_o(d/d_o)^\alpha$ with $\alpha > 1$ does σ grow with d,

$$\sigma(d) = \rho_r t_r(d) + 4\rho_b (h_o/d_o) (d/d_o)^{\alpha-1}\Delta d.$$

A more detailed numerical analysis can be performed, with the results presented in figure 5. The results have the same characteristic shape: that a membrane telescope has a mass density that decreases with increasing aperture size.

This is in distinct contrast to the scaling relationship for existing telescopes, $\sigma \propto d^{0.5}$. Thus, not only is a membrane reflector less massive to begin with, but the areal density can actually *decrease* with larger apertures if the ring and membrane are appropriately chosen. Clearly, the areal density of a telescope system can be reduced by orders of magnitude if the relatively massive supporting substrate can be minimized while maintaining the desired reflective surface.

5. Summary.

5.1. Key elements of the DART system:.

- An arrangement of clyndrical-parabolic reflectors can be made that will focus light from a distant source to a point, without any obstruction to the incident beam.
- The aberrations of such a system are dominated by coma and are similar to those found at the prime focus of a standard parabolic reflector. The diffraction limited field of view of such a system is large enough to accommodate a large format FarIR array.
- The individual reflecting surfaces can be constructed using low areal mass density membranes, with the consequence that the mass density of a complete telescope can approach 1kg/m^2.
- Low cost for large apertures.
- Well suited for passive cooling.

5.2. Recommendations for future development of the DART system:.

The remarkable simplicity of the DART design enables the rapid fabrication of a low cost, high quality telescope. A number of subsystems are in the initial stages of study and need to be aggressively pursued:

- Investigating auxiliary optics that will widen the field of view. An initial design has increased the field of view by a factor of 10.
- Identify and develop technology to produce large high quality reflective membranes. The DART prototype used commercially available metal and polymeric membranes. Scalable solutions for high quality membranes at least 10 meters wide are being investigated.
- A 5m DART could be built within one year, with sufficient funding. This would demonstrate scalability from our current prototype to a size useful for a science mission.
- Demonstrate a cooled DART system in a space environment using a low cost, near term science and technology mission.

The Wide-Field Imaging Interferometry Testbed: Progress and Plans

S. A. Rinehart[1], D. Leisawitz[2], D. Leviton[2], A. Martino[2], W. Maynard[2], L. G. Mundy[3], X. Zhang[4]

1: NRC Resident Research Associate, NASA Goddard Space Flight Center,

Mail Code 631, Greenbelt, MD 20771, rinehart@rosette.gsfc.nasa.gov

2: NASA, Goddard Space Flight Center

3: University of Maryland, NRC Senior Research Associate,

Goddard Space Flight Center

4: SSAI, NASA Goddard Space Flight Center

Abstract

We describe the technique of wide field mosaic imaging for optical/IR interferometers and present early experimental results from a laboratory instrument designed to validate, experiment with, and refine the technique. A conventional single-detector stellar interferometer operating with narrow bandwidth at center wavelength λ is limited in its field of view to the primary beam of the individual telescope apertures, or $\sim \lambda/D_{tel}$ radians, where D_{tel} is the telescope diameter. Such a field is too small for many applications; often one wishes to image extended sources. We are developing and testing a technique analogous to the mosaicing method employed in millimeter and radio astronomy, but applicable to optical/IR Michelson interferometers, in which beam combination is done in the pupil plane. An $N_{pix} \times N_{pix}$ detector array placed in the image plane of the interferometer is used to record simultaneously the fringe patterns from many contiguous telescope fields, effectively multiplying the field size by $N_{pix}/2$, where the factor 2 allows for Nyquist sampling.

This mosaic imaging technique will be especially valuable for far IR and submillimeter interferometric space observatories such as the Space Infrared Interferometric Telescope (SPIRIT) and the Submillimeter Probe of the Evolution of Cosmic Structure (SPECS). SPIRIT and SPECS will be designed to provide sensitive, high angular resolution observations of fields several arcminutes in diameter, and views of the universe complementary to those provided by HST, NGST, and ALMA.

1 Introduction

In order to fully explore the epoch of galaxy formation, high-resolution, far-infrared/submillimeter (FIR/submm) images are required, because half of the light emitted in the universe since the Big Bang is in the form of FIR/submm energy.

Such observations must also have high angular resolution to avoid confusion (multiple sources per resolution element) and probe the structure in distant objects. To achieve such high angular resolution, interferometric methods are required. However, to date, optical and infrared interferometers have had very narrow fields, making use of single detectors. In order to accurately characterize the early universe, we require broad fields of view. We must therefore develop new interferometric techniques to perform wide-field interferometric imaging.

The Wide-Field Imaging Interferometry Testbed (WIIT) is an optical testbed designed to explore imaging techniques needed for future space-based interferometers. Key elements of the design are multi-pixel detector arrays and a long delay stroke, which together enable wide-field spectral imaging. The standard output is a spectral array of images (i.e. a 3-dimensional data cube).

In Section 2, we discuss the goals of the WIIT experiment. In Section 2, we discuss the design and implementation of WIIT. Section 4 presents some of our results to date, and in Section 5 we briefly discuss some future goals for WIIT.

2   WIIT Goals

The purpose of WIIT is to develop understanding of the techniques necessary to employ wide-field imaging interferometry. There are a number of issues which we intend to address through WIIT. Some examples include

- What is the maximum practical field of view, and what limits it?

- What are the dominant error sources that limit photometric accuracy and dynamic range in the synthesized spatial-spectral data cube?

- What are the minimum tolerance requirements on various components?

- What are the short-term and long-term stability requirements?

- How can we best utilize data gathered "on the fly" by an interferometer with moving collector mirrors?

- What algorithm optimally expoits the information available in a double Fourier data set?

- Can we rely on a source or sources in the field of view for phase reference?

- What mission design considerations are suggested by our experience with WIIT?

3   The Instrument

WIIT uses a parabolic collimator to produce parallel beams from a an illuminated scene located at its focus. The parallel beam produced by the collimator enter the two collector mirrors (see Figure 1) which form the baseline of the interferometer. The baseline distance determines the angular resolution of the observation ($\theta = \lambda/2b_{max}$). The beam in one arm of the interferometer passes through a series

of fixed flat mirrors to the beam combiner, while the beam in the other arm travels to the beam combiner via a pair of fixed flats and a pair of flats on a linear motion stage. The pair of flats on the stage provide the optical delay necessary for interferometry. The beams from the two arms of the interferometer rejoin at the beam combiner; this combination is in the pupil plane, characteristic of a Michelson interferometer. Following the beam combiner, the light is brought to a focus on the detector array by a single lens. This final path is baffled by a long tube in order to reduce stray light. In a classical Michelson interferometer, the combined light is brought to a focus on a single-pixel detector. This light comes from the primary beam of the individual telescope apertures. In the WIIT, however, a detector array (CCD camera) is used to sample simultaneously many contiguous primary beams.
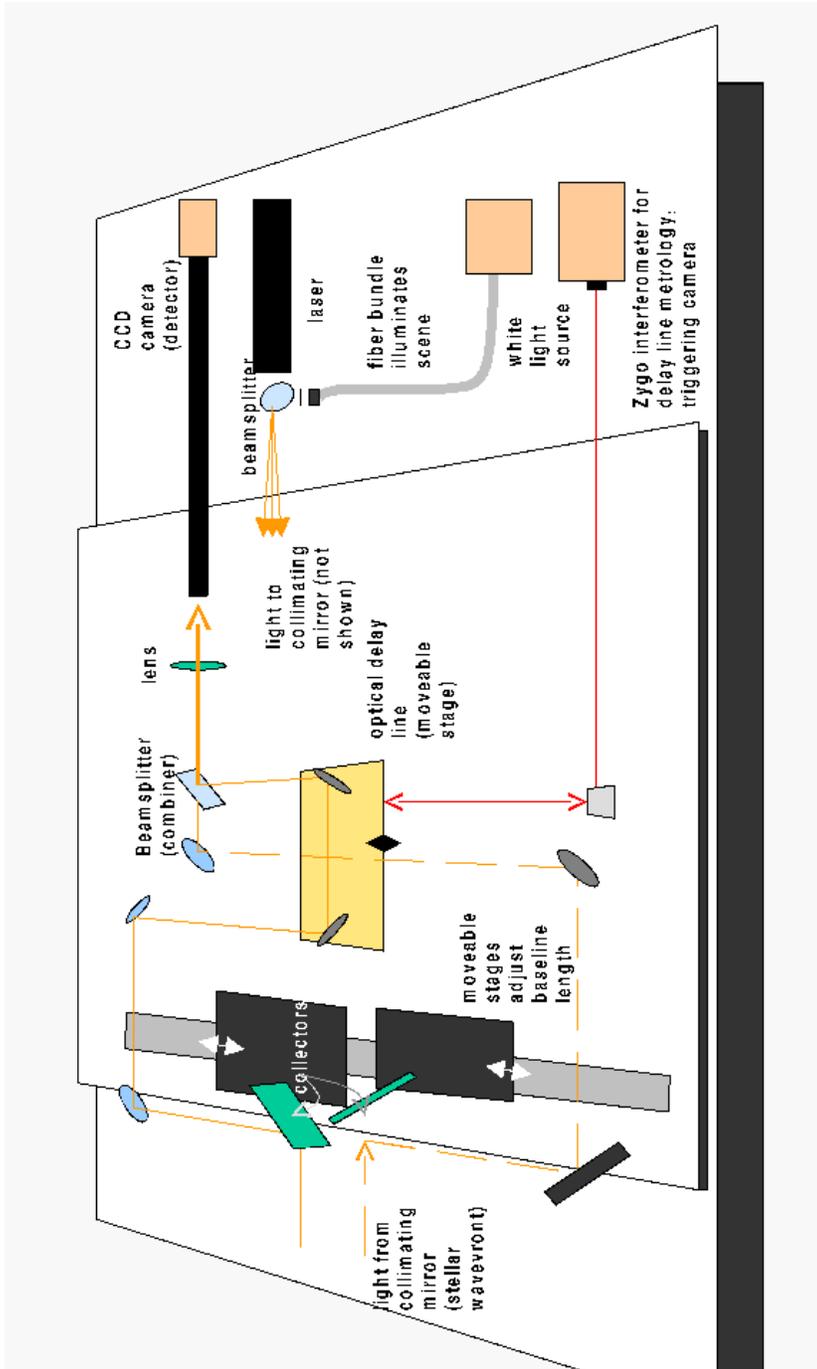
Figure 1: A schematic diagram of the Wide-Field Imaging Interferometry Testbed.

The key points of the WIIT design are:

1. Beam combination in the pupil plane.

2. Record fringe pattern in the time domain in each pixel, as in conventional Michelson.

3. Detector array sees multiple "primary beams" in parallel.

4. Add stroke to optical delay line to compensate for geometric delay across the field of view (FOV).

5. Total stroke provides both wider FOV and Fourier Transform spectroscopy (FTS).

6. Pixels Nyquist sample contiguous primary beams (analog to mosaic imaging used in radio interferometry).

3.1 Recent Developments

Accurate knowledge of the delay line (and baseline) position is critical for interpretation of data from an interferometer. The first white light fringes were detected with the WIIT in August 2001 (Figure 2). Artifacts seen in the measured interferograms suggested the presence of jitter and drift in the delay stage. Subsequently, we have worked to install an independent metrology system and couple the CCD camera exposures to the delay stage position. Recently we have installed an optical encoder system to accurately measure the absolute position of the delay line stage. We have also implemented camera triggering using the signal from the Zygo interferometer. This ties the camera exposures to the position of the delay line stage with an accuracy of roughly 10nm.

We are also in the process of implementing two housekeeping systems. First, sixteen temperature sensors are being deployed around the optical table and the interferometer. Monitoring the temperature over time will help us to understand the environment in which we operate. Second, we are installing a pair of power meters in the system. One of these will be used to measure the amount of light in the parallel beam, determining the stability of our source and providing information for accurate calibration of images. The second power meter will be placed on the second arm of the beam-combiner. This will allow us to measure the relative amounts of light traversing each arm of teh interferometer (by blocking one arm at a time). Any imbalance in the throughput could limit the achievable fringe visibility.

4 Results to Date

Version one of the testbed has been constructed and utilized. In Figure 2 we show a "white light" fringe, compared to a model calculation based on the optical properties of the system components. In Figure 3, we show a similar figure taken using a 10 nm wide bandpass filter. To date, we have measured fringes with both laser light and white light, and through a number of narrow band filters. These interferograms agree qualitatively with system models, and show that the
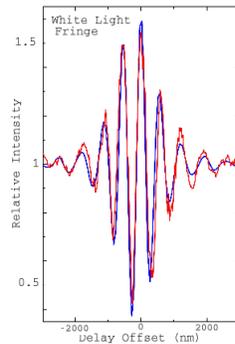
Figure 2: Measured and calculated fringes for WIIT at full bandwidth, 4500 to 6800 Angstroms extent at half intensity, 4000 to 7200 Angstroms at 10%. The red curve is the data, the blue curve is a model based on independent measurements of the source spectrum and the system spectral response function. Compison of the red and blue curve peaks suggests that the true system bandwidth is a little narrower than the model prediction. However, errors in our knowledge of the delay line stage position limited the detailed fringe comparison until recently, when the independent metrology system was introduced.
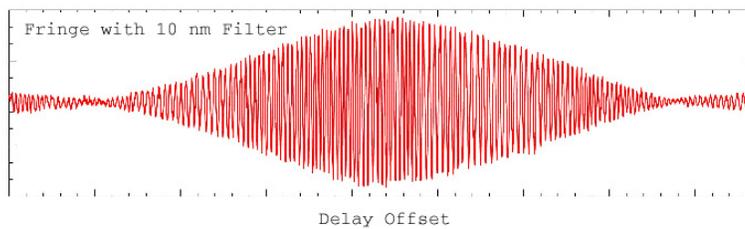
Figure 3: Measured fringes for WIIT with a 10 nm bandpass filter.

system is stable optically and mechanically over periods of weeks. We have also determined that the major limitation on the quality of our data is the uncertainty in knowledge of the position of the delay line. Environmental effects, such as air turbulence, vibrations, and temperature changes might turn out to be significant, but are much less important than jitter and drift in the delay line stage. We have taken steps to mitigate all the significant error sources.

5   Future Plans

There are several improvements underway and planned which we expect will greatly improve WIIT performance. First, when we finish the implementation of Zygo-triggered images and synchronization with the optical encoders, we will drastically improve our knowledge of delay line position. Second, our light level monitoring will provide level correction data for use during data reduction.

In addition, by using the Zygo to measure turbulence around the testbed and the temperature sensors, we will attempt to understand environmental effects. Following this, we will endeavor to provide greater isolation from these effects. With these improvements, we will attempt to systematically characterize the noise of the system and determine noise contributions from different sources within the system.

Once the system noise is well-understood, we will produce, reduce, and analyze data, building toward synthesis of increasingly complex scenes. We will begin with 1-D imaging; a single point source, then moving to more complicated scenes (including multiple point sources, extended sources, extended sources with non-uniform brightness, etc.).

Following these data, we will proceed to develop the next version of WIIT by installing source and camera rotation stages to enable 2-D imaging. We will then conduct analogous experiments, using multiple 2-D scenes of varying complexity. We will use these data to test reduction algorithms and to optimize baseline movement of filling the synthetic aperture. The same procedure will be used again for 2-D spectral imaging, using scenes which have both spatial and spectral structure.

6   Acknowledgments

Cryo-Infrared Optical Characterization at NASA GSFC

Ray Boucarut[1], Manuel A. Quijada[2], and Ross M. Henry[1]

[1]Goddard Space Flight Center, Code 551, Greenbelt, MD 20771

ray.boucarut@gsfc.nasa.gov

[2]Swales Aerospace, Beltsville, MD 20705

ABSTRACT

The development of large space infrared optical systems, such as the Next Generation Space Telescope (NGST), has increased requirements for measurement accuracy in the optical properties of materials. Many materials used as optical components in infrared optical systems, have strong temperature dependence in their optical properties. Unfortunately, data on the temperature dependence of most of these materials is sparse. In this paper, we provide a description of the capabilities existing in the Optics Branch at the Goddard Space Flight Center that enable the characterization of the refractive index and absorption coefficient changes and other optical properties in infrared materials at cryogenic temperatures. Details of the experimental apparatus, which include continuous flow liquid helium optical cryostat, and a Fourier Transform Infrared (FTIR) spectrometer are discussed.

1.0 Introduction

Optical Component testing and verification are essential aspects of optical technology development for far-infrared and sub-millimeter astronomy. The NASA Goddard Far-Infrared laboratory has historically supported optical instrument development in the atmospheric science and infrared astronomy program areas. Since the early 1980's a variety of NASA and Space science programs have been supported including the COBE (Cosmic Background Explorer), the Space Infrared Telescope Facility (SIRTF) Infrared Array Camera (IRAC) the Composite InfraRed Spectrometer (CIRS) and technology development for the Next Generation Space Telescope (NGST) (Table 1). These programs were supported through far-infrared and sub-millimeter measurements at temperature ranging from room to 1.4 K

Figure 1 shows the importance and necessity of these measurements. A temperature shift in the spectral transmittance curve of filter 1 for the SIRTF/IRAC filter is seen between 3 and 4 microns. This shift was seen in many of the filters for this instrument. Figure 2 shows the results of the spectral calibration of Goddard's KAO spectrometer used to measure the first Far-IR spectrum of Supernova SN 1987A.

2.0 Capabilities

The heart of the Far-Infrared measurement facility is the Fourier Transform Infrared Interferometer. A Bruker IFS 113V(Genzel interferometer – Figure 5) and Nicolet 8000 (Michelson interferometer) have been used. The current instrument is the Bruker (Figure 4). The Bruker has a spectral range from 250 nm up to 2.5 mm or 4 cm^{-1} to 40000 cm^{-1}. A grating

spectrometer spanning the 200 nm to 3 μm region complements the available instruments. (Perkin-Elmer Lambda 9). All of these instruments are capable of transmission and reflection measurements at variable angles of incidence. A variety of cryostats are available for cooling the samples from room to liquid helium temperatures.

3.0 Materials Studies

The Far-Infrared laboratory has been involved extensively in studying optical properties of materials at far infrared wavelengths. These studies include the measurement of the optical properties of crystal substrates and of metal mesh filters. The filter development program for the Diffuse Infrared Background Experiment (DIRBE) on the Cosmic Background Explorer Satellite (COBE) afforded the opportunity to measure the transmission of numerous materials warm and cold. Many of these materials became part of the filter stack on DIRBE.[1-3] Other materials were used as blocking filters during measurements. The results of these measurements are shown in figure 6 and 7.

Numerous studies were made on the reflectivity of various black paints. Black paints are used for suppression of stray light in optical instrument. The behavior of black paint in the far infrared is not well known. Figure 8 plots on a log-log scale the relative specular reflectance of numerous black paints measured. The spectral range of the measurements is from 1 to 1000 microns.

Additional tests have been made on high temperature superconductors. Figure 9 shows the reflectance of $Nd_{1.85}Ce_{0.15}CuO_{4-g}$ between 100 and 1000 microns. The data was taken at 8 and 25 K. The rising reflectance is an indication of the sample undergoing a transition from the normal to the superconducting state[5]. Spectra has also been taken of cyanoacetylene and cyanogens samples for comparison with measurements of the spectra of Titan's atmosphere (Figure 10). Metal mesh filters hold much promise as filters for far-infrared spectroscopy[6]. In figure 11 are plots of two different types of mesh filters: inductive and capacitive. Both types of constructions were measured for transmittance both at room and cryogenic temperatures. Figure 12 shows a comparison between theoretical models for and laboratory measurements for conductive and inductive meshes with different geometric factors.

NASA PROGRAMS INVOLVING
FAR-IR AND CRYO- OPTICS

Project	Fore Optics Temperature	Focal Plane	Wavelength
COBE	3 K	2 K	10 μm – 3 cm
IRAS	3 K – 10 K	< 3 K	8 μm – 300 μm
CLIR	30 K	10 K	2.5 μm – 25 μm
SIRTF/IRAC	1.4 K	17 K	3 μm – 10 μm
CIRS	170 K	80 K	7 μm – 1000 μm
NGST	30 K	< 8 K	0.5 – 20 μm

Table 1 . Goddard's Far infrared measurement laboratory has supported optical instrument development in a variety of infrared astronomy program areas

Transmittance of SIRTF IRAC Filter
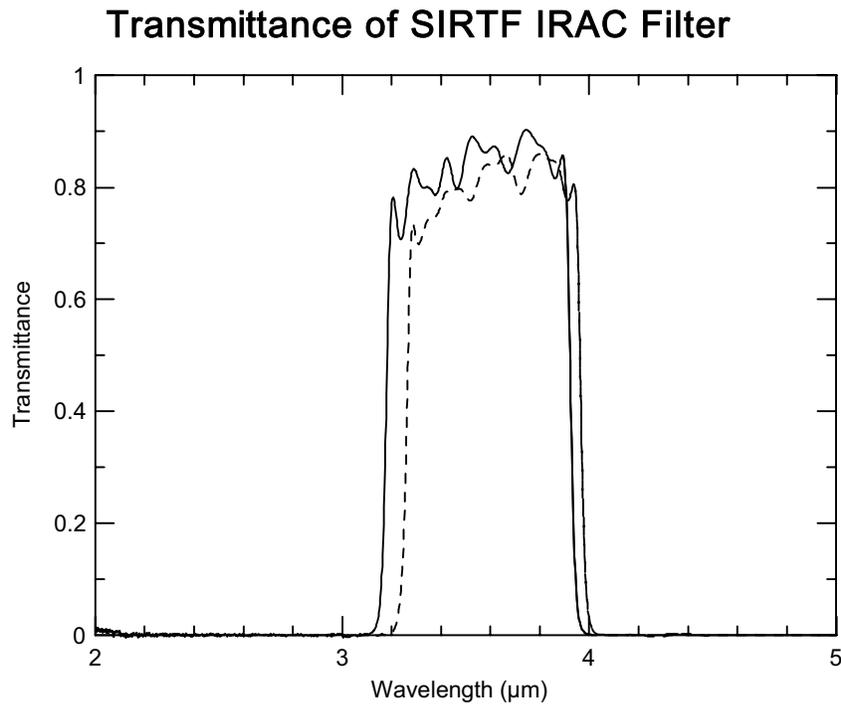
Figure 1. The transmittance of SIRTF/ IRAC Filter 1 measured at room temperature (dashed) and at 5 K (solid) and at an angle of incidence of 11 º.

Spectral Bands in the GSFC KAO Spectrometer
Used to Measure the First Far-IR Spectrum of
Supernova SN1987A
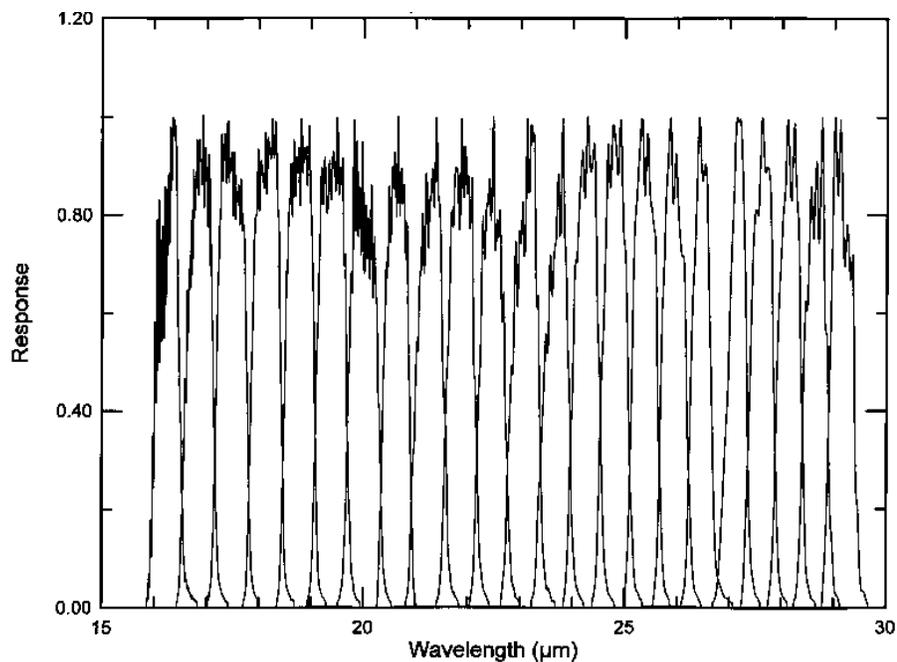
Figure 2. Spectral bands measured in the KAO Spectrometer
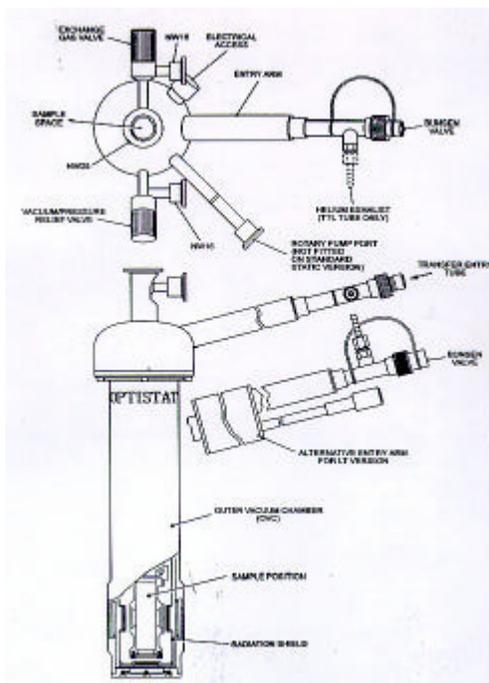
Figure 3. Diagram of Oxford Cryostat
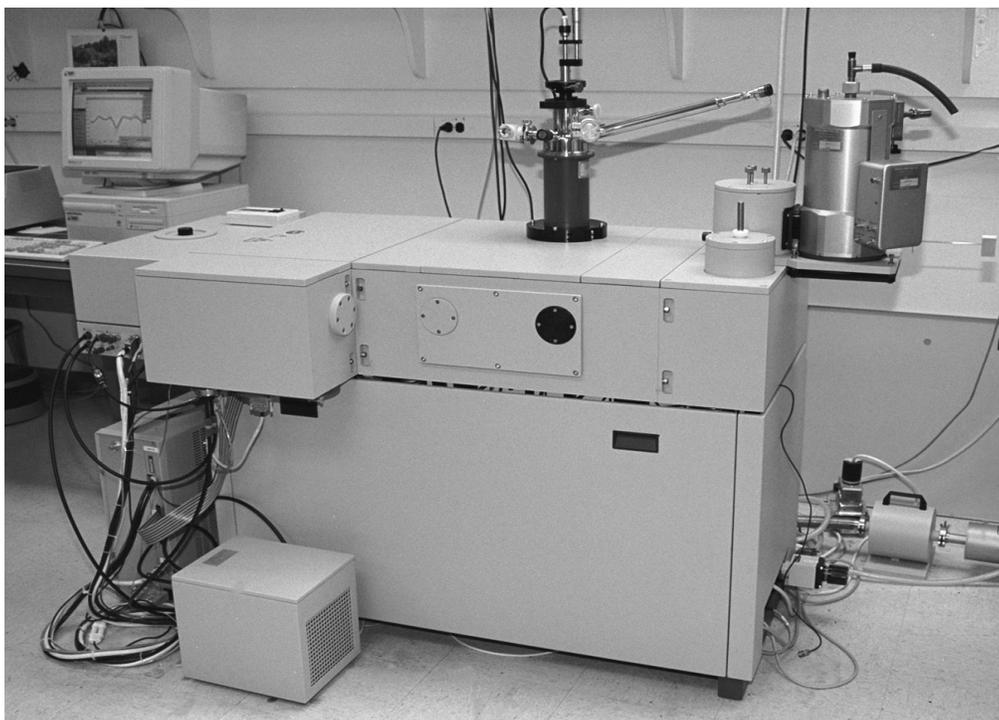
Figure 4. Bruker FTIR Spectrometer, wavelength range from 250 nm to 2 mm, Maximum resolution - 0.03 cm^{-1}
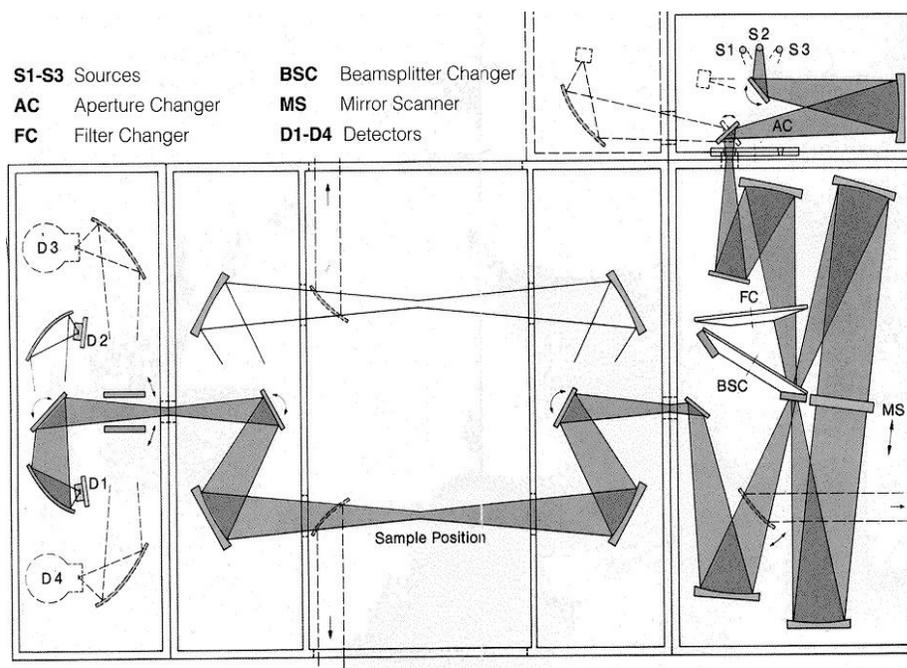
Figure 5. Layout of Genzel spectrometer
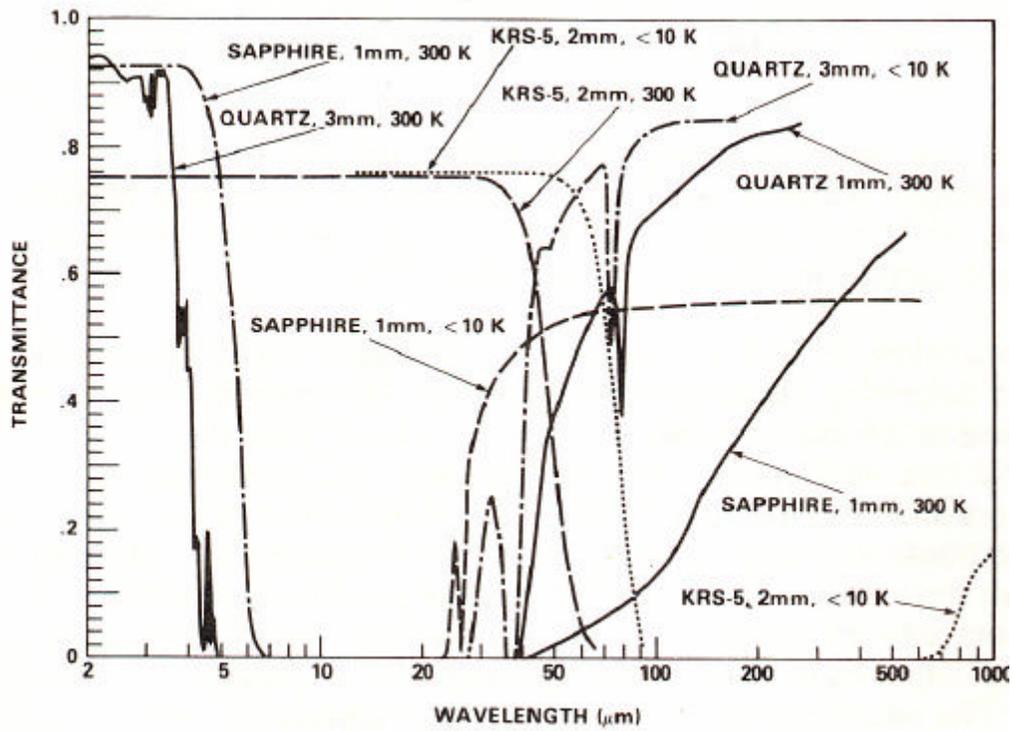
Figure 6. Transmittances of KRS-5, sapphire and crystalline quartz over the 1mm to 1000mm wavelength region at temperatures of 300K and below 10K.
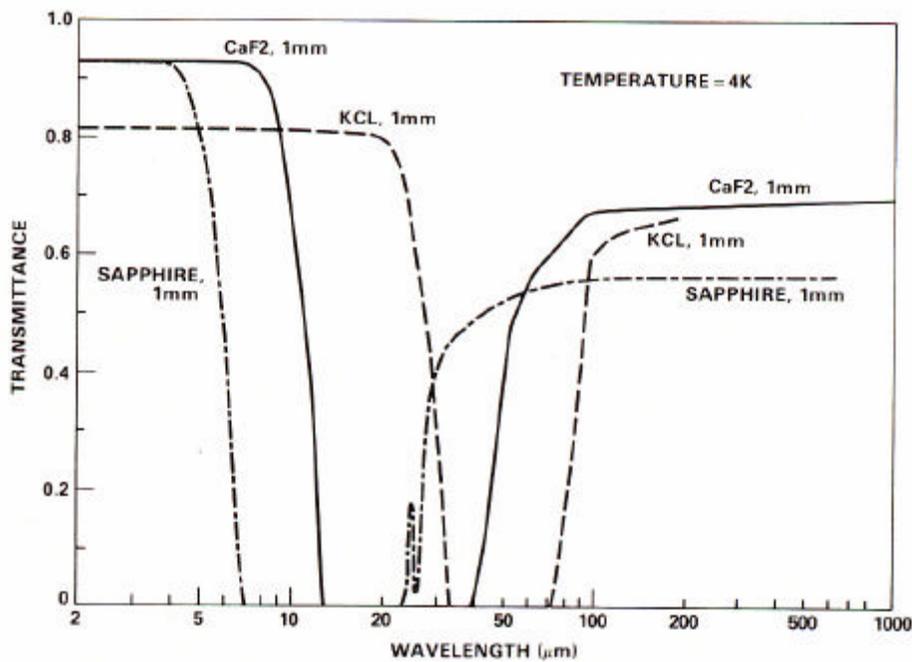
Figure 7. Transmittance of 1 mm thick sapphire, CaF. and KCl over the 2mm to 1000mm wavelength region at a temperature of 4K

Relative Specular Reflectance (%
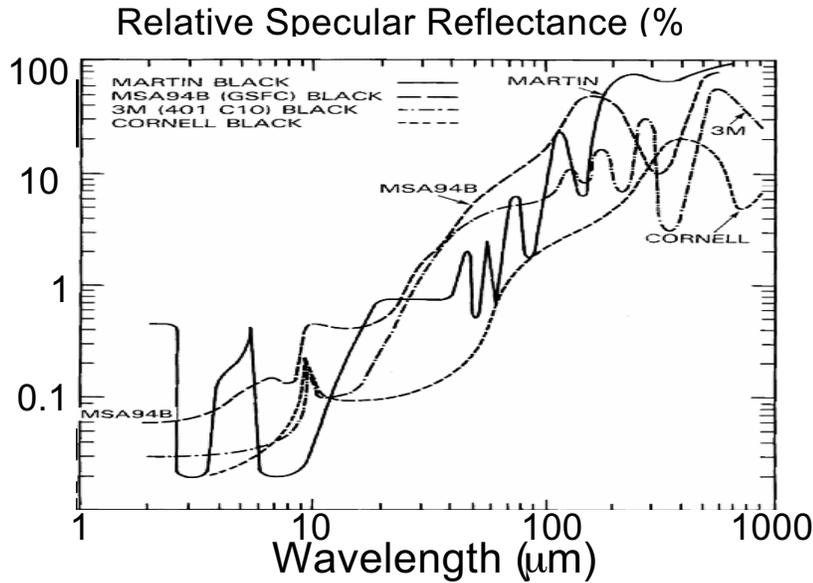
Figure 8. Relative specular reflectance of various black paints measured in Goddard's Far IR measurement laboratory. Wavelength range 1 to 1000 microns. Reflectance displayed on a logarithmic scale.

High Temperature Superconductor, $Nd_{1.85}Ce_{0.15}CuO_{4-g}$ $T_c = 21$ K
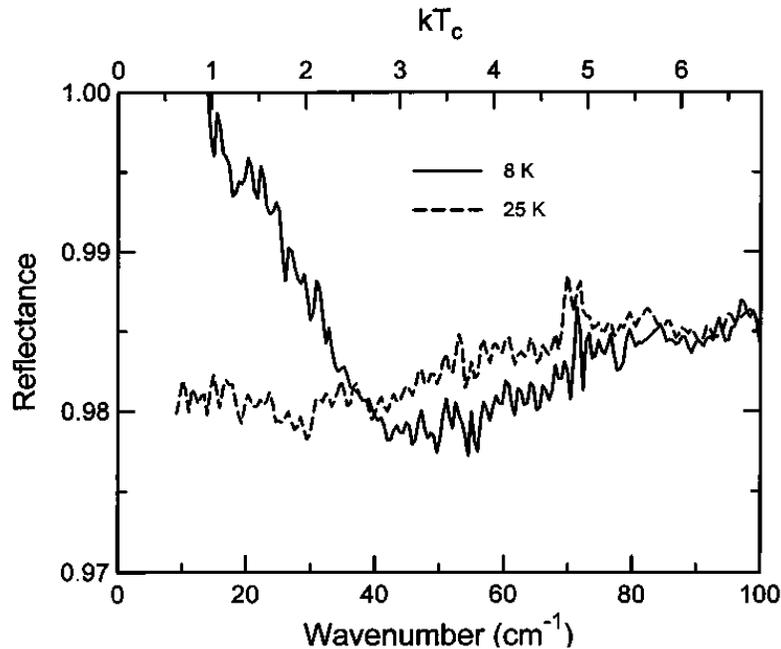
Figure 9. This figure shows changes in the reflectance as function of frequency for a high-temperature superconductor material above (25 K) and below (8 K) the critical transition temperature (Tc~ 12 K). Notice the rising reflectivity of the sample at low frequencies and for measurements done at T = 8 K. This is an indication that the sample is in the superconducting state.
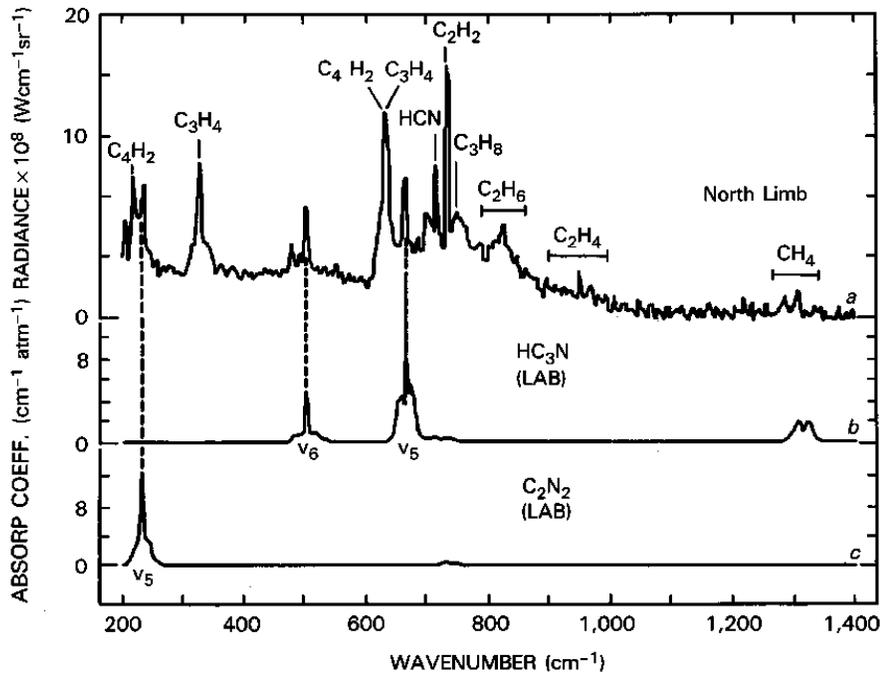
Figure 10. Voyager Titan Spectrum - Comparison of laboratory and Titan radiance spectra. The observed spectrum (a) was obtained off the limb of the planet in the northern polar cap region. The high air mass along the line of sight enhances the weak emission features, while space contamination in the field of view depresses the adjacent continuum. Laboratory spectra for cyanoacetylene (HC_3N) (b) and cynogen (C_2N_2) (c) are shown.
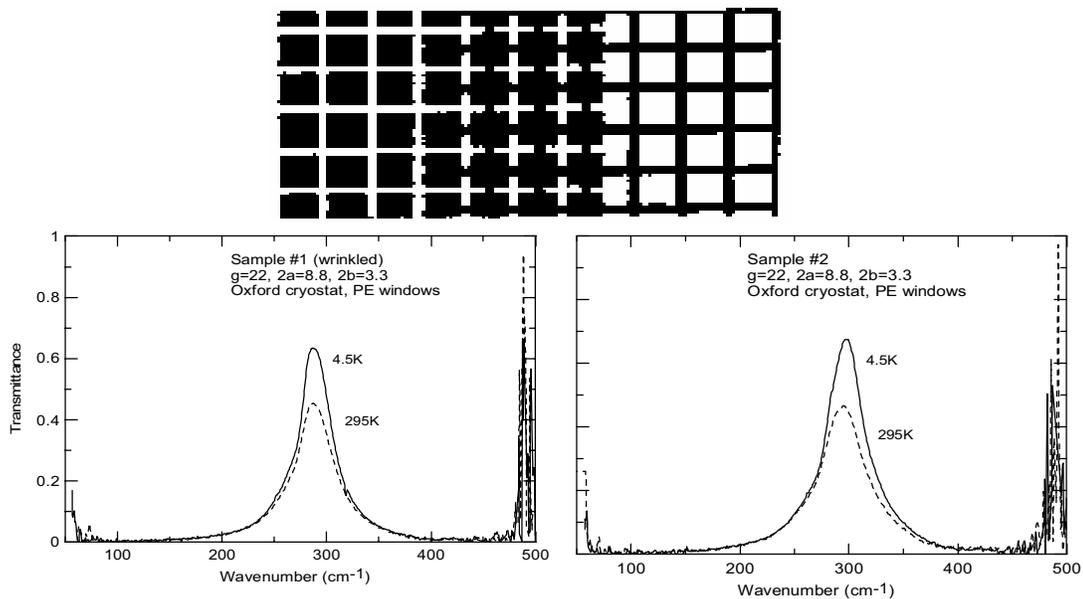
Figure 11. Production of cross shaped pattern (middle region) by superposition of a capacitive (left) and an inductive (right) grids. The dark regions represent metallic conducting areas, whereas the clear regions are dielectric insulating material.

Temperature dependence in the spectral transmittance for two inductive grids (labeled Sample #1 and #2 respectively) on a 2 μm thick polyimide substrate. The geometrical factors or grid parameters g, 2a and 2b can be defined in terms of the cross pattern shown in the middle section of previous figure: g is periodicity of the crosses, 2a is separation between crosses, and 2b is size or width of crosses
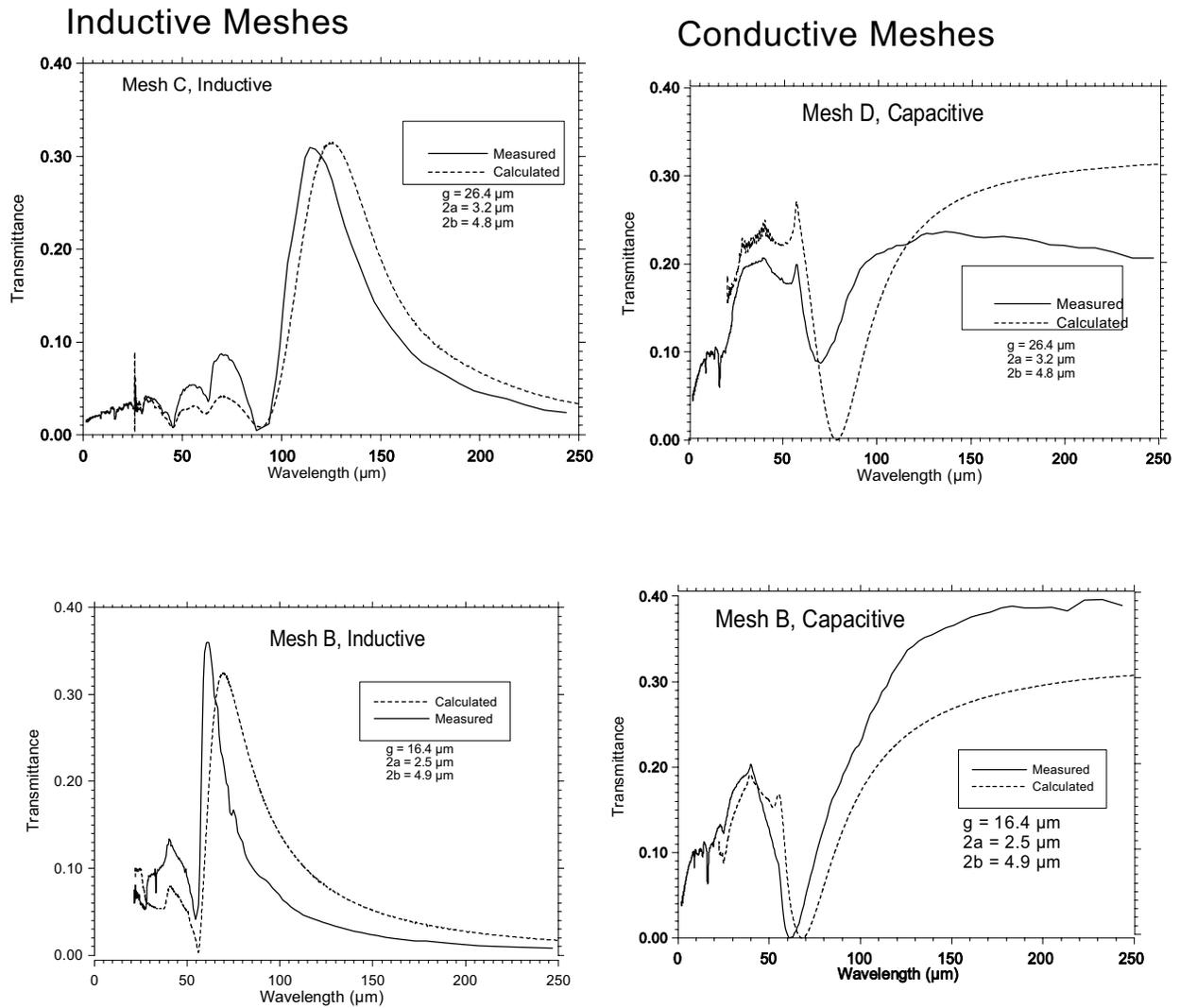
Figure 12. Inductive Meshes (Left side figures)
Comparison between the measured and calculated transmittance as a function of wavelength for two different inductive meshes, labeled C & B respectively, along with their geometrical factors g, 2a, and 2b.

Capacitive Meshes (Right side figures)
Comparison between the measured and calculated transmittance as a function of wavelength for two different capacitive meshes, labeled D & B respectively, along with their geometrical factors g, 2a, and 2b.

4.0 References

1. K. P. Stewart, M. A. Quijada, "Cryo -transmittance and reflectance of filters and beamsplitters for the SIRTF Infrared Array Camera," SPIE proceedings, 4131, 218 -227, (2000)
2. J. B. Heaney, K. P. Stewart, R. A. Boucarut, "The Cryo-Testing of Infrared Filters and Beamsplitters for the Cosmic Background Explorer's Instruments", SPIE VOl. 619, 142 - 147 (1986)
3. J. B. Heaney, R. A. Boucarut, K. P. Stewart, "Infrared filters for cryogenic Radiometers", Cryogenic Optical Systems and Instruments V, 118-130, 1993
4. K. D. Moller, R. G. Zoeller, N. G. Ugras, P. Zablocky, J. B. Heaney, K. P. Stewart and R. A. Boucarut, "Noise tube sources for the far IR and millimeter region", Applied Optics, Vol. 27, No. 10, 1988
5. K. P. Stewart, Ph. D. dissertation, University of Maryland (1997)
6. K. D. Moller, J. B. Warren, J. B. Heaney, C. Kotecki, "Cross -shaped bandpass filters for the near- and mid-infrared wavelengths regions", Applied Optics 35(31), 6210 -6215 1996

The Status of Composite Telescope Technology for Space Astronomy
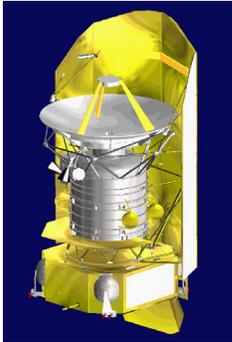
Steve Connell

(Composite Optics, Inc.)

**Second Far-IR and Sub-mm
Space Astronomy Workshop
March 7-8, 2002**

Composite Optics, Inc. • 9617 Distribution Avenue • San Diego, California 92121
Tel: (858) 621-5700 • Fax: (858) 621-5770 • E-Mail: coi@coi-world.com • Website: http://www.coi-world.com

---

Why Carbon Fiber Composites?

- **Low Mass - Herschel Primary Mirror Target Areal Density < 20 kg/m^2
  (2m CFRP, Actual < 11 kg/m^2)**

- **Low Coefficient of Thermal Expansion (CTE), Tailorable over Wide Temperature Regime**

- **Large Cost Reduction When Multiple Mirrors and/or Segments Are Made From the Same Mold**

- **Stiffness Is Roughly the Same as Glass and Thermal Conductivity Is Much Higher**

- **Design, Material, Processes Scalable to Large Apertures**
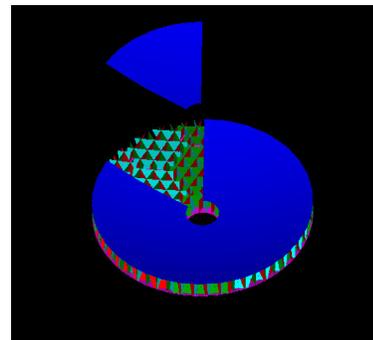
CFRP
Carbon
Fiber
Reinforced
Polymer

Comparisons of Low Expansion Materials
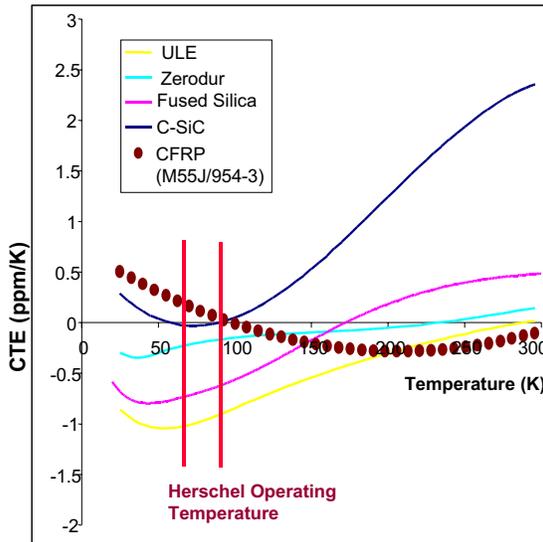
- **The CFRP Zero-CTE-Temperature Can Be Fine-Tuned Within the Cryo Regime**
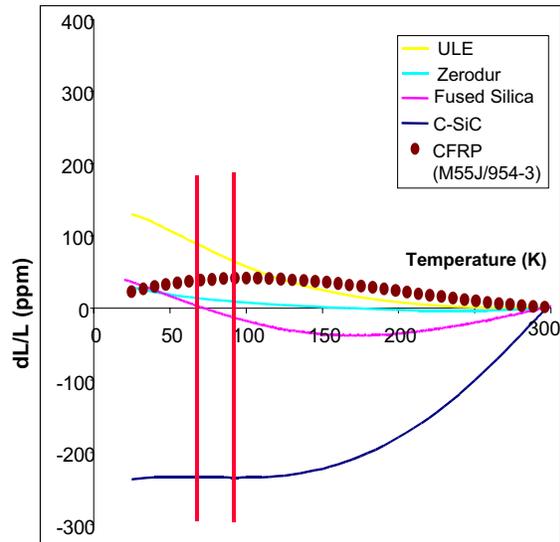
- **CFRP Exhibits Low Material Strain Over Broad Temperature Range**

NASA CFRP Mirror Development History and COI

- ◆ **CFRP Was Recommended for the 10 to 20m Segmented Large Deployable Reflector (LDR) in the Late '80s Because of the Low Mass and the Ability to Economically Reproduce Several Segments with One Mold**

- ◆ **The Precision Segmented Reflector (PSR) Program ('88 - '92) Was Funded by NASA to Develop Technology Needed for LDR**

- ◆ **JPL Worked with Hexcel to Develop CFRP Mirrors for PSR**

- ◆ **Towards the End of the PSR Program JPL Evaluated COI Mirrors**

- ◆ **Following the PSR Program CFRP Mirror Development Continued at COI with IRAD and SBIR Funds, Leading to MLS Flight Reflector**

- ◆ **1998 - 2000: JPL Funded COI to Develop CFRP Telescope for Herschel, Leading to Successful 2m Prototype**

2-Meter Lightweight Mirror Demonstrator and Metering Structure Mock-Up
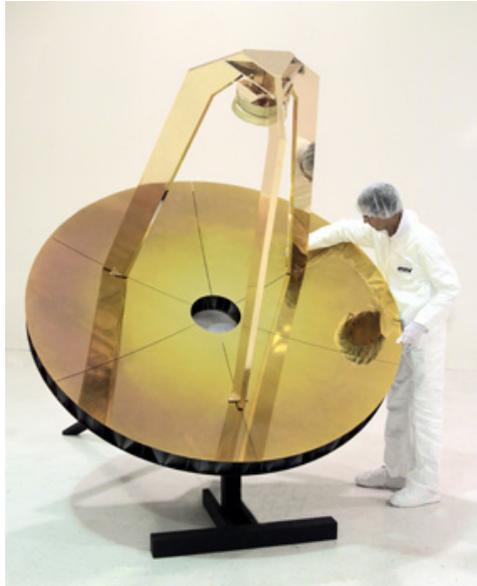
Herschel Telescope Configuration

Secondary Mirror **(M2)** with Actuator Hexapod

Primary Mirror **(M1)**

Metering Structure (Tripod)

M1 Support Flexure (3)

PLM Interface Triangle

Low Emiss/High Emiss MLI and Shroud System
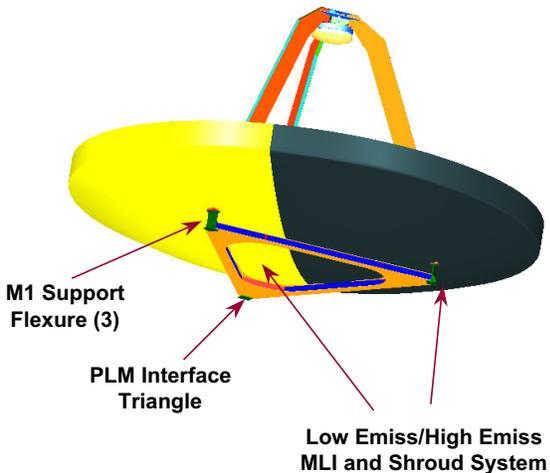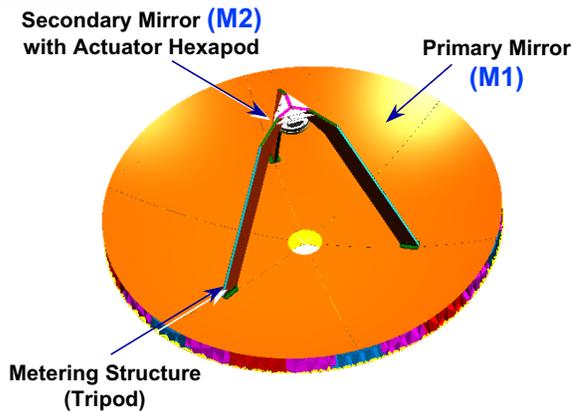
Telescope Mass Breakdown	
Primary Mirror Assembly	141 kg
PLM Interface Triangle	11 kg
Secondary Mirror Assembly	6 kg
Metering Legs	19 kg
TOTAL	177 kg
Requirement	280 kg

2m Testing Objectives

- **Measure the Figure of the 2m Mirror at Herschel Temperatures**

- **Evaluate Effect of Thermal Cycle to Cold Temperatures**

- **Induce Thermal Gradients and Understand Effects**

- **Explore Measurement Methodologies**
  - » **10.6 micron phase-shifting interferometry (baseline)**
  - » **Sub Aperture Stitching**
  - » **IR Shack-Hartmann Instrument**

Ambient 0 - G: Low Order Figure
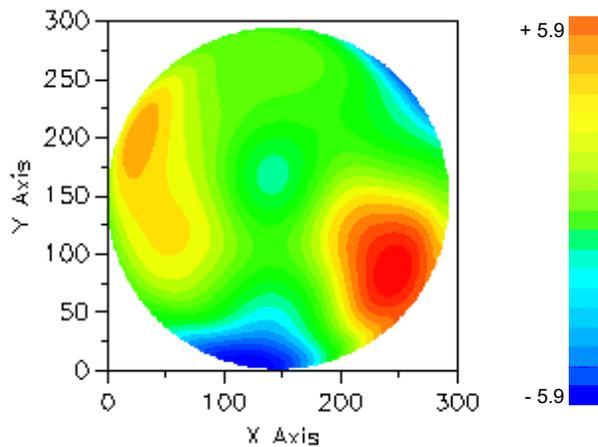
- **Reconstructed Surface Based on Z_5 through Z_{36}**
- **2.11 RMS μm ($\pm$0.34 μm)**

Evolution of Figure vs. Temp
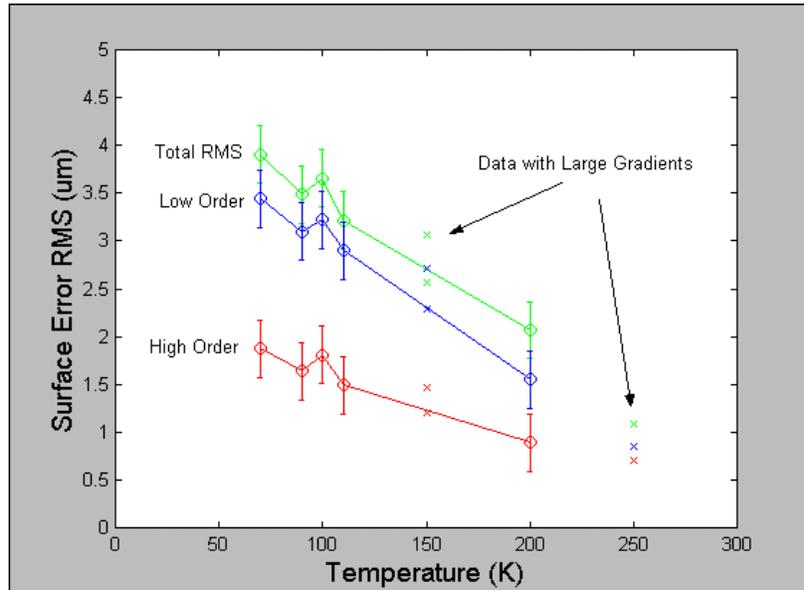
Modest, Linear Variation Over Operating Temperature

Delta Figure: 293K to 70K
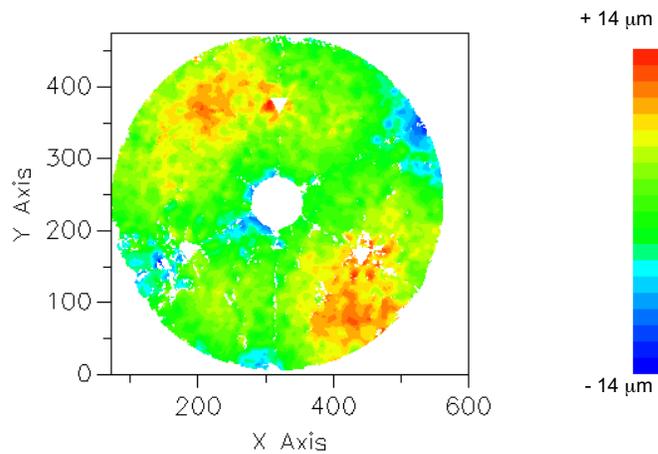
◆ **3.9 μm RMS Surface Error**

High and Low Order Figure: 293K to 70K
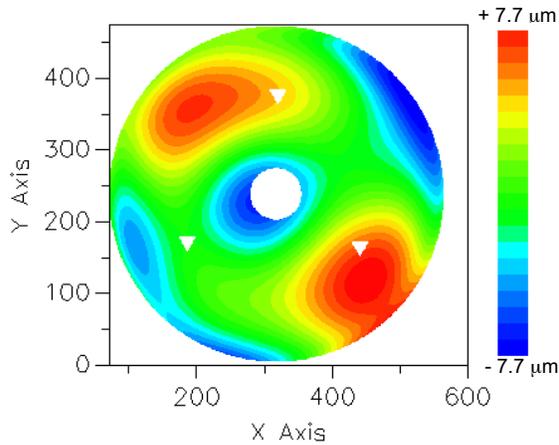
- **Z_5 through Z_{36}**
- **3.4 μm RMS**
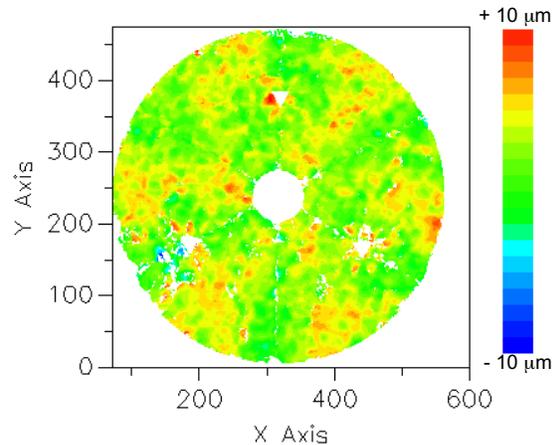
- **Residual Figure: > Z_{36}**
- **1.9 μm RMS**

Summary

- **CFRP Ideally Suited for FarIR/Submm Space Astronomy**
    - » **Low Mass, Large Aperture, Cryo-Stable**

- **CFRP a Comparatively Low-Cost Technology**
    - » **Herschel Telescope Projected Cost ~$3M/m^2 (ref. 3.5m aperture)**

- **2m CFRP Mirror Prototype Measured at Cryo**
    - » **2.1 micron As-Built Surface Error**
    - » **3.9 micron Surface Error Change at 70K**

- **Ability to Correct Predominately Low Order Errors in CFRP Mirror Has Significant Impact on Final WFE**

Future Directions & Technology Development

- **Design Refinements to Improve Cryo-Quilting**

- **Coatings to Decrease Roughness / Mask Fiber Print-Thru**

- **Polishable Coating Option**

- **Imaging Quality Technology Goal**

- **Technologies to Reduce Tooling Cost and Acquisition Schedule**

- **2-m Mirror to Fly on 2004 Balloon Borne Observatory**

For Further Information, Contact:
Dr. Steven Connell
(858) 621-5700
sconnell@coi-world.com

Session VIII

Cryogenic Technology

Cryogenic Technology: Ongoing Developments for the Next Decade

Michael DiPirro
Code 552, NASA/Goddard Space Flight Center, Greenbelt, MD 20771

To obtain optimum sensitivity a submillimeter space observatory will require low temperature mirrors (~3K) and very low temperature detectors (</~ 0.1 K). Both of these temperatures have been achieved by space cryogenic systems, but neither for a 10 year duration. Past systems used superfluid helium to provide direct cooling in the 1 to 2 K range (IRAS, COBE, IRTS, ISO) or as an upper stage for an adiabatic demagnetization refrigerator to achieve temperatures down to 0.06K (Astro-E/XRS). Boiloff vapor may be used to cool an otherwise warm telescope as in the Space InfraRed Telescope Facility (SIRTF). In SIRTF a 0.85m telescope is cooled to 5.5K by absorbing about 6 mW in the cold vapor. This residual heat is due to both radiation from a helium vapor cooled outer shield at about 20 K and from conduction through a structure mounting the cold telescope and instruments to the warm spacecraft. The boil off rate required to cool the telescope results in a 2.6 to 5 year lifetime, depending on whether other parasitic heat sources such as thermoacoustic oscillations are also present. A helium dewar results in a very heavy system to achieve 2 to 5 year lifetimes. For example it takes roughly 400 kg for XRS to achieve 0.06K for two year life with a 250K boundary temperature, and ~300 kg (including thermal shielding) for SIRTF to achieve 1.3K for 5 year life with a 35 K boundary temperature. To go to longer duration and to lower the weight, active cooling methods are required combined with more aggressive passive cooling techniques. It is possible, with some development, to provide cooling for detectors to 0.05K and telescopes and instruments to <4 K for a 10 year mission with a 100 kg system including power sources, structural support, and vacuum enclosures for critical portions of the instruments.

Passive Radiators.

Passive radiators to achieve temperatures below 40K have been built for SIRTF and to achieve temperatures down to 30K are under development for the Next Generation Space Telescope (NGST). Parasitic heat leaks and size place practical limits of about T > 25K on passive radiation cooling. Several critical issues have been identified. Exposure to hot surfaces must be severely limited. Parasitics are critical and hard to control. There are many uncertainties with any modeling including the assumed material properties such as absorptance and reflectance. Testing will be critical. Thermal/mechanical isolation and efficient cryogenic heat transport techniques will be required.

The NGST thermal models currently predicts 80-90 K on the backside of the sunshield and approximately 35K on the primary mirror. This sunshield performance is achievable but with many hurdles. The issues being worked by NGST include: deployment/stability, lifetime (meteroid damage, degradation of material and optical properties), and analytical predictions of thermal performance accounting for all degrading factors. NGST's sunshield concept consists of 4-6 separated Kapton membranes approximately 32 x 14 meters across. Heat is rejected from between layers before reaching the cold backside of the sunshield. Shield performance is dependent on shield spacing, the optical properties, and number of layers. One must balance thermal performance with deployability, packaging and mass.

One interesting thing to note is that as the mirror mass starts to dominate the warm spacecraft mass, the structural support cross section will tend toward a constant. That is, to obtain a given minimum resonant frequency, the support stiffness will depend on the spacecraft mass rather than the mirror mass. Since the thermal conductance of the support depends on its cross sectional area, this implies that conducted parasitics into a large telescope will level off as the telescope gets larger.

For a submillimeter mission a 10 degree shield (allowing for up to 10 degrees of off-normal sun angle) has a size which varies with truss length. For a 15 meter truss length the shield size would be slightly smaller than the NGST shield. This appears to be the optimal size to shield a 3m diameter mirror. See Figure A.
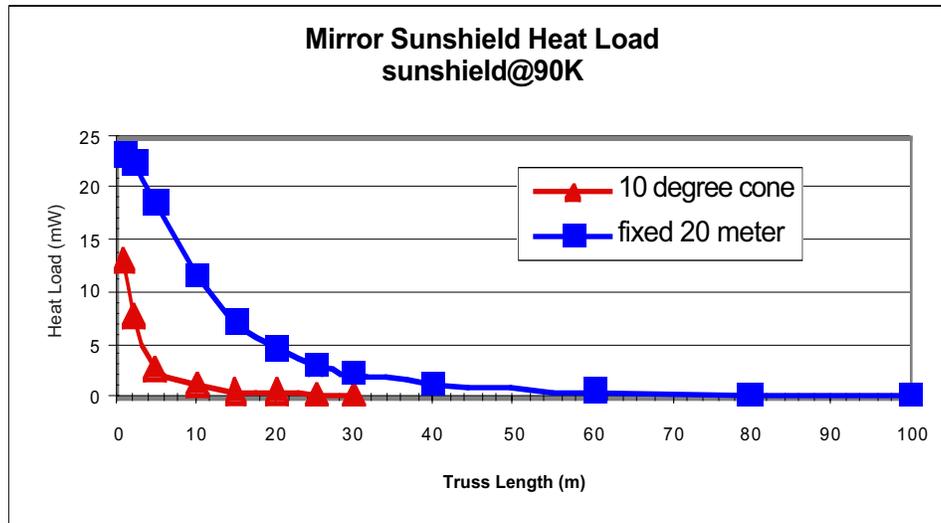
Figure A. Triangles are for sun shield diameter necessary to allow 10 degrees of telescope tilt relative to the sun. Squares represent the effect of distance from a fixed 20 meter diameter sun shield.

Cryogenic Capillary Pumped Loop (CCPL).

Capillary Pumped Loops (CPLs) are capable of transporting large amounts of heat and providing tight temperature control utilizing capillary pumping forces (no moving parts). The CCPL provides a very efficient heat transport path at low temperature. This allows cooling of remote or distributed systems with minimal temperature gradients even in the presence of a large heat flow. A CCPL can be used in a cryogenic thermal bus or as a temperature control device for cryogenic detectors. CCPLs use simple construction from copper or aluminum. They are highly flexible and suitable for miniaturization, as would be required for the low heat flows in the 2 to 4K range.

A liquid nitrogen CCPL was successfully demonstrated on STS-95 in October, 1998 tranporting 2W at 80K. Several ground demonstrations have been performed with liquid nitrogen (transporting 0.5 to 12 W in the 80 to 100K range) and liquid neon (transporting 0.25 to 3.5 W in the 35 to 40K range). The basic two-phase technology is established, the remaining questions concern operating fluid properties and parasitics. A hydrogen loop breadboard which transfers 0.5W - 5W over a 1.5 meter distance at around 20K was recently demonstrated in the lab. The loop has a quick start-up within a couple hours and very robust continuous operation for over 10 hours. An improved prototype unit is being built.

For a future submillimeter mission CCPLs could be used to cool distributed heat sources for the cold optics (liquid helium CCPL, 2-4 K) or to provide cooling to structures remote from a cryocooler or low temperature radiator (liquid neon CCPL). Development is needed for the liquid helium CCPL.

Mechanical Cryocoolers.

Space qualified cryocoolers currently achieve a low temperature of about 20K. Laboratory coolers with a reasonable thermodynamic efficiency (a few % of Carnot) and limited input power to the compressor of 200 watts or less have recently demonstrated cooling to below 10 K.. The Advanced Cryocooler Technology Development Program (ACTDP) will fund the development of cryocoolers producing 7.5 mW of cooling at 6 K and 0.25 W of cooling at 18 K for an input power of less than 150 W with a mass of less than 40 kg systems including control electronics. The first engineering units of such coolers will be ready in 2005.

<u>Stirling/J-T Coolers</u>. There are 3 versions of the Ball Aerospace multi-stage Stirling cycle cooler. The two-stage version, capable of producing 0.45 watts at 30 Kelvin for 70 watts of input power, has been qualified for flight. A protoflight model two-stage cooler has completed a life test at Goddard. Ball has proposed adding a Joule-Thomson expander stage to this cooler to achieve temperatures to 6K. Matra Marconi Aerospace has produced a high power Gifford-McMahon with a J-T stage that achieves ~4K. Unfortunately the input power required for this cooler is greater than 2kW.

<u>Pulse Tube Cryocoolers</u>. Long life, high reliability pulse tube cryocoolers for space applications have been space qualified to temperatures as low as 35 K.   The primary advantage of the pulse tube cryocooler, compared to a Stirling cycle cryocooler, is that the pulse tube does not have any moving parts at the cold end which improves reliability and reduces induced vibration at the cold tip. Also, system weight and development costs are reduced. The thermodynamic efficiency of pulse tubes has been shown to be as good or better than Stirling cycles at the same operating conditions. Thermodynamic efficiency can be improved by operating the compressor at a low temperature (down to 35 K) but additional radiator area (and therefore mass) will be required to reject heat from the compressor at this low temperature.

Exported vibration from the compressor is minimized by the opposed motion of the pistons in conjunction with an electronic feed back control loop to cancel the net momentum of the moving masses. The residual vibration levels are 0.1 N in the lateral direction since the current control schemes only control the axial vibration levels. It is possible to reduce this level of vibration but additional development is needed. Some potential areas to improve the induce vibration are: improved alignment and balancing of the opposing pistons to reduce the lateral forces, improved closed loop control of the pistons in the drive direction, and innovative techniques such as 3-axis dynamic balancing.
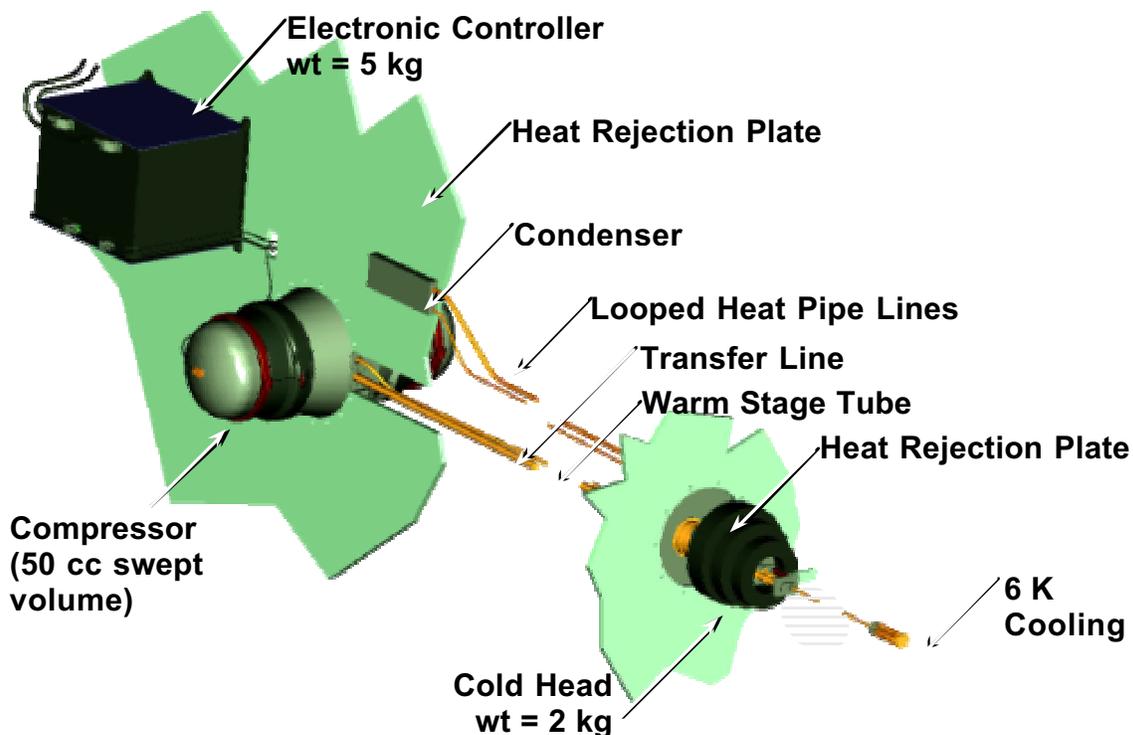
Figure B. Lockheed Miniature Low Temperature Pulse Tube Cryocooler.

Recent data from the Lockheed miniature pulse tube cryocooler program has shown that the pulse tube can be staged allowing access to low temperatures. Component tests combined with simulations have shown cooling down to 5.2 K under no load, or a cooling power of about 60 mW at 10K, for a compressor input power of only 100 W.

<u>Miniature Reverse Brayton Cooler</u>. Turbo-Brayton cooler technology has many excellent features, including essentially vibration free operation, large cooling power per unit mass and volume, high thermodynamic efficiency at low temperatures, and ease of integration. Unfortunately, it has historically not been possible to adequately miniaturize the technology for use in space.

Technical breakthroughs by Creare, Inc. have enabled this technology to be miniaturized while maintaining high thermodynamic efficiency. The cooler consists of three major components, a compressor, a counterflow heat exchanger and a turboalternator (expander). The compressor and the turboalternator use identical technology, namely high speed miniature turbines supported by gas bearings. Miniature high speed turbomachines with self acting gas bearings and low mass shafts will provide the compression and expansion functions without

Figure C. Miniature Turbo Reverse Brayton Cooler. 35K model.

vibration. Rotors of approximately 2 mm - 5 mm diameter are typical and rotational speeds are between 1,000 rev/s and 10,000 rev/s. The compressor turbine typically rotates at up to 1,000,000 RPM. A 70K version of this cooler is currently in use on the NICMOS instrument on the Hubble Space Telescope.

In order to reduce the number of compression stages and the complexity of the machines, the cryocooler must operate at relatively low pressure ratios. These conditions require very high performance recuperative heat exchangers in order to keep system input power at acceptable levels. The use of an efficient counterflow heat exchanger, instead of a regenerative heat exchanger as required by Stirling cycle coolers and pulse tube coolers, enables reverse Brayton cycle coolers to achieve extraordinary thermodynamic efficiency in the 4 to 10 Kelvin temperature range. It is the high thermodynamic efficiency of the turbo-Brayton cooler, along with its very small mass and size, that make this cooler technology so appealing for space missions with detectors and other hardware operating at or below 10 Kelvin. . For cooling loads in the range of 5 mW to 100 mW at temperatures between 4 K and 10 K, cycle power levels are expected to be between 15 W and 200 W.

Figure D. 18 K Sorption Cooler for Planck Mission. The drawing shows a number of the sorption bed compressors and condensers that make up a multistage system.

Sorption Cooler. Sorption coolers, like the turbo-Brayton cooler, use a recuperative cycle. That is, they take advantage of the returning gas to precool the warm compressed gas. They do not rely on a regenerator which must have a high heat capacity. Unlike the turbo-Brayton cooler, compression is done by an absorber, not a moving turbine. The sorption cooler can be sized for very small heat loads. The only moving parts in the refrigerator are check valves, thus the system also has little vibration. A 20K sorption cooler is under development by JPL for the Herschel/Planck mission. (See Figure D.) A breadboard unit has been tested and flight units will be produced by 2004. Development of a 6K version is under consideration for NGST. The final goal is for 10 to 20 mW of cooling power at 6 to 8 K.

Coolers for Very Low Temperature.

Dilution Refrigerator. A helium dilution refrigerator (DR) is the most common method for reaching temperatures between 0.010 and 0.300 K on the ground. The DR relies on the unique properties of liquid He-3 and He-4. Cooling is produced when He-3 atoms cross the phase boundary that exists between liquid He-3 and liquid He-4 at low temperatures. (Essentially, He-3 'evaporates' into the liquid He-4.) The principal advantages of a dilution refrigerator when operated with sorption pumps are: it is small and light, it has no moving parts to wear out, and it produces no vibration. On the ground, gravity provides the force that keeps the two helium liquids in their required places so that the cooling can happen when and where it is needed. In space this force must be replaced with weaker capillary forces that arise when the liquids are confined in porous sponges. It was found, however, that the small pores needed to control large heights of liquid on the ground are too small to allow sufficient liquid flow for effective cooling. It has been shown that it should be possible to develop a helium dilution refrigerator that will confine the liquids with capillary forces and still provide the cooling that makes the dilution refrigerator so valuable. This approach should work even better in microgravity. A shallow single-cycle version of the refrigerator that does not require large heights of liquid to be supported by capillary forces has been built tested.

Herschel/Planck will use a single shot dilution refrigerator currently under development by ESA. Models already produced have about 0.1 microwatt of cooling at 100 mK and require a 10 K cryocooler stage for precooling. These prototypes require large tanks for helium gas storage.

NASA/Ames is working on a concept for a continuously running dilution refrigerator that will produce a few microwatts of cooling at 50 mK for less weight. A 4 K cryocooler stage is required for this design. A zero g demonstration of a prototype is desirable. Figure E shows the NASA/Ames configuration to be tested on the ground. The dilute He-3 flows out of the mixing chamber into the still as before. But the He-3 gas, instead of being pumped from the still into a charcoal pump, now goes to a new chamber, the condenser, at 0.4 K, where it condenses back to a liquid and pure He-3 returns to the mixing chamber. Thus this He-3 never leaves the low-temperature region. As long as the still is heated to maintain its temperature at 0.6 K and the condenser is cooled to maintain its temperature at 0.4 K, He-3 will be continuously pumped from the still into the condenser and forced back into the mixing

Figure E. Details of the low-temperature chambers of the continuously-operating dilution refrigerator.

chamber. This continuous circulation of He-3 will produce continuous cooling in the mixing chamber where He-3 crosses the phase boundary from pure He-3 into the He-4. The condenser can be cooled by two single-cycle He-3 refrigerators, each with its own charcoal pump, and each thermally linked to the condenser by a heat switch. It also can be cooled by a continuous ADR stage running at about 0.4 K. The amount of heat dissipated in this stage is about 800 microwatts for 3 microwatts of cooling at 0.05 K.

If He-3 refrigerators are used, the charcoal pumps for the He-3 pots need to be cooled below 8 K to achieve good pumping. However, the He-3 pots need to be cooled to 2 K or below to efficiently condense He-3. If a temperature this low is not otherwise available, then a pair of He-4 pots and charcoal pumps would be added. These would condense He-4 and reject heat at 4 K, and then be pumped down to 1 K to provide a heat sink for condensing He-3 into the He-3 pots. They would be operated alternately, like the He-3 pots, to provide continuous cooling to a 1 K heat sink.

Adiabatic Demagnetization Refrigerator (ADR). An ADR uses a magnetic refrigeration cycle and consists basically of a high-field magnet, a paramagnetic material (called a "salt pill") and a heat switch. The salt pill is located in the bore of the magnet, occupying the highest field region, and the heat switch allows it to be thermally connected to or isolated from the higher temperature heat sink
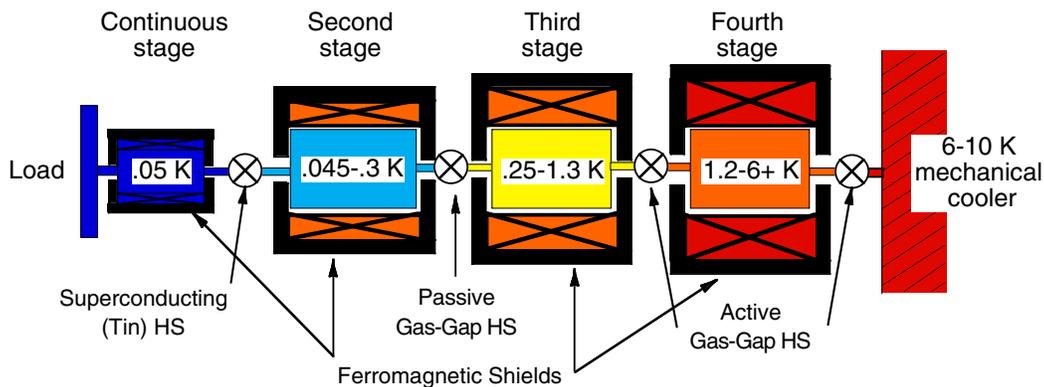
Figure F.   Schematic of a CADR. A fifth stage may be inserted between the third and fourth stages to maintain a 1.3K thermal shield and to maintain the second and third stage magnets at low temperature.

depending on which part of the thermodynamic cycle the refrigerator is in.
The continuous ADR technology being developed at GSFC under the Cross Enterprise Technology Development Program (CETDP) has currently demonstrated 1.5 microwatts of cooling at 0.035 K rising to 15 microwatts of cooling at 0.100 K dumping heat to a SFHe reservoir at 1.3 K. Development during the next two years will allow the heat to be dumped at up to 10 K as well as increase the cooling power by about a factor of 2. See Figure G. Overall efficiency of greater than 25% of Carnot is expected for a system operating between 0.05 and 10 K.

Advantages of an ADR for very low temperature cooling include: no moving parts, high thermodynamic efficiency, and simple operation. It also can be fully demonstrated on the ground; gravity plays no role in its operation. It is however, somewhat heavier than an equivalent dilution refrigerator and it may produce stray magnetic fields from which many detectors must be shielded.

 Cooling a telescope to <4 K with a 6-10 K mechanical cooler may be achieved with high efficiency by using the upper stage of an ADR and a continuous stage to maintain the low temperature. In fact, the overall efficiency of the system will improve greatly over a mechanical cooler alone due to the high efficiency of a CADR. A CADR operating from 10 K to 4 K could have an efficiency of
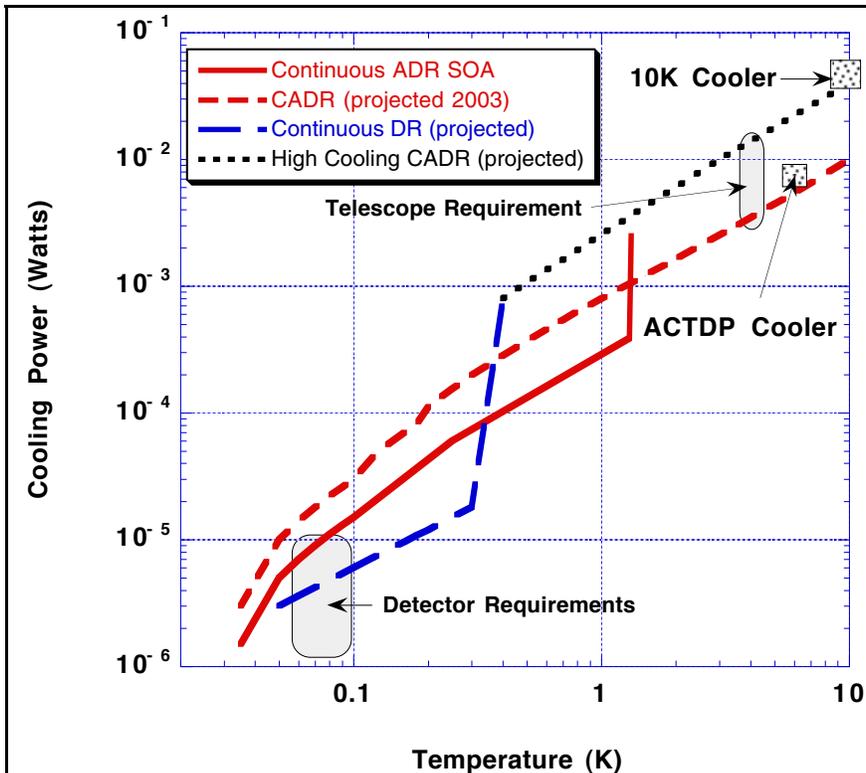
Figure G. Cooling power for continuous dilution refrigerators and continuous adiabatic demagnetization refrigerators currently under development. The heat rejection requirements are compared with the cooler to be developed under the ACTDP and for a higher power 10 K cooler. The cold mass of each of these technologies is less than 10 kg.

greater than 80% of Carnot, while a mechanical cooler's efficiency over this range decreases drastically due to the rapid change of thermodynamic properties of the helium gas working fluid.

Normal/Insulator/Superconductor (NIS) Cooling
A new technique using a superconducting analogue to thermoelectric cooling is being developed. By sweeping electrons in a normal conductor across an insulating voltage barrier and allowing the electrons to pair up in a superconductor one can achieve cooling on a small scale – large enough to be able to directly cool detectors from about 0.3 to 0.1K. In conjuction with an ADR or ^{3}He sorption cooler this could be attached to a mechanical cooler. This type of cooler for single pixels is being developed in Finland and for large arrays at Lawrence Livermore National Laboratory (LLNL) under CETDP funding. The larger LLNL NIS device will be able to cool 2 nanowatts at 0.1K while rejecting 20 nanowatts of heat at a temperature of 0.3 K. The development is expected to result in a laboratory demonstration of detector cooling by the end of 2003.

Cryo/Thermal Summary.
The cryogenic and thermal components necessary to accomplish a large long-life sub millimeter mission are currently in, or proposed for, development. If the funding is available, the necessary development will be completed in the next 10 years. Development of mechanical coolers, very low temperature coolers and improved thermal systems will significantly lower the system mass and make the mission more affordable. System level studies of how these components will work together and the optimization of various parameters like weight and complexity remain to be done.

A Continuous Adiabatic Demagnetization Refrigerator for Far-IR/Sub-mm Astronomy

Peter Shirron[1], Edgar Canavan[1], Michael DiPirro[1], Michael Jackson[1], Todd King[2] and James Tuttle[1]

[1]Code 552, [2]Code 541
NASA/Goddard Space Flight Center
Greenbelt, MD 20771

Abstract. We report on recent progress in the development of a continuous adiabatic demagnetization refrigerator (CADR). Continuous operation avoids the constraints of long hold times and short recycle times that lead to the generally large mass of single-shot ADRs, allowing us to achieve an order of magnitude larger cooling power per unit mass. Our current design goal is 10 µW of cooling at 50 mK using a 6-10 K heat sink. The estimated mass is less than 10 kg, including magnetic shielding of each stage. The relatively high heat rejection capability allows it to operate with a mechanical cryocooler as part of a cryogen-free, low temperature cooling system. This has the advantages of long mission life and reduced complexity and cost. We have assembled a three-stage CADR and have demonstrated continuous cooling using a superfluid helium bath as the heat sink. The temperature stability is 8 µK rms or better over the entire cycle, and the cooling power is 2.5 µW at 60 mK rising to 10 µW at 100 mK.

1.0 INTRODUCTION

To meet the growing need for more capable low temperature refrigerators for space astronomy missions, we are developing an magnetic refrigerator that operates continuously at low temperature [1,2]. It is a multi-stage system that uses one stage to directly cool a load, and others to periodically cascade heat up to a heat sink. Since the heat transfer process does not disrupt cooling, it can be performed on a relatively short time scale (~1 hour). This means each stage requires proportionally less refrigerant mass than a single-shot ADR [3] would need. We project that a CADR providing 10 µW of cooling at 50 mK using a 6-10 K heat sink will have a cold mass less than 10 kg.

2.0 DESIGN

The design of an ADR depends most critically on the operating temperature, the cooling power and the heat sink temperature. For future space astronomy missions using large-format, low-temperature detectors, cooling powers of up to 10 µW and temperatures as low as 50 mK are expected to be needed. Heat sink temperatures will range from about 1 K to as high as 6-10 K. This range reflects the possible pre-cooler options, from superfluid helium dewars to mechanical cryocoolers. The latter are now being baselined for a variety of upcoming missions, including Constellation-X, so it is important to begin developing low temperature coolers that can accommodate their use. The impact of such high heat rejection temperatures is actually quite
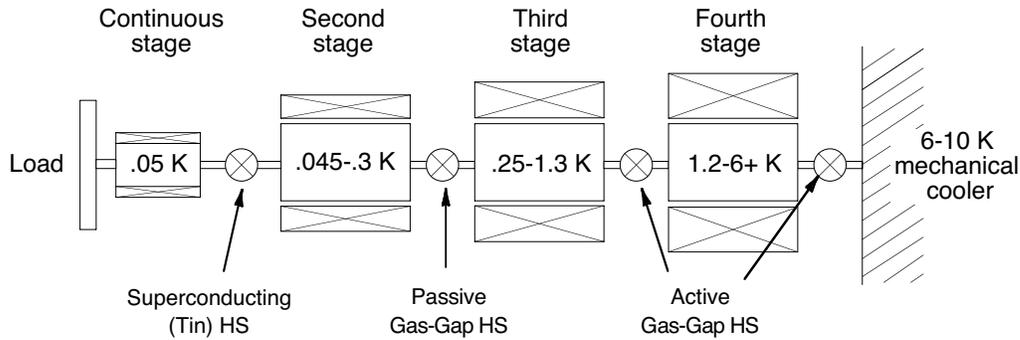
FIGURE 1. Schematic of a four-stage CADR capable of operating between 50 mK and 6-10 K.

significant since it is virtually impossible for single-stage ADRs to operate over such a wide temperature range. The multi-stage architecture of our CADR, allows us to meet this requirement with relative ease. A linear chain of ADR stages, each typically spanning a factor of 5 in temperature, can in principle operate over an unlimited temperature range.

The preliminary design for a CADR capable of operating between 50 mK and 6-10 K is shown in figure 1. Its four stages each consist of a cylindrical "salt pill" (containing the refrigerant), a magnet, magnetic shielding and a heat switch. The temperature range of each stage is determined by the properties of the heat switches. A superconducting tin switch is used between the continuous and second stages because it is one of the few types that can conduct well at temperatures below 100 mK. However its off conductance rises very rapidly with temperature, requiring the second stage to be kept below about 0.3 K at all times. Consequently the switch linking the second and third stages must be conductive at temperatures below 0.3 K. This proved challenging, but ultimately possible for a passive ^{3}He gas-gap switch [4]. The third and fourth stages reject heat through conventional getter-activated ^{3}He gas-gap switches.

To date, we have constructed a three-stage CADR and tested it with a superfluid helium heat sink. The as-built refrigerant masses and magnetic field strengths shown in Figure 2. Stages 1 and 2 use chrome potassium alum (CPA) refrigerant because its entropy capacity is higher than ferric ammonium alum (FAA) at temperatures below 50 mK. Its low ordering temperature will
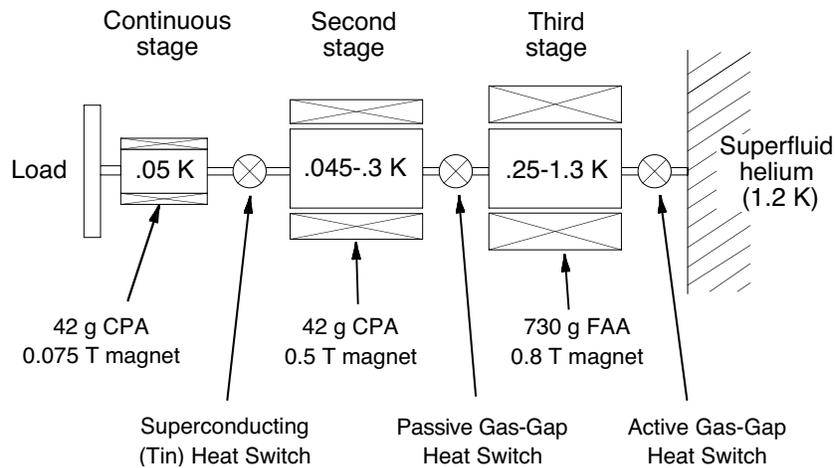
FIGURE 2. Schematic of a three-stage CADR with as-built component parameters.
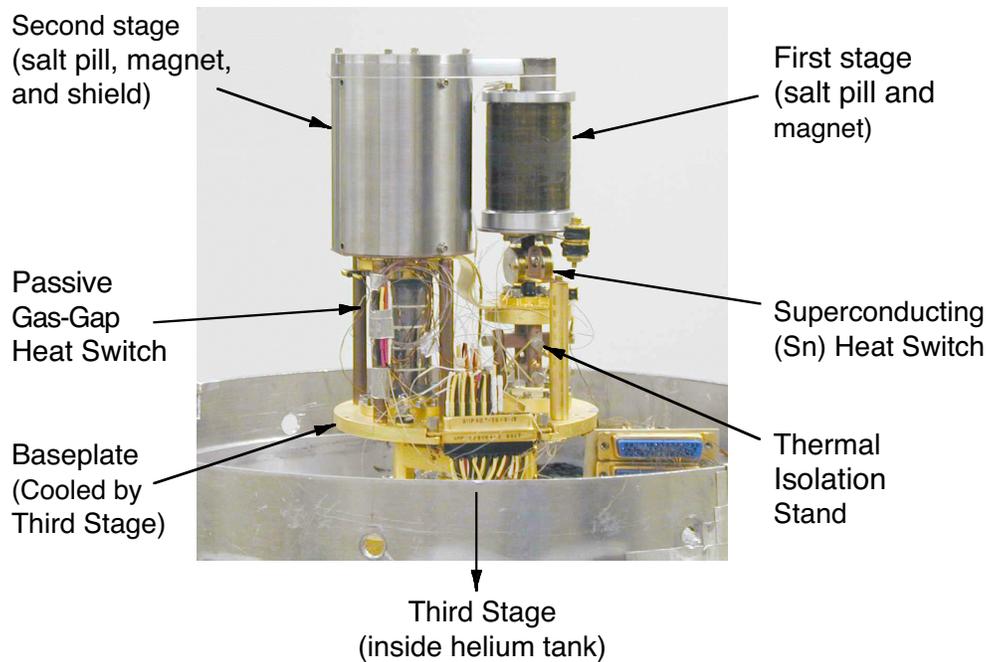
FIGURE 3. The first (right) and second (left) stages of a three-stage CADR in test configuration.

also allow the ADR to reach temperatures well below 50 mK in the event that such capability is of practical use. The third stage, using ferric ammonium alum (FAA), is an engineering unit ADR produced for the X-Ray Spectrometer mission [3], and is much larger than necessary. In the next version of the CADR, it will be replaced by an 80 g CPA salt pill, identical to the second stage salt pill, and a 1.5 T magnet. Figure 3 is a picture of the CADR in the test dewar. The third stage and its magnet are located inside the helium tank of the dewar and are therefore out of sight beyond the bottom of the picture.

3.0 OPERATION

The CADR has two operational modes. The first involves cooling down from the heat sink temperature and establishing temperature control. The second is the periodic cycling of the upper stages to maintain constant temperature.

To begin, all stages are magnetized to full field with the heat switches turned on. Starting with the third, each stage is sequentially demagnetized (after turning the appropriate heat switch off) to the low end of its operating range. The rate is not critical but should be slow enough to efficiently cool the lower stages. In the final state, the continuous stage is near full capacity and is actively cooling the load. A control routine then takes over to automatically recycle each stage as needed.

Recycling consists of demagnetizing the adjacent upper stage to a lower temperature and closing the heat switch. The temperature controller automatically magnetizes the lower stage as the upper stage absorbs its heat. When the magnetic field reaches an upper threshold, the heat switch is opened, and the upper is magnetized to the high end of its range to reject the heat to the next stage. The process is repeated until the uppermost stage rejects heat to the heat sink.

4.0 CADR PERFORMANCE

The key performance areas for a low temperature cooler are operating temperature, temperature stability, cooling power and efficiency.

4.1 Operating Temperature

To date, the CADR has operated continuously at temperatures down to 35 mK and up to 200 mK. Higher temperature operation is possible, but because of minor constraints imposed by the passive gas-gap heat switch, some software changes would be necessary in order to preserve high efficiency. These will be implemented as the need arises.

4.2 Temperature Stability

Temperature control is more of a concern for the CADR because of the need to periodically reverse heat flows into and out of the first stage. The control software minimizes the disturbances by such techniques as only turning the superconducting heat switch on and off when the temperature difference between the first and second stage is close to zero, and limiting the rate at which the heat flows can change. Another important consideration is for the temperature controllers to have fairly high bandwidth (~10 Hz). With both analog and digital controllers the CADR can achieve control that is limited only by the noise of the temperature readout. Figure 4 shows the temperature and magnetic field of the continuous stage during one cycle conducted at 100 mK. The magnetic field has the characteristic sawtooth pattern as heat is alternately absorbed and rejected, however there is no apparent correlation between temperature fluctuations and any recycling events. The noise is 8 μK rms or better throughout the cycle.

4.3 Cooling Power

The CADR's cooling power shown in Table 1. The cooling power is the maximum *sustainable* heat load that can be applied to the continuous stage. Although the CADR can tolerate larger momentary loads, when the heat load remains in excess of these values, the
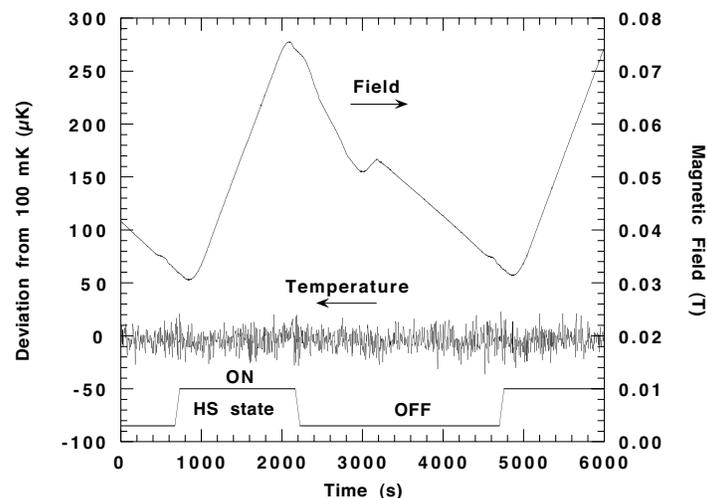
FIGURE 4. Temperature and magnetic field of the continuous stage during a complete cycle.

TABLE 1. Demonstrated Cooling Power.

Temperature (K)	Maximum Sustained Heat Load (µW)
0.10	15
0.09	13
0.08	11
0.06	7
0.05	5
0.035	1.5

continuous stage eventually runs out of cooling capacity and loses temperature control. The heat load is approximately linear in temperature above 50 mK. The non-linearity below 50 mK is due to several factors, the most important of which appear to be the rapidly increasing thermal boundary resistance between the salt and the thermal bus (which significantly reduces heat transfer efficiency), and reduced entropy capacity as the ordering temperature is approached.

4.4 Thermodynamic Efficiency

We have made careful measurements of the CADR's efficiency at both 50 and 100 mK with applied heat loads as given in Table 1. The measurement includes all parasitic effects (such as eddy current heating, magnet and magnetic shield hysteresis, heat switch and suspension system parasitics, and thermal gradients within the salt pills) except for the following: we have neglected heat that is dissipated in the third stage magnet as well as the heat applied to turn on the third stage heat switch, and we have neglected dissipation in the room temperature electronics. The reason is that these components are not yet representative of the final design, and the dissipation is consequently much higher than we eventually expect. With this in mind, we regard the measurements of 50% efficiency at 100 mK and 12% at 50 mK as being the right order of magnitude for a complete system. Granted the efficiency will decrease when the neglected terms are added, but there is also some room for improvement, for example by thermally anchoring all magnets and shields to the heat sink rather than to the third stage, as is presently the case for the second stage magnet.

5.0 SUMMARY

We have constructed a three-stage CADR that operates continuously at temperatures down to 60 mK with high cooling power. The system achieves 8 µK rms temperature stability throughout its cycling operations, and cooling powers of 2.5 µW at 60 mK rising to 10 µW at 100 mK. In the near term we will construct a fourth stage that will allow the system to operate with a 4.2 K helium bath or eventually with a mechanical cooler running at 6-10 K. We are also working on designs for suspension systems and magnetic shielding so that the system can be packaged more compactly. Our present layout for a four-stage cooler operating with a 4.2 K heat sink will fit within a 2-liter volume and weigh less than 7 kg.

6.0 ACKNOWLEDGMENTS

7.0 REFERENCES

1. Shirron, P.J., Canavan, E.R., DiPirro, M.J., Tuttle, J.G., and Yeager, C.J., "A Multi-Stage Continuous Duty Adiabatic Demagnetization Refrigerator," in *Adv. Cryo. Eng.* 45B, edited by Q.-S. Shu et al., Plenum, New York, 2000, pp. 1629-1638.
2. Shirron, P.J., Canavan, E.R., DiPirro, M.J., Jackson, M., King, T., Panek, J., and Tuttle, J.G., "A Compact, High-Performance Continuous Magnetic Refrigerator for Space Missions", presented at the 2001 Space Cryogenics Workshop in Milwaukee, WI, accepted for publication in *Cryogenics.*
3. For descriptions of one- and two-stage ADRs, see Serlemitsos, A.T., SanSebastian, M., and Kunes, E., *Cryogenics* **32**, 117 (1992), and Hagmann, C. and Richards, P.L., *Cryogenics* **34**, 221 (1994).
4. Shirron, P.J., Canavan, E.R., DiPirro, M.J., Jackson, M., Panek, J., and Tuttle, J.G., "Passive Gas-Gap Heat Switches for Use in Adiabatic Demagnetization Refrigerators," presented at the 2001 Cryogenic Engineering Conference, Madison, WI, accepted for publication in *Adv. Cryo. Eng.* 47.

Session IX

Control/Metrology/Orbit Technology

TETHERED FORMATION CONFIGURATIONS: MEETING THE SCIENTIFIC OBJECTIVES OF LARGE APERTURE AND INTERFEROMETRIC SCIENCE

Rodger E. Farley[1] and David A. Quinn[2]

Abstract

With the success of the Hubble Space Telescope, it has become apparent that new frontiers of science and discovery are made every time an improvement in imaging resolution is made. For the HST working primarily in the visible and near-visible spectrum, this meant designing, building and launching a primary mirror approximately three meters in diameter. Conventional thinking tells us that accomplishing a comparable improvement in resolution at longer wavelengths for Earth and Space Science applications requires a corresponding increase in the size of the primary mirror. For wavelengths in the sub-millimeter range, a very large telescope with an effective aperture in excess of one kilometer in diameter would be needed to obtain high quality angular resolution. Realistically a single aperture this large is practically impossible. Fortunately such large apertures can be constructed synthetically. Possibly as few as three 3-4 meter diameter mirrors flying in precision formation could be used to collect light at these longer wavelengths permitting not only very large virtual aperture science to be carried out, but high-resolution interferometry as well. To ensure the longest possible mission duration, a system of tethered spacecraft will be needed to mitigate the need for a great deal of propellant. A spin-stabilized, tethered formation will likely meet these requirements. Several configurations have been proposed which possibly meet the needs of the Space Science community. This paper discusses two of them, weighing the relative pros and cons of each concept. The ultimate goal being to settle on a configuration which combines the best features of structure, tethers and formation flying to meet the ambitious requirements necessary to make future large synthetic aperture and interferometric science missions successful.

Introduction

It follows intuition as well as optical theory that the larger the diameter of a telescope's mirror, the greater the amount of light that can be captured. That satisfies the quantity side of the equation, but as far as the quality side of the equation goes, a larger diameter mirror also implies a sharper image, which we would also intuitively expect. Here, quantity and quality have been used as euphemisms for sensitivity and resolving power. When faced with the desire for fine resolution and sufficient sensitivity to detect small dim objects (especially at long wavelengths), we find ourselves wanting impossibly large space observatories, and so we ask the question; is it possible to achieve both the quantitative and qualitative science objectives with realistically sized hardware? Fortunately, methods have been pioneered in the field of synthetic apertures, where small subapertures achieve resolutions comparable to a large mirror having a diameter equal to the spacing between the subapertures. The sensitivity is achieved by 'staring' at the target for a longer period of time, sampling different parts of the synthetic aperture. An exciting prospect is to merge the methods of a synthetic aperture with an imaging spectrometer optical system. The space science mission SPECS (Sub-millimeter Probe of the Evolution of Cosmic Structures; Mather et al. 2001) proposes to combine far-infrared interferometry in a synthetic aperture 1000 meters in diameter to produce a space platform imaging spectroscope. The heart of the instrument is a cryogenic Michelson interferometer with a stroking optical delay line. This type of optical set-up can detect spatial structures having time-invariant (during the observation), incoherent sources and has the ability to measure two basic properties:

1) Spatial brightness distribution

---

2) The spectrum in each resolution element (detector pixel)

The distance (*baseline*) between the subapertures makes the interferometer sensitive to a corresponding angular size of spatial structure. Interference fringe patterns (*interferograms*) are produced during the stroke cycle, which proceeds at a constant scan rate. The level of contrast (*fringe visibility*) between the light and dark fringes implies the amount of spatial structure (the brightness) at the angular size to which the interferometer is sensitive. Information contained in the shape of the fringe pattern corresponds to the spectrum of the object. The fringe shape is measured by varying the optical delay, and therefore during the stroke the intensity at a pixel location varies with time. This pattern is Fourier inverted to obtain the spectrum of the source. The synthetic aperture is sampled with three subapertures connected together using tethers to slowly spiral in like a record player, completely covering the plane of the aperture (*the u-v plane*). The signals and subaperture locations in the synthetic aperture plane are Fourier transformed to create a sky map image with a spectrum of a particular brightness in each pixel. The image resolution is proportional to the synthetic aperture diameter, and the spectral resolution is proportional to the length of the stroking optical delay. The subaperture diameter of 3 meters limits the field of view, but SPECS will use a mosaic technique to produce wide-field images. The data needed to produce wide-field images are obtained when individual fringe patterns fade in and out over pixels in the detector array as the geometric delay corresponding to looking off-axis is compensated by the interferometer's optical delay. This occurs as the zero-path lengths of the sources are reached with the stroking of the delay line. The synthetic aperture is fully sampled by altering the baseline such that all angular sizes of interest are 'exposed.' In this way, all the components of the Fourier decomposition are measured which are used to produce the sky image.

Spinning tethered spacecraft formations hold the promise of making a SPECS observatory and others with a similar imaging need possible. The spin produces a centripetal acceleration, which keeps a tethered arrangement tight. This opens up tethered arrangements to the imagination, where tethered creations could resemble string-art-in-space, the 'artificial-gravity' of the spin maintaining tension and shape. Slow spin rates producing sub-Newton tension forces is all that is required to maintain shape. Precision to the level of interferometry is achieved by using optical path-length corrector mechanisms to make up for the small residual motions. Feedback for the corrector mechanism could come from a source in the field of view, where the 'white light' wave front maximum is detected.

There is a stark contrast between a spinning tethered formation and a free-flying formation in the area of fuel consumption. The tethered arrangements maintain a constant angular momentum, 'reeling' in or out their tethered lengths with winch-like mechanisms to fully sample the synthetic aperture by spiraling in and out. There is no use of fuel except to minimize perturbations. Only in repointing the formation by precession is fuel consumed. By contrast, a free-flying formation sampling in the same spiral manner for a SPECS mission would consume a fuel mass that was 3 times greater than the mass of the 'empty' spacecraft, just for one observation. The thrusters would have to perform the mechanical work equivalent to that done by tension in the tethers. Clearly, free flying formations for this kind of aperture sampling are completely impractical.

Spinning tethered arrangements can fit into two major categories; those in which the component spacecrafts are connected radially using a single tether (pendulum mode), or, those which have multiple attachments, allowing triangular or other 2-D geometrical shapes to define the arrangement. Complex arrangements can be fashioned with tension-compression structures that can be utilized to create 3-D shapes. But simplicity is a goal worth striving for, and the 2-D arrangements would have a much better chance of surviving intact after the building phase, the assembly phase, and the deployment phase, not to mention launch mass constraints. Key features of a 2-D spinning space tether arrangement are the over-all light mass of the system considering the immense size, using only centripetal force and strings, and it's open-loop dynamic stability. The beauty is the simplicity.

Configurations

Maintaining a constant angular momentum for the tethered system again minimizes the need for propulsion, but one problem with tethered arrangements having only subaperture masses, is the 'ballerina' effect. As the masses move radially closer to the center, the spin rate increases by one over the square of the radius, moving the subapertures faster than the sampling time required to minimize image blur. Thus, to tame the spin speed by not using propellant, one must use counter masses in the tethered arrangement. Such is the case for the two configurations studied. They each have 3 subapertures, and 3 counter masses.

SPECS Hex

This arrangement is a pendulum type, with a central hub truss structure approximately 17 meters in radius. As the pendulum-connected subaperture reels out, the counter mass reels in with a 1:1 ratio. The radial reel-in speed required to sample the aperture plane in $\sim 10^5$ seconds, is 2.5 mm/s with an initial rotational speed of 0.0165 RPM. This produces a Coriolis acceleration in which the reeled-in element will lead the hub attachment point by an angle, and the reeled-out element will lag by the same angle. That way, the subaperture and countermass tethered elements approach each other by twice the angle in the spin plane. The central hub radius was picked so as to keep the total approach angle (lead + lag) no greater than 20 degrees. The initial radial length of the tethers is 577 m from the hub center, which is a baseline of 1000 m.

Figure 1 SPECS HEX
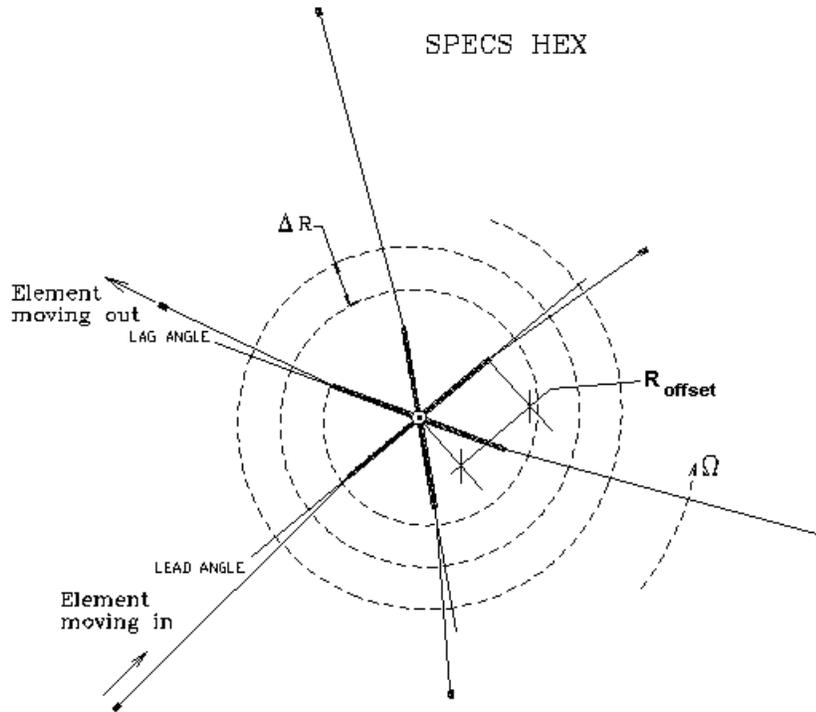
SPECS Tetra-Star

Use is made of triangles as the basic shape to give the tethered arrangement in-plane shape rigidity. The 3 counter masses are located at the apex of the outer 3 triangles, in which the two legs of each outer triangle are of a constant length tether. The base of the outer triangles form the inner triangle, where the subapertures are located in each corner. The length of this tether constitutes the interferometer baseline, and it is these three tethers which change length to spiral in and out, to fully sample the u-v plane. Each of the 3 inner tethers has a tether-winch on one end, and a passive in-line spring/damper suspension system on the other end. The triangular configuration allows passive damping to be used to great effect. For a standard observation, the reel-in rate of the inner tethers is 4.28 mm/s, with an initial rotational speed of 0.0165 RPM.

Figure 2
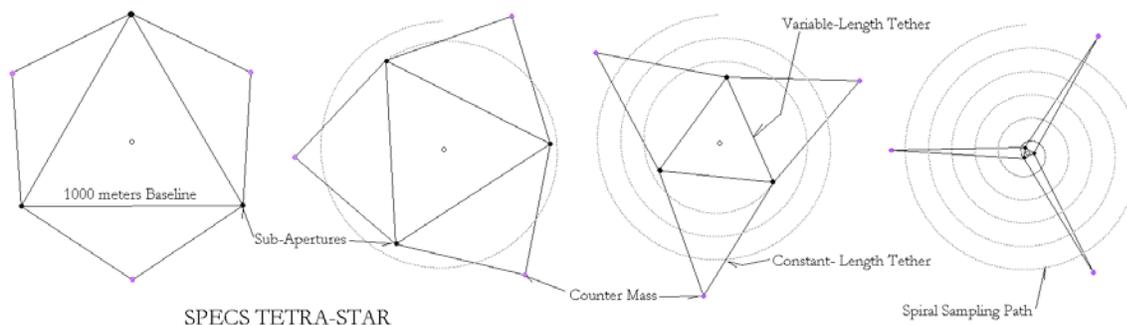
Properties

The tether material for the configurations was baselined as having a Young's modulus of 12 million psi, with an equivalent diameter of 1 mm.

The mass of the subapertures, were optimistically selected to be 500 kg each, while the counter masses were set to 150 kg. The actual SPECS subaperture is expected to be of much greater mass. It can be expected to have the following components:

1) A 3 or 4 meter diameter cryogenically cooled flat mirror.
2) Two 4 meter diameter thermally isolated / insulated stray-light baffle tubes.
3) Twelve 3-stage Adiabatic De-magnetization Refrigerators (ADR) connected to
4) Twelve cryo pump coolers, all heat sinked via cryo loop heat pipes to
5) A large thermal radiator.
6) Structure, tether mechanisms, perhaps an articulation mechanism for the mirror.
7) Sun shade with a solar array on the sunny side, the rest of the s/c in the dark.
8) Electrical system, Thruster system, Attitude control system, Communication system.

General Requirements for SPECS Tethered Configuration

Spiral sampling will cover the synthetic aperture plane without gaps down to a minimum baseline. This can be accomplished by spiraling in, spiraling out, or a combination of the two maneuvers.

The sampling baseline (length from subaperture to subaperture) will vary from a minimum of 50 meters to a maximum of 1000 meters.

Number of subapertures is 3.

The subapertures are folding-flats, having a 3 to 4 meter projected diameter.

To avoid image blur, maximum tangential velocity of the subapertures due to the spin is 1 m/s. For time efficient observations, this tip speed should be kept as close to 1 m/s during the entire sampling sequence.

Angular stability of reflecting mirror relative to the other subaperture mirrors during an observation is 1 arc-second.

Temperature of the mirrors and any part of the system in the field of view =< 4 deg K (SPECS).

Tethered bodies should have a combined [radial plus out-of-plane] displacement < 0.5 meters. At this range, the path length corrector mechanism can operate.

Knowledge of the position of the tethered subapertures: to a fraction of a wavelength, or about 10 micrometers in the case of SPECS.

The number of observations in a mission = 1500

The average re-point change between observations = 5 degrees

The total re-point angle for a mission = 7500 (5x1500) degrees

Dynamic Modeling

The equations of motion were derived using Kane's method with the aid of AutoLev, a symbolic manipulation software designed specifically for solving dynamics problems. The SPECS Hex model has 24 degrees of freedom (dof), and the SPECS Tetra Star has 18 dof. Both models used point masses to represent the subapertures and counter-mass spacecraft, and the tethers were modeled as massless springs. Considering that the tension is never greater than several Newtons, the typical tether mass is perhaps on the order of 1% of the counter mass or subaperture mass, so a massless assumption does not lead one far astray. However, this does mean that the transverse tether vibrations will not be formulated. The transverse vibrations are estimated to have a base period of 10 seconds (guitar mode), where as the spin period is between 15 and 60 minutes. The equations were numerically integrated using a 5th order Runge-Kutta with a constant time increment and animated in Visual Basic software. Even though simplifying assumptions were used, the computer code generated is extremely complex, especially the damping terms. AutoLev provides subroutines of the equations of motion in either Fortran or C. In the case of this study, the equations had to be manually altered to convert the Fortran code into Visual Basic code, with most of the effort in reformatting the line continuances.

Maneuvers

Observation scan

An observation scan entails spiraling in such a manner as to cover the u-v plane. The u-v plane is a transformation from the physical aperture plane to one non-dimensionalized by the wavelength λ. If x and y are the physical coordinates in the aperture plane, then:

$$u = (x_2 - x_1)/\lambda \quad , \quad v = (y_2 - y_1)/\lambda$$

where x_1, y_1 is the location of subaperture 1, and x_2, y_2 is the location of subaperture 2.

To physically cover the entire aperture, the spiral spacing ΔR, should be a constant. See fig. 1.

$\Delta R = D_{sub} \cdot N_{sub}$ where D_{sub} is the diameter of the subaperture, and N_{sub} is the total number of subapertures.

Spiraling occurs as the tethers are reeled-in while the constellation is rotating at the instantaneous rate Ω. The radial reel-in rate dR/dt is designated as Rdot. To make the spiral spacing a constant for efficient observations, it implies that the ratio Rdot / Ω is a constant. Indeed:

Rdot = $\Omega \cdot \Delta R / (2\pi)$. Rdot will vary during the observation, as Ω will vary with the radial position of the subapertures and counter-masses in this constant angular momentum system. The way to keep ΔR a constant, is to vary Rdot by this relationship:

$Rdot_I = Rdot_O \cdot (I_{spinO} / I_{spinI})$, where $Rdot_O$ is the initial value of Rdot, and I_{spinO} is the initial value of system spin inertia, I_{spinI} is the instantaneous system spin inertia, and $Rdot_I$ is the instantaneous value of Rdot. If the tethered system spin inertia I_{spin} is parameterized by R, the subaperture radial distance from the system center of mass, and H is the system angular momentum magnitude about the center of mass, then:

$Rdot(R) = H \cdot \Delta R / [2\pi \cdot I_{spin}(R)]$

The minimum time (seconds) it takes to complete one fully sampled spiral observation, when the rule for keeping ΔR a constant is practiced, can be calculated from this integral:

$$T_{observation} = \frac{2 \cdot \pi}{H \cdot \Delta R} \cdot \int_{R_{inner}}^{R_{outer}} I_{spin}(R)\ dR$$

Where R_{inner} is the innermost radial distance that the subapertures will be reeled to, and R_{outer} is the maximum radial position.

With the SPECS-hex configuration, the rigid hub structure requires a minimum structural radius R_{offset} to keep the Coriolis-induced angular deflections down to a required value during the observation. If the reel-in rate of the subapertures is equal to the reel-out rate of the counter-masses, then the Coriolis lead angle Ψ_{lead} will equal the Coriolis lag angle Ψ_{lag}. This means that a

subaperture and counter-mass will approach each other by the angle 2 Ψ in the aperture plane. Therefore, if Ψ_{req} is the maximum angle tolerable, then the required offset radius is:

$R_{offset} = \Delta R / (\pi \cdot Sin(\Psi_{req}))$

One can, in effect, determine the reel-in rate whereby this configuration will 'wrap itself around the axle'. If the lead and lag angles are 30 degrees, then all the tethers will converge together. For the baseline SPECS-hex configuration, that will happen when Rdot = 7.3 mm/s. This calculation demonstrates the extreme sensitivity of pendulum style tethered arrangements.

In-plane transient motion occurs with the stop-start action of the reel mechanism. This effect can be minimized with a ramp-up, ramp-down time, perhaps lasting several cycles of the axial vibration mode (1/3 to 1 Hz). In-line dampers can attenuate the residual motions to an acceptable level.

Re-pointing

One method of repointing that has been suggested (but rejected) is to reel-in all the spinning masses, rigidly hook-up, then use thrusters to de-spin. After de-spinning, then re-point, re-spin, and re-deploy. This is too complex, and too costly in terms of fuel and time. An alternative that barely uses fuel is to precess the spinning disk like a top. It is quite possible to precess a spinning constellation of tethered objects in such a way that all the objects move as though they were one rigid object. The tethers need to be a small fraction of the total spin inertia, which in this case they are. To illustrate with a mental experiment, let us consider a rigid spinning disk with an axle in the center. We take a marker and draw a random grid pattern over the disk, symbolically dividing the disk into many individual elements of various sizes. We apply a constant lateral moment to the axle, and the spinning disk precesses at a constant rate. Now, what were the forces involved to keep all the elements moving in concert with one another, at the same rate? Each element must 'feel' a perturbing force perpendicular to the disk, proportional to the mass of the element M_j in order to keep up with the other elements. Also, to keep up angularly, the force must be proportional to the radial distance from the center R_j. In addition, the applied moment on the spinning disk produces a harmonically varying force amplitude, with a period equal to the spin period. Each element has a phase angle ϕ_j corresponding to its relative angular position in the tethered arrangement.

Mathematically, the thrust required perpendicular to the spinning plane for each element j, is thus:

$Thrust_j = M_j \cdot R_j \cdot C_T \cdot sin(\Omega \cdot t - \phi_j)$, Newtons

Where Ω is the spin rate, radians/s and ϕ is the constant configuration phase angle for the element, radians. See Figure 3.

C_T is a 'specific thrust constant' with units of sec^{-2}

$C_T = \theta_p \cdot \Omega^2 / (\pi \cdot N_{revs})$

Where θ_p is the re-point angle to precess, radians.
N_{revs} is the number of revolutions in which to complete the maneuver. This does not have to be a whole number.

The precession rate $d\theta_p /dt = C_T / 2\Omega$, radians/s

Figure 3

Element phase angles in the Tetra-Star configuration

s refers to the subapertures

c refers to the counter-mass
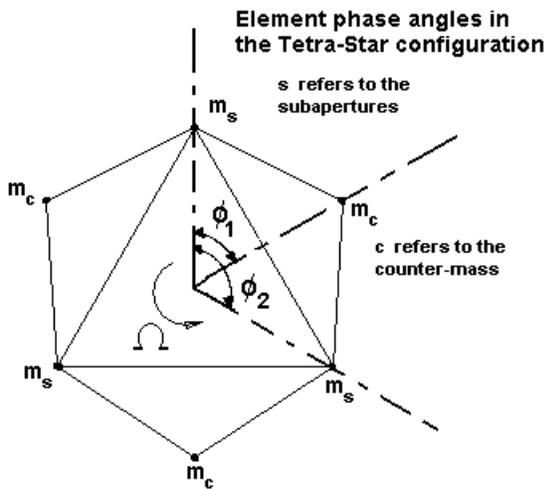
If the above set of precession rules are followed, then repointing can occur smoothly and quickly. The instantaneous spin plane normal direction is determined by averaging the vectors produced by this cross product:

$$\overrightarrow{Normal} = \frac{\sum\limits_{j=1}^{N} (\vec{R}_j - \vec{R}_{cm}) \times (\vec{V}_j - \vec{V}_{cm})}{N}$$

where cm refers to the system center of mass, and N is the total number of tethered elements. This is the instantaneous direction in which the thrusters are fired, the thrust magnitude varying harmonically.

Fuel Usage

The ratio of fuel mass to empty mass (spacecraft, less consumable fuel for repointing) equation has been derived for a tethered element, and has the following form:

$m_{fuel} / m_e = [e^{(R \cdot \Omega \cdot \theta p / (g \cdot Isp))} -1]$ where Isp is the specific impulse for the fuel, and g is the acceleration of gravity, 9.81 m/s^2. This is a simplified equation for pulse-thrusting the elements when they are at the correct location/phase. It underestimates the fuel for the real case of continuously varying thrust.

Another wonderful little equation that just slightly overestimates the fuel usage (4% over for the total mission, quite accurate for short duration burns) is:

$$M_{fuel} = \frac{4}{\pi} \cdot \frac{\theta_p \cdot \Omega}{Isp \cdot g} \cdot \left(\sum_{j=1}^{N} M_j \cdot R_j \right)$$

Where, the Mj and Rj are individual instantaneous element mass (fuel + empty) and radius in the configuration.

For a constant angular momentum system, a determination can be made as to the most efficient location for the elements in order to minimize the fuel usage per radian of re-point angle. For a single tethered element maintaining a constant angular momentum, the larger the radius R, the less fuel is consumed to tilt the spinning plane a given angle. The ratio of fuel mass to empty mass equation for constant angular momentum has the following form:

$m_{fuel} / m_e = [e^{(k \cdot \theta p / (R \cdot g \cdot Isp))} -1]$ where $k = \Omega R^2$, a constant of the initial condition. See Figure 4.

Using hydrazine as a fuel, Isp = 220 seconds, and the total re-point angle = 7500 degrees, for a <u>single</u> tethered element:

Figure 4 Single tether fuel usage to process 7500 ^{o}

Fuel mass / empty mass ratio
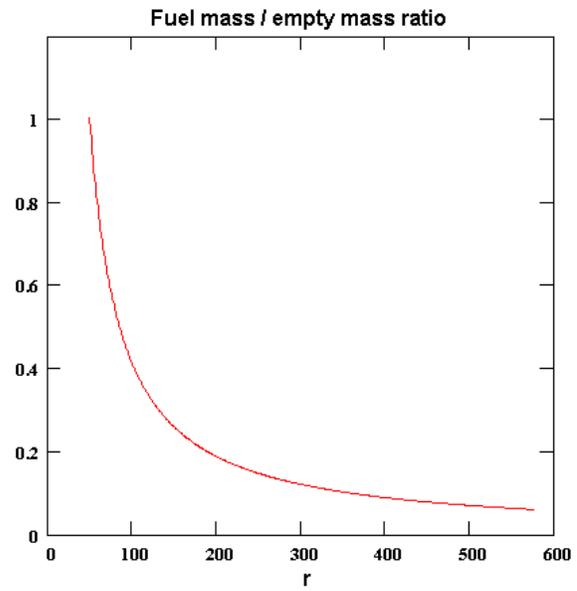
For the more complex case of a harmonically varying thrust, the just slightly over-estimating formula for fuel mass consumed is:

$$M_{fuel} = \frac{4}{\pi} \cdot \left[\frac{\theta_p \cdot H}{(Isp \cdot g)} \right] \cdot \frac{\left(\sum_{j=1}^{N} M_j \cdot R_j \right)}{\left[\sum_{j=1}^{N} M_j \cdot \left(R_j \right)^2 \right]}$$

where H is the angular momentum of the tethered system.

For a configuration like the Tetra Star, the change of radial positions between the subapertures and counter-masses progresses in a non-linear fashion, and a system determination of the fuel usage must be made. Figure 5 addresses this case.

Fuel mass consumed for 7500 degrees of precession for the Tetra-Star configuration, for various inner tether lengths at constant system angular momentum and 2100 kg initial system mass:

Figure 5 Tetra Star mission re-point fuel consumption.

The fuel mass to total mass fraction varies from 7.5% to 9.5% to accomplish the total mission-repointing requirement with the Tetra-Star configuration. If the baseline length during repointing is kept at 1000 meters, then 165 kg of fuel is consumed by the subapertures and counter masses combined, out of an initial system mass of 2100 kg. The worst configuration occurs at a baseline length of 330 meters, where 205 kg of fuel are required.

Simulation Results

The preliminary open-loop dynamic motion results for the Tetra-Star configuration are presented here. The simulations are separated into 2 categories: the observation mode, and the re-point mode. During the observation mode, the reel-in simulations determined the tension, tangential speed, coverage of the synthetic aperture plane (and thus u-v plane coverage), in-plane transient motions, and the total elapsed time for an observation. The initial conditions for the simulations did not include tether pre-stretch due to centrifugal force (in real life they would). This created an unrealistic initial transient 'breathing mode' in the tethered arrangement. However, this tested and illustrated the effectiveness of the passive damping shock struts at the tether's terminal attachment point. To spiral from a baseline of 1000 meters to 50 meters took 34.3 hours for an observation, exactly as calculated from the formula developed.

Mass of mission fuel to repoint

Fuel mass, kg

$M_{fuel}(L)$

L
Tether sub-aperture baseline, meters

Observation Mode Results

Figure 6: Sub-aperture tangential speed as a function of time.

Tangential Speed m/s

1.0 m/s

0.1 m/s intervals

10,000 second intervals
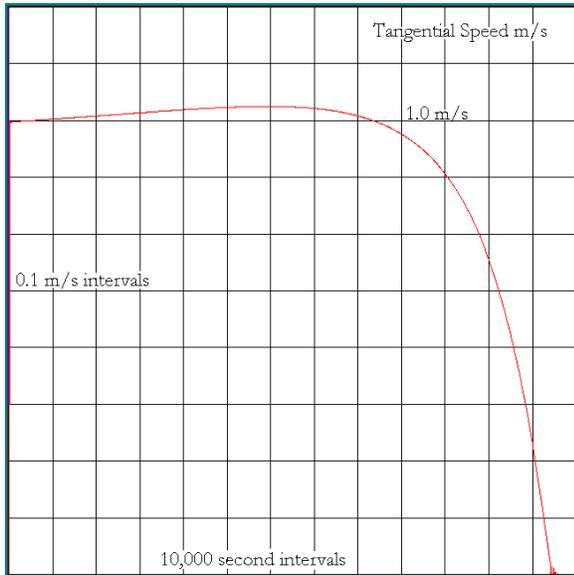
Figure 8: Sub-aperture tether stretch transient.

TETHER STRETCH vs TIME

10 mm

70 sec per division
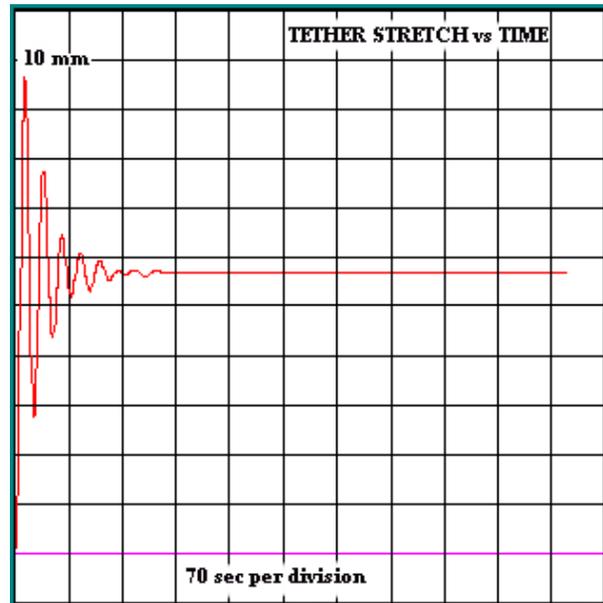
Figure 7: Sub-aperture tether tension as a function of time. Tether material stress is less than 200 psi.

2.0 Newton

Baseline Tether Tension, Newtons

Stop transient

0.1 Newton force intervals

10,000 second intervals
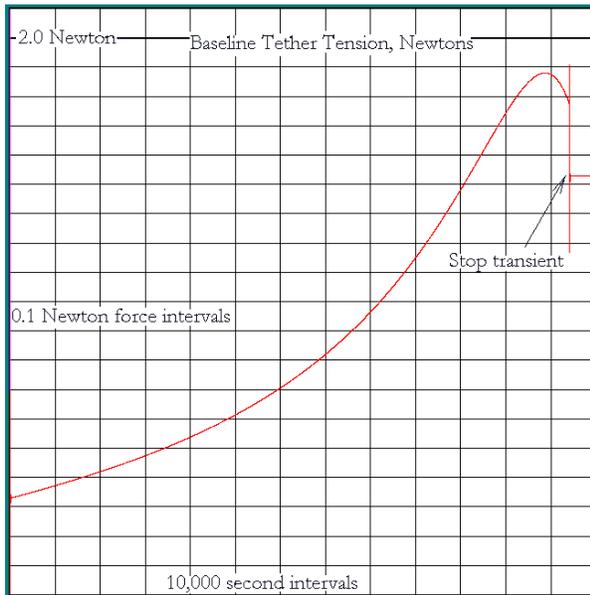
Figure 9: Sub-aperture tether stretch as a function of time.

TETHER STRETCH vs TIME

9 mm stretch maximum

10000 sec per division

Repointing Mode Results

The simulated results of repointing the Tetra-Star configuration worked exactly as the closed form solutions predicted, both in precessed angle, fuel consumed, and verifying the smoothness of the maneuver. The repointing maneuver requires the use of a variable thrust, sub-Newton propulsion system with a magnitude of 0.1 Newton. The sinusoidal thrust profile can be approximated step-wise if the resolution of the thrust steps is of the order of 0.001 Newton. A trade to consider, is of precessing the tethered constellation in a rougher, pulse-like manner and have the control system expend fuel to dampen the out-of-plane disturbances after the maneuver is completed. The simulations show, however, that a smooth re-point maneuver is possible and stable, driven in a completely open-loop fashion with variable thrust propulsion.

Future studies

Improvements in the modeling will include the use of bodies, not just point masses, and the modeling of tethers having mass with sufficient grid points to capture the first 2 transverse vibration modes. This will allow subtle out-of-plane motions to be studied. Also the inclusion of a control system with sensors and actuators to determine if 1 arc second of stability between subapertures is possible. Sensitivity of the parameters will be investigated, and the effect of transverse tether vibrations will be noted. If the transverse vibrations are problematical, then devices to dampen them actively will be sought.

Future configurations will focus on rigidly-shaped triangular arrangements, as the pendulum style configuration is overly sensitive to in-plane motions, though it is not ruled out by any means.

On the practical side, the tether material design will require investigation for maximizing the life of the tether, to account for environmental deterioration and micrometeoroid damage. Preliminary layouts and designs should be attempted for the winch mechanism, terminal damping shock strut, and the sub-aperture spacecraft itself for a more realistic mass estimate. A full evaluation of the system must include all of the closed-loop control systems which affect spacecraft attitude, mirror pointing, and wave-front control, as well as interior and exterior perturbation sources. The deployment sequence for a tethered observatory should be investigated. For example, a possible deployment scenario could include building, launching, and assembling of components in low earth orbit (LEO), then final launching to the L2 Lagrange point. When the stowed assembly is in L2 orbit, then the deployment and initialization of the tethered constellation can proceed. Perhaps the International Space Station will discover that its forte will be the assembly of large spacecraft, whose large components were built on the ground and boosted to LEO in the usual fashion. The mission design must be carefully conceived and scrutinized, as risk will be a palpable factor (the technique being so new), and must be an equal partner with all other parameters when mission trade studies are exercised. A technology validation mission will be needed.

Last Words

Spinning tethered constellations combined with synthetic aperture techniques can make impossibly large space-platform observatories come true. But nothing comes for free, as it will still have some complexities regarding optics, controls, and deployments to work out. Compare that to the rigid, massive alternatives. Surely this is an enabling technology that will revolutionize space-platform observation.

Formations of Tethered Spacecraft
As Stable Platforms for
Far IR and Sub-mm Astronomy

Marco B. QUADRELLI
Fred Y. HADAEGH
Michael SHAO
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA 91109-8099
marco@grover.jpl.nasa.gov

Enrico C. LORENZINI
Harvard-Smithsonian Center for Astrophysics
60 Garden St., Cambridge, MA 02138

ABSTRACT In this paper we describe current research in tethered formations for interferometry, and a roadmap to demonstrating the required key technologies via on-ground and in-orbit testing. We propose an integrated kilometer-size tethered spacecraft formation flying concept which enables Far IR and Sub-mm astronomy observations from space. A rather general model is used to predict the dynamics, control, and estimation performance of formations of spacecraft connected by tethers in LEO and deep space. These models include the orbital and tethered formation dynamics, environmental models, and models of the formation estimator/controller/commander. Both centralized and decentralized control/sensing/estimation schemes are possible, and dynamic ranges of interest for sensing/control are described. Key component/subsystem technologies are described which need both ground-based and in-orbit demonstration prior to their utilization in precision space interferometry missions using tethered formations. Defining an orbiting formation as an ensemble of orbiting spacecraft performing a cooperative task, recent work has demonstrated the validity of the tethering the spacecraft to provide both the required formation rigidity and satisfy the formation reconfiguration needs such as interferometer baseline control. In our concept, several vehicles are connected and move along the tether, so that to reposition them the connecting tether links must vary in length. This feature enables variable and precise baseline control while the system spins around the boresight. The control architecture features an interferometer configuration composed of one central combiner spacecraft and two aligned collector spacecraft. The combiner spacecraft acts as the formation leader and is also where the centralized sensing and estimation functions reside. Some of the issues analyzed with the model are: dynamic modes of deformation of the distributed structure, architecture of the formation sensor, and sources of dynamical perturbation that need to be mitigated for precision operation in space. Examples from numerical simulation of an envisioned scenario in heliocentric orbit demonstrate the potential of the concept for space interferometry.

KEYWORDS: Tethered Spacecraft, Formation Flying, Dynamics, Control, Pointing, Retargeting, Astrophysical Detectors, Variable Baseline

1. INTRODUCTION

NASA's future Earth and Space science missions involve formation flying of multiple coordinated spacecraft. Several space science missions (e.g., Terrestrial Planet Finder [1], Terrestrial Planet Imager, Starlight, LISA, SPECS) include distributed instruments and a large phased array of lightweight reflectors and antennas, and long variable baseline space interferometers. A collection of collectors and combiner/integrator spacecraft will form a variable-baseline optical space interferometer for a variety of science applications. Formation

flying spacecraft must conform to extremely stringent control and knowledge requirements. The control system for space interferometry, for example, must provide precision station-keeping from coarse requirements (relative position control of any two spacecraft to less than 1 cm, and relative attitude control of 1 arcmin over a large range of separation from a few meters to tens of kilometers) to fine requirements (nanometer relative position control, and .01 milliarcsec relative attitude control).
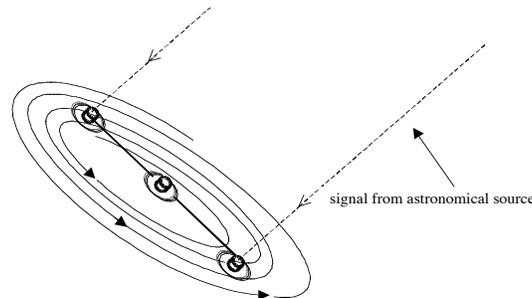
Figure 1. Tethered Interferometer Operation during Source Observation

Apertures of kilometric size are realized by connecting two or more light collecting spacecraft by means of one or more tethers. The advantage of using the tethers is that a variable controllable baseline can be achieved by reeling the tethers in or out, with a much smaller fuel consumption for reconfiguring the spacecraft as compared to the case of separated spacecraft in formation, in which on-board thrusting is continuously required. The idea of connecting the spacecraft to each other by means of a lightweight deployable tether is particularly attractive because: a variable baseline for interferometric observations can be achieved by deploying or retracting the tether; the coverage of the observation plane can be done continuously by spinning the whole system; the high levels of propellant consumption currently demanded by the ACS (Attitude Control System) of separated spacecraft in formation can be dramatically reduced by clever tension control of the interconnecting tethers; and two-dimensional and three-dimensional architectures can be constructed. Figure 1 depicts a configuration of a tethered interferometer in heliocentric orbit currently being considered by a joint JPL-Smithsonian Astrophysical Observatory (SAO) research study ([2]). These spinning tethered configurations are stable and can be tested in LEO. However, to mitigate the thermal dynamics ensuing in the system at each terminator crossing, a near polar sun-synchronous orbit would be preferred. Additional off-the-shelf ACS and tether deployer technology could be used at a relatively low cost. This would make a LEO demonstrator of a tethered formation for space interferometry possible in the near term. Rather than directly analyzing the feasibility and dynamical behavior of a complex tethered interferometer such as the proposed SPECS (Submillimeter Probe of the Evolution of Cosmic Structure), we propose to deal with the simpler architecture of Figure 1 as a precursor to SPECS, but which has the essential features required for its operation: variable baseline, spin dynamics, tether dynamics, multiple light collectors and a central light combiner.

In this paper, first, we describe the drivers and constraints for tethered formations designed for space interferometry, and explain why tethered formations can play a major role in space interferometry. Next, we discuss the approach to predicting the performance using dynamics models and control, sensor and estimation models required by a tethered system. Next, we identify the key sensing/control authority levels which space interferometry demands of tethered

formations, and outline a roadmap for ground testing and in-orbit testing of key tethered formation component technology.

2. DRIVERS AND CONSTRAINTS OF TETHERED FORMATIONS

There are several potential drivers and constraints that affect a tethered interferometer system design [2]:

- *pointing stability:* The pointing direction of the interferometer is required to be held within one arc minute (at 1 km baseline) with respect to the line of sight throughout the period of an observation.
- *distance collectors-combiner:* The distances collector1-combiner and collector2-combiner must never differ more than 10 cm from each other.
- *minimum and maximum tether tension:* For a tether to be controlled at the cm level a minimum tension of about 100mN is required so that inner residual tensions and hysteresis phenomena can be limited. Moreover a higher tension is an asset for the stability of the interferometer subjected to solar pressure, whereas depending on the diameter of the tether the tension should be at least one order of magnitude less than the material yield tension.
- *maximum tangential velocity:* The minimum number of photons of the observed source to be collected at a certain baseline length and orientation provides a limit for the maximum tangential velocity of the end mirrors. This velocity should be of the order of 1-5 m/s, provided that sufficiently large mirrors and advanced photon detection systems are employed.
- *boresight with respect to the Sun:* The angle between the anti-Sun direction and the boresight axis must be kept under 20-30 degrees to prevent the solar radiation noise from degrading the measurement.
- *u,v plane coverage:* The Fourier plane would need to be fully sampled from short lengths to 1000m baseline and as rapidly as possible. The high-resolution area (from 100 m to 1000 m baseline) is scientifically the most important.
- *fuel consumption:* The thrusting maneuvers should be reduced as much as possible. The ideal solution would be to keep the magnitude of the angular momentum constant throughout the observations and be able to fulfill all the requirements.
- *survivability:* The tether has to be able to survive in a micrometeoroids environment with high probability (more than 95%) for a 4-5 years mission.

3. WHY TETHERED FORMATIONS CAN PLAY A MAJOR ROLE IN SPACE INTEROMETRY

The building block of the formation is a tether connecting two (or more) telescopes on a line. Our teams at SAO and JPL have analyzed in details the orbital perturbations acting on a linear formation in heliocentric (Earth trailing) orbit and the resulting dynamics for the last one-year [2]. The conclusions of our study are that the contributions of disturbances associated with the tether dynamics forced by external perturbations to the overall pointing and relative positioning of the formation are negligible when compared to the effect of the same perturbations acting on the satellites. A steady-state pointing (of less than 1 arcmin) and positioning accuracy

requirements (of less than 1 cm) specified for the free-flying formations can be met by a tethered configuration in heliocentric orbit. Figure 2 shows the pointing angular errors of a 1-km-baseline tethered system formed by two collectors and a central combiner on a line. This figure was derived for a specific initial orientation that drives the out-of-plane hard but not the in-plane. For other initial orientations, the in-plane angle is more perturbed than the out of plane but the overall pointing errors of the tethered interferometer in heliocentric orbit are always below 1 arcmin over periods of many months without requiring any overall attitude formation control during the observations.

A comparative analysis of the perturbations acting on the spacecraft of a configuration such as the TPF one vs. those associated with the tether itself indicates that the satellite sun shields contribute 99% of the relevant environmental perturbation forces while the tether only contributes 1%. In conclusions, the contribution of the tether to the formation errors is negligible when compared to the effect of the perturbations acting on the satellites sun shields. Geometric and/or optical asymmetries of the sun shields will produce the lion share of the differential-mode noise components that will impact the control of the free-flying formation. The tethered configuration is actually more robust than the free-flying formation at tolerating those effects because thanks to the possible higher spin rates, it has a higher angular momentum and greater stability.
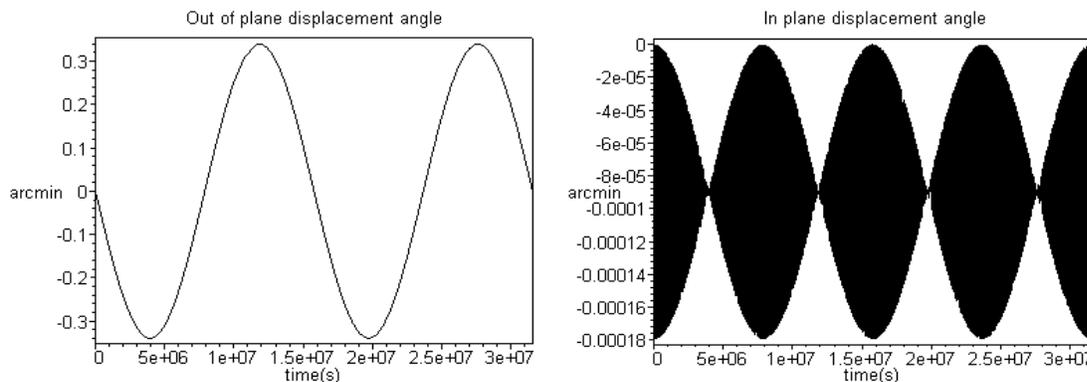
Figure 2 Pointing angular errors of tethered interferometer in solar orbit (over 1 year). Baseline length = 1km; rotation period = 2 hr 20 min; symmetric sun shields on collectors

Another important point is that besides being very small the noise brought about by environmental perturbations acting on the tether is a common-mode type of noise, that is, it alters equally the optical path lengths of the interferometer and consequently it does not require differential corrections of the optical path lengths. Differential-mode noise can be produced by retargeting maneuvers that excite odd modes of lateral tether vibrations. The amplitudes of these modes are proportional to the retargeting speed and are strongly limited by the tether tension. Retargeting maneuvers of the tethered interferometer and techniques for damping out those modes will be one the subject of our future research. Finally, the spectral content of a tether for TPF in the length range 100 m to 1 km is at low frequency. Natural longitudinal (i.e., stretching) modes are readily damped out by material damping or simple tether attachment damping devices and they are not a concern. Natural lateral modes have first-harmonic frequencies in the range 0.1 Hz to 0.03 Hz assuming a 1-hr rotation period as a reference. The frequency of external perturbations acting on the tether are also very low, appearing at one or twice the rotation frequency and orbital frequency. This low frequency content points to the fact that the

decreasing-amplitude higher-order harmonics should not be a problem for the fine control system of delay lines which is typically designed for a 1 kHz frequency range.

Consider an interferometer configuration in which the four in-line collectors of TPF could be connected by a light (a few kilograms) tether with a relatively simple mechanization while leaving the combiner free flying. In this case a very large portion of propellant can be saved for station-keeping the four collectors during observations (only the combiner needs to be propelled). We have estimated, based on the geometry of TPF, that the propellant for station keeping can be reduced by a factor of 6.7 in a tethered collector formation with respect to the free-flying configuration. Because of the lower propellant consumption, the spin rate of TPF could be increased from the present 8 hours to, let us say, 2 hours or even 1 hour (as indicated in [1]) by enabling the observation of a larger set of target stars for planets search. Figure 3 shows the propellant required exclusively for planets detection for the free-flying TPF and the four-tethered-collector configuration as a function of the rotational period. Planet detection (for which propellant estimates are available from the TPF study [1]) accounts only for a portion of the total propellant expenditure of TPF. Imaging astrophysical sources, which requires continuous covering the u-v plane, can also be readily accomplished in a tethered formation. The tether simply removes the limitations imposed on TPF by the propellant consumption which limits the observation spin rate (that is the number of targets) and builds more flexibility into the mission operation by adding a propellant-free actuation capability for baseline reconfiguration

(a)　　　　　　　　　　　　　　　(b)
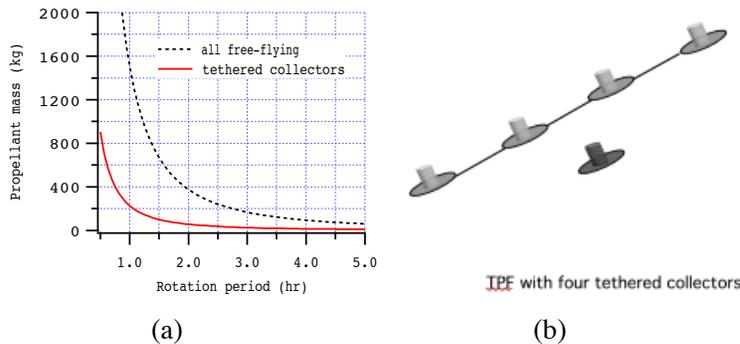
Figure 3 (a) preliminary estimate of cumulative propellant (adapted from Ref. [1] required for planet detection only for a TPF with all free flying elements and a TPF with four tethered collectors as depicted in (b).

(a)　　　　　　　　　　　　　　　(b)
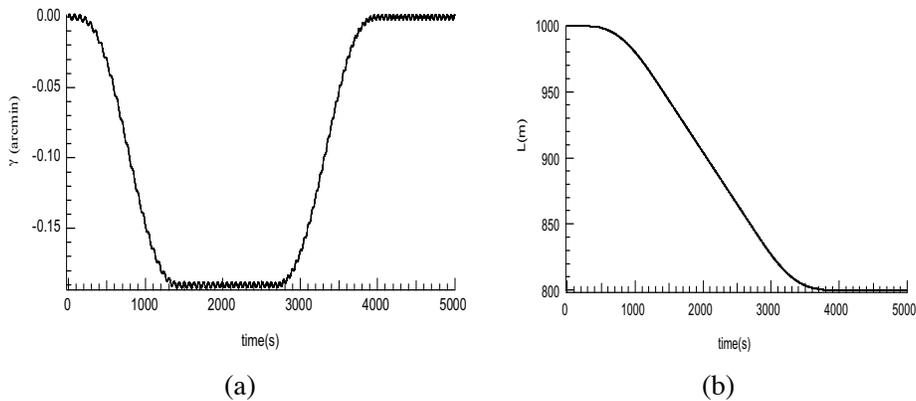
Figure 4 (a) spacecraft bearing error during a reconfiguration maneuver with (b) the baseline length varying from 1000m to 800m.

and u-v plane coverage. Similar considerations would apply in comparing the fuel performance of a tethered vs. non-tethered SPECS interferometer. Reconfiguring the baseline from one length to another can also be accomplished by reeling in (or out) the tether with the use of energy and almost no propellant. Figure 4 shows the simulated dynamics of a baseline reconfiguration from 1000 m to 800 m (with a perfect actuator) and the associated bearing angle error (produced by Coriolis forces) of one collector with respect to another. The tether, if attached off the center of mass, produces substantial restoring torques that stabilize the bearing angle of the spacecraft.
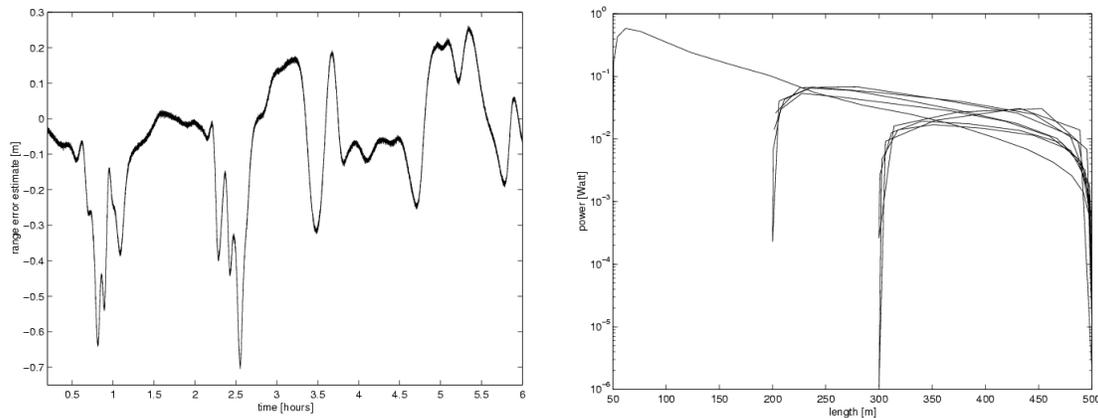
Figure 5. (a) Error between true and estimated range. (b) Tether deployer power as a function of system's length.

Figure 5a shows shows that the error between the true and estimated range between a collector and the combiner is of the order of tens of centimeters, consistent with the expected performance of the Autonomous Formation Flying radio-frequency metrology sensor. Additional laser metrology will be needed for higher precision metrology. Figure 5b shows a tether deployer power utilization as a function of the baseline length in the range of 10^{-2} Watt, provided most of the reconfiguration required for U-V plane coverage occurs at intermediate baselines. Figure 6a shows the trace of coverage in the normalized U-V plane achieved with the baseline reconfiguration program depicted in Figure 6b. From Figure 6 one can determine the advantage of the tethered system in providing high density U-V plane coverage. Figure 7a shows the tether temperature profile as a function of the position in the orbit, suggesting that ways to isolate the tether will be required to avoid thermal perturbations. Figure 7b shows the interferometer baseline rate to be in the range of cm/s for spin rates of the order of 0.01 rpm. In conclusions, the results of our analysis for a tethered formation in heliocentric orbit indicates that the steady-state dynamics of the tether forced by the environmental perturbations is small and its effect on the pointing and separation of the formation are well within the specified requirements. Moreover, tether dynamics produce low-frequency noise that should be handled readily by the broad-banded fine control system of proposed interferometers such as SPECS and TPF.
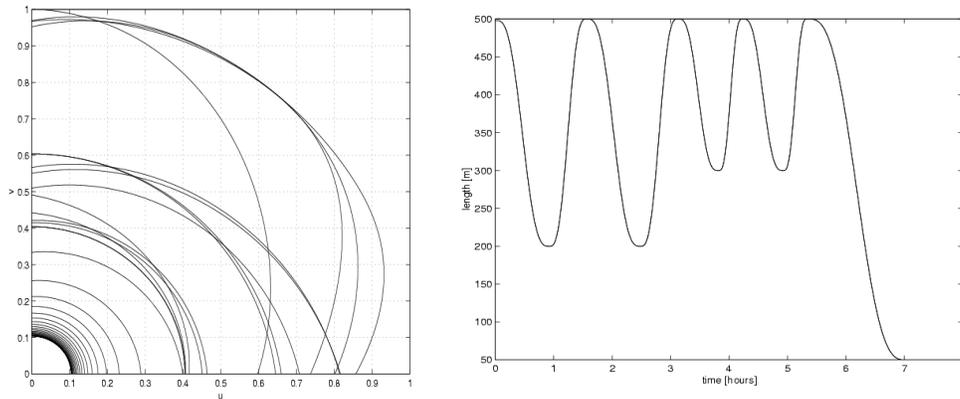
Figure 6. (a) U-V plane coverage trace. (b) Interferometer variable baseline program to cover U-V plane.
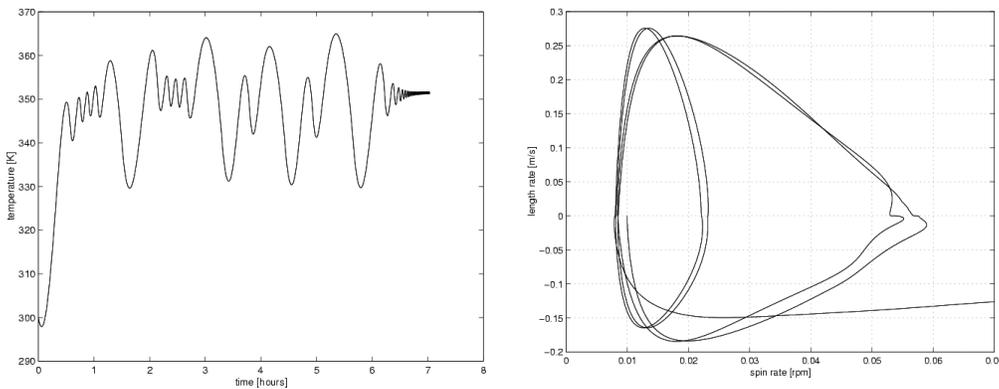
Figure 7. (a) Tether temperature profile. (b) Interferometer baseline rate vs. spin rate.

4. SYSTEM ARCHITECTURE

4.1 Modeling and Dynamics

From a dynamical standpoint, a formation of tethered spacecraft is characterized by a wide dynamic range (from less than 1 Hz in the spacecraft dynamics to KHz in the operation of the instrument synthesized by the formation), and by spatial scales ranging from sub-micron to kilometers. The formation can be thought of a *virtual truss* ([3], [4], [5]) in which the stiffness and dissipation levels of the connecting links are dictated by the control action on the relative sensing and actuation between two or more neighboring spacecraft. The dynamic model of this virtual truss suffers from undesired deformation modes caused by sensor noise, actuator non-linearity, dynamic uncertainties, and environmental disturbances.

4.2 Sensing/Estimation

Figure 8 depicts various sensing/estimation schemes required by tethered formations. Formation Estimation plays a key role in formation flying control of distributed spacecraft. In order to fully appreciate the complexity of the formation estimation problem, consider the illustration in Figure 8, which depicts four possible architectures for information exchange for a formation of four spacecraft. The arrows denote the relative state measurement made by the spacecraft located at the tip of the arrow. For the simplest case (A) each member of the formation uses only the relative state with respect to a designated master. In the second case (B), a centralized solution is the only possible architecture. Architecture (C) allows any member of the formation to make, visibility permitting, relative state measurements with respect to any other member. In architecture (D) the master and another member of the formation, labeled Reference, form a "baseline". The Reference receives information only from the Master, while all other spacecraft in the formation use relative states with respect to the Master and the Reference. A particular mechanization of information exchange will directly impact the quality of the formation estimate and therefore the quality of formation control.
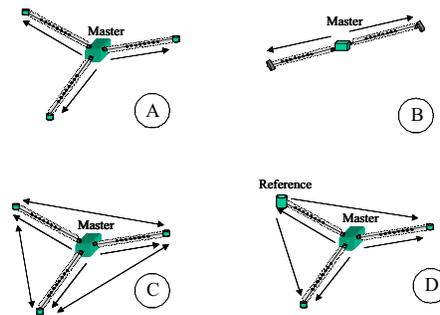
Figure 8. Four Possibilities of Making Relative State Measurements

4.3 Commander/Controller

The observation cycle is at least of two types: Stop and Stare observing mode (in which the configuration is brought to a halt with zero relative velocity between spacecraft before any observations are attempted), and Observe on the Fly mode (in which fringe measurements on astronomical targets can be made while the spacecraft are moving). There are at least four internal dynamics control modes in the system when working as an interferometer [4]:
- **Attitude Rigidity Control Mode**. This mode is used for fine pointing and stabilization only.
- **Spin/Despin Control Mode**. This mode is used to modulate the rotational spin rate of the system about its center of mass.
- **Tether Deployment/Retrieval Control Mode**. This mode is used to change the baseline of the interferometer or to control the baseline finely for corrections at the centimeter level or less.
- **Retargeting Mode**. This mode is used when the tethers are retracted into the collector spacecraft, and the whole system is repointed to a different target before the whole sequence of u-v plane coverage begins for the new target, and involves a precession maneuver.

4.4 Sensing/control authority levels for interferometry

Figure 9 depicts a block diagram of sensor and control levels for a tethered interferometer spacecraft with similar goals to the Terrestrial Planet Finder (imaging, nulling, planet finding). We may identify four levels of control and sensing authority with different bandwidths and precision:

1. Level 1: is the formation global control residing on the light-combining spacecraft. Ground communication links, inertial pointing, and inertial guidance for the formation are commanded from this location (0.1Hz, meters, arc-sec).

2. Level 2: refers to the control/sensing by each collector spacecraft for purposes of baseline stabilization (0.01-0.1 Hz, sub-cm and arc minute).

3. Level 3 refers to control authority enabled by RF and optical links to stabilize the metrology loop (KHz, micrometer, sub arc-sec).

4. Level 4 refers to tracking/nulling operations involving the maximum precision level of the interferometer during observation (KHz, nm and sub arc-sec).

The focus of the technology to be demonstrated in a potential LEO flight is on Level 1 and 2, although a demonstrator in deep space would be able to demonstrate Level 3 control and estimation technology.

5. TECHNOLOGY IN NEED OF DEVELOPMENT FOR FUTURE PRECISION TETHER APPLICATIONS

Several important technologies have already been demonstrated in-orbit during at least 16 tethered spacecraft flights:
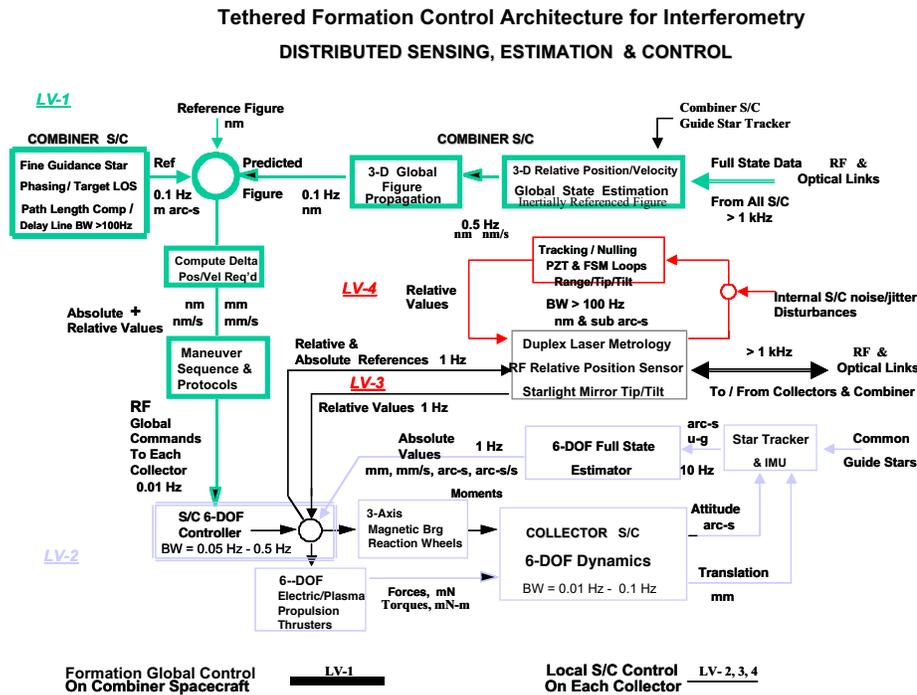
Figure 9. Sensing/Control authority for TPF class tethered interferometer spacecraft.

1. Controlled deployment, with accurate control along local vertical: +/- 4 degrees (this was achieved with a simple spool-out-only deployer; a much higher accuracy can be achieved with a reel-in and out deployer).
2. Controlled stationkeeping , allowing for long term orbit and attitude dynamic prediction.
3. Long-term (>5 years) survival and dynamic stability of a 4-km tether in LEO at a 1000km-altitude, high-inclination orbit (i.e., the dirtiest region in LEO).
4. Sizeable current flow in both directions (for boost and deboost applications) of conductive tethers.

Several technologies need further development before autonomous and reliable precision applications of tethered spacecraft can be made.
1. Controlled tethered system retargeting strategies to different sources in the sky.
2. Precision stationkeeping.
3. Disturbance rejection and vibration abatement of tether dynamics caused by transient maneuvers.
4. Very smooth reeling in and out of tether suitable for precision baseline control.

These objectives can be accomplished with ground testing and in-orbit validation of the following key technologies:
1. Active/passive control of tether attachment point, via movable hinge or movable boom or via dissipative flexures or joints;
2. autonomous on-board control logic for reliable deployment and retrieval at specified tether length and tension profiles;
3. crawler technology, to enable distributed arrangements of tethered vehicles on a very long tether or multiple tethers;
4. accurate metrology between adjacent tethered vehicles which does not suffer from scattered illumination from the intervening tether.

6. CONCLUSIONS

In this paper, we have described the models currently being used at JPL for dynamics analysis, control, and estimation, of tethered formations in deep space and in LEO intended as precursor demonstrators of SPECS. We have identified the features needed by a tethered interferometer in space in order to qualify as a system capable of imaging, nulling, planet detection and general far-IR and sub-mm space science. The key technologies which need to be pursued and developed in order to achieve these goals have also been described.

Acknowledgments

REFERENCES

[1] *Terrestrial Planet Finder*, Beichman ,C.A., Wool, N.J., Lindensmith, C.A., eds., JPL Publication 99-3 (Jet Propulsion Laboratory: Pasadena, California, 1999).

[2] Quadrelli M.B., Hadaegh F. Y., Lorenzini E. C., and Bombardelli, C.: *Precision Tethered Formations for LEO and Space Interferometry Applications*, presented at the 16th International Symposium of Spaceflight Dynamics, Pasadena, CA, December 2001.

[3] Hadaegh, F. Singh, G. Quadrelli M, and Shields, J.: *Modeling and Control of Formation Flying Spacecraft in Deep Space and Earth Orbits*, Terrestrial Planet Finder Technology Interchange Symposium, April 27, 2000.

[4] Quadrelli M.: *Modeling and Dynamics Analysis of Tethered Formations for Space Interferometry*, presented at the AAS/AIAA Spaceflight Mechanics Meeting, Santa Barbara, CA, February 11-14, 2001.

[5] Quadrelli M., Hadaegh, F.Y., Singh, G., and Shields, J.F.: *Some observations on the Dynamics and Control of Spacecraft Formations*, presented at the IFAC Conference on Aerospace Controls, Bologna, Italy, September 2-7, 2001.

New Frontiers in Spacecraft Orbit Design

Pini Gurfil[*] and N. Jeremy Kasdin[†]

Mechanical and Aerospace Engineering Department, Princeton University

Princeton, NJ 08544

Abstract

Utilizing the Sun-Earth spatial elliptic three-body problem (ER3BP) framework, this paper presents new families of geocentric orbits and isolated distant quasiperiodic orbits which do not orbit the Lagrangian equilibrium points. The main driver for this study is the need to find practical orbits that remain within a bounded distance from Earth, thus allowing high data-rate communication while ensuring safe operational environment far from thermal perturbations and visual occultations as well as Earth's magnetic and radiation fields. The orbit characterization procedure is performed using a novel approach. Optimal initial conditions are found using niching genetic algorithms, which render global optimization while permitting several optimal or sub-optimal solutions to co-exist. This approach yields a diverse probing of the state-space of the ER3BP. Stability of the orbits is determined using the notion of practical stability. The effect of solar radiation pressure and the Moon's gravitational perturbation are simulated, showing that the orbits are not significantly affected. This feature implies that no station-keeping is required. Optimal direct transfer trajectories from Low Earth orbit are briefly examined, showing that insertion into the characterized orbits may be performed using modest energetic requirements.

Introduction

Over the next few decades, the National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA) plan a variety of deep-space science missions ranging from mid-IR interferometers (such as ESA's Darwin [1]) to visible-light coronagraphs (such as NASA's Terrestrial Planet Finder [2]). Sophisticated telescopes and technologies will be developed that will collect and process the desired information about the origin and development of galaxies, stars, planets, and the chemical conditions necessary to support extra-terrestrial life. The majority of NASA's current efforts are focused on developing space-based observatories through a series of precursor missions, such as the Space Infrared Telescope Facility (SIRTF); first generation missions, such as Starlight, Space Interferometry Mission (SIM) and the Next Generation Space Telescope (NGST); and second generation missions, such as the Terrestrial Planet Finder (TPF).

---

[*] Research Associate. Phone: 609-258-6909. Fax: 609-258-6109. Email: pgurfil@princeton.edu
[†] Assistant Professor. Phone: 609-258-5673. Fax: 609-258-6109. Email: jkasdin@princeton.edu

These missions pose many orbit design challenges. Astrophysical observatories must be separated from Earth by a large distance to avoid stellar occultation, thermal perturbations and environmental effects such as the magnetic and radiation fields. On the other hand, it is required that the spacecraft remain bounded at some pre-specified distance to maintain high data rate communication and to enable servicing and replenishment capabilities. It is also desirable to eliminate or minimize station-keeping, minimize launch and transfer energy and to maximize the allowable launch windows.

To satisfy the mentioned constraints, several viable trajectories have been proposed. For example, for the NGST mission, a halo orbit around the Sun-Earth collinear Lagrangian libration point L_2 has been considered [3]. For the TPF mission, L_2, Earth-trailing and Earth drift-away orbits have been examined [4]. Starlight and SIRTF have planned arrested drift-away trajectories [5] where a single impulsive velocity change is used to keep the spacecraft within a rough formation relative to Earth.

The framework usually adopted for practical deep-space orbit analysis and design is that of the planar or spatial circular restricted three-body problem (CR3BP) with the Sun as the large primary and the Earth as the small primary. Investigation of the motion near the CR3BP collinear Lagrangian equilibrium points has yielded many families of trajectories. Lyapunov orbits obtained in the planar case [6] bifurcate into a spatial family of Lissajous orbits [7], which generate the well-known halo orbits as a particular case when perturbations are ignored and specific initial conditions are selected [8].

In addition, investigations of the motion near the small primary using the CR3BP framework with Hill's simplified formulation [6] has yielded numerous families of periodic trajectories about the small primary [9]-[11]. In the Sun-Earth case, some of these orbits have been referred to as distant retrograde orbits (DROs) of the Earth [12]. These periodic orbits are characterized by a bounded distance to Earth and modest energetic requirements. Some of these orbits are stable, thus eliminating the need for station-keeping.

Although the use of the CR3BP has proven fruitful, this approach constitutes an approximation because Earth's orbital eccentricity about the Sun is neglected. When the eccentricity is taken into account, the CR3BP becomes the elliptic restricted three-body problem (ER3BP). The incorporation of the eccentricity renders the analysis more involved. For example, in the ER3BP the location of the collinear libration points is pulsating with respect to Earth and there is no closed-form for the non-autonomous integral of motion. Omitting the influence of the eccentricity causes secular position and velocity errors that are particularly substantial when keeping in mind that deep-space orbits are usually planned for prolonged mission lifetimes.

A wealth of literature exists on the ER3BP. Numerous works have considered derivation of periodic orbits in the planar case [13]-[16], periodic orbits in the spatial case with a critical mass ratio of 0.5 [17] or a critical eccentricity of 1 [18], approximate integrals of motion [19]-[20], specialized regularization [21], and the dynamics of libration points

[22]. To the best of our knowledge, there has not been a systematic derivation of Earth-centered trajectories in the ER3BP based on a practical set of mission constraints.

This work is aimed at filling this gap by performing a systematic characterization of families of such orbits. We do not restrict the work to finding periodic orbits but rather we explore orbits that do not violate the pre-specified mission constraints such as minimum and maximum distance from Earth and stability. Energetic feasibility is then studied via examination of optimal transfer trajectories. The result is a real-world, well-established orbit design that may constitute an attractive alternative to the currently used or proposed orbits. By utilizing the non-restrictive framework of the ER3BP, it is guaranteed that these orbits would be very close to those simulated using the complete ephemeris and avoid the secular errors resulting from the use of the CR3BP simplification.

Synthesis of the proposed orbits is carried out using a novel approach. We perform a stochastic yet structured constrained search of initial conditions that maximize a given objective function. The stochastic search is performed using genetic algorithms (GAs), which are an optimization technique based on the Darwinian principle of the survival of the fittest [23]. GAs have been causing an increasing interest in the space community due to their potentially fascinating applications in this field (see e.g. [24], [25] and the references therein).

The GA optimization is performed for various lower and upper bounds on initial conditions. This enables the characterization of 14 families of trajectories. Each family contains topologically distinct orbits, which offer promising features for deep-space missions in general and space-borne observatories in particular. The different families are then categorized according to various criteria to yield Distant Retrograde Orbits, Distant Prograde Orbits, Earth-Return Orbits and Delayed Escape Orbits.

After orbit characterization, we study stability using the notion of practical stability. It is shown that most of the trajectories are practically stable. We also briefly examine disturbances on the nominal orbits. It is shown that solar radiation pressure and the Moon's gravitational disturbance do not constitute a significant error source for most of the orbits examined. This result implies that no station-keeping is required.

Equations of Motion

In most past research, a barycentric rotating-pulsating coordinate system has been used to model the ER3BP [6]. Usually, this coordinate system is suitable to model the problem and yields a fruitful characterization of diverse families of trajectories. However, in this study we have adopted a slightly different rotating coordinate system, which we found to be particularly useful for design of Earth-centric orbits. The basic notion is to choose the origin of the coordinate system at the center of Earth, as depicted in Figure 1. In this frame, denoted by $\hat{x}, \hat{y}, \hat{z}$, the $\hat{x}$-axis is directed radially outward along the local vertical, the $\hat{y}$-axis lies along the direction of Earth motion and the $\hat{z}$-axis is normal to the

ecliptic to complete the Cartesian right-hand setup. It should be stressed that this coordinate system is also useful from the engineering standpoint, since the orbit determination process can be straightforwardly performed with the position and velocity vectors measured relative to Earth.

Using the above coordinates, the following equations of motion in the rotating frame are derived (see [27] for details):

$$x'' = 2y' + (1 + e\cos\theta)^{-1}\left[x - \frac{\mu x}{\left(x^2 + y^2 + z^2\right)^{3/2}} - \frac{(1-\mu)(x+1)}{\left[(x+1)^2 + y^2 + z^2\right]^{3/2}} + 1 - \mu\right] + \hat{d}_x \qquad (1)$$

$$y'' = -2x' + (1 + e\cos\theta)^{-1}\left[y - \frac{\mu\, y}{\left(x^2 + y^2 + z^2\right)^{3/2}} - \frac{(1-\mu)y}{\left[(x+1)^2 + y^2 + z^2\right]^{3/2}}\right] + \hat{d}_y \qquad (2)$$

$$z'' = -z + (1 + e\cos\theta)^{-1}\left[z - \frac{\mu\, z}{\left(x^2 + y^2 + z^2\right)^{3/2}} - \frac{(1-\mu)z}{\left[(x+1)^2 + y^2 + z^2\right]^{3/2}}\right] + \hat{d}_z \qquad (3)$$

Optimal Orbit Characterization

The discussion in the previous section reveals the complex nature of the spatial ER3BP. Due to the inherent linearized instability of periodic and quasi-periodic orbits of the two degrees-of-freedom Hamiltonian system rendered by the ER3BP [28], solutions may be very sensitive to initial conditions. Thus, a single guess of initial conditions aimed at deriving an optimal (in some sense) solution may result in either divergence or convergence to some undesirable local extrema without revealing practical and realizable solutions. Hence, the use of traditional optimization methods (such as the gradient search and the simplex method) to search for initial conditions might fail; the search needs to be *global*, i.e. performed over some given range of permissible initial conditions.

Genetic algorithms (GAs) offer a convenient solution to this problem due to their specialized features compared to traditional optimization procedures: First, they work with a coding (usually binary) of the parameter set ("strings", in the GA terminology) instead of the parameters themselves. Second, they search from a population of initial conditions instead of a single point. Third, they use only objective function ("fitness") evaluations instead of derivatives or other auxiliary information. Fourth, they use probabilistic transition rules instead of deterministic rules.

The GA most commonly used in engineering applications is the so-called Simple GA. To perform the evolutionary search, the Simple GA uses the operators of crossover, reproduction and mutation. Crossover is used to create new solution strings ("children")

from existing strings ("parents"). Reproduction copies individual strings according to the objective function values. Mutation is an occasional random alteration of the value of a string position, used to promote diversity of solutions.

Although Simple GAs are quite robust in finding the global optimum [23], they suffer from two main drawbacks: First, a convergence to a local optimum is possible due to the effect of premature convergence, where all individuals in a population become nearly identical before the optima has been located; second, convergence to a single optimum does not reveal other optima which may exhibit attractive features. To overcome these problems, modifications of Simple GAs have been considered. These modifications are called niching methods, and are aimed at promoting a diversity of solutions for multi-modal optimization problems. In other words, instead of converging to a single (possibly local) optimum, niching allows for a number of optimal solutions to co-exist and lets the designer choose the appropriate one. The niching method used throughout this study is Deterministic Crowding [26]. According to this method, individuals are first randomly grouped into parent pairs. Each pair generates two children by application of the standard genetic operators. Every child then competes against one of his parents. The winner of the competition moves on to the next generation.

The orbit search procedure using Deterministic Crowding GAs begins with the definition of the dynamic parameter optimization problem and the definition of orbital stability.

Definition of the Objective Function, Constraints and Parameters

The constrained dynamic optimization problem of finding geocentric orbits of the ER3BP can be formulated in a variety of ways. The resulting orbits depend on three main factors: The definition of the objective function, the search bounds on the initial conditions and the minimum Earth approach constraints. Different families of periodic and quasi-periodic orbits may be found for different formulations of the optimization problem. This notion is utilized to characterize the optimal orbits. For a given objective function, the lower and upper search bounds on initial conditions are changed in each optimization set, so that a diverse exploration of feasible orbits is obtained.

A few candidate formulations of the objective function were examined. In each formulation, several combinations of lower and upper search bounds on initial conditions were used. The value of the initial true anomaly, θ_0, was set to either 0 or π. Using θ_0 as an additional optimization parameter did not seem to be beneficial to the characterization process. After examining various objective function formulations, we found that the formulation that yielded the most significant results is as follows:

Find the optimal set of initial conditions, $\mathbf{x}_0^*$, satisfying

$$\mathbf{x}_0^* = \arg\max_{x_0 \in \mathbf{I}} \frac{1}{\left(r_{max} - r_{min}\right)^2 + 1} \tag{4}$$

where $r_{\min} \triangleq \min\limits_{\theta \in [\theta_0, \theta_f]} \|\mathbf{r}\|$, $r_{\max} \triangleq \max\limits_{\theta \in [\theta_0, \theta_f]} \|\mathbf{r}\|$, and $\mathcal{X}$ denotes a hyper-rectangular domain defined by the upper and lower bound on initial conditions:

$$\mathcal{X} \triangleq \left\{ \mathbf{x}_0 : \left(\mathbf{x}_0\right)_{\min} \leq \mathbf{x}_0 \leq \left(\mathbf{x}_0\right)_{\max} \right\} \tag{5}$$

The objective function (4) was maximized subject to (5), the dynamic constraints of the equations of motion, with $\hat{d}_x = \hat{d}_y = \hat{d}_z \equiv 0$, and the following inequality constraint:

$$r_{\min} > R_E \tag{6}$$

where R_E is the mean radius of Earth. The constraint (6) was used to rule out collision trajectories with Earth. Note that the objective function (4) is normalized to attain values between 0 and 1, which facilitates the optimization procedure.

Throughout the optimization, we have used the GA constants given Ref. [27]. The mission lifetime selected was 1 year for all the optimization sets. Although a typical mission lifetime may be longer, the 1-year integration time considerably reduces the computer time required to perform the various optimization sets. After obtaining the optimal initial conditions, the integration is extended to 5 years to simulate the orbits using a more realistic mission lifetime and to study their stability in the sense defined shortly.

Practical Orbital Stability

For given initial conditions, an orbit is said to be practically stable (PS) if the spacecraft has not gone beyond some pre-specified distance from Earth at a given true anomaly (time) interval. Otherwise, the orbits is practically unstable (PUS). After orbit characterization, which is performed for 1 year, the integration is extended to 5 years (represents a typical mission lifetime) using the optimal initial conditions and the drift of the vehicle from Earth is examined. If the distance from Earth exceeds 0.1 AU (about 15 million km) the orbit is considered to be practically unstable.

It should be noted that PS as used here is stability in a loose sense. In order to study stability in the presence of disturbances and uncertainties in initial conditions, a more rigorous approach should be adopted which is beyond the scope of this paper. Nevertheless, we address the issue of disturbances via simulation.

Orbit Characterization Results

The 12 GA optimization sets yielded 14 families of orbits. 3 families, named A, B, and C, were obtained from optimization set 1, while families D-N were obtained from sets 2-12, respectively.

Due to the use of a niching GA, each optimization set may yield a few distinct solutions, which represent the global, and possibly a number of local, optima of the objective function within the search bounds. In this context, "family" will be used to distinguish solutions that are topologically different (in the configuration space) from one another. Within a certain family, a few solutions having different quantitative features may co-exist. Thus, the GA optimization may results either in different families of orbits, different orbits belonging to the same family, or both. The term "type" will be used to categorize the orbit according to the following metrics: planarity or spatiality; the direction of the initial velocity; closeness of Earth approach; and practical stability.

A practically stable (PS) planar orbit with $x_0 > 0$ and $y_0' < 0$ will be called a Distant Retrograde Orbit (DRO), in consistency with the literature [12]. A PS spatial orbit having the same characteristics will be called a 3D DRO. A PS planar orbit with $x_0 > 0$ and $y_0' > 0$ will be called a Distant Prograde Orbit (DPO) [12]. A PS spatial orbit having the same characteristics will be referred to as a 3D DPO. A planar or a spatial orbit characterized by zero initial velocity, which usually results in a close Earth fly-by, will be termed an Earth-Return Orbit (ERO) or a 3D ERO, respectively. EROs may be used as transfer trajectories to DROs or DPOs and hence their importance in the characterization process. If an orbit is practically unstable (PUS), as revealed by an extension of the integration to 5 years, it will be referred to as a Delayed Escape Orbit (DEO) in the planar case or a 3D DEO in the spatial case, regardless of any other terminology.

The complete description of the 14 families of orbits, A-N, can be found in Ref. [27]. For brevity, we present here only a few families. In each figure, the top panel describes results for a 1-year optimization, and the bottom panel presents extension of the orbits to 5 years.

Quantitative features of representative orbits of the different families, namely minimum and maximum distance (r_{min}, r_{max}) from Earth, initial distance form Earth (r_0), initial velocity (v_0), minimum and maximum distance from the Sun (ρ_{min}, ρ_{max}), orbit type and stability, are summarized in Table 1. Note that in this table we slightly abused the unit conventions, so that distances from Earth are given in kilometers and the distance from the Sun is given in AUs.

The top panel in Figure 2 depicts three orbits belonging to Family A. These are classical quasi-periodic quasi-elliptic simple DROs, with 2 crossings of the xz plane per year. All three orbits described in the top panel satisfy

$$\dot{y}_0 = -1.9967x_0 \approx -2x_0$$

Interestingly, the condition $\dot{y}_0 = 2x_0$ for the existence of simple quasi-periodic DROs was discovered by Henon [9] in Hill's modification of the CR3BP, a much more simplified problem than the ER3BP. The bottom panel of Figure 2 depicts an extension of the middle orbit (top panel) to 5 years. The 5-year orbit consists of an ellipse librating in the cross-track direction. Clearly, this DRO is PS. Examining the quantitative features of this orbit, given in the first row of Table 1, shows its practicality and usefulness.

Figure 3 describes Family B of quasi-periodic DROs. For clearness of presentation, only a single orbit representing this family is depicted. This family is different from Family A as the number of xz plane crossings is 4 per year. The bottom panel verifies that this orbit is PS.

Figure 4 depicts yet another DRO, representing Family C. This family consists of quasi-circular orbits with 8 crossings of the xz plane per year. This orbit is particularly useful, as it remains within almost a constant distance of 1 million km from Earth even for a 5-year mission lifetime.

Figure 5 depicts Family D, consisting of DPOs. The representing orbit depicted in the top panel is characterized by 10 xz plane crossings per year. An extension of the solution to 5 years reveals that this orbit is PS and remains within a constant distance from Earth. The motion may be interpreted as a librating ellipse about the Earth in the along-track sense.

Figure 6 depicts Family E, which consists of DPOs having 8 xz plane crossing per year. This quasi-periodic orbit remains PS for 5 years. Interestingly, the shape of this orbit somewhat resembles halo and quasi-halo orbits around the Lagrangian libration points. Practically speaking, this orbit is particularly useful due to its relatively close Earth approach (see Table 1 for details).

Figure 7 shows another family of DROs, Family F. This family, described by a cross-track and along-track librating ellipse, is fundamentally different from the previous 3 families of DROs (A, B and C) due to the relatively close Earth approach (224,900 km, see Table 1), which can be utilized to facilitate maintenance, replenishment and possibly replacement of a damaged spacecraft. This orbit is characterized by 12 crossings of the xz plane per year and it is PS as the bottom panel of Figure 7 shows.

The top panel in Figure 8 depicts three 3D DROs belonging to Family J. These are classical 3D quasi-periodic DROs, with 2 crossings of the xz plane, 2 crossings of the xy plane and 2 crossings of the yz plane per year. All three orbits described in the top panel satisfy, similarly to the planar case,

$$\dot{y}_0 = -1.987x_0 \approx -2x_0$$

The bottom panel of Figure 8 depicts an extension of the middle orbit (top panel) to 5 years. Clearly, this 3D DRO is PS. The maximum out-of-ecliptic deflection of this orbit

is about 0.04 AU, which may be utilized to reduce to the noise sensitivity of the space-borne observatory due to the interference of the interplanetary zodiacal cloud [25].

Figure 9 depicts Family K of 3D DROs. The representing orbit is characterized by a quasi-circular projection in the *xy* plane and a small out-of-ecliptic deflection (maximum 0.0009662 AU, about 145,000 km). Within 1 year, there are 8 *xz* plane crossings, 8 *xy* plane crossings and 8 *yz* plane crossings. This orbit is PS. Examination of Table 1 shows that this orbit is characterized by less than 10% difference between the maximum and minimum distances from Earth during a 1-year mission, a very attractive feature in terms of real-life missions.

Figure 10 presents a representative orbit of distinct isolated small periodic orbits which do not librate about the Lagrangian equilibrium points. This new finding is a subject of an ongoing research.

Effect of Disturbances

A Complete description of sensitivity to disturbances may be found in [27], where it is shown that due to the robustness of the orbits to disturbances, no station-keeping is required. Unlike the widely used halo orbits, which are extremely sensitive to disturbances due to their inherent instability, the orbits discussed above do not drift considerably from their nominal values due to disturbances.

Summary and Conclusions

This study provided much encouragement for the use of ER3BP orbits for deep-space missions in general and space-borne observatories in particular. The use of Deterministic Crowding genetic algorithms as the parameter optimization procedure has resulted in a fruitful probing of the complex dynamics of the ER3BP. The newly found orbits require modest insertion energy and are stable in the loose sense of practical stability. They also exhibit good robustness to disturbances, which imply that no station-keeping is required.

References

[1] ESA's Darwin Web Site, http://sci.esa.int/home/darwin/index.cfm

[2] NASA's Origins Web Site, http://origins.jpl.nasa.gov

[3] Lubow, S., "Controlling the NGST Orbit", http://icarus.stsci.edu/~lubow/ngst/orbits.html

[4] The TPF Science Working Group, *Terrestrial Planet Finder*, NASA/JPL Publication 99-3, May 1999.

[5] NASA's SIRTF Web Site, http://sirtf.caltech.edu/.

[6] Szebehely, V., *Theory of Orbits*, Academic Press, New York, 1967.

[7] Gomez, G., Massdemont, J. and Simo, C., "Lissajous Orbits around Halo Orbits", *Advances in the Astronautical Sciences*, Vol. 95, 1997, pp. 117-134.

[8] Howell, K. C., Barden, B. T., and Lo., M. W., "Application of Dynamical Systems Theory to Trajectory Design for a Libration Point Mission", *The Journal of the Astronautical Sciences,* Vol. 45, No. 2, April-June 1997, pp. 161-178.

[9] Henon, M., "Numerical Exploration of the Restricted Problem. V. Hill's Case: Periodic Orbits and Their Stability", *Astronomy and Astrophysics,* Vol. 1, 1969, pp. 223-238.

[10] Goudas, C. L., "Three Dimensional Periodic Orbits and Their Stability", *Icarus*, Vol.2, 1963, pp. 1-18.

[11] Henon, M., "Vertical Stability in the Restricted Problem. II. Hill's Case", *Astronomy and Astrophysics*, Vol. 30, 1974, pp. 317-321.

[12] Ocampo, C. A., and Rosborough, G. W., "Transfer Trajectories for Distant Retrograde Orbits of the Earth", AAS Paper 93-180.

[13] Shelus, P. J., and Kumar, S. S., "Some Periodic Orbits in the Elliptic Restricted Problem of Three Bodies", *The Astronomical Journal*, Vol. 75, No. 4, 1970, pp. 315-318.

[14] Benest, D., "Elliptic Restricted Problem for Sun-Jupiter: Exsitence of Stable Retrograde Satellites at Large Distances", *Astronomy and Astrophysics,* Vol. 13, 1971, pp. 157-160.

[15] Cors, J. M., Pinyol, C., and Soler, J., "Periodic Solutions in the Spatial Elliptic Restricted Three-Body Problem", *Physica D*, Vol. 154, 2001, pp. 195-206.

[16] Gomez, G., and Olle, M., "Second-Species Solutions in the Circular and Elliptic Restricted Three-Body Problem. II. Numerical Explorations", *Celestial Mechanics and Dynamical Astronomy,* Vol. 52, No. 2, 1991, pp. 147-166.

[17] Olle, M., and Pacha, J. R., "The 3D Elliptic Restricted Three-Body Problem: Periodic Orbits Which Bifurcate from Limiting Restricted Problems", *Astronomy and Astrophysics,* Vol. 351, 1999, pp. 1149-1164.

[18] Kallrath, J., "Bounded Orbits in the Elliptic Restricted Three-Body Problem", *Celestial Mechanics*, Vol. 43, No. 1-4, 1988, pp. 399-408.

[19] Contopoulos, G., "Integrals of Motion in the Elliptic Restricted Three-Body Problem", *The Astronomical Journal,* Vol. 72, No. 5, 1967, pp. 669-673.

[20] Vrcelj, Z., and J. H. Kiewiet, "An Invariant Relation in the Elliptic Restricted Problem of Three Bodies", *The Astronomical Journal,* Vol. 83, Nu. 5, 1978, pp. 514-521.

[21] Ries, J. G., "Numerical Investigation of Depletion in the Outer Asteroid Belt", *Icarus*, Vol.121, 1996, pp. 202-206.

[22] Howell, K. C., and Hiday-Johnston, L. A., "Time-Free Transfers between Libration-Point Orbits in the Elliptic Restricted Problem", *Acta Astronautica*, Vol. 32, No. 4, 1994, pp. 245-254.

[23] Goldberg, D., E., *Genetic Algorithms in Search, Optimization and Machine Learning,* Addison-Wesley, Reading, MA, 1989.

[24] Hartman, J. W., Coverstone-Carroll, V. L., and Williams, S. N., "Optimal Interplanetary Spacecraft Trajectories via a Pareto Genetic Algorithm", *The Journal of the Astronautical Scienses,* Vol. 46, No. 3, July 1998, pp.267-282.

[25] Gurfil, P., and Kasdin, N. J., "Optimal Out-of-Ecliptic Trajectories for Space Borne Observatories", to be published in *The Journal of the Astronautical Science*.

[26] Mahfoud, S. W., *Niching Methods for Genetic Algorithms*, Illinois Genetic Algorithms Laboratory Report No. 95001, University of Illinois at Urbana-Champaign, 1995.

[27] Gurfil, P., and Kasdin, N. J., "Practical Geocentric Orbits in the Spatial Elliptic Restricted Three-Body Problem", submitted to *The Journal of the Astronautical Sciences*.

[28] Meyer, K. R., and Hall, G. R., *Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*, Springer-Verlag, New York, 1992.

Figure Captions

Figure 1: Definition of coordinate systems.

Figure 2: Family A, simple Distant Retrograde Orbits, satisfying $\dot{y}_0 \approx -2x_0$. (a) Three members of Family A. (b) Middle orbit in (a) extended for 5 years.

Figure 3: Family B, Distant Retrograde Orbits. (a) Representing orbit for 1 year. (b) Orbit extended for 5 years.

Figure 4: Family C, Distant Retrograde Orbits. (a) Representing orbit for 1 year. (b) Orbit extended for 5 years.

Figure 5: Family D, Distant Prograde Orbits. (a) Representing orbit for 1 year. (b) Orbit extended for 5 years.

Figure 6: Family E, Distant Prograde Orbits. (a) Representing orbit for 1 year. (b) Orbit extended for 5 years.

Figure 7: Family F, Distant Retrograde Orbits with an along-track and a cross-track librating ellipse. (a) Representing orbit for 1 year. (b) Orbit extended for 5 years.

Figure 8: Family J, simple Three Dimensional Distant Retrograde Orbits, satisfying $\dot{y}_0 \approx -2x_0$. (a) Representing orbit for 1 year. (b) Orbit extended for 5 years.

Figure 9: Family K, quasi-circular Three Dimensional Distant Retrograde Orbits. (a) Representing orbit for 1 year, having an out-of-ecliptic displacement of 145,000 km. (b) Orbit extended for 5 years.

Figure 10: Isolated displaced orbit exists fat from the equilateral Lagrangian equilibrium points.

Figures
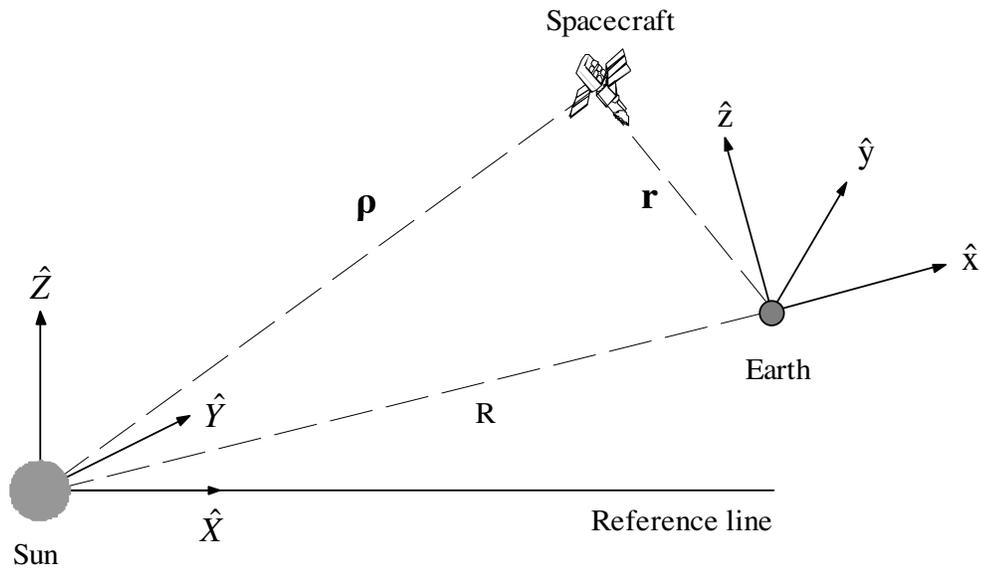
Figure 1
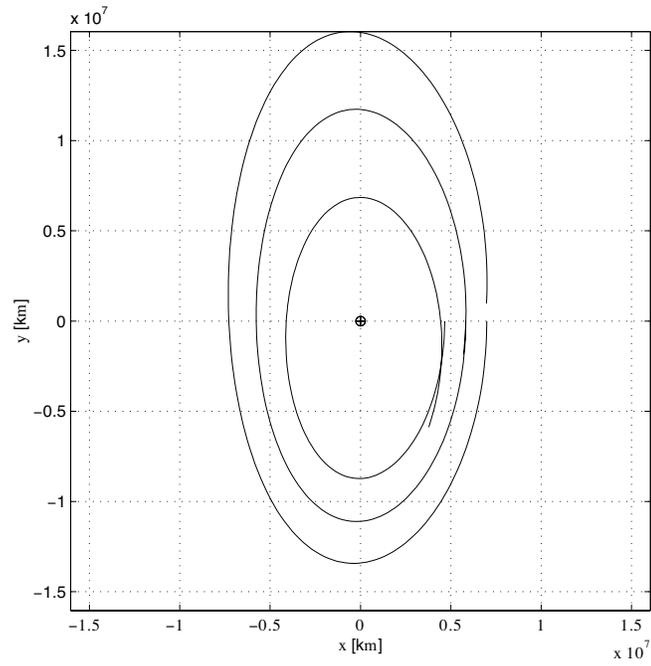
(a)
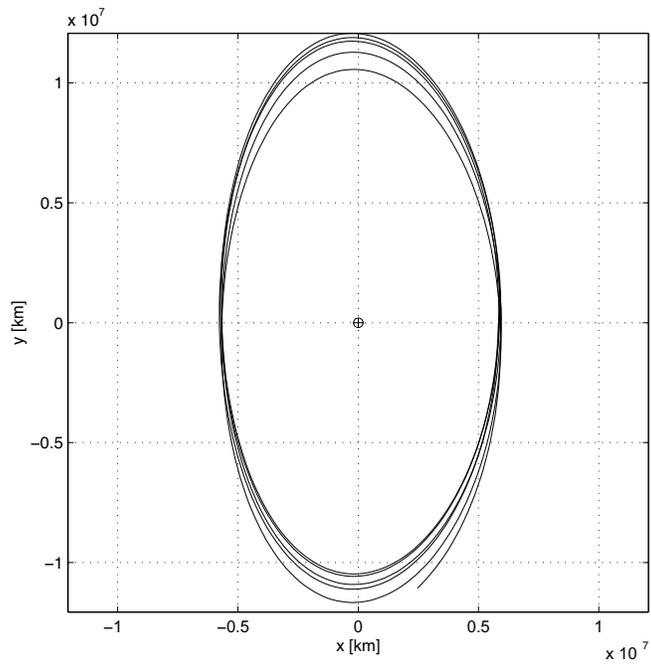
(b)
Figure 2
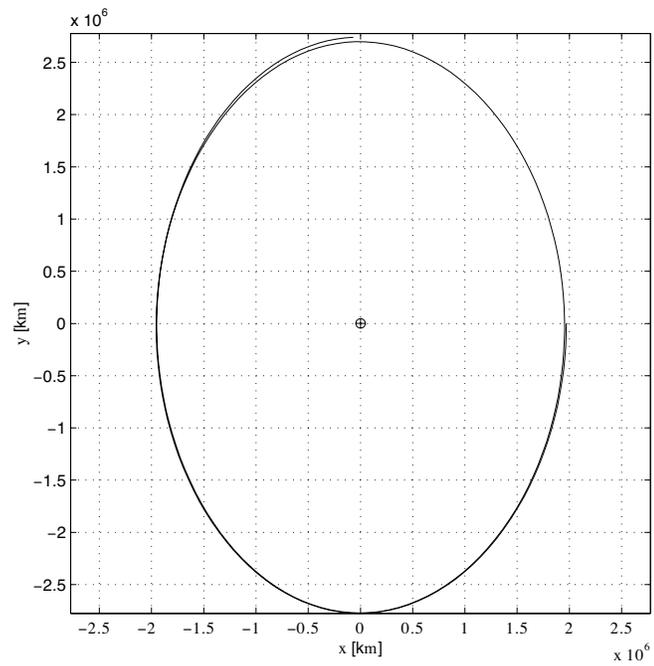
(a)
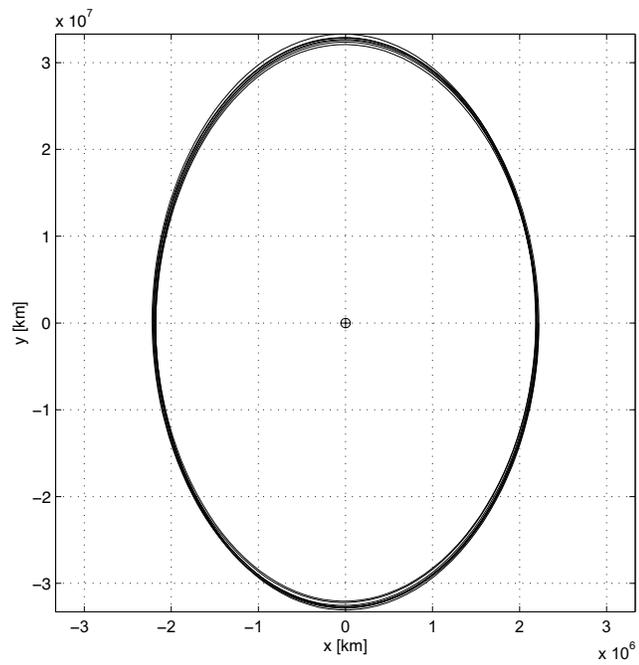
(b)
Figure 3
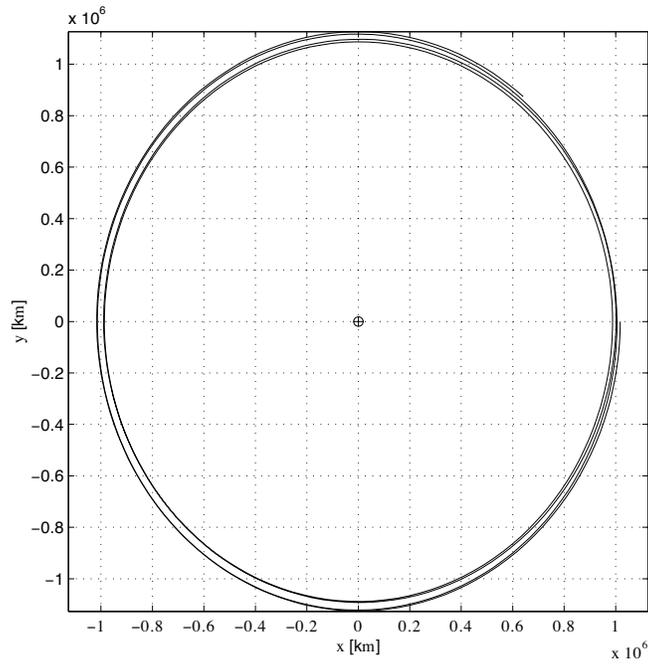
(a)
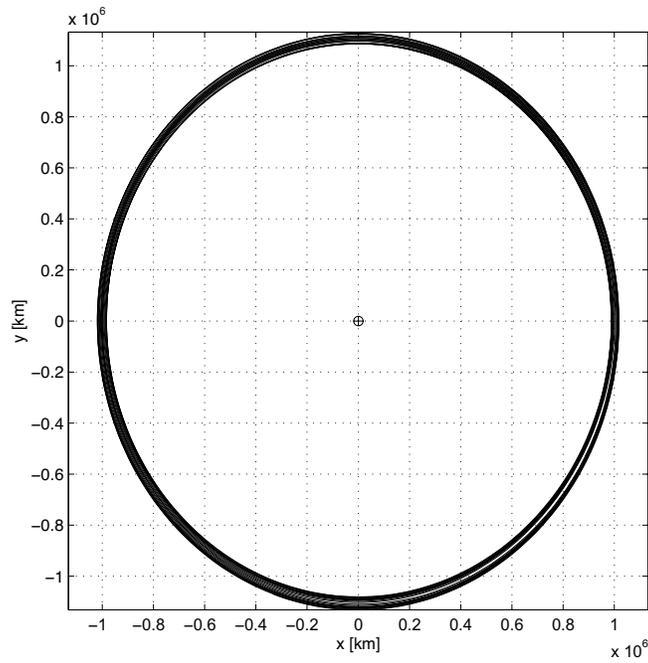
(b)
Figure 4
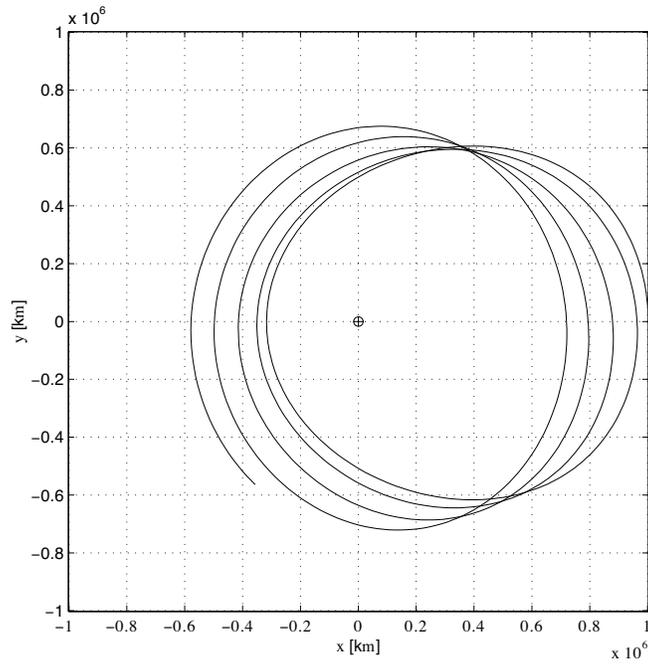
(a)
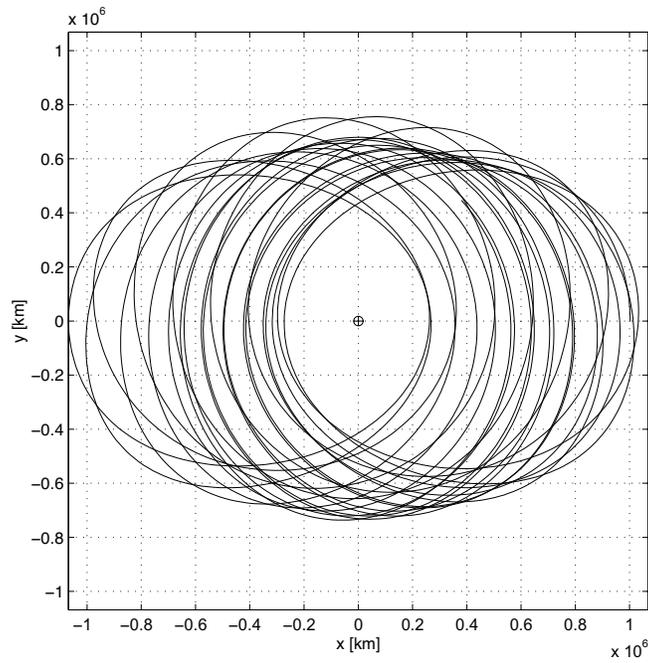
(b)
Figure 5
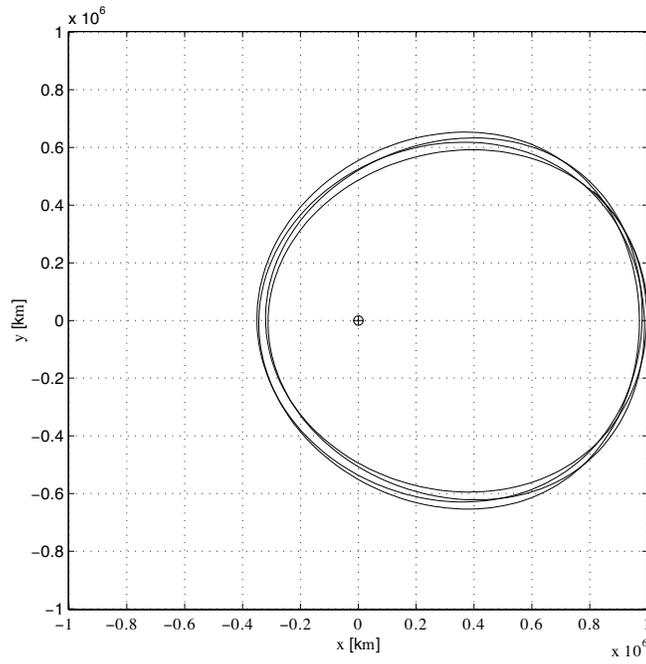
(a)
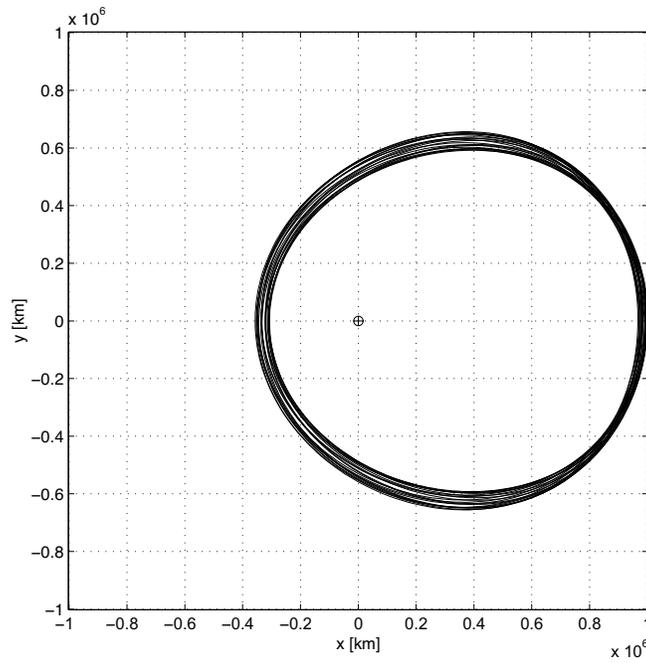
(b)

Figure 6
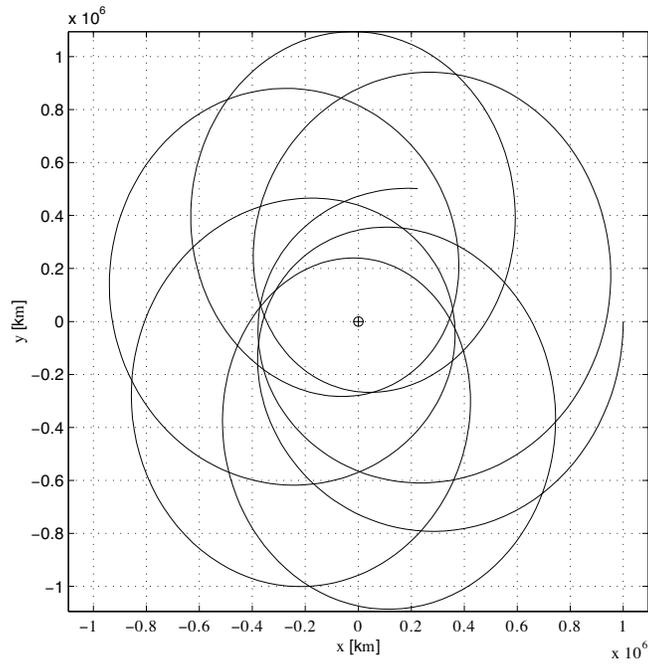
(a)
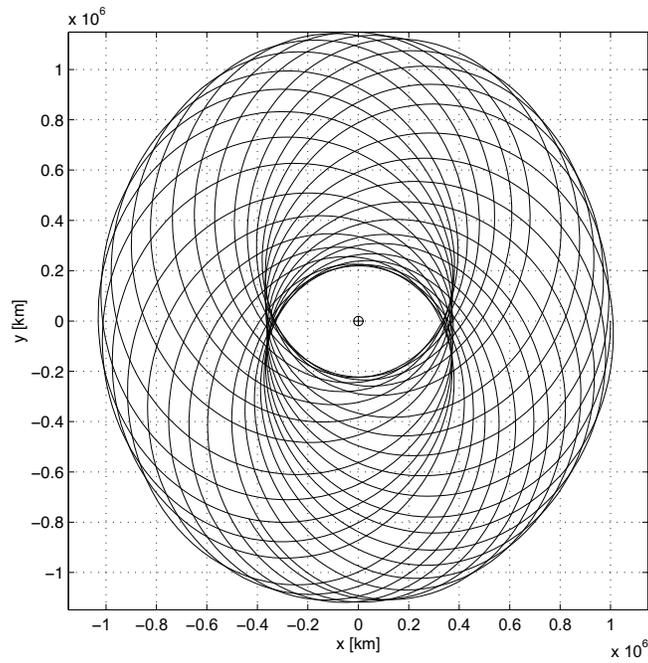
(b)
Figure 7
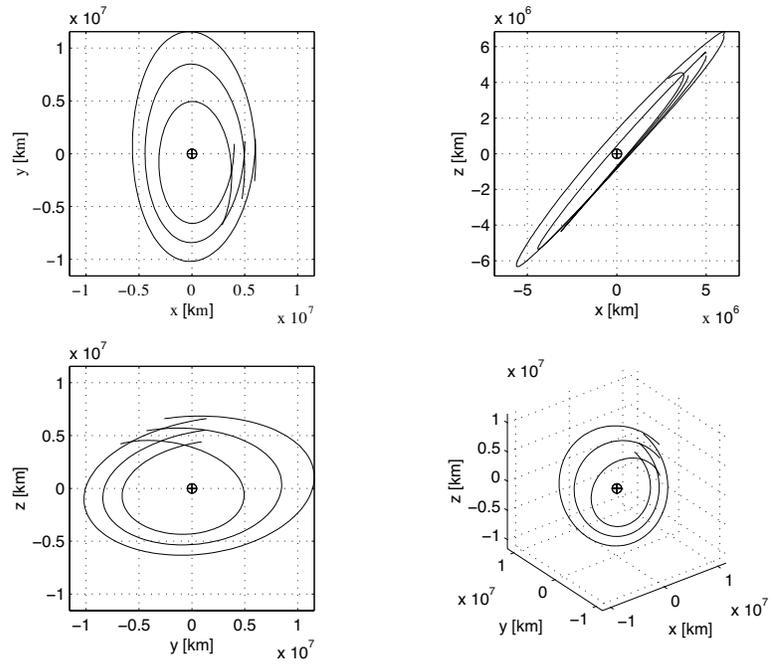
(a)
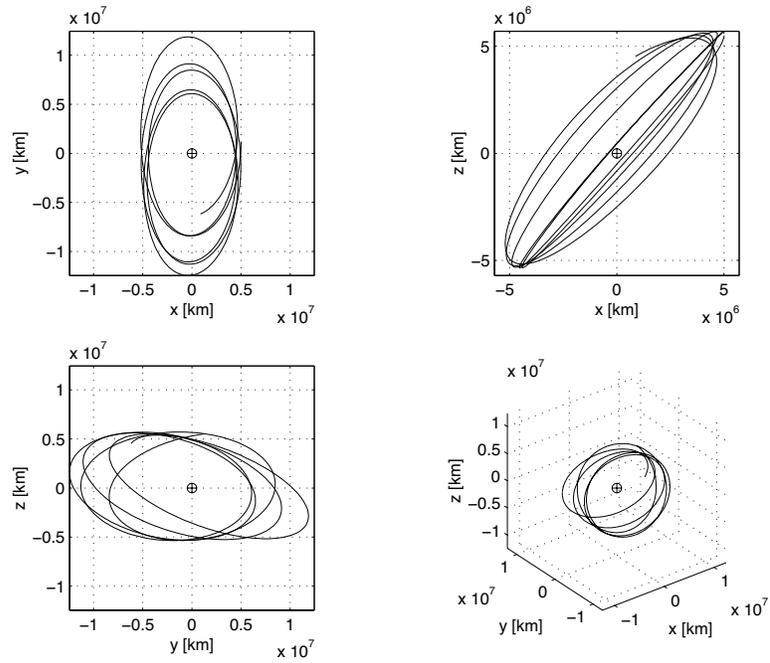
(b)
Figure 8
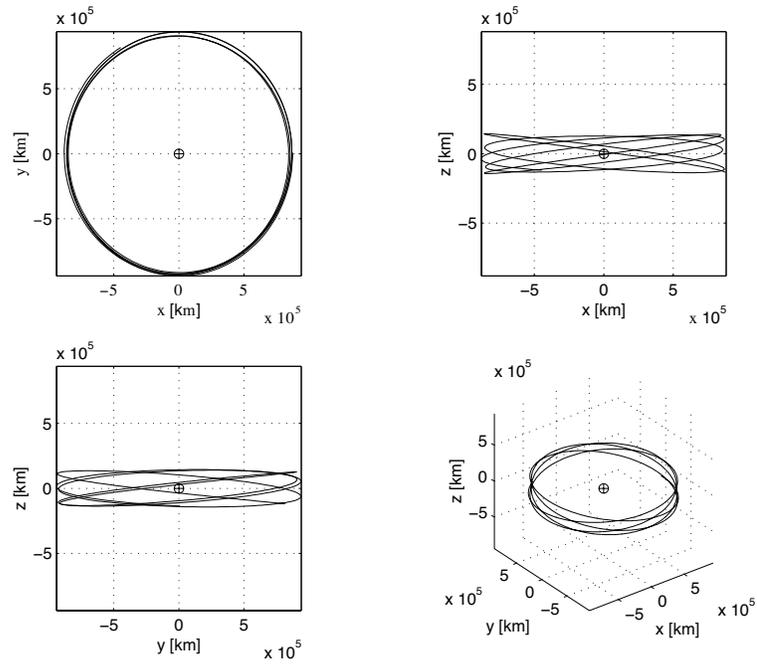
(a)
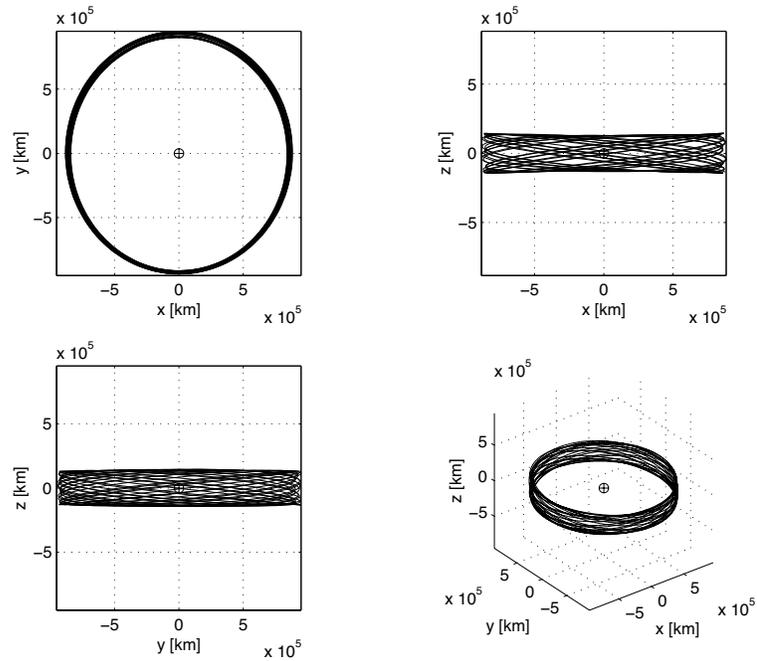
(b)
Figure 9
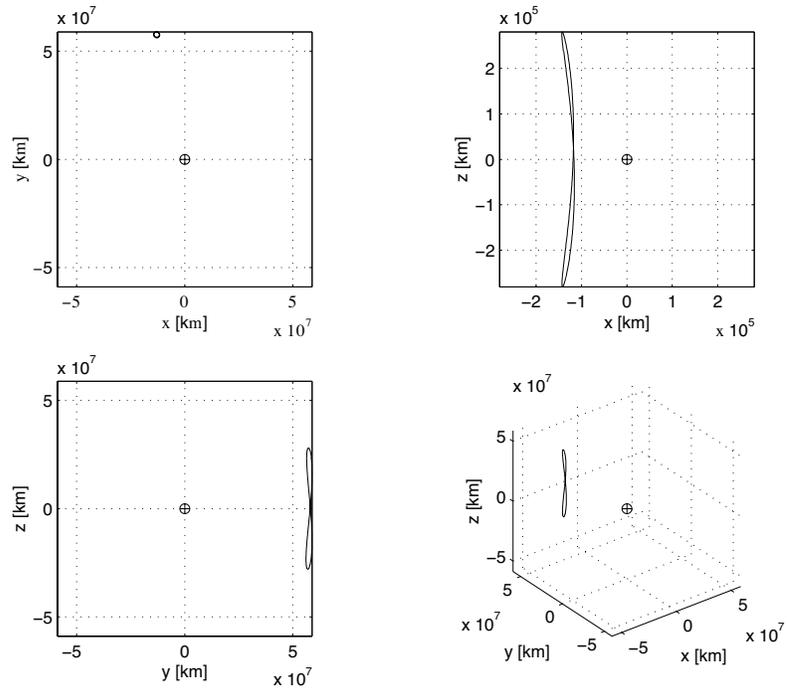
(a)
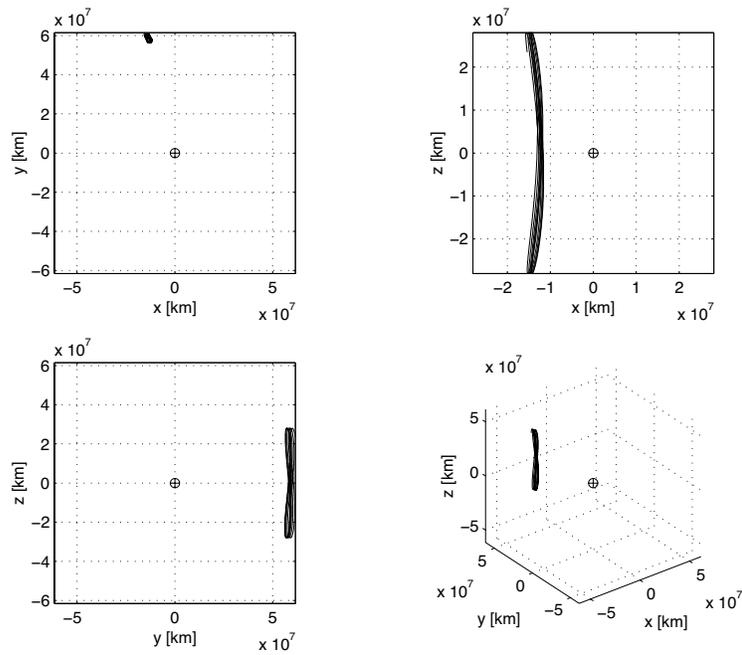
(b)
Figure 10

Table 1: Quantitative features of orbits representing each family of optimal solutions. Results are given for a 1-year mission. Stability assessment obtained from extension to 5 years.

Representing Orbit of Family:	r_{min} [km]	r_{max} [km]	r_0 [km]	v_0 [km/s]	ρ_{min} [AU]	ρ_{max} [AU]	Type	Stability
A	5,769,577	11,740,892	5,825,955	2.3557	0.97794	1.0229	DRO	PS
B	2,191,130	3,270,091	2,218,402	1.0718	0.96888	1.0268	DRO	PS
C	986,996	1,126,563	1,017,578	0.8792	0.97898	1.0232	DRO	PS
D	317,151	1,002,197	1,002,197	0.3252	0.97963	1.022	DPO	PS
E	311,893	1,002,197	1,002,197	0.3145	0.98384	1.0234	DPO	PS
F	224,900	1,138,236	1,000,000	0.6865	0.97707	1.0234	DRO	PS
G	11,532	1,084,448	1,000,000	0	0.98388	1.0163	ERO	PS
H	15,883	750,000	750,000	0	0.98346	1.0165	ERO	PS
I	2,328,213	3,962,617	3,177,032	0.8503	0.98753	1.0335	DEO	PUS
J	6,892,060	8,510,975	7,527,807	2.0339	0.98732	1.0187	3D DRO	PS
K	854,531	939,889	883,956	0.8838	0.97779	1.021	3D DRO	PS
L	755,223	1,662,761	1,638,694	0.25619	0.98392	1.0216	3D DEO	PUS
M	17,282	1,035,854	1,000,221	0	0.98233	1.017	3D ERO	PS
N	409,758	1,521,047	1,521,047	0	0.98051	1.0175	3D DEO	PUS

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE April 2004	3. REPORT TYPE AND DATES COVERED Conference Publication

4. TITLE AND SUBTITLE
New Concepts for Far-Infrared and Submillimeter Space Astronomy

5. FUNDING NUMBERS
Code 685

6. AUTHOR(S)
Edited by Dominic Benford and David Leisawitz

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS (ES)
Goddard Space Flight Center
Greenbelt, Maryland 20771

8. PEFORMING ORGANIZATION REPORT NUMBER
2003-03082-0

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS (ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
CP-2003-212233

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Unclassified–Unlimited
Subject Category: 89
Report available from the NASA Center for AeroSpace Information,
7121 Standard Drive, Hanover, MD 21076-1320. (301) 621-0390.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The "Second Workshop on New Concepts for Far-Infrared and Submillimeter Space Astronomy" aimed to highlight the groundbreaking opportunities available for astronomical investigations in the far-infrared to submillimeter using advanced, space-based telescopes. Held at the University of Maryland on March 7-8, 2002, the Workshop was attended by 130 participants from 50 institutions, and represented scientists and engineers from many countries and with a wide variety of experience. The technical content featured 17 invited talks and 44 contributed posters, complemented by two sixperson panels to address questions of astronomy and technology. A summary of the Workshop can be found in "Proceedings of Far-IR, Sub-MM & MM Detector Technology Workshop", (J. Wolf, J. Farhoomand & C.R. McCreight, eds.), NASA/CP-211408, 2003, under the title "New Concepts for Far-Infrared and Submillimeter Space Astronomy" (paper #i-02, by D.J. Benford).

14. SUBJECT TERMS
far-infared, space, astronomy, NASA

15. NUMBER OF PAGES
585

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL