

Near Infrared Spectroscopic Imaging of NGC 7023
and
Charge Trapping in InSb Photodiode Arrays

by

Steven L. Solomon

Submitted in Partial Fulfillment
of the
Requirements for the Degree
Doctor of Philosophy

Supervised by
Professor William J. Forrest

Department of Physics and Astronomy
The College
Arts and Sciences

University of Rochester
Rochester, New York

1998

Dedication

This thesis is dedicated to Julie and Joshua,
the two brightest stars in the sky.



We've looked and looked, but after all where are we?

Do we know any better where we are,
And how it stands between the night to-night
And a man with a smokey lantern chimney?
How different from the way it ever stood?

—Robert Frost, from *The Star-Splitter*

Curriculum Vitae

The author was born in New Rochelle, New York on 5 September, 1957. He attended Colgate University from 1975–1977, and 1983–1984 and the University of Rochester from 1985–1986 for undergraduate studies in Physics. He began graduate studies in Physics and Astronomy at the University of Rochester in the Fall of 1986. He received the Rush Rhees Fellowship from the University of Rochester in 1986 and a NASA Graduate Student Researcher’s Fellowship from 1991 to 1993. He pursued his research in astrophysics under the direction of Professor William J. Forrest and received the Master of Arts in 1989.

Acknowledgements

Julie, the road's behind us now. I can't begin to express my appreciation for hanging in through the frustrations of grad school, the vagaries of life, a job and this albatross. I marvel at your strength, patience and fortitude. It's been a far longer road than I ever envisioned and I can't imagine having a better companion, friend and lover than you. You are everything to me and I love you.

Josh, you light up my life. Daddy's "fesis" is done now, let's play!

Mom & Dad, thanks for things far too numerous to list here, but topping the list are love, encouragement, and the joy of learning. Dad, I miss you profoundly and am sorry you aren't here to see the end of this road with me.

JP, I don't even know how to thank you. You're a wonder as a teacher, role model and friend (even if you don't know how to make chicken soup!!). You have touched my life on so very many levels that I can't begin to make the list.

Bill...weeeeeeeell, there's the moon! Thank you for teaching me to be a good scientist and observer, and for showing me how to make detectors sing. I will never forget the way you count magnitudes of extinction on a cloudy night. Your perseverance and thoroughness have left an indelible mark on me as both a scientist and a human being.

There have been many people who made my time in graduate school rewarding, frustrating, educational and just plain fun. Fun is good. There were many twisted times and individuals without whom I would have gone gobbling to the fount...James for brotherhood, being there, teaching and the zucchini...Scott and Paula, comrades in the endeavors to learn and understand, and for the mountaintop madness ...Mikey for simply being the lunatic you are and for your succulent you know whats...Hao for being a peegootoe (or however you spell it) and for walking the nonexistent line between physicists and engineers with me...Gilles for ze life, and for being so very wrong about humankind...Didier for the joyful silliness and for learning me about the night sky... Debby for the great parties and your indomitable spirit...Larry for teaching, exploring and enjoying every step...Betty and Barb for being

my grad school moms and for their smiles... Pat for showing me how to cut threads on a lathe...Jack Dodd and Tony Aveni for their encouragement and example...Dave Kaplan for, well, whatever you call it...Tira for the romps and companionship...and NOVA & PBS for starting this whole mess in the first place.

I am grateful for the support received from the University of Rochester in the form of a Rush Rhees Fellowship for the academic year 1986-1987. I would also like to thank the NASA Graduate Researcher's Program for support during the period 1991-1993.

Thesis Abstract

We have performed broadband and narrowband spectroscopic infrared imaging of the reflection nebula NGC7023 over the wavelengths 1.2–3.8 μm . The dust composition varies widely across the face of the nebula. Both the 3.3 μm and 3.4 μm dust emission features are bright and variable across the nebula. The bright filaments in NGC7023 1) are redder than the ambient nebula, 2) are self shielding, 3) have a 3.29 μm line to continuum ratio of ~ 40 , 4) show asymmetric spatial profiles and 5) exhibit a 3.4 μm /3.29 μm emission feature ratio of one half. A simple Hubble law analysis shows that scattered light is insufficient to account for this surface brightness. We have investigated the far infrared–3.29 μm feature luminosity correlation in the context of NGC 7023. HD200775, the central star of NGC7023, is the dominant radiative excitation source in the nebula. Of the seven stellar objects observed in the frame (excluding HD200775), two were found to be late-type main sequence stars and the remainder were hypothesized to be protostellar objects.

We have demonstrated experimentally and theoretically that the latent image phenomenon in astronomical InSb photodiode infrared focal plane arrays passivated with SiO_x can be explained by charge trapping at the front surface of the detector array. The type of front surface passivant and the electrical state of the surface both play crucial roles in determining the magnitude of this effect; SiO_x exhibits significantly more image latency than does Si_3N_4 . Behaviorally, the latent images are consistent with a modified Shockley, Read, Hall trapping model. The trap energies and cross sections were determined from experimental data; they are consistent with hole traps in the SiO_x that lie within approximately 50 \AA of the SiO_x /InSb interface. Quantum mechanical tunneling was proposed as the mechanism for populating and depopulating the traps, and the molecular species In_2O_3 was advanced as a plausible candidate for the traps responsible for latent images in SiO_x passivated InSb.

Table of Contents

1. Introduction	1
2. Observations	3
2.1 Instrumental	3
2.2 Observations	5
2.3 Calibration	7
2.4 Data Reduction	12
3. NGC 7023	17
3.1 Introduction	17
3.2 Structure in NGC 7023	32
3.3 The Nebular Spectrum	41
3.3.1 The 3.29 μ m Emission Feature	41
3.3.2 The 3.4 μ m Emission Feature	49
3.4 Broadband Nebular Emission	61
3.5 FIR Correlation	68
3.6 Stellar Objects	76
3.6.1 Star D	81
3.7 Conclusion	84
4. Charge Trapping in InSb Photovoltaic Infrared Detector Arrays	87
4.1 Introduction	87
4.2 Indium Antimonide Infrared Detector Arrays	88
4.3 Surface Passivation	91
4.4 Latent Images & The Physics of Trapped Charge	99
4.5 Experimental Procedure	103
4.6 Trap Statistics Under Illumination	109
4.7 Derived Trap Parameters	114

4.8 Trapping Mechanism.....	120
4.9 Latency Dependence on Operating Conditions	124
4.9.1 Latency Dependence on Gate Bias	124
4.9.2 Latency Dependence on Photoflux	126
4.10 Surface Calculations.....	131
4.11 Conclusion.....	138
5. Bibliography.....	139
6. Appendix A - Optical Distortion	145
7. Appendix B - Detailed Calculation of Inferred Spectrum	147
8. Appendix C - Detailed Calculation of Surface Potential	148

List of Figures

Figure 3.1 - Broadband J image of NGC7023. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 97.7"×97.2" (E/W×N/S).	20
Figure 3.2 - Broadband H image of NGC 7023. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 106.0"×101.1" (E/W×N/S).	21
Figure 3.3 - Broadband K' image. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 102.1"×102.1" (E/W×N/S).	22
Figure 3.4 - CVF 3158 image. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 67.0"×62.4" (E/W×N/S).	23
Figure 3.5 - CVF 3291 image. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 104.4"×102.4" (E/W×N/S).	24
Figure 3.6 - CVF 3367 image. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 61.7"×61.7" (E/W×N/S).	25
Figure 3.7 - Broadband L' image. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 62.5"×62.5" (E/W×N/S).	26
Figure 3.8 - K' Frame with various apertures, cuts, etc. overlaid. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 102.1"×102.1" (E/W×N/S).	27
Figure 3.9 - HK' color temperature map, emissivity index $n=-1.75$. Prior to calculating the map, the K' frame was first convolved with a gaussian kernel to match the resolution (PSF) of the H frame and then demagnified to match the platescale of the H frame.	28
Figure 3.10 - Cut across the western filament. The data have been background subtracted, and the cut is 2.5" wide approximately 6W17S of star D. All the data have been normalized for ease of comparison. The curves are as follows: solid curve - 3.29 μ m emission, dotted line - K' band, dashed line - H band, dot-dashed line - J band.	34
Figure 3.11 - Cut across the eastern filament. The data have been background subtracted, and the cut is 2.5" wide directly north of star H. All the data have been normalized for ease of comparison. The curves are as follows: solid curve - 3.29 μ m emission, dotted line - K' band, dashed line - H band, dot-dashed line - J band.	34
Figure 3.12 - Cuts from HD200775 to star D. The plots, from top to bottom, are K', H, J, and 3.291 μ m. The width of each of the plotted vectors is 8 pixels ($\sim 2''$, perpendicular to the direction of the cut). The vertical axis is in μ Jy and the horizontal axis is in arcsec (from HD200775). The 3.291 μ m frame has been transformed by $(F/80)^{-3}$, where F is the signal flux in Janskies, in order to display it on the same plot as the broadband data.	37
Figure 3.13 - N/S cuts through star D. The plots, from top to bottom, are K', H, J, and 3.291 μ m. The width of each of the plotted vectors is 8 pixels ($\sim 2''$, perpendicular to the direction of the cut). The vertical axis is in μ Jy and the horizontal axis is in arcsec. The 3.291 μ m frame has been transformed by $(F/80)^{-3}$, where F is the signal flux, in order to display it on the same plot as the broadband data.	38
Figure 3.14 - Cuts from star H to star D. The plots, from top to bottom, are K', H, J, and 3.291 μ m. The width of each of the plotted vectors is 8 pixels ($\sim 2''$, perpendicular to the direction of the cut). The vertical axis is in μ Jy and the horizontal axis is in arcsec. The 3.291 μ m frame has been transformed by $(F/80)^{-3}$, where F is the signal flux, in order to display it on the same plot as the broadband data. Star H is on the left side of the graph at 0" and star D is on the right at 39". Note that star H was not detected in the 3.291 μ m measurement.	40
Figure 3.15 - Observed spectra (log Janskies) for nebular apertures indicated. Note that the spectra have been scaled per the legend in order to facilitate comparison.	42

- Figure 3.16 - The line to continuum ratio along the vector joining HD200775 to star D. The methodology employed for calculating this parameter ($l:c(3.29\mu\text{m})$) is discussed in the text. The parameters of the cut are identical to those of Figure 3.12. The plot has been smoothed by applying a $3''$ gaussian kernel (in the direction of the cut) to return weighted averages of the raw data.46
- Figure 3.17 - The line to continuum ratio along the vector joining star H to star D. The methodology employed for calculating this parameter ($l:c(3.29\mu\text{m})$) is discussed in the text. The parameters of the cut are identical to those of Figure 3.12. The plot has been smoothed by applying a $3''$ wide gaussian kernel (in the direction of the cut) to return weighted averages of the raw data. Star H is at the zero point of the x axis and star D is $38''$ from star H. Note that star H was not detected in the $3.291\mu\text{m}$ frame.....47
- Figure 3.18 - The inferred spectra for the filaments. The observations are also shown in the figure (data points with error bars). The eastern filament is plotted on the bottom and the western is on top. The spectra have been calculated from the observed data as described in the text - see eq (3.4). The vertical axis is plotted logarithmically in Janskies and the horizontal axis is in μm . The eastern filament's flux has been multiplied by ten in order to separate the plots. The continuum has been fit by forcing a 1500K blackbody through the K' band observation.55
- Figure 3.19 - The inferred spectrum for the $1'\times 1'$ frame in the immediate neighborhood of the filaments. The observations are also shown in the figure (data points with error bars). The spectrum was calculated from the observed data as described in the text - see eq (3.4). The vertical axis is plotted logarithmically in Janskies and the horizontal axis is in μm . The continuum has been fit by forcing a 1500K blackbody through the K' band observation.55
- Figure 3.20 - The inferred spectrum for the location 30W20N. The observations are also shown in the figure (data points with error bars). The spectrum was calculated from the observed data as described in the text - see eq (3.4). The vertical axis is plotted logarithmically in Janskies and the horizontal axis is in μm . The continuum has been fit by forcing a 1500K blackbody through the K' band observation.56
- Figure 3.21 - Combined FIR spectrum for NGC7023. The data points with error bars are the observations by Casey (1991) and Whitcomb. *et al.* (1981) discussed in the text. Casey's observations have been normalized to Whitcomb's beamsize ($1'$). The solid line is a spectrum of the form $vB_v(T)$ where B_v is the Planck function evaluated for $T=45\text{K}$ and forced through the $100\mu\text{m}$ observation.72
- Figure 3.22 - The locations of the IR emission peaks relative to HD200775. The black dot is star D, the \times and $+$ are the 55 and $125\mu\text{m}$ peaks, respectively (Whitcomb *et al.*, 1981), the square and (upper) diamond are the 100 and $200\mu\text{m}$ peaks (Casey, 1991), and the open circle is the $3.291\mu\text{m}$ peak. The beamsize of both sets of FIR observations is $1'$75
- Figure 3.23 - Color-color plot showing the [J-H] versus [H-K] colors for various components in NGC 7023. The diamonds and circles are the stellar sources and nebular positions, respectively. The heavy curves on the left side of the plot are the colors for ZAMS dwarfs, giants and supergiants respectively. The straight line with crosses is the reddening vector for the interstellar extinction curve of Draine (1989); the crosses are spaced such that adjacent crosses indicate $A_v=1$. The straight line with pluses corresponds to Planckian emission where the pluses run from 2000 to 10000K in 1000K increments (2000K is at the upper right).79
- Figure 3.24 - Spectrum of HD200775. A 17000K blackbody curve has been fit to the V band observation (square, Altamore, *et al.*, 1980), and is shown by the line on the plot. The broadband observations (Sellgren, 1984) are plotted as circles. All the observations have been corrected for $A_v=1.5$80
- Figure 3.25 - The spectra of HD200775 and star D (scaled by 70x), shown with squares and circles, respectively. The short wavelength data for HD200775 (all the points shortward of $2.5\mu\text{m}$) are from Sellgren (Sellgren, Werner and Dinerstein, 1983, Sellgren *et al.*, 1985) and the long wavelength data are from this work.....81

- Figure 4.1 - Typical detector geometry with electrical connections indicated. The photons are incident from the bottom of the diagram (referred to as the backside). The detector is biased across V_{bias} and V_{detsub} while V_{gate} is used to control the potential of the surface.90
- Figure 4.2 - Schematic of the InSb/passivant system (after Sze, 1982). Both the signs and the spatial locations of the various passivant charges are for illustrative purposes only, and do not represent the actual InSb passivation configuration.93
- Figure 4.3 - The fraction of the incident photoflux transmitted into the bulk of the InSb as a function of depth from the photo-incident side at 10K according to equations (4.1) and (4.2). The solid line is the fraction transmitted at an incident wavelength of $2\mu m$, the dotted line is that at $3\mu m$, and the dashed line at $5\mu m$. Note that this plot does not include surface reflections.95
- Figure 4.4 - Charge induced in the InSb surface as a function of surface potential. Note that the sign of the surface charge in the accumulated state is actually negative, and has been displayed with inverted sign here for compactness. The point where the surface charge goes to zero is the flatband condition ($\psi_{surface}=0$). The pictures at the bottom of the graph show the state of both the metallurgical depletion region as well as the induced depletion region at the surface.102
- Figure 4.5 - Typical latent image decay grayscale. The vertical bar frame is composed of 16 distinct rectangular subframes, each separated in time by 5 seconds so that the total time from the top to the bottom of the image is 80 seconds. The top image was acquired approximately 1 second after the photon source was removed from the field of view.104
- Figure 4.6 - A schematic of the circuitry associated with a single detector (photodiode). The direction of the photocurrent is indicated. The MOSFET directly above the detector is a switch that resets the detector to its maximum reverse bias at the start of the integration time (also see Figure 4.1 for a physical representation of the detector). The reset switch is open (non-conducting) while integrating photocharge on the capacitances of the detector and the source follower gate.106
- Figure 4.7 - A latent image decay for FPA84 at 6K after exposure to $2.2\mu m$ photons. The circles are the measured data after removal of the light source. Each is a five second integration on five second centers after shuttering the array. The solid line is a two time constant fit following the methodology described in the text ($\tau_1\sim 4sec$, $\tau_2\sim 37sec$), and the dashed line represents a single time constant fit to the data ($\tau\sim 19sec$).115
- Figure 4.8 - A comparison of a latent image decay for FPA84 (circles) with ASM244 (crosses) at 10K, showing that the decay time constants during the first 50 seconds are the same. The data have been normalized to show the similarities of the time constants. The illustrated decay was after exposure to $2.2\mu m$ photons.116
- Figure 4.9 - Plot of equation 11 from which trap parameters were extracted for FPA84. The top dotted line is the slow trap and the bottom is the fast trap. The slopes yield energies of 2.0 ± 0.7 and 2.4 ± 0.2 meV below the valence band edge for the fast and slow traps, respectively. Note the large error bars on the fast trap data.118
- Figure 4.10 - Energy band diagram for the gated area of the bulk InSb in thermal equilibrium (not to scale). All energies are in electron volts (eV) and are relative to the vacuum. E_C and E_V are the conduction and valence band energies respectively. The location of the trap within the oxide is much closer to the InSb surface than is shown here. The bandgap of In_2O_3 is indicated (3.5eV). The electric field across the oxide due to the applied gate bias will lower the conduction band of the In_2O_3 relative to the InSb valence band.121
- Figure 4.11 - WKB tunneling transmission factor for a 3.4eV barrier. The parameters used in the calculation are given in the text.122
- Figure 4.12 - Dark current vs. gate voltage for FPA84 at 12K. The dark current has been normalized to its maximum value at $V_{gate} = -1.3V$, where it measured to be $150 e^-/sec$. The measurement was performed by subtracting a 1 second integration from a 101 second integration. The increase in the dark current from $V_{gate} = -2.5V \rightarrow -2.8V$ is not clearly visible on the plot, but its magnitude increases from near zero to $15 e^-/sec$125

- Figure 4.13 - Magnitude of the saturated latent image (as measured by the strength of the first latent image in a decay) vs. V_{gate} . The data is for FPA84 at 12K, with an integration time of 5 seconds. The vertical axis has been normalized to the maximum value observed at $V_{\text{gate}} = -1.3\text{V}$, which was $7700 e^-$. The regions corresponding to the InSb surface are indicated at the top of the plot.126
- Figure 4.14 - Signal (e^-) in the first latent image, after the light source was removed, as a function of incident flux for FPA84 at 6.2K. The solid line is the best fit obtained from equations (4.7) and (4.15).129
- Figure 4.15 - Surface solution for n-type substrate assuming $T=12\text{K}$, $V_{\text{gate}} = -2.3 \rightarrow V_{\text{surf}} = -265\text{mV}$. The solid curve is the solution to Poisson's equation (equation (4.19)); the flatband condition occurs where $q\phi_{\text{surf}}/kT=0$, depletion and inversion of the surface occur to the left of this point and accumulation to the right. Note that the sign of the induced charge for the accumulated case is actually negative, its absolute value has been plotted here. The dotted curve represents the solution to the voltage drop equation (equation (4.21)). The intersection of the two curves represents the surface solution, i.e. under what condition the surface exists for a given gate bias.....134

List of Tables

- Table 2.1 - Summary of observations for NGC 7023 at WIRO. The column labelled λ and $\Delta\lambda$ are the central wavelength and the FWHM of each of the filters, respectively, Δt is the integration time/frame, NCD is the number of coadds, sample mode is the sampling technique used for that observation. NSAMP is the number of sample pairs per frame for the case of Fowler sampling and is 1 for S-G mode. The airmass is given for the middle of the observing period.6
- Table 2.2 - Photometric and astrometric calibrations. The standard stars used for each wavelength are indicated in the last column. The CVF platescale was only calibrated at 3.291 μm , the platescale for the remaining CVF wavelengths are taken to be the same as that at 3.291 μm9
- Table 2.3 - Sensitivity limits for the observations of NGC 7023. The standard deviation was obtained from the final (mosaic) images presented here. We take 3σ as our detection limit, though 1σ is displayed here. The limit in the final images varies over each image due to the nature of the mosaics; different total numbers of frames make up different portions of each image. The observed outer value quoted in the table represents the standard deviation in the outer portions of each final image and so is an upper limit to the peak sensitivity (typically seen in the middle of the frame). The observed central value is the minimum observed noise in the center portion of the mosaiced frame, the noise measured here is an upper limit due to the fact that this calculation of the standard deviation includes actual signal variation in the confused center region of the image. The final column is the total time, in seconds, spent observing NGC 7023.11
- Table 3.1 - Observed magnitudes for the stellar objects in NGC7023. The data given is from three independent sets of observations. SS is the data presented here, FP is Piché (1992), and KS is Sellgren (1983). The first three columns indicate the observer's identification key. Note that Sellgren's star j is a close visual binary, including both stars b and c. The fourth and fifth column indicate the offset (in arcseconds) from HD200775. The remaining columns provide the observed magnitudes in each of the broadband filters observed here. The data for the L" observation of HD200775 was taken from Sellgren (1985). Note that there are slight differences in the filter bandpasses between each of the observers listed here. The photometric accuracies are as follows: SS ± 0.05 , FP ± 0.07 , KS ± 0.05 . The stars listed with asterisks beside their identification are noted to be variable (see discussion in 3.6).29
- Table 3.2 - Nebular aperture photometry for NGC 7023. F_v is given in mJy, B_v is given in mJy/square arcsec. The aperture used for the photometry is given parenthetically (beam diameter in arcseconds) beneath the nebular position in the first column. KS represents Kris Sellgren's photometry and has been interpolated from published graphical spectrum. The position of the knot is 14W23N. The positions indicated with a dash indicate incomplete or no data available. The column labelled l:c(3.29) is the line to continuum ratio (see equation (3.2)) for the 3.29 μm emission feature. The final column labelled l:c(3.29)/l:c(3.37) provides a measure of the magnitude of the 3.4 μm emission feature; see the text in section 3.3.2 for a discussion of the relationship between the 3.367 μm observation and the 3.4 μm feature. Unless otherwise indicated, the photometric errors are $<10\%$30
- Table 3.3 - Filamentary and total frame photometry for NGC 7023. F_v is given in mJy, B_v is given in mJy/square arcsec. The aperture used for the photometry is given parenthetically (beam area in square arcseconds) beneath the nebular position in the first column. The positions indicated with a dash in the "Big Frame" indicate that our observations are incomplete here. The column labelled l:c(3.29) is the line to continuum ratio (see equation (3.2)) for the 3.29 μm emission feature. The final column labelled l:c(3.29)/l:c(3.37) provides a measure of the magnitude of the 3.4 μm emission feature; see the text in section 3.3.2 for a discussion of the relationship between the 3.367 μm observation and the 3.4 μm feature. Unless otherwise indicated, the photometric errors are $<10\%$31

Table 3.4 - The measured and inferred width for a section of each of the filaments. The locations measured were approximately 6" west and 17" south of star D for the western filament and directly north of HD200775 for the eastern. The methodology employed to determine the inferred width is discussed in the text. The data presented in this table was determined from Figure 3.10 and Figure 3.11.	35
Table 3.5 - $F_{3.4}/F_{3.3}$ emission line ratios. The ratios were derived from the calculated spectrum of NGC 7023 for the various components listed. The inferred spectral values were calculated using equation (3.4) assuming $FWHM_{3.4\mu m}=0.028\mu m$ and $FWHM_{3.29\mu m}=0.042\mu m$. The total strength in the line was calculated by integrating the calculated spectrum over the linewidth, assuming a gaussian profile for each of the features.	60
Table 3.6 - Peak surface brightness in mJy/square arcsec at three locations in NGC 7023. Offset is the average angular offset between HD200775 and the nebular position in arcsec. The theoretical value for the maximum reflected light, $Max B_{ref}$, is calculated from equation (3.6) assuming unity albedo and using the observed infrared fluxes for HD200775.	62
Table 3.7 - The ratio of reflected to total light (S_{ref}/S_{total}) from Sellgren <i>et al.</i> , (1992). [J-H] and [H-K] colors as measured in this work are given in the final two columns for comparison.	63
Table 3.8 - The $L_{FIR}/L_{3.29}$ ratio in a variety of IR bright galaxies. References for these data are: (1) Moorwood (1986) , (2) Dennefeld and Désert (1990), (3) Satyapal, 1995, (4) Mizutani <i>et al.</i> (1989). An asterisk beside a reference indicates that the ratio has been corrected for extinction at 3.29 μm	68
Table 3.9 - Total FIR luminosity of NGC7023. The methodology used to determine the FIR luminosity is described in the text (see also Figure 3.21). The reference column refers to: (1) Whitcomb <i>et al.</i> , (1981), (2) Casey, (1991).	72
Table 3.10 - Broadband magnitudes for star D over a 13 year baseline. References are as follows: (1) Sellgren, 1983, (2) Piché, 1992, (3) this work.	81
Table 4.11 - Summary of key arrays used to study latent images. The size column refers to the number of pixels as rows \times columns, doping is the net carrier concentration of the InSb substrate in cm^{-3} , front passivant is the type of front surface passivation material, and thickness is the measured thickness of the InSb detector material. ASM243 & 244 were experimental, composite arrays with both square and octagonal pixels of two different sizes (76 μm pitch), with regions of gapped gate electrodes. The 60 μm pixel size is in boldface type because it is the pixel geometry used in our investigation of latent images on ASM244.	106
Table 4.12 - Trap parameters for FPA84 from fitting equation (4.15) to the latent image decays. The fitting scheme was iterative, as described in the text. The error in the short time constant is considerably larger than that of the long time constant. The column labelled number of measurements is the number of decays for which time constants were determined and averaged to yield the fast and slow time constants listed in this table.	118
Table 4.13 - Total number of traps inferred from saturated latent image data in conjunction with the fluxes that generated the saturation level for 5 SCAs. The two arrays listed with asterisks are gateless devices, while the remaining three have gates.	128
Table 4.14 - Summary of surface state with and without passivation trapped charge. The trapped charge ($6.8\times 10^9 cm^{-2}$, the saturated value) was added to the passivation fixed charge for the purposes of calculation after the fixed charge was determined. All calculations were performed for $T=12K$, $\Phi_{ms}=-0.1V$, and $Q_{it}=0.1\times Q_{ss}$. The rows are as follows: state refers to the state of the InSb surface: accumulated, flatband, depleted or inverted, Φ_s is the surface potential, E_{fij} , W_{fij} and Q_{fij} (fij=field induced junction) are the average electric field strength, width and fixed charge, respectively, in the field induced depletion region, Q_s , Q_{sn} and Q_{sp} are the total, electron, and hole areal charge density, respectively, induced in the InSb surface, and E_{ox} is the average electric field strength in the oxide.	137

1. Introduction

Large area, infrared imaging arrays have revolutionized infrared astronomy. They represent an enormous increase in the capability to perform accurate, simultaneous photometric and astrometric observations, and they provide an incomparable tool for morphological studies and searches for new objects. The number of detector elements (pixels) in an integrating array maps directly to increased observing efficiency, and increased sensitivity. The utility an image presents over aperture photometry is substantial; aperture positions relative to stars, etc. are unambiguous in an image, whereas aperture measurements are prone to pointing inaccuracies and require extreme care in their calibration. Arrays are also ideal for investigation of small scale structure that is typically averaged out by larger beam observations and they excel in comparison with single element detectors for mapping extended sources at high spatial resolution.

We investigate the nebular excitation and morphology of NGC 7023, a nearby reflection nebula with a young stellar cluster embedded in a molecular cloud, via both broadband and spectroscopic infrared imaging. We have imaged the central region of the nebula over the wavelength region from $1.2\mu\text{m}$ – $3.8\mu\text{m}$ including the $3.29\mu\text{m}$ emission feature, a tracer of star formation and the attendant ultraviolet radiation fields. Chapter 2 discusses the observations: the camera system and telescope, the specifics of the observational program employed to investigate NGC 7023, as well as the data reduction methodology used to generate the final images and the photometric and astrometric calibrations. In chapter 3, we discuss our observations of NGC 7023. The nebula exhibits complex spatial structure, evident in the images we present. The strength and morphology of the $3.29\mu\text{m}$ emission feature varies considerably over the nebula, and we investigate this variation and its excitation. The $3.4\mu\text{m}$ emission feature (sometimes present in astrophysical objects which exhibit the $3.29\mu\text{m}$ feature) is inferred to be present as well, and we consider the excitation of this feature in the context of a current model obtained from the literature. We identify the stellar-like objects observed in NGC 7023 as both late-type main sequence stars and protostellar objects.

Finally, chapter 4 examines a manifestation of charge trapping observed in photovoltaic indium antimonide imaging arrays used for astronomy. The phenomenon evidences itself as “image memory,” or “image latency” and can present significant difficulty for astronomical observations. Experiments were designed and carried out in order to investigate this effect and a model is presented in which the image latency is hypothesized to result from charge trapping at the surface of each pixel. We propose a specific impurity species as the entity responsible for the charge trapping, which should enable a reduction in the magnitude of image latency via appropriate changes in the device fabrication and processing.

2. Observations

2.1 Instrumental

All observations were made with the third generation University of Rochester IR Array Camera, which houses a cryogenically cooled 256×256 indium antimonide (InSb) photodiode focal plane array from Santa Barbara Research Center. The detector array used for these observations was FPA48, a 256×256 pixel array with 30μm pixels (additional characteristics are listed in table 4.1). An individual detector's geometrical configuration can be found in Figure 4.1. The camera housing consists of a double jacketed liquid nitrogen/liquid helium dewar capable of cooling the detector array to 4.2K. A fiberglass printed circuit board holds the detector array behind a cooled filter wheel with both broadband filters ($d\lambda/\lambda \approx 20\%$), and two 2% spectral resolution circular variable filters (CVF), covering the wavelength ranges from 1.9 to 2.8μm and 2.5 to 4.5μm. The spectral resolution of the CVF is limited by a cooled aperture (Lyot stop) placed just above it. Matching the f beam of the dewar to that of the telescope determines the diameter of the Lyot stop according to $d_{ls} = fl / f\#_{telescope}$ where d_{ls} is the diameter of the Lyot stop, $f\#_{telescope}$ is the f number of the telescope and fl is the focal length of the camera's re-imaging optics. All the observations discussed here were performed with a Lyot stop diameter of 2.35mm on an f/30 telescope, resulting in a CVF spectral resolution on the order of 2% (see equation (3.3)). The camera optics consist of a nearly achromatic Si/CaF₂ doublet which also serves as a vacuum window. The optics serve two purposes: to transfer the telescope focus onto the detector array, and to image the telescope secondary mirror onto the plane of the Lyot stop. The purpose of the latter is to exclude radiation from outside the telescope beam from spilling onto the detectors.

The sensitivity of the array camera is ultimately limited by the detector dark current and the read noise of the multiplexer and drive electronics (for the non-background limited case). The detector dark current was measured to be $<1 \text{ e}^-/\text{sec}$ at 30K. The detector referred noise was measured to be

$12e^-$ at 29K using Fowler sampling (Fowler & Gatley, 1990, 1991) with 64 sample pairs. The camera controller system is capable of a variety of sampling techniques; the sampling methods used in the observations presented here were Fowler sampling and signal minus ground (hereafter, S-G). The choice of sampling mode for a given observation is determined by the dominant noise source (electronics or photon), as well as the type of astronomical object being imaged (Garnett & Forrest, 1993). Fowler sampling is a non-destructive detector readout technique whereby the detector output is sampled many times at both the beginning and the end of the integration period, and the resulting sample pairs are averaged and then subtracted in order to increase the signal to noise ratio. The signal minus ground sampling technique is a single measurement of the detector's output at the end of the integration time, referenced to analog ground. This allows for the most rapid readout of the detector array and is used in cases where the background is so large as to require a very short integration time, typically used for long wavelength ($\lambda > 3\mu\text{m}$), broadband observations. Latent images were not seen in the observations presented here (see section 4 for a discussion of latent images).

2.2 Observations

The observations of the reflection nebula NGC 7023 presented here were obtained at the Wyoming Infrared Observatory on the nights of 30 September and 2 October, 1992. The detector array was operated over a temperature range of 28–30K. Details of the observational program (on-chip integration time, coadded readouts/frame, etc.) can be found in Table 2.1. Broadband and CVF observations of NGC 7023 were made between 1.13–1.16 airmasses¹. CVF observations of the 3.29 μ m emission feature in NGC 7023 were made at 3.291 μ m and at the neighboring wavelengths of 3.158 and 3.367 μ m. The CVF used for these observations has a measured resolution ($d\lambda/\lambda$) of 2.0% for the equivalent width (EW)² and 1.9% for the full width at half maximum (FWHM)³. For all observations, pairs of on-source frames were alternated with single background frames, switching between feature and continuum for roughly every four source frames. The background beam used for subtraction (from the source frame) was observed by slewing the telescope 120" to the W of any given source beam. The background frames were acquired under the same operating conditions as the image frames. All the observations presented here were obtained in Fowler sampling mode except for the L" acquisitions which were acquired in S-G mode. This was necessary because very large thermal backgrounds from the telescope and optics required that fast integrations be used, which in turn imposes a minimal number of samples per integration time.

¹ Airmass is defined to be the length of the atmospheric column through which one is observing, normalized to that when looking at the zenith. Thus, by definition, the air mass is equal to one when looking at the zenith, and is greater than one when observing anywhere but the zenith.

² Equivalent width, for an emission line, is defined as the spectral width (typically in wavelength units) that an emission line would have with a unity amplitude, a perfectly rectangular passband and the same integrated area as the observed emission profile has. Equivalent width is given by:

$$EW \equiv \int_0^{\infty} \frac{F_{\lambda}^{\text{line}} - F_{\lambda}^{\text{contin}}}{F_{\lambda}^{\text{contin}}} d\lambda$$

³ The FWHM of an emission line is defined as the spectral width (typically in wavelength units) of an emission line where the amplitude of the emission is one half that of its peak value.

Date	Filter	λ (μm)	$\Delta\lambda$ (μm)	Δt (millisec)	NCD	Sample Mode	NSAMP	Airmass
30 Sep 92	K'	2.158	0.385	5000	4	Fowler	8	1.13
"	CVF	3.160	0.060	500	64	Fowler	1	1.14
"	CVF	3.291	0.063	500	64	Fowler	1	1.14
"	CVF	3.367	0.064	500	64	Fowler	1	1.14
"	L''	3.810	0.63	60	512	S-G	1	1.16
2 Oct 92	J	1.245	0.292	10000	2	Fowler	8	1.12
"	H	1.639	0.290	5000	4	Fowler	8	1.13

Table 2.1 - Summary of observations for NGC 7023 at WIRO. The column labelled λ and $\Delta\lambda$ are the central wavelength and the FWHM of each of the filters, respectively, Δt is the integration time/frame, NCD is the number of coadds, sample mode is the sampling technique used for that observation. NSAMP is the number of sample pairs per frame for the case of Fowler sampling and is 1 for S-G mode. The airmass is given for the middle of the observing period.

2.3 Calibration

Standard stars of known brightness were observed in order to perform the photometric calibration; these stars are listed in Table 2.2. Photometric calibration was determined by calculating the counts (analog to digital converter bit count, hereafter ADU) for zero magnitude (CFZM), which is defined by:

$$\text{CFZM}(\lambda) = \frac{10^{0.4 \cdot m(\lambda)} \cdot \text{counts}}{\Delta t \cdot A} \quad (2.1)$$

where $m(\lambda)$ is the apparent magnitude of the standard star, counts are the signal counts for the standard star, Δt is the integration time and A is the electronic gain of the camera system (in this case, $A = 20$). Once CFZMs are determined for each wavelength from observations of standard stars, equation (2.1) can be inverted to solve for m , the apparent magnitude of the desired object. Alternately, the spectral flux density in Janskies ($1 \text{ Jy} = 10^{-26} \text{ Watt cm}^{-2} \text{ Hz}^{-1}$) can be determined according to the following equation:

$$F_{\text{Jy}} = \frac{\text{JyFZM} \cdot \text{counts}}{\text{CFZM} \cdot \Delta t \cdot A} \quad (2.2)$$

where JyFZM are the Janskies for zero magnitude, and the remainder of the variables in equation (2.2) are as in equation (2.1). The adopted JyFZM for the broadband filters have been derived from Cohen, *et al.*, (1992) and Tokunaga (1986) via interpolating to the central wavenumbers of our filters, and for the CVF observations they have been derived from the aforementioned references in combination with model atmospheres calculated for α Lyra (Kurucz, private communication) and are given in Table 2.2. These values are all based on α Lyra being defined as the zero of the magnitude scale.

The point spread function (FWHM) was determined by fitting a radially symmetric gaussian function to the radial brightness profile of both the standard stars used for the photometry as well as for the stars observed in NGC 7023. The standard star used for the CVF calibration (κ And) was recorded with nearly the same integration time immediately after the images of NGC 7023 were acquired. A

broadband calibration source, HD203856, was also measured both before and after the observations of NGC 7023. The FWHM for the CVF data was found to be $1.0 \pm 0.1''$ (440 AU^4 at a distance of 440 parsec⁵, the distance of NGC 7023); the adopted values are given in Table 2.2. The platescale at the telescope for all observed wavelengths was determined using the stars in the Trapezium. Two dimensional gaussian profiles were fit to the stellar images to locate their centers and then known positional values for the stars were used to determine the platescale at each wavelength. The position angle of the array was measured to be $0.5 \pm 0.1^\circ$. The positions of the 4 Trapezium stars were used to determine the magnitude of any optical distortion (see appendix) present in the images due to the camera system's refractive optics. The distortion was found to be negligible (<1 pixel at the corner of the array), so no correction was performed. This result is in agreement with the distortion expected from ray-tracing calculations. A discussion of optical distortion, and any corrections applied to an image due to its presence can be found in the appendix. Table 2.2 summarizes both the photometric and astrometric calibrations adopted.

⁴ An astronomical unit (AU), is the mean distance between the earth and the sun, 1.5×10^{13} cm (Allen, 1976).

⁵ A parsec (pc) is an astronomical distance unit corresponding to 3.09×10^{18} cm.

λ (μm)	Filter	PS ("'/pixel)	FWHM (")	JyFZM	CFZM	Standard Stars
1.245	J	0.223	1.42	1630	17926	HD203856
1.639	H	0.234	1.36	1060	22188	HD203856
2.158	K	0.239	1.0	660	14949	GI811.1, HD3029
3.158	CVF	–	1.0	336	263	κ And
3.291	CVF	0.241	1.0	331	190	κ And
3.367	CVF	–	1.0	318	259	κ And
3.810	L"	0.244	1.4	250	3225	κ And

Table 2.2 - Photometric and astrometric calibrations. The standard stars used for each wavelength are indicated in the last column. The CVF platescale was only calibrated at 3.291 μm , the platescale for the remaining CVF wavelengths are taken to be the same as that at 3.291 μm .

Table 2.3 shows the sensitivity limits for the camera/telescope combination. Assuming that the observations are photon noise limited, we expect the standard deviation (in counts) for any given background subtracted frame to be given by:

$$\sigma_{\text{adu}} = \frac{\sqrt{\text{counts} \cdot \alpha}}{\alpha} \cdot \frac{\sqrt{2}}{\sqrt{\text{NCD}}} \cdot \frac{1}{\sqrt{N_{\text{frames}}}} \quad (2.3)$$

where counts is the average number of background counts, α is the number of electrons per ADU, NCD is the number of coadds for the frame under consideration, N_{frames} is the number of frames averaged into the mosaic image under consideration and the $\sqrt{2}$ accounts for the background subtraction. The conversion factor is determined by the capacitance of the detector array and the voltage gain in the electronics, and was measured using the signal vs. noise squared method (Fowler, *et al.*, 1987). For the observations presented here, α was determined to be 10 e⁻/ADU. The “theoretical” values predicted by equation (2.3) agree with the standard deviations obtained from a random sampling of frames from the observations to within 10% for both the CVF and the L" data. At the shorter (non-thermal) wavelengths the observations may have included a non-negligible read noise component, so using

equation (2.3) to calculate the expected noise provides a lower limit to the standard deviation. Table 2.3 gives the sensitivity (3σ) obtained from the final mosaic images, along with the relevant observing parameters. Given a final, reduced frame, the sensitivity is a function of the position within that image because each final image consists of many individual background subtracted frames mosaiced together, hence there are differing numbers of frames contributing to a given region within the final image. The two columns listed as observed outer and observed central in the table are the measured 3σ value in the outer and central portions of the frame, respectively. There are fewer frames in the outer regions of each image than in the central regions, so the former will have lower sensitivities (*i.e.* are noisier). The standard deviation measured in the central region of a given frame is an upper limit to the actual sensitivity because these regions include real signal variations in the confused center region of the images. The agreement with the theoretical calculations is seen to be good.

λ (μm)	3σ (mJy/pix)			Itime (ms)	Numbe of frames	NCD	total time (sec)
	observed outer	observed central	predicted				
1.245	2.3E-3	1.6E-3	1.3E-3	10000	21	2	420
1.639	2.4E-3	1.8E-3	1.5E-3	5000	18	4	360
2.158	2.7E-3	2.1E-3	2.1E-3	5000	20	4	400
3.158	0.39	0.30	0.22	500	4	64	128
3.291	0.37	0.27	0.25	500	22	64	704
3.367	0.44	0.41	0.37	500	4	64	128
3.810	0.19	0.18	0.21	60	4	512	123

Table 2.3 - Sensitivity limits for the observations of NGC 7023. The standard deviation was obtained from the final (mosaic) images presented here. We take 3σ as our detection limit, though 1σ is displayed here. The limit in the final images varies over each image due to the nature of the mosaics; different total numbers of frames make up different portions of each image. The observed outer value quoted in the table represents the standard deviation in the outer portions of each final image and so is an upper limit to the peak sensitivity (typically seen in the middle of the frame). The observed central value is the minimum observed noise in the center portion of the mosaiced frame, the noise measured here is an upper limit due to the fact that this calculation of the standard deviation includes actual signal variation in the confused center region of the image. The final column is the total time, in seconds, spent observing NGC 7023.

2.4 Data Reduction

The camera system's response to integrated photocurrent is intrinsically non-linear. The detectors are reverse biased at the start of each integration and the photocurrent (plus any dark current) is integrated on both the detector and source follower capacitances, causing the detectors to debias (see chapter 5, figure 7 for the circuit schematic). The charge capacity of the detectors is limited by the difference between the reset voltage and a threshold voltage drop from the voltage rail of the unit cell source follower in combination with the capacitance according to the relation $Q=CV$, where Q is the integrated charge, C is the capacitance and V is the voltage generated across the capacitance by the integrated charge. We assume that the detectors can be approximated as abrupt p-n junctions; we have shown this approximation to be valid via comparing laboratory linearity measurements with the abrupt junction approximation, as well as via detector thickness analysis (Garnett, private communication). The non-linearity arises because the detector capacitance is a function of the bias according to:

$$\begin{aligned} Q(V) &= (V_{\text{bias}} - V) \cdot (C_{\text{det}} + C_{\text{fix}}) \\ &= (V_{\text{bias}} - V) \cdot \left(\frac{2C_o}{\sqrt{1 + \frac{V}{V_{\text{bi}}}}} + C_{\text{fix}} \right) \end{aligned} \quad (2.4)$$

where Q is the total integrated charge, V_{bias} is the detector bias at the start of the integration, V is the detector bias the end of the integration, C_{det} is the detector capacitance, C_{fix} is the fixed capacitance (the source follower MOSFET gate capacitance plus any additional stray capacitances), C_o is the zero bias detector capacitance, and V_{bi} is the built-in voltage of the detectors (234mV at 30K with negligible temperature dependence over the operating range).

The non-linearity referred to in equation (2.4) is corrected by applying a linearizing algorithm (in software) to all raw frames on a pixel by pixel basis prior to subsequent image processing. Based on the observed non-linearity as a function of signal, S , we assume that the linearized signal is given by an equation of the form:

$$S'_{i,j} = S_{i,j} \cdot \left[1 + A_{i,j} S_{i,j} \right] \quad (2.5)$$

where $S_{i,j}$ is the measured signal in pixel i,j , $S'_{i,j}$ is the linearized signal for that pixel and $A_{i,j}$ is a matrix determined by requiring that two signal frames (measured signal, preferably at a constant flux level, but not so large as to saturate the array output) acquired at very different integration times be equal to the ratio of the integration times. The following equation expresses the methodology used to uniquely calculate the A matrix:

$$\frac{S(\Delta t_1) \cdot (1 + A \cdot S(\Delta t_1))}{S(\Delta t_2) \cdot (1 + A \cdot S(\Delta t_2))} = \frac{\Delta t_1}{\Delta t_2} \quad (2.6)$$

where Δt_1 and Δt_2 are the integration times for the frames used to calculate the A matrix, and there is an implicit i,j subscript on both A and S since there is a unique A value for every pixel on the array. The derived A matrix was tested for a variety of different integration times to ensure that it successfully linearizes the measured signal to $\leq 3\%$ over a certain “useful” range. We define this range as a fraction of the saturation voltage of the entire array, where saturation is defined to be the point at which the array output no longer responds to increases in either photon flux or integration time. This fraction was empirically chosen to be 0.85 and is given by:

$$\text{SatFract} = \frac{S_{3\%}}{S_{\text{sat}}} \quad (2.7)$$

where S_{sat} is the saturation signal, $S_{3\%}$ is the signal at the integration time where the uncorrected data deviates from the A map fit by more than three percent, and there is an implicit i,j subscript on all the terms since there is a unique Satfract value for every pixel on the array. Thus, if the signal in a given pixel exceeds this fraction, the error in the linearity correction will exceed 3%. The observations presented here (see Table 2.1) were acquired in two different modes: Fowler and S–G. The linearity correction differs from one mode to the next, so that it was necessary to obtain linearization data for each of the observing modes.

The standard star frames were processed in the same manner as the NGC 7023 frames. Temporally weighted background frames were subtracted from each of the image frames; the background frames were acquired under the same operating conditions as the image frames. The frames were then divided by a flat field (the average of several dome frames with the average of several sky frames subtracted) to account for pixel to pixel variations in response and multiplied by the scalar average of the flat field. Dead, noisy and saturated pixels were flagged and replaced via bilinear interpolation using the four nearest neighbors. A correction for electronic crosstalk was applied to the L' frames because they were read out faster. Relative spatial offsets between frames were determined by a cross-correlation routine for the brightest point-like source in the frames. The frames were then simultaneously shifted and coadded to make the final mosaic image. In order to complete the final K' mosaic frame, it was necessary to add 15–30 ADU/pixel offsets to several of the frames in order to get the backgrounds to match. This corresponds to 7–13 μ Jy, or 3-5 times the 3σ sensitivity limit in the K' band.

Observations with relatively less total time spent on source (less total integration time, fewer frames and lower signal to noise ratios - see Table 2.3) may have inaccurate photometric zero points (hereafter, pzp), *i.e.* there can be a global (spectral flux density) offset in a given frame. This is because our pzp relies on the background subtracted signal averaging to zero over many frames, thus if there are not enough frames in the final image, a photometric offset can result. Specifically, the final 3.158 μ m and 3.367 μ m CVF frames as well as that at L'' are each comprised of approximately 2 minutes worth of total observing time apiece, and may therefore be subject to this error. One method to determine the accuracy of our pzp is by comparison with other observers. Sellgren (1984) observed the location 30W20N in a 10.5'' beam using both a CVF as well as with broadband filters, and we use these observations as an independent check on our pzp. An additional method useful for the observations presented here is to spatially locate the minimum flux region(s) in a frame with relatively higher signal to noise (the K' image, for example) and then verify that the signal in the low signal to noise frame under consideration is also a minimum at the same location. Ideally, this minimum will be zero signal,

thereby allowing us to subtract a constant value from each pixel in the frame with lower signal to noise until its photometry in the same aperture yields null flux. We consider each of the wavelengths listed as having low integration times in turn in the following discussion, and we discuss and determine their pzp.

Our final 3.158 μm frame yielded spectral flux densities considerably in excess of those observed by Sellgren (1984). We determine the “true” value of the continuum at 3.158 μm from Sellgren’s spectrum for the location 30W20N by averaging her measurements (7 different wavelengths) covering the range 2.901–3.187 μm . Given Sellgren’s 1% spectral resolution and a 42nm FWHM for the 3.29 μm emission feature (Tokunaga, 1991), her 3.187 μm observation will contain $\sim 10^{-5}$ of the flux in the 3.291 μm feature, thereby validating this choice of wavelength for the continuum observation. In this manner we obtain $34 \pm 10 \text{ mJy}$ for the 3.158 μm continuum at 30W20N in a 10.5” beam diameter. The standard deviation (10mJy) is the dominant error source in our pzp-corrected 3.158 μm data. We adjust all of our 3.158 μm observations by subtracting a constant 188 $\mu\text{Jy/pixel}$ (3236 $\mu\text{Jy/square arcsec}$) from each pixel in the frame. For comparison, the flux error (1σ) in our 3.158 μm frame for any given 10.5” beam is 5.0mJy, which amounts to 3 $\mu\text{Jy/pixel}$ (or 58 $\mu\text{Jy/square arcsec}$), so we see that the error in pzp is many times larger than the sensitivity of our measurements. The minimum spectral flux density in the K’ frame occurs at the location 70N (see Table 3.2), where the photometry yields $2.7 \pm 0.1 \text{ mJy}$ in a 10.5” aperture. The 3.158 μm flux is also a minimum at this location, and while the 3.158 μm observation is slightly higher than that at 60N, they are both well within 1σ of one another.

The 3.367 μm frame requires no pzp correction. Comparison of our photometry for the 30W20N position with that of Sellgren (1984) yields consistent measurements; we find $71 \pm 10 \text{ mJy}$, while Sellgren obtained $78 \pm 10 \text{ mJy}$. The L” observations require additional consideration owing to the uncertainty in Sellgren’s observations. She presents both broadband as well as CVF measurements in this spectral region, and the two sets of observations do not match very well. Her broadband photometry at 3.8 μm yields $56 \pm 4 \text{ mJy}$, while the CVF photometry indicates flux densities

approximately 25% higher. Averaging her CVF measurements over the bandpass of our broadband filter with equal weighting for each wavelength results in $72 \pm 8 \text{ mJy}$ for the 30W20N aperture. We find $80 \pm 2.3 \text{ mJy}$ for this aperture, and conclude that our L” broadband pzp does not require adjustment.

3. NGC 7023

3.1 Introduction

NGC 7023 is a visual reflection nebula in the constellation Cepheus, at the northern end of a $30' \times 60'$ dark molecular cloud, approximately 440 pc distant. It is prominent in the Palomar Sky Survey photographs, in a region largely free of other nebulosity. NGC 7023 lies between a highly obscured region $\sim 7'$ to the east and a “hole” in the cloud $\sim 5'$ to the west. The cloud has a FIR light to mass ratio $\sim 5 L_8 / M_8$, typical of active star formation regions (Fuente *et al.*, 1992). The central illuminating star is HD200775, a Herbig Be star (thought to be the high mass analog of a TTauri star⁶), is classified as a B3e star with an effective temperature of $\sim 17,000\text{K}$ and a strong IR excess between 1 and $20 \mu\text{m}$ (Whitcomb *et al.*, 1981). This IR excess is attributed to thermal processes: free-free (bremsstrahlung) and free-bound (hydrogen) emission from the ionized stellar wind as well as emission from hot circumstellar dust (Milkey & Dyck, 1973, Strom, Strom & Grasdalen, 1975). HD200775 has a strong stellar wind, as evidenced by a well-developed P Cygni profile in MgII (Grady, Imhoff and Bjorkman, 1988). Radio observations of HD200775 are consistent with free-free emission in a spherically symmetric, fully ionized stellar wind (Skinner *et al.*, 1990). Estimates for the stellar mass loss range from $2\text{--}8 \times 10^{-7} M_8 \text{ year}^{-1}$. CO observations show HD200775 to be driving a bipolar outflow with lobes approximately aligned in the NE-SW direction extending $4'$ from HD200775 in each direction (Watt *et al.*, 1986b).

Since B stars do not generate large HII regions within reflection nebulae, the radiative cross sections of the dust grains will dominate the equations of radiative transfer in the nebula. The radius of a (pure hydrogen) classical Strömngren sphere in the first approximation is given by:

⁶ TTauri stars are low mass protostars embedded in dense clouds of gas and dust that show infrared excess, stellar winds, and rapid variability.

$$r = \left[\frac{3}{4\pi} \frac{N_L}{n_H^2 \alpha_B} \right]^{1/3} \quad (3.1)$$

where n_H is the hydrogen number density, α_B is the recombination coefficient for hydrogen and N_L is the ionizing stellar luminosity (Osterbrock, 1989). Taking $\log(N_L)=43.69$ (Panagia, 1973) for a B3 ZAMS star, $\alpha_B=2.6 \times 10^{-13} \text{ cm}^3 \text{ sec}^{-1}$ (for $T=10^4 \text{ K}$, Osterbrock, 1989) and $n_H=10^4 \text{ cm}^{-3}$ (calculated from VLA radio observations of Skinner *et al.*, 1990) we find $r \sim .002 \text{ pc}$ which corresponds to approximately $1''$ at a distance of 440pc. As the radiation field from the central star expands, both r^{-2} and selective extinction will attenuate the radiative energy density. The wavelength dependence of the reflection nebula's appearance is therefore a reflection of the wavelength dependence of the grain properties.

A simple visual inspection (see Figure 3.1–Figure 3.6) shows that the nebula is nonuniform, consisting of knots, filaments and clumps of varying surface brightness. We have observed 11 stars in the H band image (Figure 3.2) including HD200775, and labeled them with the letters A–K in the direction of increasing right ascension. Stellar photometry was calculated from the aperture containing 90% of the total light after the sky background was subtracted from the signal in the aperture. Sky background levels were determined from the mean signal in a 3–5 pixel wide annulus ($0.7''$ – $1.2''$ in the K' band), and the backgrounds thus chosen were verified by plotting the radial signal from the stellar object out to the location of the annulus to verify that the background selected was reasonable. The placement of the background annulus varied with the ambient structure of the particular star being measured, typically being in the range 15–20 pixels ($3.6''$ – $4.8''$ in the K' band) from the edge of the photometric aperture. In cases where the surrounding emission did not permit the use of an annulus to determine the sky background, the background was estimated by plotting the signal out to a radius of approximately 15 pixels and using the minimum range obtained. The observed magnitudes for these stars are given in Table 3.1, where they are compared with other observers' data.

The two filamentary structures have been labeled the western and the eastern filament in the direction of increasing right ascension, *i.e.* the filament directly to north of HD200775 is termed the eastern filament. We note that the western filament may actually be two separate filaments whose inclusion as a single entity is an artifact of our perspective. Table 3.2 and Table 3.3 provide the nebular photometry for various locations in the frame as well as the filaments. Photometry for the filaments was obtained by first magnifying or demagnifying each image to achieve a common platescale across all wavelengths (chosen to be the K' band platescale), and then convolving each of the images with a radial gaussian kernel in order to obtain the same resolution in each frame. The polygonal apertures used to acquire the filamentary photometry were defined by the half peak flux contour for the K' frame. The same polygonal aperture (one for each filament) was then overlaid on each of the images and the total signal inside the aperture was determined (no background subtraction was performed). The positioning of the polygonal aperture was verified by using a contour plot with the polygon overlaid.

Figure 3.1 - Broadband J image of NGC7023. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 97.7"×97.2" (E/W×N/S).

Figure 3.2 - Broadband H image of NGC 7023. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 106.0"×101.1" (E/W×N/S).

Figure 3.3 - Broadband K' image. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 102.1"×102.1" (E/W×N/S).

Figure 3.4 - CVF 3158 image. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 67.0"×62.4" (E/W×N/S).

Figure 3.5 - CVF 3291 image. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 104.4"×102.4" (E/W×N/S).

Figure 3.6 - CVF 3367 image. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 61.7"×61.7" (E/W×N/S).

Figure 3.7 - Broadband L" image. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 62.5"×62.5" (E/W×N/S).

Figure 3.8 - K' Frame with various apertures, cuts, etc. overlaid. North is at the top and east is to the left. The grayscale is displayed linearly from XX to YY (black to white). The frame is 102.1"×102.1" (E/W×N/S).

Figure 3.9 - HK' color temperature map, emissivity index $n=-1.75$. Prior to calculating the map, the K' frame was first convolved with a gaussian kernel to match the resolution (PSF) of the H frame and then demagnified to match the platescale of the H frame.

Stellar ID		Position		[J]		[H]		[K]		CVF		[L']		
SS	FP	KS	E/W	N/S	SS	FP	KS	SS	FP	KS	3.158	3.291	3.367	SS
HD*	-	a	0	0	-	-	6.09	-	-	5.42	3.92	3.86	-	(3.37)
A	61	-	-68.3	0.5	13.07	12.82	-	12.56	12.04	-	-	12.5	-	-
B*	60	j	-58.8	51.5	11.17	11.09	10.67	10.17	10.11	9.62	-	9.21	-	-
C*	57	j	-54.9	51.0	12.97	12.15	-	11.80	11.02	-	-	10.31	-	-
D*	52	k	-36.6	48.9	11.30	11.27	10.64	10.33	10.36	9.80	9.86	9.24	9.26	8.93
H*	46	-	0.3	38.2	14.76	14.16	-	13.89	13.47	-	-	>11.12	>12.28	>12.8
J	42	m	25.1	69.3	-	12.18	12.13	11.42	11.37	11.37	-	10.84	-	-
K	41	-	26.7	19.7	-	15.25	-	15.04	14.57	-	-	>12.5	-	-

Table 3.1 - Observed magnitudes for the stellar objects in NGC7023. The data given is from three independent sets of observations. SS is the data presented here, FP is Piché (1992), and KS is Sellgren (1983). The first three columns indicate the observer's identification key. Note that Sellgren's star j is a close visual binary, including both stars b and c. The fourth and fifth column indicate the offset (in arcseconds) from HD200775. The remaining columns provide the observed magnitudes in each of the broadband filters observed here. The data for the L' observation of HD200775 was taken from Sellgren (1985). Note that there are slight differences in the filter bandpasses between each of the observers listed here. The photometric accuracies are as follows: SS ± 0.05 , FP ± 0.07 , KS ± 0.05 . The stars listed with asterisks beside their identification are noted to be variable (see discussion in 3.6).

	J		H		K		CVF		L''		$\frac{l:c(3.29)}{l:c(3.37)}$			
	SS	KS	SS	KS	SS	KS	3.158	3.291	3.367	SS		KS		
30N	F_v	10.4	6.6	15.6	11.4	18	17.4	29±15	194	61±9	49±5	48	9.8	4.1
(10.5'')	B_v	0.12	0.076	0.18	0.132	0.21	0.201	0.34	2.23	0.704	0.57	0.55	n/a	n/a
60N	F_v	5.3	–	5.6	–	4.9	5.5	18.6±15	211	62.2±9	13±3	16	42.1	3.6
(10.5'')	B_v	0.061	–	0.065	–	0.057	0.074	0.21	2.44	0.716	0.15	0.18	n/a	n/a
70N	F_v	3.7	2.4	3.7	2.8	2.7	–	22.2±15	212	18±7	< 7.2±3	–	77.5	13.7
(10.5'')	B_v	0.042	0.028	0.043	0.032	0.031	–	0.26	2.45	0.21	< 0.08	–	n/a	n/a
30W 20N	F_v	11.7	7	14.7	12.7	18.4	18.7	(34)±15	244	(71.3)±9	(80)	56	12.3	4.3
(10.5'')	B_v	0.14	0.081	0.17	0.147	0.21	0.21	(0.39)	2.8	(0.82)	(0.92)	0.65	n/a	n/a
Knot	F_v	36.7	~30	49	~39	63.8	~60	(52.8)±31	464	(199)±19	(197)	–	6.3	3.0
(16'')	B_v	0.18	0.149	0.24	0.194	0.32	0.30	(0.30)	2.3	(0.99)	(0.98)	–	n/a	n/a

Table 3.2 - Nebular aperture photometry for NGC 7023. F_v is given in mJy, B_v is given in mJy/square arcsec. The aperture used for the photometry is given parenthetically (beam diameter in arcseconds) beneath the nebular position in the first column. KS represents Kris Sellgren's photometry and has been interpolated from published graphical spectrum. The position of the knot is 14W23N. The positions indicated with a dash indicate incomplete or no data available. The column labelled l:c(3.29) is the line to continuum ratio (see equation (3.2)) for the 3.29 μ m emission feature. The final column labelled l:c(3.29)/l:c(3.37) provides a measure of the magnitude of the 3.4 μ m emission feature; see the text in section 3.3.2 for a discussion of the relationship between the 3.367 μ m observation and the 3.4 μ m feature. Unless otherwise indicated, the photometric errors are <10%.

		J	H	K	CVF			L''	l:c(3.29)	$\frac{l:c(3.29)}{l:c(3.37)}$
		3.158	3.291	3.367						
W filament	F_v	20.2	24.9	31.5	117±24	1,049	540	185	32.3	2.0
	B_v	0.13	0.17	0.21	0.80	6.9	3.6	1.2	n/a	n/a
E filament	F_v	11.4	14.7	17.9	97±16	572	282	104	31.0	2.1
	B_v	0.12	0.15	0.18	1.0	5.9	2.9	1.1	n/a	n/a
Big Frame	F_v	989	1062	1130	–	14,481	–	–	11.8	–
	B_v	0.13	0.14	0.15	–	1.9	–	–	n/a	n/a
Small Frame	F_v	306	363	395	755±490	10,044	3,995	1414	24.4	2.7
	B_v	0.08	0.09	0.10	0.20	2.6	1.0	0.37	n/a	n/a

Table 3.3 - Filamentary and total frame photometry for NGC 7023. F_v is given in mJy, B_v is given in mJy/square arcsec. The aperture used for the photometry is given parenthetically (beam area in square arcseconds) beneath the nebular position in the first column. The positions indicated with a dash in the “Big Frame” indicate that our observations are incomplete here. The column labelled l:c(3.29) is the line to continuum ratio (see equation (3.2)) for the 3.29 μ m emission feature. The final column labelled l:c(3.29)/l:c(3.37) provides a measure of the magnitude of the 3.4 μ m emission feature; see the text in section 3.3.2 for a discussion of the relationship between the 3.367 μ m observation and the 3.4 μ m feature. Unless otherwise indicated, the photometric errors are <10%.

3.2 Structure in NGC 7023

The most striking aspect in the images presented here are the filamentary emission patterns seen to the immediate northern and western vicinity of HD200775. The morphology of the nebular emission appears to be chaotic, suggesting a turbulent environment. Both filaments show considerable structure. Morphologically, we observe that the western filament takes the form of an arc whose center of curvature appears close to the position of star D. Similarly, the arc of the eastern filament appears centered on star H, with a faint extension to star J. These observations suggest a physical association; perhaps stars D and H have winds interacting with the ambient medium to create the filaments, or perhaps they have expelled some portion of their outer envelopes. These scenarios do not address the source of the filamentary emission *excitation*, they are morphologically suggestive in nature only. We also note that the filaments and the knot are aligned in the NE-SW direction of the bipolar outflow observed in NGC 7023 (Watt *et al.*, 1986), Figure 3.8 illustrates the direction of the bipolar outflow overlaid on the K' frame. Whether the filamentary structure is related to the bipolar outflow is unknown, but the alignment evident in the grayscale is indicative of a connection between the two. Additionally, there is a roughly semi-circular knot of high surface brightness concentric with the western filament (whose outer arc is roughly centered on star D), about half-way between the filament and HD200775. We call this the knot; it appears approximately 20'' wide in the E/W direction and 10'' high in the N/S direction, roughly centered at the position 14W23N (relative to HD200775).

The spatial width of the filaments acts as a tracer of the densities of the emitting components. The filamentary morphology and environs are somewhat confused and disordered, so a suitable, undisturbed region was selected to investigate the width of the filaments as a function of wavelength. We chose a region approximately 6W17S of star D for the cut across the western filament (see Figure 3.8 for exact location), and directly north of star H for the eastern filament. Both cuts were 2.5'' wide in order to smooth out the variations due to noise. The western filament appears wider at shorter wavelengths, while the eastern filament is wider at longer wavelengths. This may be seen in

Figure 3.10 and Figure 3.11, where we have plotted the average signal in a 2.5'' wide cut across the filaments at J, H, K', and 3.29 μ m. To facilitate visual comparison of the spectra in a single plot, the cuts have been background subtracted and normalized. The measured width of the filament in the J band in Figure 3.10 is almost 1'' wider than at the longer wavelengths in the western filament and 0.5'' narrower in the eastern. Table 3.4 lists the measured widths determined in Figure 3.10. The spatial resolution (PSF, see Table 2.2) at a given wavelength will also affect this measurement, so we have removed this factor by assuming that the resolution adds in quadrature with the actual width of the filament and inferring the real width of the filaments by subtracting, in quadrature, the spatial resolution from the measured value for the filamentary width. The resolution is determined by both the angular extent of the seeing disk as well as the instrumental resolution of our camera and the telescope; both are folded into the PSF measured in Table 2.2. For the western filament, Table 3.4 shows that the width of the filament monotonically decreases with increasing wavelength, which we take as evidence that the species responsible for the short wavelength continuum emission (which includes the grains responsible for scattering the light from HD200775) is more diffuse than the longer wavelength emitting grains. The smallest measured FWHM occurs in the emission line at 3.29 μ m, which indicates that the emitting species is less spatially extended than the dust responsible for the shorter wavelength light. We hypothesize that the ambient grains are shielding the 3.29 μ m feature carriers from the exciting radiation field. In Figure 3.12 and Figure 3.13, cuts through different portions of the western filament, we observe similar behavior regarding the wavelength dependence of the filamentary FWHM. As shown in Figure 3.11 and Table 3.4, the eastern filament is more diffuse than the western in the broadband frames, which indicates that the western filament is observed edge-on while the eastern filament is being seen at a larger angle relative to the line of sight, *i.e.* we observe more of the eastern filament's "surface," than we do of the western filament.

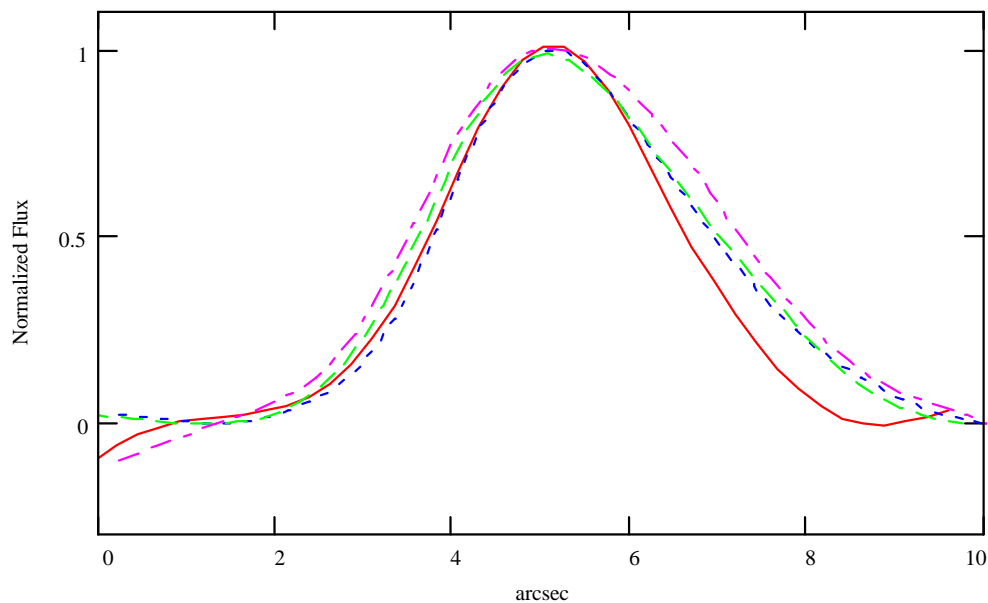


Figure 3.10 - Cut across the western filament. The data have been background subtracted, and the cut is $2.5''$ wide approximately $6W17S$ of star D. All the data have been normalized for ease of comparison. The curves are as follows: solid curve - $3.29\mu\text{m}$ emission, dotted line - K' band, dashed line - H band, dot-dashed line - J band.

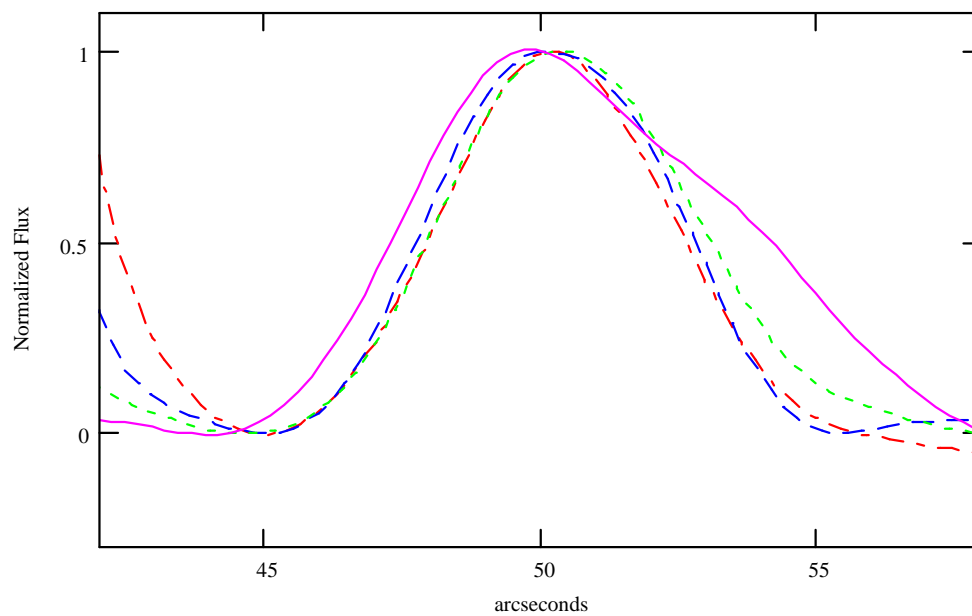


Figure 3.11 - Cut across the eastern filament. The data have been background subtracted, and the cut is $2.5''$ wide directly north of star H. All the data have been normalized for ease of comparison. The curves are as follows: solid curve - $3.29\mu\text{m}$ emission, dotted line - K' band, dashed line - H band, dot-dashed line - J band.

Wavelength	Western Filament FWHM (arcsec)		Eastern Filament FWHM (arcsec)	
	Measured	Inferred	Measured	Inferred
J	3.7	3.4	4.8	4.6
H	3.4	3.1	5.2	5.0
K	3.2	3.0	5.2	5.1
3.291μm	2.9	2.7	7.0	6.9

Table 3.4 - The measured and inferred width for a section of each of the filaments. The locations measured were approximately 6" west and 17" south of star D for the western filament and directly north of HD200775 for the eastern. The methodology employed to determine the inferred width is discussed in the text. The data presented in this table was determined from Figure 3.10 and Figure 3.11.

The spatial profile of a portion of the western filament as well as the knot is shown in Figure 3.12, where we have plotted an 8 pixel wide ($\sim 2''$, perpendicular to the direction of the cut) cut from HD200775 to star D in the J, H and K' bands as well as the 3.291 μ m CVF frame. HD200775 is at the left side of the plot and star D is on the right side (the stellar peaks are off scale in order to display the nebular emission). The peak at 24" (from HD200775) is the knot and the peak at approximately 46" is the western filament. There is a slow rise in the 3.29 μ m emission strength starting approximately 10" from HD200775 and continuing to the western filament. There is a decline on the northern side of the filament, though not evident in this plot because of star D's proximity. Figure 3.13 plots an 8 pixel wide ($\sim 2''$, perpendicular to the direction of the cut) N/S cut directly through star D starting at a declination approximately 16" south of HD200775. The first increase in signal evident at approximately 10" is a relatively bright emission region that appears to be an extension of the knot (see broadband grayscale frames, Figure 3.1–Figure 3.3), and the second hump at approximately 50" is the southern end of the western filament. Both figures show several common effects which we shall address in the following discussion: 1) the nebular emission is very red, 2) the profile across both filaments is asymmetric; there is a steeper gradient in surface brightness on the side of HD200775, 3)

the profile of the knot is similar to the filaments in that there is a steep brightness gradient on the HD200775 side, while there is a more gradual change in brightness on the side of star D, 4) the knot is very weak (if present at all) at $3.291\mu\text{m}$, and 5) the eastern filament is more diffuse (spatially extended) than is the western filament. These observed properties are illustrated by these different cuts through the nebula.

Figure 3.12 clearly shows that the $3.291\mu\text{m}$ emission feature is not correlated with the knot, while the knot is unmistakably present in the broadband cuts. The lack of evidence for feature emission in the knot does not mean that there cannot be a diffuse background of $3.29\mu\text{m}$ emission, rather, it shows that there is no enhanced feature emission associated with the knot, from which we conclude that the dust composition in the knot is distinct from that in the filaments. We hypothesize that the knot's surface brightness is largely due to scattered light from classical dust grains, and that the dearth of small grains in this region may be the result of destruction by the radiation field from HD200775, leaving only larger grains. The small grain destruction may have resulted from the proximity of HD200775 in combination with a lack of shielding, thereby exposing the small grains in the knot to a relatively brighter and harder radiation field. Using the Hubble law (equation (3.6)), with a unity grain albedo, we find that the *maximum* reflected surface brightness from HD200775 in the K band is almost a factor of 3 larger than that observed (see Table 3.6), indicating that the K band surface brightness in the knot can probably be explained by reflected light. In addition, the broadband colors of the knot and HD200775 are very similar; $[J-H]=0.67$ and 0.78 for HD200775 and the knot, respectively, and $[H-K]=0.80$ for both HD200775 and the knot. We suggest that the J,H, and K' band fluxes from the knot are due to reflected light. Evidently, either the small grains in this region have been destroyed by the intense UV flux from HD200775 or they have been ejected by a stellar wind.

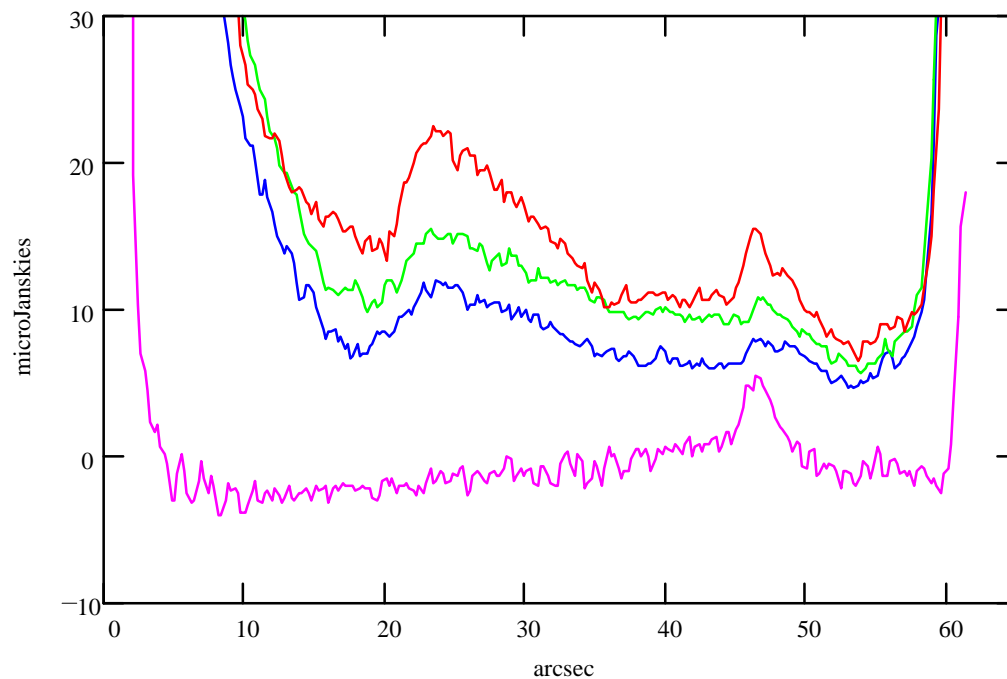


Figure 3.12 - Cuts from HD200775 to star D. The plots, from top to bottom, are K', H, J, and 3.291 μ m. The width of each of the plotted vectors is 8 pixels ($\sim 2''$, perpendicular to the direction of the cut), and the average signal (across the width of the cut) is plotted here. The vertical axis is in μ Jy and the horizontal axis is in arcsec (from HD200775). The 3.291 μ m frame has been transformed by $(F/80)-3$, where F is the signal flux in Janskies, in order to display it on the same plot as the broadband data.

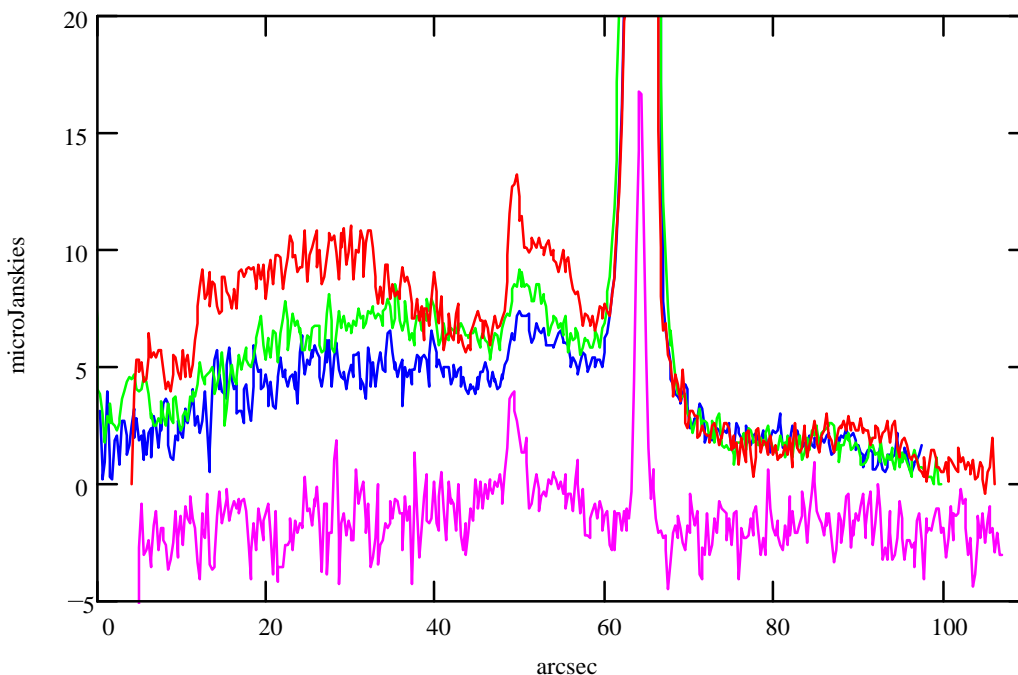


Figure 3.13 - N/S cuts through star D. The plots, from top to bottom, are K', H, J, and 3.291 μ m. The width of each of the plotted vectors is 8 pixels ($\sim 2''$, perpendicular to the direction of the cut), and the average signal (across the width of the cut) is plotted here. The vertical axis is in μ Jy and the horizontal axis is in arcsec. The 3.291 μ m frame has been transformed by $(F/80)^{-3}$, where F is the signal flux, in order to display it on the same plot as the broadband data.

Figure 3.14 shows an 8 pixel wide ($\sim 2''$, perpendicular to the direction of the cut) cut from star H to star D; the eastern filament is traversed from approximately 11'' to 17'' from star H and the western filament from roughly 22'' to 27''. Note that we have not detected star H in the 3.291 μ m CVF measurement (see Table 3.1 for the upper limit to star H's photometry at 3.291 μ m). From this plot we can see that the profile of the western filament is very similar to that observed in Figure 3.12 and Figure 3.13. The eastern filament is considerably more diffuse than its western counterpart. Its spatial profiles do not show the same abrupt rise and gradual decline in signal (in either direction) perpendicular to the filamentary arc. This can be explained by projection effects; the eastern filament may appear more diffuse because we are not seeing it edge-on, or because it has more spatial extent parallel to the line-of-sight compared with the western filament. The 3.29 μ m cut is spatially narrower

than the J,H and K' emission for the western filament, which we take as evidence for self-shielding of the filamentary interior.

An asymmetric brightness profile is observed in both of the filaments as well as the knot in moving from HD200775 towards star D (see Figure 3.12–Figure 3.14). The rapid rise on the side of HD200775 and a subsequent slow decline is suggestive of an expanding medium interacting with the ambient molecular cloud. We observe this phenomenon edge-on, so that the filaments appear brighter because of projection effects. There are three possible explanations for this observation in the context of NGC 7023; a weak shock front resulting from a stellar wind's interaction with the ambient cloud, excess pressure from an HII region producing a shock that propagates into the surrounding molecular cloud or an expanding envelope from one or more of the stars. Given that some form of mechanical interaction is driving the abrupt density variation, the UV flux from HD200775 can then produce the emission profiles observed, as is typical for photodissociation regions (PDR, Tauber *et al.*, 1994). An alternate possibility for the excitation is that the energy from the shock itself is exciting the emission. The source of the mechanical energy as evidenced by the western filament is most likely star D because it is well centered on the arc of the filament. Star D is a protostellar object (discussed in section 3.6), and as such, is likely to have outflows and winds associated with it. It is also possible that the well known stellar wind associated with HD200775 is interacting with the surrounding cloud to produce the observed morphology, or that the observed structures result from HD200775's bipolar outflow driving into the ambient medium. The direction of the bipolar outflow is aligned along the filamentary axis which suggests, for this scenario, that the mechanism confining the outflow also has a hand in the structure of the filaments. Imaging observations of molecular hydrogen emission should be undertaken in this region in order to assess the possible presence of shock excitation associated with the filamentary emission.

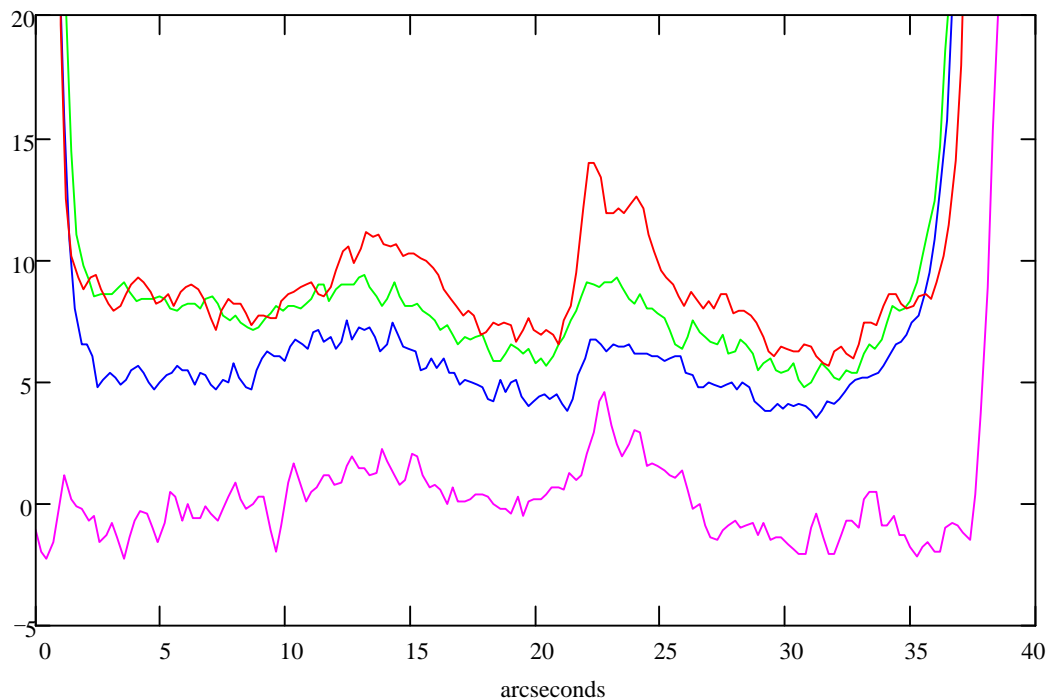


Figure 3.14 - Cuts from star H to star D. The plots, from top to bottom, are K', H, J, and $3.291\mu\text{m}$. The width of each of the plotted vectors is 8 pixels ($\sim 2''$, perpendicular to the direction of the cut), and the average signal (across the width of the cut) is plotted here. The vertical axis is in μJy and the horizontal axis is in arcsec. The $3.291\mu\text{m}$ frame has been transformed by $(F/80)-3$, where F is the signal flux, in order to display it on the same plot as the broadband data. Star H is on the left side of the graph at $0''$ and star D is on the right at $39''$. Note that star H was not detected in the $3.291\mu\text{m}$ measurement.

3.3 The Nebular Spectrum

3.3.1 The 3.29 μ m Emission Feature

We present the observed NIR spectrum of various components of NGC 7023 in Figure 3.15; these data have been scaled in order to facilitate their comparison in a single plot. The most notable characteristic present in these spectra is the presence of a very bright emission feature at 3.29 μ m. The morphology of the filaments in the 3.29 μ m frame closely follows that of the K' broadband frame; where the K' frame is bright, the emission feature is similarly bright. This suggests that the carriers of the 3.29 μ m emission feature are coextant with the species responsible for the broadband, continuum emission in the K' band (primarily transiently heated small grains), and further, that the density variations follow one another within the filaments. The primary differences between the broadband and CVF frames are the presence of the knot in the former, the absence of stars H and A in the CVF frames and that the western filament is spatially narrower in the CVF frames, while the eastern filament is wider.

The 3.29 μ m dust emission feature is a member of the “unidentified IR band” (UIB) feature family, whose other primary members are found at 6.2, 7.7, 8.6, and 11.3 μ m (Gillett *et al.*, 1973, Russell *et al.*, 1977a). These emission features have been observed in Galactic objects (planetary nebulae, HII regions and reflection nebulae) containing sources of UV photons, and in HII region galaxies (starbursts) and LINERs, but rarely in QSOs or AGNs. The emission feature at 3.29 μ m is associated with the NIR continuum, and is theorized to be a link in the chain of dust chemistry responsible for the NIR emission in many galactic objects and galaxies. [It may be that the species responsible for the 3.29 \$\mu\$ m feature are the same as those that generate the continuum, or that they are related to one another through dust evolution via the attendant chemistry.](#) The major sources of interstellar dust are evolved, mass losing stars: red giants (includes giants, supergiants and AGB stars), planetary nebulae and their immediate precursors (hot post-AGB stars), as well as novae and

supernovae. Mass loss from hot stars is not known to produce dust. Carbon rich planetary nebulae and post AGB stars are both known to eject the $3.29\mu\text{m}$ feature carrier into the interstellar medium (Whittet, 1992). It is likely that carbon rich red giants also produce the feature carriers, but since these stars are cool (and hence, have a dearth of UV photons) their spectra do not exhibit the UIB features.

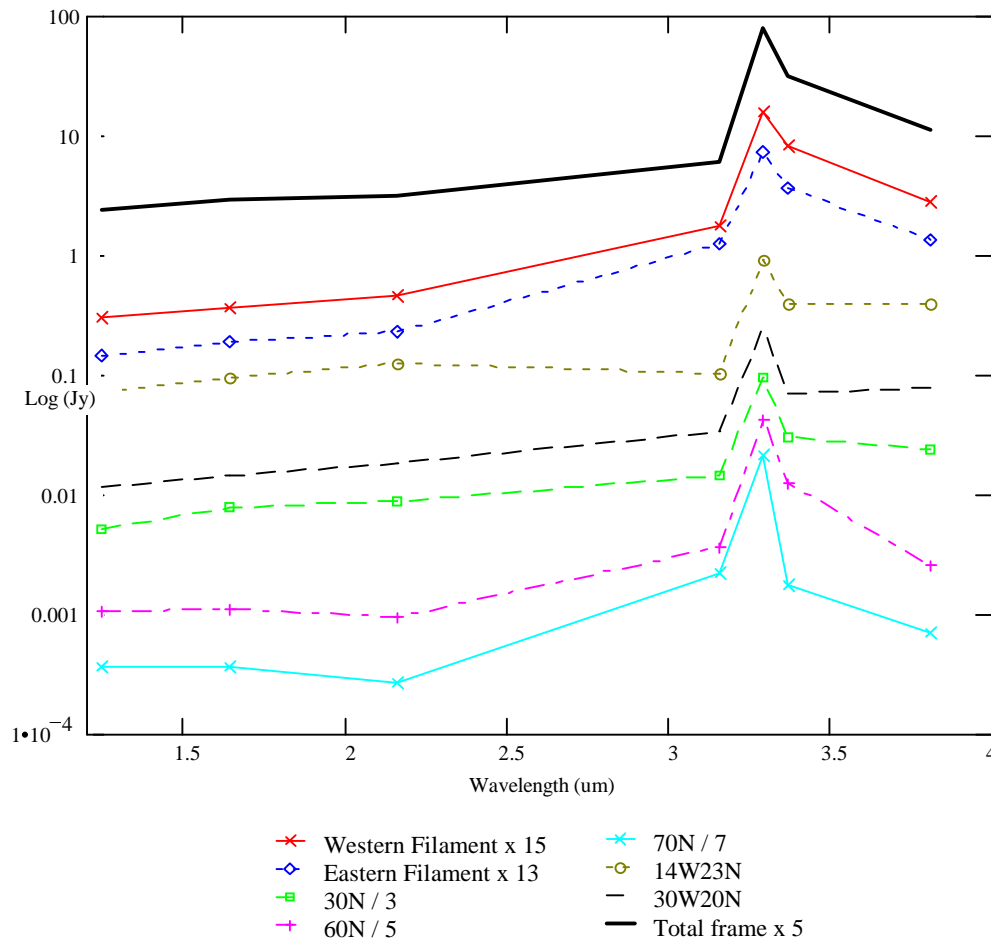


Figure 3.15 - Observed spectra (log Janskies) for nebular apertures indicated. Note that the spectra have been scaled per the legend in order to facilitate comparison.

While the $3.29\mu\text{m}$ emission feature is widely attributed to the $\nu=1\rightarrow 0$ fundamental C–H stretch mode, the feature carrier’s molecular identity has not yet been unambiguously determined. The proposed carriers fall into two main groups: free-flying molecules and solid state dust grains. The leading contenders are all aromatic species: polycyclic aromatic hydrocarbon (PAH) molecules and PAH clusters (Léger & Puget, 1984, Allamandola *et al.*, 1985), hydrogenated amorphous hydrocarbons

(HAC, Duley & Williams, 1981) and quenched carbonaceous compounds (QCC, Sakata *et al.*, 1984). The fundamental building block for each of these materials is the aromatic carbon ring, consisting exclusively of carbon and hydrogen. The dust grain hypotheses (HAC, QCC) invoke aromatic species loosely bonded to a larger silicate or graphitic core. The various models differ in the amount of hydrogenation, ionization, side-group species, and in the degree to which the aromatic rings are coupled. Allamandola, *et al.* (1987) argue that there is likely a spectrum of free flying molecules coexistent with larger HAC grains in most astronomical objects; the free flying molecules producing the narrow emission features while the carbon particles contribute the broad, underlying component.

The electronic excitation resulting from the absorption of a UV photon is redistributed in the vibrational modes of the carrier on timescales short compared to the IR fluorescence time (~ 0.1 s). There are effectively two de-excitation modes competing with IR emission: 1) bond dissociation and/or ionization, and 2) coupling to the phonon bath where the energy is re-radiated in the FIR due to its low equilibrium temperature. The first mechanism is effective for both the solid state as well as the free flying molecule models while the second is only effective for the solid state model (Léger & d'Hendecourt, 1987).

Tokunaga *et al.* (1991) have shown convincingly that there are two types of $3.29\mu\text{m}$ emission feature profiles: type 1 is associated with PNs and HII regions and type 2 with pre-main-sequence and evolved stars (stellar sources). There are examples of each type that both display and do not display the $3.4\mu\text{m}$ emission feature. The two types differ in their spectral widths and the location of the peak wavelength. The distinction between the two types is not clearly defined since there is at least one object (HD 44179) that possesses a type 2 spectrum near its exciting star and a type 1 spectrum farther away.

The ratio of the flux in an emission line to that in the associated continuum (the line to continuum ratio, or l:c) is an important observational parameter that must be predicted by any theory purporting to explain the UIBs. In the optically thin limit, it is directly proportional to the ratio of the number of emitting entities in the line to that in the continuum. Sellgren has observed that this ratio

ranges over 3–10 in a wide variety of astronomical objects (in large aperture observations), suggesting that the line and the continuum emitting species are somehow related. The magnitude of this ratio also provides a “reliability” metric for the emission line observation under discussion; i.e. if the value of the line to continuum ratio is small, then the presence of the emission feature will be less reliable because (for the same signal to noise ratio), the error in the $3.29\mu\text{m}$ observation may be such as to make it indistinguishable from the continuum, thereby requiring higher signal to noise in order to improve the reliability. Given 10% error in our photometry, high confidence in the presence of line emission will require that the line to continuum ratio be >0.4 (assuming equal errors in both the line and the continuum observations). This derives from the algebraic inequality that requires the continuum signal plus twice the error in the continuum measurement be less than the signal in the line minus twice the error in the line: $F_{\text{cont}} + 2\sigma_{\text{cont}} < F_{\text{line}} - 2\sigma_{\text{line}}$, i.e. the 2σ error bars cannot overlap. The observed line to continuum ratio is defined by:

$$\begin{aligned} l:c(3.29\mu\text{m}) &\equiv \frac{F(3.29\mu\text{m}) - F(K')}{F(K')} \\ &= \frac{F(3.29\mu\text{m})}{F(K')} - 1 \end{aligned} \tag{3.2}$$

where $F(\lambda)$ is the flux at the wavelength λ . We have three observations that can potentially be used to provide a measure of the continuum level at $3.29\mu\text{m}$: the K' and L' bands, and the $3.158\mu\text{m}$ CVF frames. In every location plotted in Figure 3.15, the K' band flux is the minimum of the three wavelengths we consider for the continuum, and appears to be a reasonable approximation for the continuum. The L' band observation is not necessarily a good choice for the continuum because it is often elevated in the presence of the $3.29\mu\text{m}$ and $3.4\mu\text{m}$ emission features, and may be associated with a broad emission pedestal seen in the presence of UIBs. The $3.158\mu\text{m}$ frame would be an excellent choice for the continuum, but for the poor signal to noise ratio as well as the incomplete spatial coverage in our data set. Spectra obtained by Sellgren (1984) corroborate this choice for the continuum, the K' band underestimates the continuum (as determined by interpolating between

Sellgren's measured K and M band measurements) by approximately 20%, which will lead to ~20% larger values for $l:c(3.29\mu\text{m})$ than would be the case for a "true" measurement of the continuum point at $3.29\mu\text{m}$ for large values of the line to continuum ratio.

The magnitude of the $l:c(3.29\mu\text{m})$ in the filaments is extremely high, the peak flux in the line being close to two orders of magnitude larger than that of the continuum. This ratio is plotted in Figure 3.16 along the vector connecting HD200775 to star D. The vector is $3''$ wide (perpendicular to the direction of the cut), in order to increase the signal to noise ratio. The maximum value of 38 is attained at the western filament, $46''$ from HD200775. We also note that there is no $3.3\mu\text{m}$ feature emission within $13''$ of HD200775, consistent with Sellgren's findings (Sellgren, 1984). The strength of the line to continuum ratio is seen to steadily rise starting approximately $13''$ from HD200775 and continue rising until the peak is reached within the western filament, whereupon it declines until it reaches a secondary maximum $6''$ from star D. This behavior is similar to that of the strength of the feature emission itself. Sellgren *et al.* (1996) have measured the line to continuum ratio in over 20 different reflection nebulae, and have found it to be between 6 and 9 for aperture sizes of $6.2''$ – $10.5''$, independent of exciting star's temperature over the range 6800 – 33000K (note that these are large aperture measurements, compared with our $1''$ psf). Our measurement of this ratio yields 12.3 for the location 30W20N (see Table 3.2), somewhat higher than that obtained by Sellgren. The reason for the discrepancy lies primarily in our choice for the continuum (the K' band); adopting the $3.158\mu\text{m}$ observation for the continuum results in $l:c(3.29)=6.5$, in agreement with that measured by Sellgren *et al.* (1996).

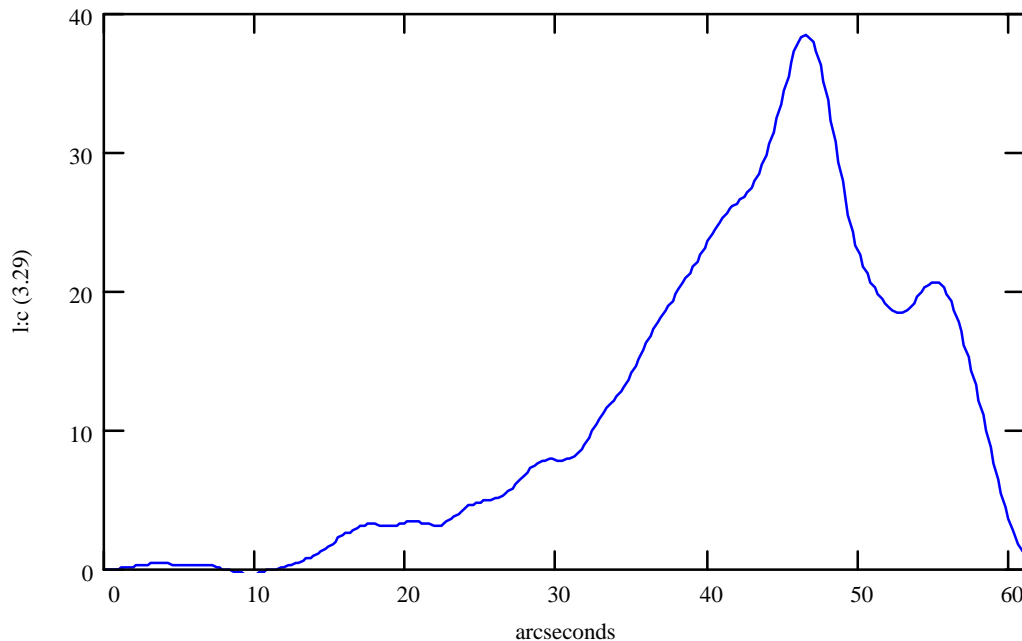


Figure 3.16 - The line to continuum ratio along the vector joining HD200775 to star D. The methodology employed for calculating this parameter ($l:c(3.29\mu\text{m})$) is discussed in the text. The parameters of the cut are identical to those of Figure 3.12. The plot has been smoothed by applying a $3''$ gaussian kernel (in the direction of the cut) to return weighted averages of the raw data.

Outside of the filamentary region in NGC 7023, our observations agree with those of Sellgren in large apertures: it is the filaments and the immediate neighborhood of star D that show an anomalously high line to continuum ratio. This is where the utility of imaging arrays is superior to aperture measurements, enabling variations across complicated structures such as NGC 7023 to be easily observed and analyzed. We propose that this ratio is larger in the filaments due to self-shielding, which is to say that the interior of the filaments is shielded from UV photons that destroy and/or alter the carriers of both the feature and the continuum emission. Figure 3.17 plots a similar vector from the location of star H (not detected in the $3.291\mu\text{m}$ observation) to star D: the vector is $3''$ wide (perpendicular to the direction of the cut) in order to increase the signal to noise ratio. The grayscale (Figure 3.5) shows that the eastern filament along this cut is not well defined, but rather appears to merge into the background flux, which has a high $l:c(3.29\mu\text{m})$ value as well. As we move from the location of star H along this cut, $l:c(3.29\mu\text{m})$ increases steadily from 10 to 34 at in the eastern filament

(15" from star H), then increases to its peak value of 38 in the western filament. The $l:c(3.29\mu\text{m})$ value measured in the western filament in this cut is consistent with that measured in other cuts at different locations along its length, indicating that the filament is uniform and suggesting that it is a coherent structure. Beyond the western filament, $l:c(3.29\mu\text{m})$ drops by a factor of two and rises again 5" from star D before falling to less than one. This clearly shows that there is $3.29\mu\text{m}$ feature emission associated with star D, whose nature will be discussed in sections 3.3.2 and 3.6.

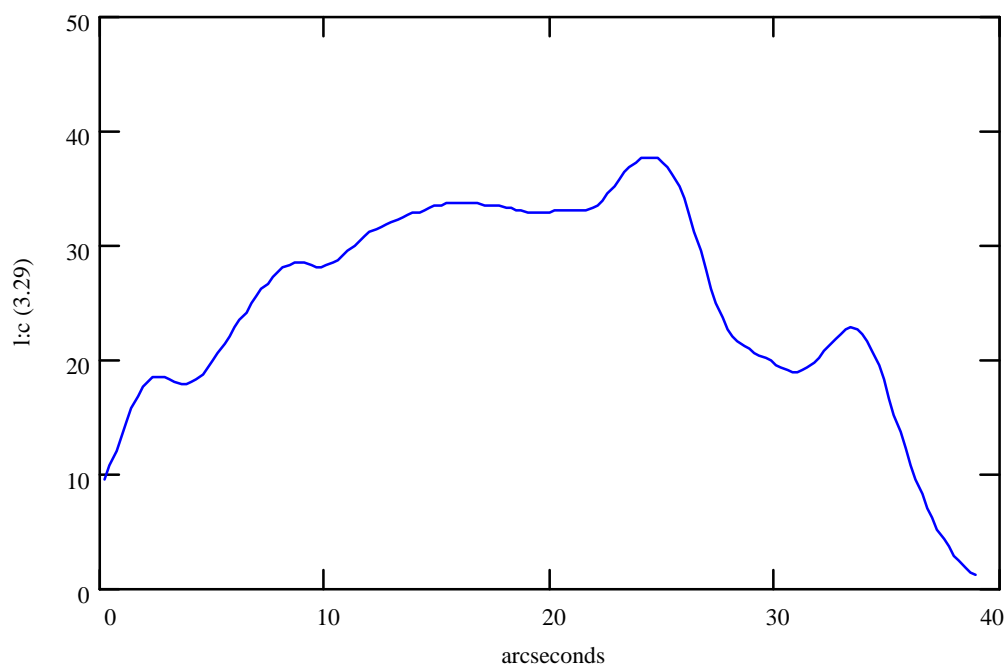


Figure 3.17 - The line to continuum ratio along the vector joining star H to star D. The methodology employed for calculating this parameter ($l:c(3.29\mu\text{m})$) is discussed in the text. The parameters of the cut are identical to those of Figure 3.12. The plot has been smoothed by applying a 3" wide gaussian kernel (in the direction of the cut) to return weighted averages of the raw data. Star H is at the zero point of the x axis and star D is 38" from star H. Note that star H was not detected in the $3.29\mu\text{m}$ frame.

We argue that the continuum is dominated by small particle emission, and that the variation in $l:c(3.29\mu\text{m})$ is real (it is much larger than our observational uncertainties). There are several possibilities that can account for this variation. The K' band flux may be a combination of mechanisms (e.g. small particle emission, dust scattering, atomic and/or molecular emission lines), which would

raise the continuum level, thereby decreasing the value measured for $l:c(3.29\mu\text{m})$. It may also be that there is no causal relationship between the emitting species for the $3.3\mu\text{m}$ feature and its underlying continuum, either because they are different entities, or because their chemistry varies with environment. The large values measured for the line to continuum ratio as well as its variability over the face of the nebula (see Table 3.2 and Table 3.3) indicate that the species responsible for the continuum is distinct from that producing the line emission. Assuming an intrinsic line to continuum ratio of 40 (the value measured in the filaments), along with Sellgren's measured range $3 < l:c(3.29\mu\text{m}) < 9$, requires that 75%–90% of the observed continuum be attributable to mechanisms other than that hypothesized to be associated with the $3.29\mu\text{m}$ feature. We conclude that the continuum emission at $3.29\mu\text{m}$ is not correlated with the $3.29\mu\text{m}$ feature emission feature.

It is possible that the variations in $l:c(3.29\mu\text{m})$ are attributable to variations in the measured continuum in the K' band due to line emission. Table 3.2 shows that all our measurements which are related to the continuum (J and H bands, as well as both the $3.158\mu\text{m}$ and the $3.81\mu\text{m}$ observations) track the K' data, which indicates that the latter does not contain any sizable fraction of its flux resulting from line emission. Furthermore, the only sources of near infrared continuum emission are scattered light and non-equilibrium thermal emission by small grains, and since the polarization measurements (Sellgren *et al.*, 1992) show that scattered light is not dominant in the K' band, we conclude that the K' band emission is dominated by small particle, continuum emission over most of this region (except close to HD200775). Hydrogen, the most ubiquitous element in the Universe, can contribute to our K' band observations through both atomic and molecular emission. We have shown that the size of a Strömgren sphere surrounding HD200775 will be extremely small ($\sim 1''$), thereby ruling out any atomic contribution to the K' data. Molecular hydrogen, via $v=1\rightarrow 0$ S(1) line emission at $2.122\mu\text{m}$ does not contribute any appreciable flux to our K' band measurements; Sellgren (1986) did not detect any Q branch⁷ fluorescent H₂ emission in NGC 7023 at the 30W20N position in a $12''$ beam,

⁷ The sum of the 1-0 Q(1) at $2.413\mu\text{m}$, 1-0 Q(2) at $2.424\mu\text{m}$ and 1-0 Q(3) at $2.407\mu\text{m}$ emission lines.

and was able to place an upper limit on any H₂ contribution to the K' band signal as <10%. From these arguments, we conclude that the K' observations are dominated by continuum emission far from HD200775, and hence, that the variations observed in I:c(3.29μm) are real variations.

The fact that NGC 7023 is a well known PDR coupled with theoretical calculations (Black and Dalgarno, 1976) suggest that there should be H₂ emission from NGC 7023 given the UV flux from HD200775, so it is likely that internal nebular (UV) extinction is responsible for its non-detection by Sellgren, consistent with our conclusions here. More importantly, Sellgren only looked for molecular hydrogen at the single location (30W20N) in a large aperture beam. The existence of H₂ is most likely to be detected in the filaments and their immediate surroundings, since it has been observationally associated with filamentary emission in the Orion Bar (Sellgren, 1981, Sellgren, Tokunaga and Nakada, 1990) where the 3.29μm feature is observed to be spatially sandwiched between the ionized atomic hydrogen and the molecular emission components. The bar in Orion is believed to be a shock front associated with an expanding ionization front excited by the stars of the Trapezium. According to well established theory (Spitzer, 1978, Osterbrock, 1989) a shock front is expected to expand ahead of the ionization front. In the case of NGC 7023, the ionization front is very close to HD200775 (see equation (3.1)). A similar geometry is also observed in NGC 2023 (Gatley *et al.*, 1987), where the 3.29μm feature is also observed to lie just inside the H₂ emission. Lemaire *et al.* (1996) have observed filamentary H₂ emission at 2.121μm (v=1-0 S(1)) and 2.033μm (v=1-0 S(1)) in NGC7023. The spatial structure of the H₂ emission correlates strongly with the filaments observed here, and supports our contention that the filaments are self-shielding.

3.3.2 The 3.4μm Emission Feature

Just longward of the 3.29μm feature there is a pair of weaker emission bands at 3.4 and 3.51μm often observed in objects exhibiting the UIBs. These have been variously attributed to: (1)

combination and overtone bands involving C–C stretching and/or C–H bending modes (Allamandola *et al.*, 1989), (2) hot band emission from the aromatic C–H stretch (excited states of PAH's: $\nu=2\rightarrow 1$ transition, Allamandola *et al.*, 1985), and (3) aliphatic $-\text{CH}_2$ or $-\text{CH}_3$ sidegroups attached to the periphery of a PAH molecule (Duley and Williams, 1981, Jourdain de Muizon *et al.*, 1990). Recent laboratory spectra indicate that the aliphatic sidegroups are the dominant contributor to the $3.4\mu\text{m}$ feature (Joblin 1992, Joblin *et al.*, 1995). Observations (this work, Joblin *et al.*, 1996) show high levels of excitation ($F_{3.4}/F_{3.29}$ ratios) in reflection nebulae that are inconsistent with the $3.4\mu\text{m}$ feature resulting from hot band transitions (see below). For the case of hot band emission, Schutte *et al.* (1993) have performed theoretical calculations indicating that the range of the integrated line ratio, $F_{3.4}/F_{3.29}$, varies from 0.2 for benzene (C_6H_6) to 0.11 for coronene ($\text{C}_{24}\text{H}_{12}$) before photodissociation dominates IR emission. Higher levels of excitation will lead to rapid dissociation of the C-H bond before infrared emission can occur. We have estimated the strength of this ratio in NGC 7023 and found that it falls well outside this range (see Table 3.5). Both features have been observationally correlated with non-ionizing UV radiation, and conversely, have not been observed in environments with hard radiation fields.

The $3.4\mu\text{m}$ emission feature sits atop a broad emission pedestal extending approximately from $3.2\text{--}3.6\mu\text{m}$ (Geballe *et al.*, 1985, Joblin *et al.*, 1996). This broad continuum and associated substructure are well correlated with the presence of both the $3.4\mu\text{m}$ and $3.29\mu\text{m}$ emission features, *i.e.* all objects exhibiting this pedestal show both line features (deMuizon *et al.*, 1987). It is likely that all three of the mechanisms listed in the previous paragraph contribute to this broad pedestal. The plateau is roughly triangular with a peak spectral flux density at $3.35\mu\text{m}$ of approximately 50% of the peak spectral flux density in the $3.4\mu\text{m}$ feature, and it decreases linearly with wavelength, reaching the continuum at $3.60\mu\text{m}$. The observed widths of the 3.4 and $3.51\mu\text{m}$ features are approximately 1.2–1.5 times smaller than that of the $3.29\mu\text{m}$ feature; typical observed linewidths are $0.042\mu\text{m}$ and $0.028\mu\text{m}$ for the $3.29\mu\text{m}$ and $3.4\mu\text{m}$ emission features, respectively (Tokunaga *et al.*, 1991, Joblin *et al.*, 1996).

The ratio of the 3.4 μm to the 3.29 μm emission features has been the subject of some study recently (Geballe *et al.*, 1989, Bregman *et al.*, 1993, Joblin *et al.*, 1996). Geballe *et al.* (1989) have observed that the $F_{3.4}/F_{3.29}$ ratio increases with distance from the exciting star when going from the ionized to the neutral region in Orion (across the “bar”). Similarly, Joblin *et al.* (1996) have shown that the 3.4 μm to 3.29 μm ratio is strongly anti-correlated with the intensity of the UV radiation field in the reflection nebulae NGC 2023, NGC 1333 SVS3 and the Red Rectangle as well as across the Orion Bar PDR. Changes in the $F_{3.4}/F_{3.29}$ ratio probably indicate a chemical change in the emitting species since factors which affect the intensities in each band, such as column density of the emitting species (assuming that the carriers are correlated) will cancel out of the ratio. Joblin *et al.* (1996) propose a chemical evolution model in which methylated sidegroups ($-\text{CH}_3$) on PAH molecules are the dominant contributor to the 3.4 μm emission feature. In their model, the number of $-\text{CH}_3$ groups decreases over time (as a function of the radiation field) owing to the process of uni-molecular photo-decomposition, *i.e.* the loss of a peripheral $-\text{CH}_3$ sidegroup via photon absorption. The timescale for this loss is determined by the radiation field and its temporal evolution. This model predicts $F_{3.4}/F_{3.3}$ ratios as large as 2 in regions protected from photons with photodissociative energies; such high values have been observed in proto-planetary nebulae⁸ (Geballe *et al.*, 1992, Joblin *et al.*, 1996).

The spectra of the eastern and western filaments in Figure 3.15 clearly show evidence for the 3.4 μm emission feature in NGC 7023, as do many of the aperture spectra in NGC7023 (see Figure 3.15). Our 3.367 μm long wavelength “continuum” bandpass contains a portion of the 3.4 μm feature’s wavelengths, and the observed signal at this wavelength is elevated with respect to the continuum as defined by the broadband observations. However, blindly plotting the $F_{3.367}/F_{3.29}$ ratio can be misleading because this ratio will be unity in the case where there are no emission features: one must look at the data carefully to discern the presence of the feature(s) before relying on this ratio to draw

⁸ A planetary nebula is a hot stellar core (typically $\sim 100,000\text{K}$) with a surrounding shell of emitting gas and dust. Towards the end of a low mass star’s lifetime, it ejects its outer layers, exposing the hot core, which heats and ionizes the expanding shell.

conclusions regarding the presence of the emission feature(s). Imaging arrays again prove their utility in this exercise, it is a simple matter to discern the presence of the emission feature in an image. The observations presented here are unambiguous, the only emission clearly evident in the 3.367 μm frame above any diffuse emission is from the region immediately surrounding the filaments, the filaments themselves and star D. Table 3.2 and Table 3.3 show the ratio $I:c(3.29)/I:c(3.367)$ for the nebular apertures and filaments discussed here. The observed $F_{3.367}/F_{3.29}$ ratio in NGC 7023 varies widely from 0.1 to 0.5 for the various locations presented in the tables. While the calculated quantity $I:c(3.367)$ does not represent the “true” value of the line to continuum ratio for any 3.4 μm emission feature present, the relative variations in this quantity indicate that there is a real as well as large variation in the ratio of the 3.29 μm to the 3.4 μm feature. In the following, we determine the magnitude of the 3.4 μm emission feature from our 3.367 μm observations.

In order to infer the value of the 3.4 μm emission from our 3.367 μm observation we must deconvolve our filter’s spectral characteristics to estimate of the “true” spectrum. The CVF passband is approximately gaussian in nature with a $\text{FWHM}_{3.367} = 0.059\mu\text{m}$ ($\text{FWHM} = 2.355\sigma$, where σ is the standard deviation of a gaussian distribution), which means that any 3.4 μm feature flux contained in the 3.367 μm frame will be attenuated as a result of the filter’s transmission alone. To determine the average transmission of the CVF as a function of wavelength we convolve the passband of the CVF (a gaussian) with a circular aperture function (the Lyot stop in wavelength space):

$$T_{\text{cvf}}(\lambda, \lambda_o) = \frac{2}{\pi r_{\text{ls}}^2} \int_{-r_{\text{ls}}}^{r_{\text{ls}}} \sqrt{r_{\text{ls}}^2 - x^2} e^{-\frac{1}{2}\left(\frac{\lambda - \lambda_o - x}{\sigma}\right)^2} dx \quad (3.3)$$

where r_{ls} is the radius of the Lyot stop, x is the dimension in the Lyot stop along which the CVF transmission characteristics change with wavelength, λ_o is the wavelength of the CVF in the center of the aperture, and σ is the standard deviation of the filter (provided by OCLI as a FWHM and converted to σ assuming a gaussian profile). The flux incident on the detector when the CVF bandpass is

centered on λ_o will be given by integrating the (normalized) product of equation (3.3) and the sum of the profiles approximating the emission features (3.29 μm , 3.4 μm and plateau) over all wavelengths:

$$F_{\text{obs}}(\lambda_o) = \Phi_{\text{obs}}(\lambda_o) \cdot \text{EW}(\lambda_o) \\ = \int_0^{\infty} \left(A e^{-\frac{(\lambda - \lambda_{3291})^2}{0.361 \cdot \text{FWHM}_{3291}^2}} + B e^{-\frac{(\lambda - \lambda_{3400})^2}{0.361 \cdot \text{FWHM}_{3400}^2}} + C \cdot \text{plateau}(\lambda) + \Phi_{\text{cont}} \right) \frac{T_{\text{cvf}}(\lambda, \lambda_o)}{T_{\text{cvf}}(\lambda_o, \lambda_o)} d\lambda \quad (3.4)$$

where $\Phi_{\text{obs}}(\lambda_o)$ includes both the observed feature as well as the continuum emission, EW is the equivalent width of the CVF centered at λ_o , A and B are the peak spectral flux densities of the 3.291 μm and 3.4 μm emission features respectively, λ_{line} and $\text{FWHM}_{\text{line}}$ are the two features' central wavelengths and full widths at half maxima, respectively, and Φ_{cont} is the continuum flux density. The plateau emission, $\text{plateau}(\lambda)$ is taken to be a triangular function with a peak spectral flux density $C = 0.5B$ at 3.35 μm that goes linearly to zero flux at 3.60 μm and 3.2 μm . The 3.29 μm emission feature is assumed to have a $\text{FWHM} = 0.042\mu\text{m}$ (Tokunaga *et al.*, 1991), while that of the 3.4 μm feature is taken as $\text{FWHM} = 0.028\mu\text{m}$ (Joblin *et al.*, 1996). The continuum flux density was modeled as a graybody forced to go through the K' broadband observation data point. The temperature of the graybody was chosen to yield agreement with the fraction of scattered light in the H band at 30W20N as measured by Sellgren *et al.* (1992, where we have used the middle of the $S_{\text{ref}}/S_{\text{total}}$ range found in Table 3.7). Equation (3.4) represents 2 equations (one for each of the CVF positions 3.291 μm and 3.367 μm) in two unknowns (A and B). Solving for A and B yields the inferred spectrum (the bracketed portion of equation (3.4)). This same formalism was used to show that the 3.29 μm feature contributes 4–5% to the 3.367 μm observation for each filament. The final, inferred spectra thus consist of four components: a graybody continuum, the pedestal and the two emission features.

The spectral data reduced in this manner are presented in Figure 3.18, Figure 3.19, Figure 3.20 for the eastern and western filaments, the 1'x1' frame in the immediate environment of the filaments (referred to as the "small frame" in Table 3.3), and the location 30W20N, respectively. Table 3.5 summarizes the results. The inferred spectra clearly show a strong 3.4 μm emission feature in the

filaments. From Table 3.5 we observe that the peak $F_{3.4}/F_{3.29}$ ratio is very high in the three calculated spectra. In contrast, the locations 30N and 70N both show a $3.367\mu\text{m}$ flux that indicates weaker (if present at all) $3.4\mu\text{m}$ line emission (see Figure 3.15). The continuum spectrum for the filaments alone is fairly well fit by a $1500\pm 100\text{K}$ graybody, and we have assumed the same temperature for the $1'\times 1'$ frame and the position 30W20N. The $3.158\mu\text{m}$ observations in the filaments do not agree with the 1500K thermal continuum (it is higher by a factor of 3–4), which may indicate that the “true” continuum is brighter than what we have inferred here. The fact that the filaments are redder than both their immediate surroundings as well as the $1'\times 1'$ frame indicates either that they are cooler than their surroundings or that they scatter less light. The inferred spectra show that the fraction of the total emission resulting from thermal emission processes in the H band is $\sim 70\%$ in the filaments. Sellgren *et al.* (1985) have also detected the $3.4\mu\text{m}$ emission feature in NGC7023 at the position 30W20N, as evidenced by the spectrum contained within that reference. The two sets of observations agree to within 5 percent.

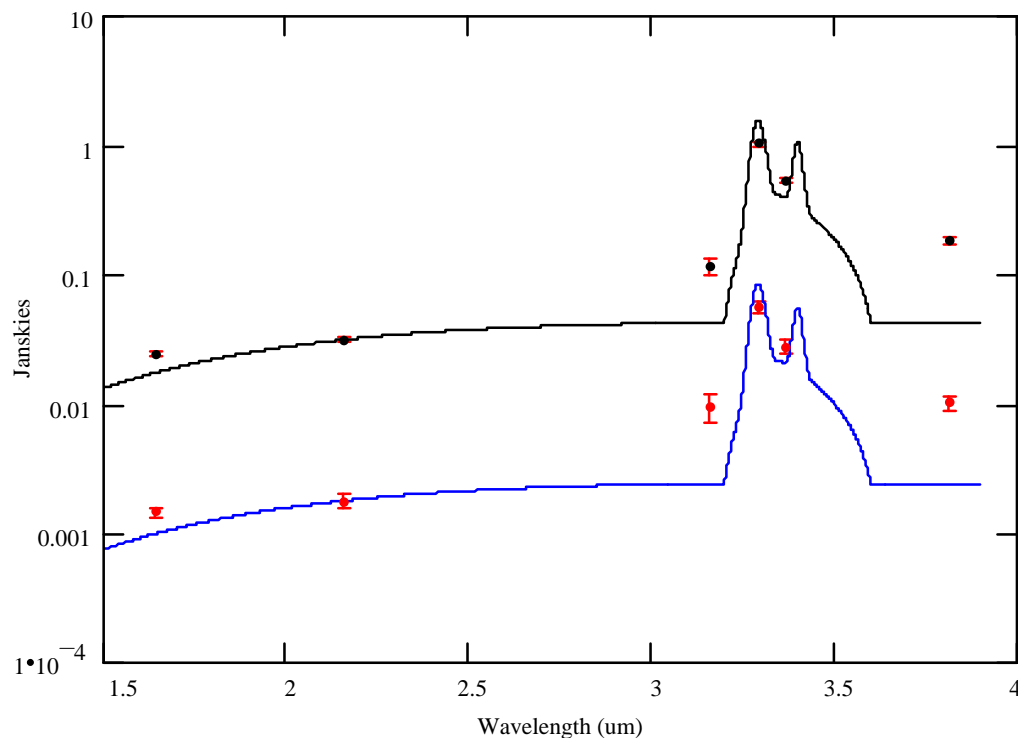


Figure 3.18 - The inferred spectra for the filaments. The observations are also shown in the figure (data points with error bars). The eastern filament is plotted on the bottom and the western is on top. The spectra have been calculated from the observed data as described in the text - see eq (3.4). The vertical axis is plotted logarithmically in Janskies and the horizontal axis is in μm . The eastern filament's flux has been divided by ten in order to separate the plots. The continuum has been fit by forcing a 1500K blackbody through the K' band observation. Note that the 3.158 μm observation is an upper limit as given by the measured flux in the frame shown in Figure 3.4.

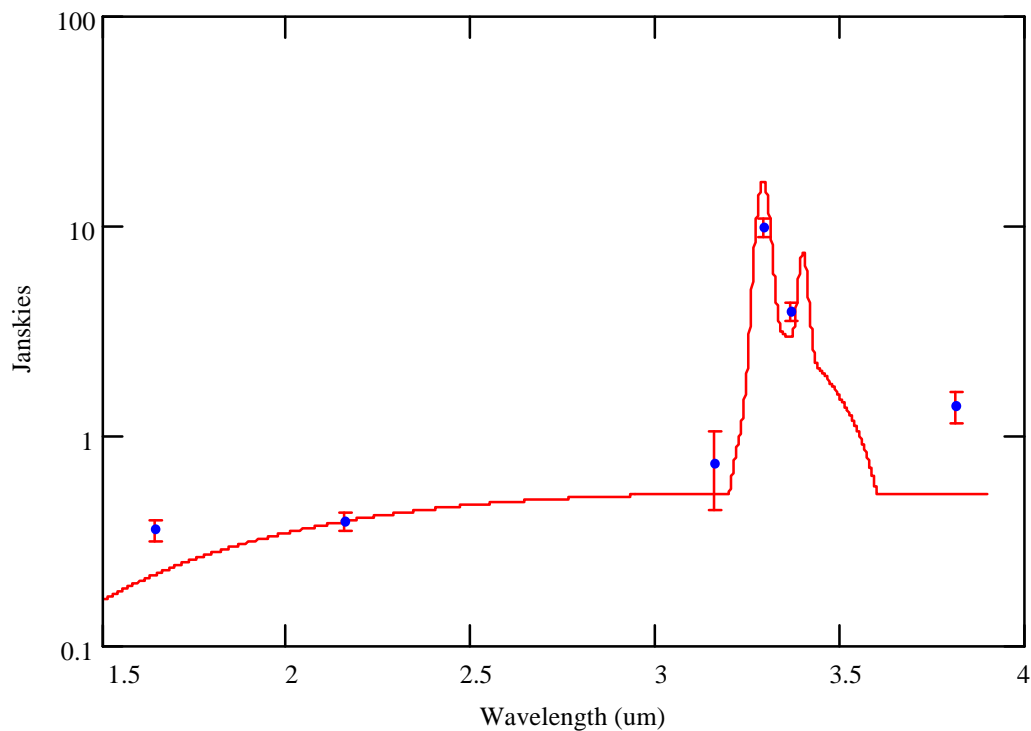


Figure 3.19 - The inferred spectrum for the 1'x1' frame in the immediate neighborhood of the filaments. The observations are also shown in the figure (data points with error bars). The spectrum was calculated from the observed data as described in the text - see eq (3.4). The vertical axis is plotted logarithmically in Janskies and the horizontal axis is in μm . The continuum has been fit by forcing a 1500K blackbody through the K' band observation.

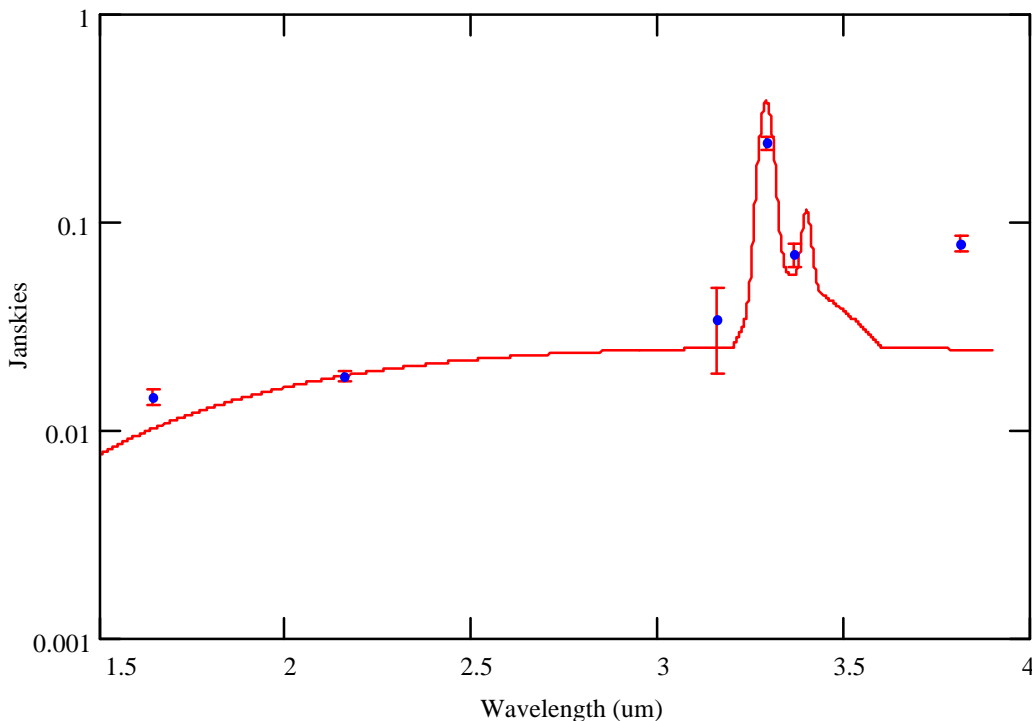


Figure 3.20 - The inferred spectrum for the location 30W20N. The observations are also shown in the figure (data points with error bars). The spectrum was calculated from the observed data as described in the text - see eq (3.4). The vertical axis is plotted logarithmically in Janskies and the horizontal axis is in μm . The continuum has been fit by forcing a 1500K blackbody through the K' band observation.

According to the model of Joblin *et al.* (1996), the large $F_{3.4}/F_{3.29}$ ratios we infer for NGC 7023 imply a relatively low UV intensity compared to most of the regions observed in the Orion Bar, NGC 1333, NGC 2023, and the Red Rectangle. Their model predicts the peak spectral flux density ratio, $F_{3.4}/F_{3.29}$, as a function of the radiation field. The intensity of the UV field which excites these features is denoted G , which is the UV flux integrated from 6eV to 13.6eV. In the galactic plane near the sun, the UV flux is $G_0 = 1.6 \times 10^{-3} \text{ erg cm}^{-2} \text{ s}^{-1}$ (Habing, 1968). Based on the (inferred) strength of the $3.4\mu\text{m}$ emission feature relative to the $3.29\mu\text{m}$ feature, Joblin *et al.*'s model predicts $G = 10G_0$ for the whole frame and $< 100G_0$ for each of the filaments. Independent observations of HD200775 in the ultraviolet enable us to compare the model predictions with measured values. Adopting Witt and Cottrell's (1980) extinction corrected UV fluxes for HD200775, we estimate $G \sim 3200G_0$ at the position of the filaments, assuming that their projected distance is the true distance. This is clearly much larger

than the $<100G_0$ estimated above. A flux of $3200G_0$ would, according to the model, lead to $F_{3.4}/F_{3.29}$ ratios in the range 0.17–0.25, less than half of the inferred values found in the last column of Table 3.5. There are slight differences between our simple model for the plateau and that of Joblin *et al.* At $3.4\mu\text{m}$, their plateau flux level is approximately half the flux in the $3.4\mu\text{m}$ feature (similar to our model), and decreases smoothly to the continuum at $3.15\text{--}3.2\mu\text{m}$, while ours goes to zero at $3.2\mu\text{m}$. This will have the effect of increasing their $F_{3.4}/F_{3.29}$ ratio relative to ours since it will decrease the magnitude of their $3.29\mu\text{m}$ feature by the amount of the plateau at $3.29\mu\text{m}$. These differences are on the order of 10%, and cause the discrepancy between our observations and their model to increase slightly.

The large differences between our observations and the model predictions suggest that either the UV radiation field has been incorrectly inferred, or more likely, that other parameters such as extinction, geometry and density variations play important roles in determining the $F_{3.4}/F_{3.29}$ ratio (assuming that Joblin, *et al.*'s model is correct). There are three possible sources for UV extinction of the flux from HD200775: circumstellar, nebular and interstellar. The UV corrected flux used to compare our inferred $F_{3.4}/F_{3.29}$ ratio with Joblin, *et al.*'s model is a line-of sight extinction, and so provides an upper limit to the UV radiation field. The extinction in the plane of the galaxy (including scattering) is $A_V \sim 2$ magnitudes/kpc on average (Whittet, 1992), which results in $A_V \sim 0.9$ at the distance of NGC 7023. The line-of-sight extinction to this region has been estimated to be $A_V = 1.4\text{--}1.8$ by a variety of observations (Altamore *et al.*, 1980, Witt & Cotrell, 1980, Casey, 1991), which means that there is a sizable extinction component internal to the nebula. A circumstellar component of the extinction will radiate in the $1\text{--}10\mu\text{m}$ spectral region (corresponding to equilibrium dust temperatures in the range $300\text{--}3000\text{K}$), while any nebular component will be apparent in the FIR emission. HD200775 is known to possess such an infrared excess (Altamore *et al.*, 1980, Hartman, Kenyon, and Calvet, 1993), which indicates a dusty disk or shell surrounding HD200775. We determine the relative contribution of this circumstellar envelope to the extinction by comparing the observed IR flux from HD200775 with the UV flux. The ratio of the $12.6\mu\text{m}$ luminosity to that in the optical+UV is ~ 0.2

(Altamore *et al.*, 1980), which implies that the circumstellar extinction comprises a relatively small fraction of the total extinction internal to NGC 7023. Furthermore, if most of the nebular extinction were due to a circumstellar envelope of some sort, we would expect the same radiation field, and hence, according to the model of Joblin, *et al.*, the same $F_{3.4}/F_{3.29}$ ratio at a given distance in all directions. The fact that the filaments exhibit a different ratio than elsewhere in the nebula also argues against a circumstellar shell as the dominant extinction source.

If the model of Joblin, *et al.* is correct, then we require a UV field which is a factor of 30 lower than that which we have inferred from the observations of Witt and Cotrell (1980). Adopting the line-of-sight extinction, $A_V=1.5$, as an upper limit to the internal nebular extinction, we find $A_{0.2\mu\text{m}}=4.2$, or a factor of 49 using the mean extinction curve of Whittet (1992), which is consistent with value required to bring the UV field into agreement with the model. In accord with the preceding paragraph's discussion, we conclude that nebular extinction is likely the cause for our large discrepancy with the model, though we cannot determine the magnitude of the UV radiation field with enough accuracy to validate the model for NGC 7023.

One possible source for the nebular extinction is indicated by the morphology; the knot's position is aligned in such a manner as to very effectively shield the western filament as described. This is also consistent with the knot's lack of emission above the background level at $3.29\mu\text{m}$ (see Figure 3.12). This possibility is ruled out by the following two observations: 1) the extent of the knot is insufficient to provide the shielding for the eastern filament, and 2) the knot will not provide selective extinction, rather it will attenuate the UV field in such a manner as to reduce the brightness of the filaments across all wavelengths.

As a result of self-shielding in the filaments, the interior of the filaments will not only have a lower incident UV flux than the edges, but the incident spectral distribution will also be altered (depending on the optical properties of the extinguishing dust). The fact that the western filament is spatially narrower in the $3.29\mu\text{m}$ feature provides evidence that the feature carriers are more spatially localized than the transiently heated dust grains responsible for the continuum emission, which is

consistent with the interiors of the filaments being shielded from UV flux which might otherwise photodissociate the UIB carriers. This observation also supports the argument that the species responsible for the 3.29 μm line and the NIR continuum are distinct and uncorrelated with one another. We propose that the harder UV photons responsible for the destruction of the 3.4 μm feature carriers (which, according to the model of Joblin *et al.*, amounts to removing a $-\text{CH}_3$ group from a PAH) are absorbed by the filamentary envelope, thereby selectively shielding the interior from these photons and hence, resulting in large values of the $F_{3.4}/F_{3.29}$ ratio at the interior of the filaments. This dynamic process exposes fresh 3.4 μm feature to the UV flux carriers continually. Joblin *et al.*'s model argues for an evolutionary sequence in which the $F_{3.4}/F_{3.29}$ declines with time as a result of the reduction in number of $-\text{CH}_3$ sidegroups on PAHs due to photochemical erosion, but this is a static case in the sense that it assumes there are no new 3.4 μm carriers being introduced to the radiation field, whereas in the filaments, there are new carriers being exposed to continuously (until they have all been destroyed), as the "onion skin" of the filament is peeled away by UV photons. According to Joblin *et al.*, photons with energies between 4 and 4.85eV will uniquely remove the methyl sidegroup from a PAH molecule. The UV observations of HD200775 result in $290G_0$ or $7 \times 10^{10} \text{ cm}^{-2} \text{ s}^{-1}$ worth of such photons from HD200775 at the location of the filaments, or a loss rate of $3.5 \times 10^{-5} \text{ s}^{-1}$. We have assumed that the dehydrogenation cross section for PAHs ($\sigma = 5 \times 10^{-16} \text{ cm}^2$, Geballe *et al.*, 1989) is comparable to that for methyl sidegroups in this calculation. These observations provide an important constraint on the destruction timescale for the 3.4 μm carriers, and should prove useful in choosing between theoretical models.

Joblin's observations show a single data point for NGC 2023 with an enhanced $F_{3.4}/F_{3.29}$ ratio (relative to his model) that is still smaller than the peak values found in the filaments of NGC 7023. NGC 2023 is a reflection nebula similar to NGC 7023 in that it shows a comparable NIR spectrum, filamentary structure in both the broadband NIR and the 3.29 μm emission feature (Solomon, unpublished data), with somewhat more energetic excitation conditions ($T_{\text{star}} = 22,500 \text{ K}$). Given these resemblances, it not surprising that the two objects show comparable results vis à vis the 3.4 μm feature.

location	3.29 μ m		3.4 μ m		F _{3.4} /F _{3.29}
	Peak (Jy)	Integr. W/cm ² $\times 10^{-18}$	Peak (Jy)	Integr. W/cm ² $\times 10^{-18}$	Peak
West	0.674	0.844	0.314	0.256	0.47
East Filament	1.225	1.534	0.619	0.504	0.50
30W20N	0.341	0.422	0.063	0.049	0.18
Small Frame	14.336	17.772	4.996	3.898	0.35

Table 3.5 - F_{3.4}/F_{3.3} emission line ratios. The ratios were derived from the calculated spectrum of NGC 7023 for the various components listed. The inferred spectral values were calculated using equation (3.4) assuming FWHM_{3.4 μ m}}=0.028 μ m and FWHM_{3.29 μ m}}=0.042 μ m. The total strength in the line was calculated by integrating the calculated spectrum over the linewidth, assuming a gaussian profile for each of the features.

3.4 Broadband Nebular Emission

It is well known that reflection nebulae are bluer than their exciting stars (due to dust preferentially scattering blue photons). NGC 7023 is a decidedly red object (see Figure 3.5), indicating that much of the near infrared surface brightness is emission rather than light reflected from HD200775. In the absence of any extinction between the exciting star and the nebula and the observer, the reflected component of the nebula's surface brightness (for the case where the nebula is optically thin) will be given by:

$$B_{\text{ref}} = \frac{F_* \omega \tau H(g, \phi) \sin^2 \phi}{4\pi\theta^2} \quad (3.5)$$

where B_{ref} is the reflected nebular surface brightness, F_* is the measured flux from the star, ω is the albedo of the scattering grains, τ is the optical depth in the nebula, $H(g, \phi)$ is the scattering phase function which contains both the scattering angle ϕ and the phase asymmetry parameter $g = \langle \cos\theta \rangle$, and θ is the angular offset between the star and the nebula (Sellgren *et al.*, 1992). Note that this measure of surface brightness is independent of the distance to the nebula. For the case where there is extinction between the exciting star and the nebula (as is certainly the case for NGC 7023), equation (3.5) provides an upper limit. The surface brightness as given by equation (3.5) reaches a maximum for the case when $\phi=\pi/2$, $\tau=1$ and $H(g, \phi)=1$ (values of $\tau > 1$ do not exceed $B_{\text{ref}}^{\text{max}}$ due to internal nebular extinction).

$$B_{\text{ref}}^{\text{max}} = \frac{F_* \omega}{4\pi\theta^2} \quad (3.6)$$

Equation (3.6) provides an upper limit to the reflected surface brightness if we take $\omega=1$. Table 3.6 shows the maximum reflected contribution to the surface brightness at several positions in the nebula for this case. The broadband nebular surface brightness of the filaments (J, H and K') as determined using equation (3.6) is only consistent with reflected light if we assume an average grain albedo of 1. More realistic estimates for ω are on the order of 0.1 in this spectral region (Draine, 1989)

which means that the broadband emission from the nebula is larger than the light available from HD200775 for scattering. This indicates either sizable non-scattering (*i.e.* emission) contributions to the surface brightness of NGC 7023 at these wavelengths, or the existence of an excitation source in addition to HD200775, or both. The sum of the IR luminosities of the remaining stars visible in the frame is insufficient to provide the flux required to attribute the near infrared (NIR) emission solely to reflected light (see Table 3.1).

Location	Offset (")	J	Max B _{ref}	H	Max B _{ref}	K	Max B _{ref}	[J-H]	[H-K]
Knot	27	0.25	0.66	0.28	0.79	0.37	1.03	0.78	0.80
30W20N	36	0.14	0.37	0.17	0.44	0.21	0.57	0.71	0.76
Eastern Filamen	50	0.18	0.19	0.20	0.23	0.28	0.30	0.74	0.73
Western Filamen	48	0.15	0.21	0.19	0.25	0.24	0.32	0.69	0.77

Table 3.6 - Peak surface brightness in mJy/square arcsec at three locations in NGC 7023. Offset is the average angular offset between HD200775 and the nebular position in arcsec. The theoretical value for the maximum reflected light, Max B_{ref}, is calculated from equation (3.6) assuming unity albedo and using the observed infrared fluxes for HD200775.

Infrared polarization, which is a characteristic of scattered light, has been measured in the nebula by (Sellgren *et al.*, 1992) in 10.5" beams in the nebula at positions where we have observed the color temperature to be hotter than in the filaments. These observations indicate that the emission from these regions is not dominated by light reflected from HD200775, rather Sellgren attributes the bulk of the observed NIR flux to small, transiently heated particle emission. Table 3.7 shows the fraction of the total surface brightness Sellgren attributes to reflected light in 10.5" diameter apertures at several locations in the nebula, along with the colors measured for the same apertures. There are two components making up the NIR surface brightness of the nebula with different energy distributions: polarized scattered starlight and unpolarized small particle emission. The fact that the [H-K] color

becomes more blue as the observations move to the north of HD200775 is consistent with Sellgren's calculations that there is a larger reflected component to the north. The colors qualitatively follow Sellgren's inferred ratio of reflected to total light ($S_{\text{ref}}/S_{\text{total}}$) in that larger [H-K] values (redder colors or lower color temperatures) clearly correlate with smaller fractions of reflected light. The shorter wavelengths (the J band and to a lesser degree the H band) contain larger fractions of light scattered from HD20075 due to scattering from small dust grains, and hence exhibit bluer colors.

Offset	$S_{\text{ref}}/S_{\text{total}}$			Colors	
	J	H	K	[J-H]	[H-K]
30W 20N	0.26–1.00	0.11–0.42	0.04–0.17	0.71	0.76
30N	0.21–1.00	0.07–0.35	0.05–0.22	0.91	0.67
60N	–	–	0.13–1.00	0.53	0.37
70N	0.22–1.00	0.18–0.82	–	0.48	0.17

Table 3.7 - The ratio of reflected to total light ($S_{\text{ref}}/S_{\text{total}}$) from Sellgren *et al.*, (1992). [J-H] and [H-K] colors as measured in this work are given in the final two columns for comparison.

A unique temperature can be calculated from the Planck function given the brightness of an object at two wavelengths, this temperature is referred to as the color temperature. Note that the color temperature determined in this way does not distinguish between reflected and emitted light so that regions dominated by light reflected from HD200775 will show higher color temperatures than those whose brightness is derived exclusively from dust emission due to the high temperature of HD200775. It is determined by solving the following transcendental equation for T_c , the color temperature:

$$\frac{F_{\lambda_1}}{F_{\lambda_2}} = \frac{\epsilon_{\lambda_1} B_{\lambda_1}(T_c)}{\epsilon_{\lambda_2} B_{\lambda_2}(T_c)} \quad (3.7)$$

where F_λ is the spectral flux density, ϵ_λ is the grain emissivity and B_λ is the Planck function. In order to create the color temperature frame, the component images were scaled to the same resolution (FWHM) and platescale prior to making the maps by first convolving the K' frame (smaller PSF) with a gaussian

kernal, and then “demagnifying” via bilinear interpolation. We approximate the true grain emissivity with a power law such that $\epsilon \propto \lambda^n$, where $n=-1.75$ (Draine & Lee, 1984, Draine, 1989). The resultant grayscale image is shown in Figure 3.9

The emissivity of the grains is dependent on their properties and hence, a function of wavelength. Consistent with Sellgren’s large beam observations (Sellgren, 1984), we find that the [H–K] colors (and hence, the color temperature) remain roughly constant with distance from HD200775, except for the filamentary regions as well as to the north of HD200775. Between approximately 10” from HD200775 and star D the color temperature ranges from 1400 to 1500K (assuming an emissivity law proportional to $\lambda^{-1.75}$). This constancy is not compatible with equilibrium grain heating and emission, as will be discussed later in this section (see equation (3.10)) from which we expect the grain temperature to fall with distance from the exciting source. Table 3.7 shows that northward of the eastern filament the broadband colors become bluer (the locations 60N and 70N are the bluest of our nebular observations of NGC 7023). This may provide an indication of additional excitation sources in the region, or that this region contains a larger fraction of scattered light from HD200775 due to varying dust composition. The filamentary colors are *redder* than the surrounding nebulosity, indicating lower color temperatures (see Figure 3.9) by approximately 100–200K). While the average value for the H–K’ color in the filaments is approximately 0.75 (corresponding to a color temperature of 1400K), the maximum H–K’ color found in the reddest regions of the filaments is approximately 1 (corresponding to a color temperature of 1200K), which provides additional evidence for self-shielding of the filamentary interior. The locations where the colors are reddest are also the regions where the 3.29 μ m emission is the brightest. We argue that the spatial distribution of the dust grain population (classical and very small grains) gives rise to variations in the observed color temperature of the nebula; the light from the bluer regions is dominated by photons reflected from the central star, while the filaments are dominated by dust emission with redder colors. This scenario qualitatively explains some of the polarization measurements (see Table 3.7).

If the NIR emission were due to equilibrium grain heating, we expect the dust temperature to fall with distance from the exciting star according to the energy balance equation $P_{\text{abs}} = P_{\text{rad}}$, which states that the power absorbed and radiated by the grains are equal when thermal equilibrium prevails. The power absorbed by the grain is given by:

$$\begin{aligned} P_{\text{abs}} &= \pi a_{\text{gr}}^2 \int Q_{\text{abs}}(a_{\text{gr}}, \lambda) F(\lambda) d\lambda \\ &\approx \pi a_{\text{gr}}^2 \langle Q_{\text{abs}}(a_{\text{gr}}) \rangle \frac{L_{\text{src}}}{4\pi r^2} \end{aligned} \quad (3.8)$$

where a_{gr} is the grain radius, $Q_{\text{abs}}(\lambda)$ is the grain absorption efficiency (defined as the ratio of the absorption to geometrical cross sections), and $\langle Q_{\text{abs}} \rangle$ is its Planck averaged value, $F(\lambda)$ is the flux incident on the grain, L_{src} is the source luminosity and r is the distance from the source to the grain. The power radiated by the grain is given by:

$$\begin{aligned} P_{\text{rad}} &= 4\pi a_{\text{gr}}^2 \pi \int Q_{\text{em}}(a_{\text{gr}}, \lambda) B(\lambda, T) d\lambda \\ &= 8\pi^2 a_{\text{gr}}^2 Q_{\text{em}}(a_{\text{gr}}, \lambda_o) \lambda_o \frac{(kT)^{4+n}}{h^{3+n} c^{2+n}} \Gamma(4+n) \zeta(4+n) \end{aligned} \quad (3.9)$$

where $Q_{\text{em}}(\lambda)$ is the grain emissivity, $B(\lambda, T)$ is the Planck function, T is the grain temperature, ζ is the Riemann zeta function and Γ is the gamma function (Abramowitz & Stegun, 1965). We have assumed a power law dependence for the emissivity such that $Q_{\text{em}}(\lambda) = Q_{\text{em}}(\lambda_o) (\lambda_o/\lambda)^n$. Setting equation (3.8) equal to equation (3.9) and solving for the temperature yields:

$$T(r, L_{\text{src}}, a_{\text{gr}}, n) = \left[\frac{\langle Q_{\text{abs}}(a_{\text{gr}}) \rangle L_{\text{src}}}{Q_{\text{em}}(\lambda_o, a_{\text{gr}}) r^2} \frac{h^{n+3} c^{n+2}}{32\pi^2 \lambda_o^n k^{n+4} \Gamma(n+4) \zeta(n+4)} \right]^{\frac{1}{4+n}} \quad (3.10)$$

The optical constants used to calculate the equilibrium temperatures from equation (3.10) were obtained from Draine & Lee (1984). We calculated the Planck averaged value of the absorption efficiency using the temperature of HD200775 (17000K) and obtained $\langle Q_{\text{abs}} \rangle = 0.67$ and 1.3 for silicate and graphite grains with radii of $0.1\mu\text{m}$, respectively. We use values of $Q_{\text{em}}(\lambda_o = 3.3\mu\text{m}) = 0.014$ and

0.067 for silicate and graphite grains (of the same radius), respectively. Equation (3.10) predicts temperatures in the range 70–80K at the distance of the filaments for a $4000L_8$ excitation source (the extinction corrected luminosity of HD200775) and an emissivity index $n=1$. We note that the temperature calculated from equation (3.10) is only weakly dependent on ratio of the grain optical constants; an increase (decrease) in this ratio by an order of magnitude results in an increase (decrease) in the calculated temperature by a factor of 1.6 (0.6). These temperatures are clearly too low to account for the NIR emission observed in NGC 7023. Sellgren *et al.*, (1985) published spectra of NGC 7023 from 4–13 μm and found that the continuum level was much flatter than expected from a single temperature graybody, but that the data could be fit with a range of grain temperatures from 1000–1500K, in agreement with the color temperatures we have calculated (see, for example Figure 3.19). In addition, as noted, equilibrium grain heating should show a temperature dependence on distance from HD200775 proportional to $r^{-2/n+4}$, whereas the near infrared colors across NGC 7023 remain remarkably constant between HD200775 and star D. Interestingly, the colors become bluer as we traverse the vector directly northward of HD200775 (see Table 3.7). In this case, the reflected component is also seen to increase with distance from the central star from the polarization measurements, indicating either the presence of another source of scattering photons towards the north or varying dust composition in this direction.

Several potential mechanisms for the excess IR emission have been investigated in the literature and found wanting. They include equilibrium grain emission, dust feature emission, and hydrogen free-free emission. The widely accepted conclusion for the NIR continuum emission is that it consists mainly of non-equilibrium heating of small grains/large molecules (radii~3–100 \AA) during which the random absorption of a single UV photon is followed by a large temperature excursion and subsequent thermal decay of the grains. The broad NIR continuum thus results from a large number of small grains undergoing temperature transients in this manner. The individual grains are not in radiative equilibrium; rather the grain temperatures are wildly fluctuating (with the average temperature corresponding to the equilibrium value).

3.5 FIR Correlation

The total flux in the 3.29 μ m emission feature has been correlated with a variety of emission processes: atomic hydrogen (Br γ - Mizutani *et al.*, 1989), molecular hydrogen ($v=1 \rightarrow 0$ S(1), Mouri *et al.*, 1990), and most interestingly, the total far infrared dust luminosity. This latter correlation has been quantified for a number of IR luminous galaxies; the empirical correlation is given as $L_{\text{FIR}}/L_{3.29} \approx 1400$ where L_{FIR} and $L_{3.29}$ are the far infrared and 3.29 μ m feature luminosities, respectively (Dennefeld and Désert, 1990). This relation was derived for reddened objects (uncorrected for extinction at 3.29 μ m), and so represents an upper limit to the expected ratio. Theoretical considerations derived from the observed luminosity of PAHs per hydrogen atom embedded in the solar neighborhood interstellar radiation field (ISRF, Mathis *et al.*, 1983) support this relationship; it is supposed to scale (linearly) with the energy density of the radiation field. Satyapal. *et al.* (1994) have measured the extinction in the nuclear region of M82 using IR hydrogen recombination lines and hence derive an extinction corrected value for the correlation as $L_{\text{FIR}}/L_{3.29} \approx 1340$. They have also determined extinction-corrected values for NGC3690A and NGC3690A/B of 1840 and 1550, respectively (Satyapal, 1995). Table 3.8 provides this ratio for several galaxies.

Object	$L_{\text{FIR}}/L_{3.29}$	Type	note/ref
NGC253	2330	starburst galaxy	(1),(2)
M82	1340	starburst galaxy	(3)*
NGC3690A	1840	galactic merger	(3)*
NGC3690B/C	1550	galactic merger	(3)*
NGC4194	1390	galactic merger	(2),(4)

Table 3.8 - The $L_{\text{FIR}}/L_{3.29}$ ratio in a variety of IR bright galaxies. References for these data are: (1) Moorwood (1986) , (2) Dennefeld and Désert (1990), (3) Satyapal, 1995, (4) Mizutani *et al.* (1989). An asterisk beside a reference indicates that the ratio has been corrected for extinction at 3.29 μ m.

If the $L_{\text{FIR}}/L_{3.29}$ correlation is, as has been observed, constant to within a factor of ~ 1.5 across both a wide range of excitation conditions as well as varying spatial extents, then there is likely a relationship between the emitting species or in the chemistry leading to their creation. If a roughly constant $L_{\text{FIR}}/L_{3.29}$ ratio is representative of the *global* physics of astronomical objects (Table 3.8 shows that this appears to be the case), there will necessarily be additional constraints on the emitting grains and their properties. Dennefeld and Desert (1990) suggest that coextensivity of the $3.29\mu\text{m}$ feature carriers with the FIR emitting dust is the primary requirement for this correlation to hold, and that the abundance of the $3.29\mu\text{m}$ feature carrier relative to the total dust composition is constant. While Figure 3.22 shows that the location of the longest wavelength FIR emission peak is farther from the exciting source, the bulk of the FIR luminosity is derived at shorter wavelengths where the emission peaks are spatially coincident. Given that the $3.29\mu\text{m}$ feature is excited by UV/optical photons, while the FIR grain emission is observed over a much wider range of excitation conditions, we propose that the following additional conditions are met for the correlation to hold: 1) the radiation field exciting the $3.29\mu\text{m}$ emission feature is not so hard as to destroy a sizable fraction of the $3.29\mu\text{m}$ carriers, 2) the $3.29\mu\text{m}$ carriers are optically thin to the radiation exciting the NIR emission, 3) the nebula is optically thick to the wavelengths responsible for exciting the FIR emission. The last condition permits the $L_{\text{FIR}}/L_{3.29}$ correlation to hold for the case where the two grain populations are not coextensive; as long as most of the photons that heat the FIR grain population are absorbed in the nebula, we expect that the relative constancy of the populations will then make the correlation hold.

Before proceeding to calculate this ratio for NGC 7023 we consider the completeness of our observations and whether there is additional $3.29\mu\text{m}$ emission outside the volume of space we have observed. The feature emission drops to our detection limit within $10''$ of HD200775 to both the east and south. Additionally, Sellgren (1985) found no $3.29\mu\text{m}$ feature emission in a $12''$ beam $120''$ to the north of HD200775 (with her detection limit of 32 mJy or $0.28 \text{ mJy arcsec}^{-2}$). The full $3.29\mu\text{m}$ frame ($100'' \times 100''$, see Table 3.3) shows extended diffuse emission over most of the region observed to

the north of the filaments in the range $0.8\text{--}1.7\text{ mJy arcsec}^{-2}$ ($50\text{--}100\mu\text{Jy/pixel}$), corresponding to a total flux over the frame of 6.4 Jy which is about 45% of the total $3.291\mu\text{m}$ emission in the frame. The diffuse emission is not uniform over the frame, it peaks along the vector running through and parallel to the long axis of the filaments ($\text{NE} \leftrightarrow \text{SW}$) and falls on either side towards the NW and SE; it goes to zero near HD200775 and to the NW of the filaments. We believe that this extended emission is real since our $3.29\mu\text{m}$ measurements agree with those of Sellgren (1983, 1984) at several widely spread positions in the nebula. The presence of this diffuse, extended emission suggests that our observations may not be complete, *i.e.* that we have not measured all the $3.29\mu\text{m}$ emission from NGC 7023.

The total emission in the $3.291\mu\text{m}$ band is larger than what we have observed due to the finite spectral width of the emission feature as well as the spectral resolution of our CVF. Assuming that the $3.29\mu\text{m}$ feature is gaussian in nature with $\text{FWHM}=0.042\mu\text{m}$ (type 1 spectrum - Tokunaga *et al.*, 1991), we can integrate the spectra derived from equation (3.4) to obtain the total flux in the emission line(s). Excluding the stellar emission (stars D and H), the total observed *nebular* $3.29\mu\text{m}$ emission in the region containing the filaments (see Figure 3.5 and Table 3.3, small frame, approximately $1'\times 1'$) is 10.044 Jy , from which we infer a peak emission in the feature of 14.032 Jy (see Table 3.5). The total integrated flux in the $3.29\mu\text{m}$ feature is then $1.74\text{E-}17\text{ watt cm}^{-2}$ which corresponds to a luminosity of $1.05L_8$. Using observed FIR data we now compute the FIR luminosity in order to compare the $L_{\text{FIR}}/L_{3.29}$ ratio for NGC 7023 with that of other observers.

The FIR spectrum of NGC 7023 peaks between $60\mu\text{m}$ and $100\mu\text{m}$, corresponding to dust temperatures in the $30\text{--}50\text{K}$ range (note the agreement with the calculations from equation (3.10). Whitcomb *et al.* (1981) mapped NGC 7023 at $55\mu\text{m}$, $125\mu\text{m}$ ($50''$ FWHM beam diameters) and $400\mu\text{m}$ ($1'$ beam FWHM) from the ground, and also made airborne measurements at 140 , 170 , 200 and $300\mu\text{m}$ with a $1.7'$ beam. From these observations, they measure $F_{\lambda>40\mu\text{m}} = 1400L_8$ in a roughly $4'$ diameter region centered on HD200775. Additional photometry, based on KAO and IRTF observations centered on the FIR emission peak at 100 , 160 , and $200\mu\text{m}$ with a $53''$ beam (Casey, 1991) give the

total FIR emission in NGC 7023 as $F_{\lambda > 30\mu\text{m}} = 587L_{\odot}$ for a roughly $2' \times 3'$ angular area. There is ample evidence for nonuniform, centrally peaked emission given the nebular morphology evident in the grayscale images presented here. In order to accurately assess the FIR correlation, given the probable incompleteness of our $3.29\mu\text{m}$ observations, it is necessary to compare the same areas on the sky. Casey's (1991) KAO observations at $100\mu\text{m}$, $160\mu\text{m}$ and $200\mu\text{m}$ were all made with $53''$ gaussian beams, which is slightly smaller than the size of our frame. We have synthesized a FIR spectrum for NGC 7023 using both observers' data, shown in Figure 3.21, using the observations of Casey and Whitcomb *et al.* where we have scaled Casey's observations to the beam size of Whitcomb *et al.* ($1'$ beam) centered on the FIR peak. A spectrum of the form $\nu B_{\nu}(T)$ has been overlaid on the data, where B_{ν} is the Planck function, $T=45\text{K}$ and the spectrum was forced through the $100\mu\text{m}$ observation. The spectral fit is seen to be excellent for wavelengths beyond about $30\mu\text{m}$, consistent with the emissive properties expected for amorphous, layer-lattice materials such as carbon compared with graphite, silicon carbide and other crystalline dielectrics which exhibit spectral indices that follow λ^{-2} (Tielens and Allamandola, 1987). The excess emission at $25\mu\text{m}$ relative to the thermal fit (νB_{ν}) in Figure 3.21 is attributed to the pervasive emission of nonequilibrium dust grains, well known to dominate much of the IR luminosity at wavelengths less than $25\mu\text{m}$ in objects where the UIBs are observed, as well as in the galactic cirrus. Integrating the νB_{ν} spectrum from $30\text{--}400\mu\text{m}$, we find $230L_{\odot}$ worth of luminosity for $1'$ beam. Table 3.9 summarizes the FIR luminosities determined for NGC7023, in addition to the calculated $L_{\text{FIR}}/L_{3.29}$. We are unable to confirm or refute the correlation without additional observations of a larger area than we have observed. If there is no additional $3.29\mu\text{m}$ feature emission outside the area which we have observed, then the correlation observed globally in galaxies appears to be valid in NGC 7023, we obtain $L_{\text{FIR}}/L_{3.3}=1333$. In this case, the FIR and $3.29\mu\text{m}$ emission are not coextant, the $3.29\mu\text{m}$ carriers are more spatially localized than those responsible for the FIR emission. On the other hand, if there is more $3.29\mu\text{m}$ feature emission than is contained in our observation area, then the

values of this ratio are considerably smaller than that observed globally in galaxies, showing that the correlation does not hold for the region containing the peak 3.29 μm and FIR emission in NGC7023.

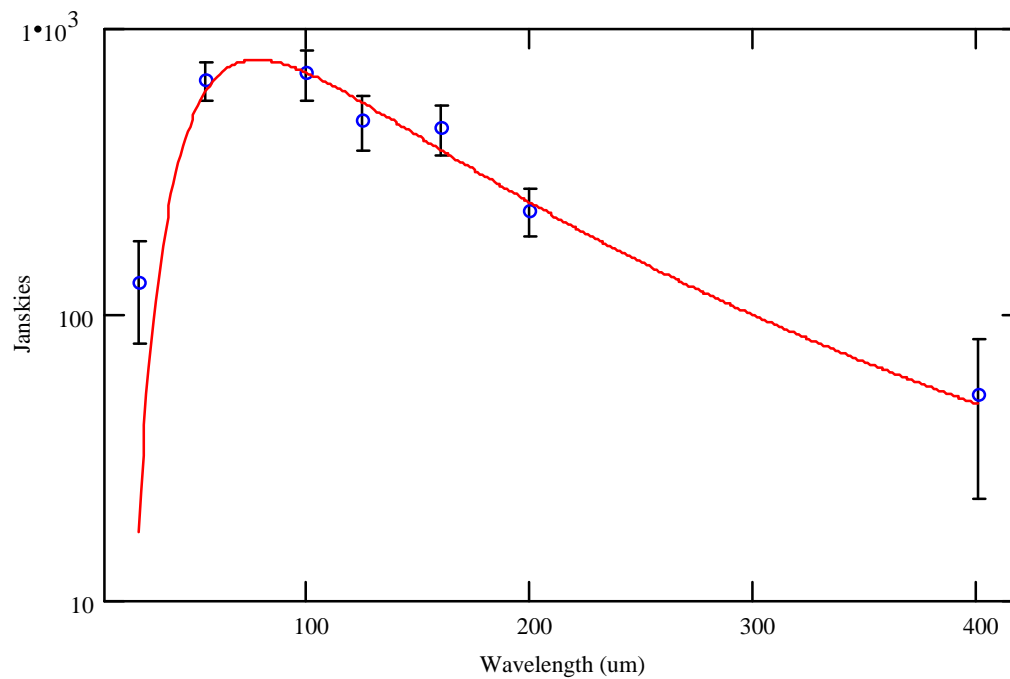


Figure 3.21 - Combined FIR spectrum for NGC7023. The data points with error bars are the observations by Casey (1991) and Whitcomb. *et al.* (1981) discussed in the text. Casey's observations have been normalized to Whitcomb's beamsize (1'). The solid line is a spectrum of the form $\nu B_{\nu}(T)$ where B_{ν} is the Planck function evaluated for $T=45\text{K}$ and forced through the 100 μm observation.

Wavelength (μm)	Observation area	$L_{\text{FIR}} (L_{\odot})$	$L_{\text{FIR}}/L_{3.29}$	L_{FIR} Method	Ref.
>40	~4' diameter	1400	1333	125 μm obs	(1)
>30	~2'x3'	587	559	spline	(2)
>30	~1'x1'	230	220	$\nu F_{\nu}(45\text{K})$ fit	(1),(2)

Table 3.9 - Total FIR luminosity of NGC7023. The methodology used to determine the FIR luminosity is described in the text (see also Figure 3.21). The reference column refers to: (1) Whitcomb *et al.*, (1981), (2) Casey, (1991).

The ISRF may also be contributing to the excitation in NGC7023, so we will now calculate the distance from HD200775 where the ISRF is comparable to the diluted radiation field from the star.

The classical grains responsible for the FIR emission are expected to follow equation (3.10) since their heat capacity is large enough for them to remain in thermal equilibrium with the exciting radiation field. This grain population contains a mixture of both graphite and silicate grains; the emissivity of graphite is more important below $8\mu\text{m}$, while that of silicate grains will dominate in the $8\mu\text{m}$ – $50\mu\text{m}$ spectral region, and the two components contribute equally beyond $50\mu\text{m}$ (Draine and Lee, 1984). The steady state temperatures for graphite and silicate grains exposed to the solar neighborhood ISRF are 19K and 15K, respectively (for $0.1\mu\text{m}$ grains, optical constants from Draine and Lee, 1984). When the grain temperatures in NGC7023 reach this temperature regime, the ISRF will become important in the energetics establishing the observed grain temperatures.

This calculation is complicated by the presence of internal nebular extinction. The existence of IR excess in the spectrum of HD200775 (see section 3.6) implies circumstellar extinction, and the morphology of the nebula also strongly suggests that there is internal extinction, both of which will dilute the radiation field from HD200775 as it expands and heats the nebular dust. The simplest approach to the problem is to use the measured extinction, $A_V=1.5$, for the nebular extinction. Alternately, we estimate the magnitude of this “internal” nebular extinction by the ratio of the FIR luminosity of the nebula to the total luminosity of HD200775, tantamount to assuming that the all of the extinction results in the heating of classical grains that emit in the FIR. For the FIR luminosity, we adopt $1400L_\odot$ (see Table 3.9), which when combined with the stellar luminosity of $4000 L_\odot$ indicates that 35% of the total extinction is internal to the nebula. The resultant extinction determined in this way yields $A_V=0.5$, consistent with the estimate made in section 3.3.2 where we calculated the average line of sight extinction in the plane of the galaxy to be $A_V=0.9$. The interstellar radiation density due to stellar illumination and scatter is $7\times 10^{-13} \text{ erg cm}^{-3}$ (Allen, 1976). In order to determine the distance at which the diluted radiation field from HD200775 matches that of the ISRF we solve the following equation:

$$\begin{aligned} \text{ISRF} \times c &= \frac{L_*}{4\pi r^2} \\ &= \frac{4\pi r_*^2 \pi \int B_\lambda(T_*) 10^{-0.4A(\lambda)} d\lambda}{4\pi r^2} \end{aligned} \quad (3.11)$$

where r is the distance being solved for, B_λ is the Planck function, $A(\lambda)$ is the extinction law, and L_* , r_* and T_* are the luminosity, radius and effective temperature of HD200775, respectively. Adopting $r_* = 7.2R_\odot$, $T_* = 17000\text{K}$ and $A(\lambda) = A_V (0.55\mu\text{m}/\lambda)$ yields distances of 0.74pc (5.9') and 1.5pc (12') for $A_V = 1.5$ and 0.5, respectively. These distances are well beyond the region of NGC7023 which we have observed here. From this analysis, we conclude that HD200775 dominates the nebular grain heating in NGC7023.

The spatial locations of the FIR peaks (relative to HD200775) can be seen in Figure 3.22, where we have also indicated the location of the 3.291 μm peak and that of star D, and HD200775 is at the origin. The spatial position of the 3.291 μm emission peak was determined by integrating the signal flux in a 1' area in order to match the apertures used by Casey and Whitcomb. From this figure, we see that the 3.291 μm peak corresponds closely to the location of the 55 μm , 100 μm and 125 μm peaks, lending empirical credence to a correlation between the FIR and the 3.291 μm emission. We note that the FIR emission peaks move further from HD200775 with increasing wavelength (decreasing temperature) as is expected for thermal equilibrium emission of dust heated by HD200775. Figure 3.22 shows that from the NIR out to 125 μm , the emission peaks are practically coincident. This is the spectral region that contains the bulk of the FIR luminosity, while the 200 μm emission peak, 30' further north, contains less than 10% of total FIR luminosity.

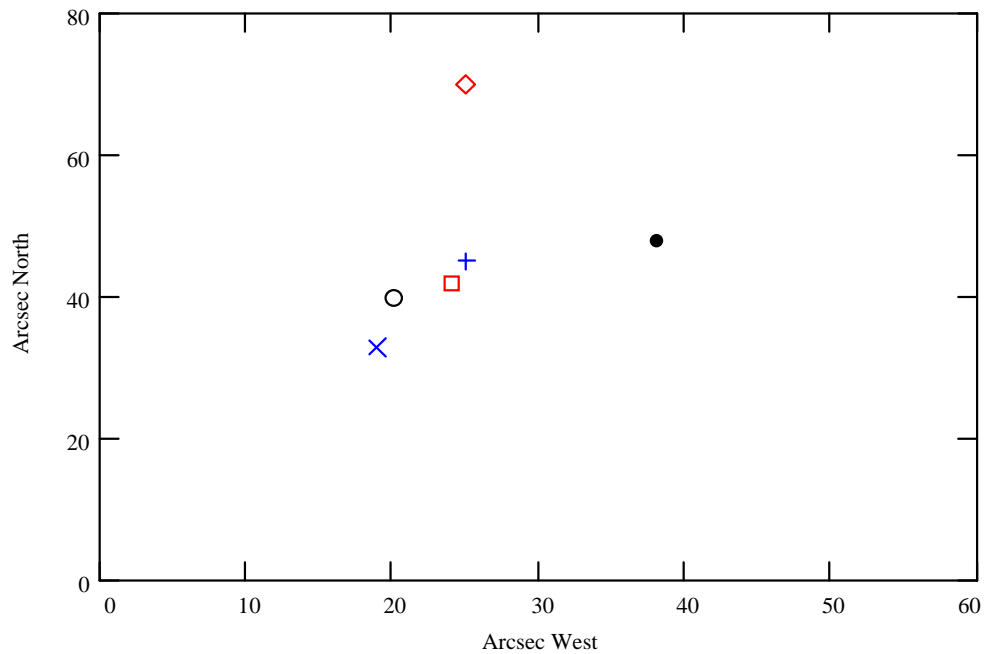


Figure 3.22 - The locations of the IR emission peaks relative to HD200775. The black dot is star D, the \times and + are the 55 and 125 μm peaks, respectively (Whitcomb *et al.*, 1981), the square and (upper) diamond are the 100 and 200 μm peaks (Casey, 1991), and the open circle is the 3.291 μm peak. The beamsize of both sets of FIR observations is 1', and the positional errors are $\sim 10''$.

3.6 Stellar Objects

The observed infrared morphology of NGC 7023 is driven by the energy sources within the nebula (the stellar objects) interacting with the ambient medium (the molecular cloud). Given that HD200775 is a young stellar object, it is reasonable to expect that there will be other similar types of objects in the nebula as well since star formation is known to occur in clusters within molecular clouds. Sellgren (1982) searched a region approximately $5' \times 7'$ centered on HD200775 for stellar objects and found 30 such objects, of which 11 were brighter than $[K]=10.5$ (restricted in order to obtain a complete and reasonably uniform sample). Based on both optical as well as infrared photometry, 6 out of the 11 stellar objects were found to be variable (of which we have observed 3: at least one of stars B and C as well as stars D and J). There are additional indications of protostellar activity as evidenced by the presence of both excess infrared emission and hydrogen line emission from several of the stars. All the variable stars were found to be those that were close to HD200775, the central and brightest star of the nebula. The spatial clustering also argues that the stars observed are associated with the molecular cloud. In this section, we will attempt to reveal the nature of the stellar objects exciting the nebula in order to gain a better understanding of the excitation conditions in NGC 7023. We will also investigate star D as a possible cause of the filamentary morphology, which is suggested by the nebular geometry.

The spectra of both star D and HD200775 (see Figure 3.25) most closely resemble that of a protostellar type II or IID object (Wilking, 1989, Zinnecker *et al.*, 1992) since both objects appear to exhibit considerable IR excess. Neither is red enough to be a class I object. The protostellar classification scheme refers to the stage of accretion: Class I - infall onto star and accretion disk, Class IID - clearing at the poles, infall along the equator and outflow along the poles, Class II - no infall, excess IR emission from stellar heating of disk, Class III - naked star, no circumstellar material. In classes II and III all the light from the object derives from the star itself, while classes I and IID luminosity comes, at least in part, from the disk. NGC 7023 exhibits a bipolar molecular outflow (Watt

et al., 1986), lending support to the type IID classification if the outflow is associated with either of the two stars under discussion.

Figure 3.23 is a plot the [J-H] vs [H-K] colors for the stellar objects in the frame as well as the filaments and several nebular positions compared to the unreddened colors of stars. Also included are the reddening vector (determined using the interstellar extinction curve from Draine, 1989) and the colors for thermal sources of various temperatures. The infrared colors of stars A and K are compatible with their being K3V and M3V stars, respectively. These are the only stellar objects in the frame which are clearly consistent with main sequence stars. The colors of stars J and H require extinction more than an order of magnitude in excess of that observed in order to bring them onto the main sequence near stellar type M0V. Furthermore, both stars appear to be variable (see Table 3.1, and Sellgren, 1983), adding evidence that they are not a main sequence objects. Similarly, the colors of stars B, C and D indicate that they are not main sequence objects either. Identification of stars B, C, D, J and H as evolved giant or supergiant stars would place them well out of the galactic plane, also an unlikely scenario. The variability and location of these stellar objects implies membership in the young cluster, and we propose that they are pre-main sequence object. Our photometry for stars E, F and G is not sufficiently accurate to draw conclusions regarding their nature.

We note the similarity in colors between HD200775, the filaments and the positions 30W20N and 14W23N. Comparison of the nebular spectra presented in Figure 3.15 with the spectrum of HD200775 presented in Figure 3.25 clearly shows that similar colors do not necessarily imply that the nebular light is scattered from HD200775, or that the dust composition is similar in these regions. If the NIR colors of HD200775 were intrinsic then we would expect the filamentary colors to be bluer than the star for the case of scattered light because scattering is more efficient at shorter wavelengths. Since the infrared colors of HD200775 are due to dust in its immediate vicinity (rather than being intrinsic to the star, see discussion in 3.6) and the light from the filaments is mostly nonequilibrium emission (at least for wavelengths longer than $1.6\mu\text{m}$), we conclude that the circumstellar dust around HD200775 is dissimilar to that contained in the nebular positions plotted in Figure 3.23. The nebular

positions 30N and 60N fall very close to the blackbody colors at 2300K and 3500K, respectively. The 60N aperture has significantly bluer colors compared with all the other nebular positions shown. We take this as evidence that scattering is the dominant source of light from this location which, in turn, shows that the dust composition varies considerably over the nebula. This is consistent with the polarization measurements of Sellgren *et al.*(1992) shown in Table 3.7. It is also possible that there is an additional excitation/scattering source to the north, perhaps star H or star J. The location 14W23N is within the knot discussed in sections 3.4; its colors are only slightly redder than those of the filaments. The weakness of the $3.29\mu\text{m}$ emission feature in the knot suggests the possibility that the feature carriers are destroyed in the vicinity of HD200775, leaving small grain emission and scattering to provide the surface brightness observed.

All observations of NGC 7023 point to HD200775 as the primary source of radiative excitation in the nebula. HD200775 is classified as a Herbig Be star, which is thought to be an intermediate mass, young, pre-main sequence star (Herbig, 1960, Strom *et al.*, 1972, Finkenzeller & Mundt, 1984). As a class, they are frequently found to be variable, though there is no evidence for variability in HD200775 (Sellgren, 1983). Almost all of the flux from HD200775 lies in the UV and visible portion of the spectrum (Casey, 1991). The energy distributions of many Herbig Be stars show excess IR emission in the $2\text{--}20\mu\text{m}$ spectral region, well above that expected from either the intrinsic photosphere or an HII region in combination with free-free emission. HD200775 exhibits such an IR excess, which is believed to arise from either a surrounding dusty envelope (Hartmann *et al.*, 1993) or a circumstellar disk (Hillenbrand *et al.*, 1992, Lada & Adams, 1992) that reprocesses the stellar radiation. Figure 3.24 shows a 17000K blackbody curve (to approximate the stellar photosphere) fit to the extinction corrected V band ($0.54\mu\text{m}$) observation of HD200775 (Altamore *et al.*, 1980), in addition to the broadband IR observations presented here. From this figure, we see that the $2\text{--}4\mu\text{m}$ emission observed is well above that of the stellar photosphere, consistent with excess emission due to dust in a circumstellar envelope or disk heated by HD200775. The total luminosity of HD200775 (corrected for $A_V=1.5$) is approximately $4000L_\odot$, and the $2\text{--}20\mu\text{m}$ luminosity of approximately $150L_\odot$.

(Whitcomb *et al.*, 1981). This is consistent with an optically thick disk covering 4% of the 4π sr solid angle around HD200775. Fitting a spline to the spectrum of HD200775 in Figure 3.25 and integrating over the 2–4 μ m region, we obtain 43 L_8 worth of power in this bandpass alone, which is 30% of that observed over the entire 2–20 μ m spectral region.

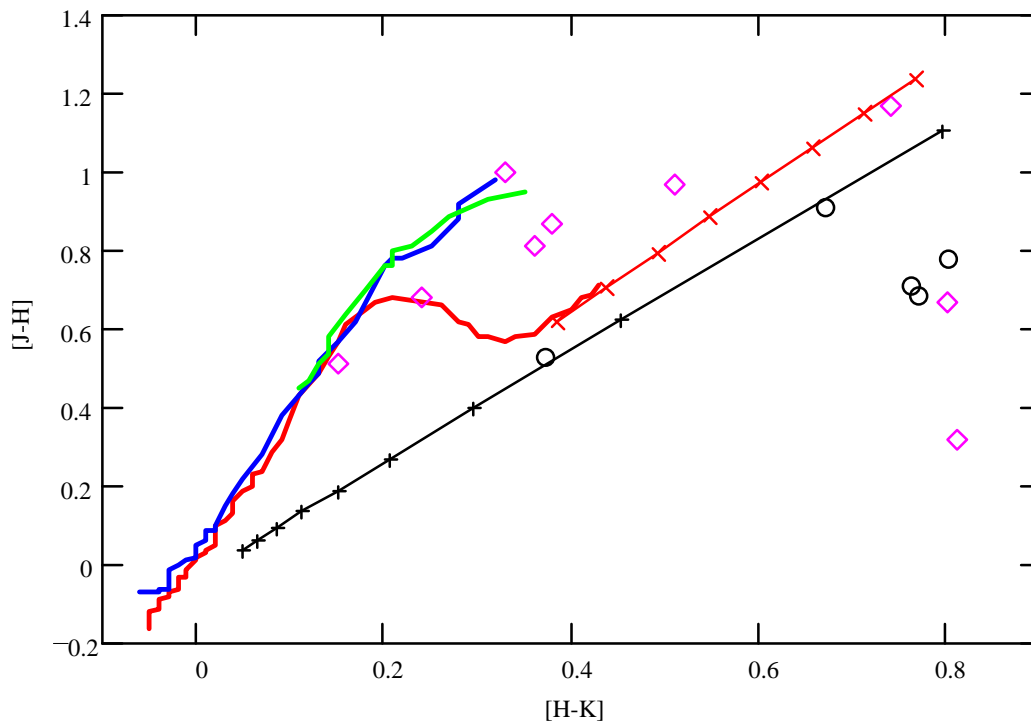


Figure 3.23 - Color-color plot showing the [J-H] versus [H-K] colors for various components in NGC 7023. The diamonds and circles are the stellar sources and nebular positions, respectively. The heavy curves on the left side of the plot are the colors for ZAMS dwarfs, giants and supergiants respectively. The straight line with crosses is the reddening vector for the interstellar extinction curve of Draine (1989); the crosses are spaced such that adjacent crosses indicate $A_V=1$. The straight line with pluses corresponds to Planckian emission where the pluses run from 2000 to 10000K in 1000K increments (2000K is at the upper right).

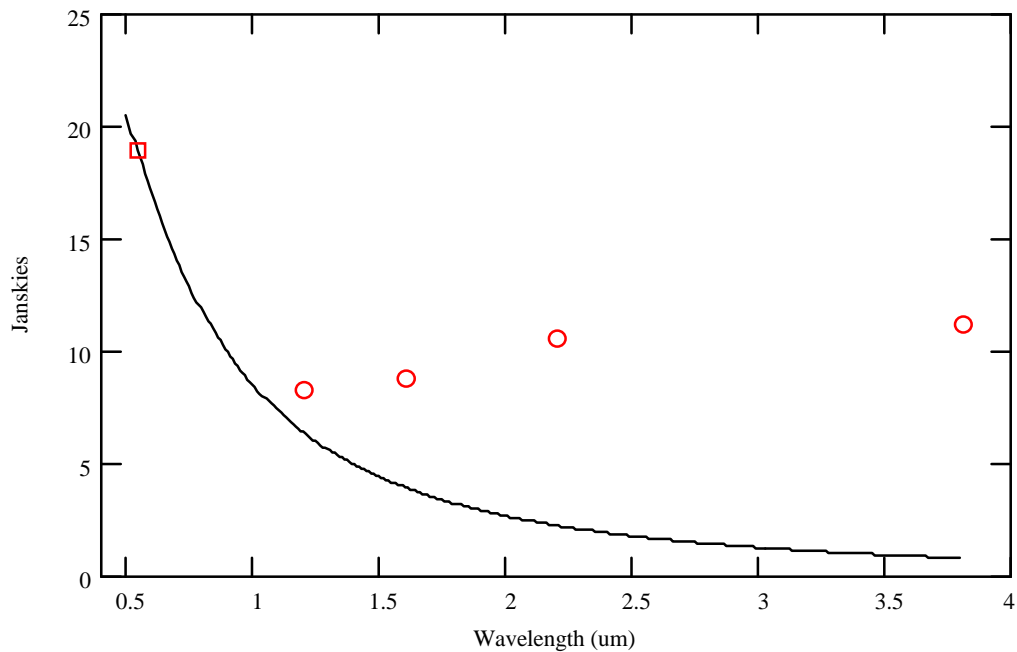


Figure 3.24 - Spectrum of HD200775. A 17000K blackbody curve has been fit to the V band observation (square, Altamore, *et al.*, 1980), and is shown by the line on the plot. The broadband observations (Sellgren, 1984) are plotted as circles. All the observations have been corrected for $A_V=1.5$.

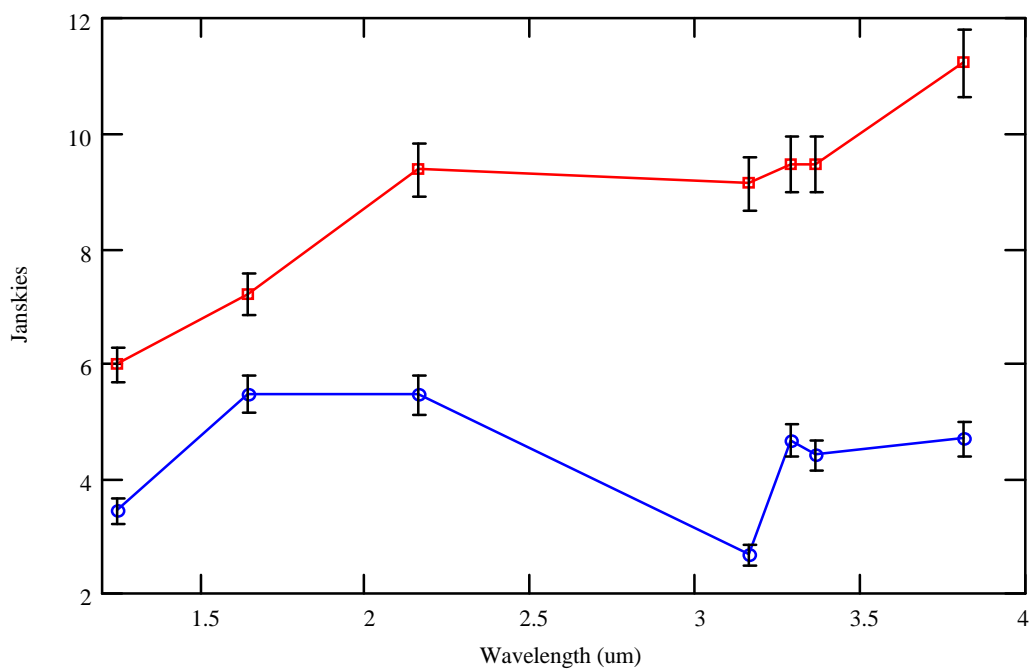


Figure 3.25 - The spectra of HD200775 and star D (scaled by 70x), shown with squares and circles, respectively. The short wavelength data for HD200775 (all the points shortward of $2.5\mu\text{m}$) are from Sellgren (Sellgren, Werner and Dinerstein, 1983, Sellgren *et al.*, 1985) and the long wavelength data are from this work.

3.6.1 Star D

In the following section we discuss and determine the nature of star D. Repeating the Hubble law analysis found in section 3.4, we find that the maximum possible reflected contribution from star D to the broadband emission in the filaments is less than 10% of their observed surface brightness, consistent with HD200775 being the primary source of radiative excitation in the filaments. There is strong photometric evidence that star D is variable, per Table 3.10. Variability is a common feature among both protostellar objects and evolved stars. The latter are very bright objects; a typical late type red giant ($10^3 L_{\odot}$ and $T_{\text{eff}} = 3000\text{K}$) has an absolute K band magnitude of -4.1 (Allen, 1976, Korneef, 1983) and would have to be over 6 kpc distant in order to appear at the magnitude we observe given the extinction measured to the region. This distance would place it well out of the galactic plane. The variability alone strongly suggests that star D is a member of the NGC 7023 cluster, and that star D is in an active phase of its life.

Year	Beam	[J]	[H]	[K]	Reference
1979	11"	10.64 ± 0.09	9.80 ± 0.05	9.36 ± 0.06	(1)
1991	10.5"	11.27 ± 0.15	10.36 ± 0.15	9.83 ± 0.15	(2)
1992	11"	10.99 ± 0.05	10.10 ± 0.05	9.62 ± 0.50	(3)

Table 3.10 - Broadband magnitudes for star D over a 13 year baseline. References are as follows: (1) Sellgren, 1983, (2) Piché, 1992, (3) this work.

We observe weak, but significant $3.29\mu\text{m}$ line emission associated with star D, evidence for UV emission from the very near vicinity of the star. The $3.29\mu\text{m}$ emission feature is observationally associated with soft UV radiation fields, and has not been associated with stars later than approximately type F. Its presence rules out the possibility that star D is either a late type main sequence object, or a

red giant star since neither of these stellar types produce a hard enough radiation field to excite the feature carriers. Using the 3158 μm continuum observation, we find 28 mJy worth of peak line emission from star D, and a peak line to continuum ratio of 0.7. The spectrum shows evidence for the 3.4 μm emission feature as well. There are no such features in the spectrum of HD200775 (see Figure 3.15).

Figure 3.16 shows the 3.29 μm line to continuum cut from HD200775 to star D (on the far right hand side of the plot). The decrease in the l:c(3.29 μm) as we approach star d indicates that star d is affecting the dust in its very near vicinity. Figure 3.12 clearly shows that the reason for this decline is that the continuum flux increases as we near star D, while the 3.29 μm flux remains roughly constant. The two possible sources for this increase are thermal emission from dust heated by star D's radiation field or light scattered from star D. The 3.29 μm line to continuum ratio around star D is too small and the errors associated with the 3.158 μm observation are too large to warrant further, detailed modelling; future observations should be performed in order to confirm the magnitude of the emission feature(s) and the continuum associated with star D.

The preceding data and discussion clearly indicate that there is a dusty nebula surrounding star D. We consider the possibility that HD200775 is responsible for exciting the emission in the immediate vicinity of star D. In the following analysis, we neglect any extinction between HD200775 and star D, which will yield the maximum contribution HD200775 can have to the nebular surface brightness around star D. In order for stars D and HD200775 to contribute equal fluxes at a given distance from star D, the following equality must be met:

$$\begin{aligned} \frac{L_{\text{HD200775}}}{4\pi \cdot \theta_{\text{HD} \rightarrow \text{D}}^2} &= \frac{L_{\text{star D}}}{4\pi \cdot \theta_{\text{star D}}^2} \\ \rightarrow \frac{L_{\text{star D}}}{L_{\text{HD200775}}} &= 2.2 \times 10^{-4} \end{aligned} \quad (3.12)$$

where the L's are the luminosities of the respective stars, $\theta_{\text{HD} \rightarrow \text{D}}$ is the angular distance between HD200775 and star D (50'') and $\theta_{\text{star D}}$ is the angular half width of the nebula surrounding star D

(0.75"). If the ratio given in equation (3.12) is larger than 2.2×10^{-4} , then star D will contribute more than half of the excitation energy to its local nebula. Assuming that the two stars experience similar line of sight extinctions, we use their respective V magnitudes (Piché, 1992, Altamore *et al.*, 1980) to estimate the luminosity ratio:

$$\begin{aligned} [V_{\text{star D}}] - [V_{\text{HD200775}}] &= 8.0 \\ \Rightarrow \frac{L_{\text{star D}}}{L_{\text{HD200775}}} &= 6.3 \times 10^{-4} \end{aligned} \tag{3.13}$$

Equation (3.13) indicates that star D contributes approximately 75% of the V band flux to its nebula. The V magnitudes also determine that the contributions from the two stars will be equal 1.3" from star D. In addition, we note that the surface brightness of Star D's nebula in both the broadband filters as well as the CVF frames is considerably higher than elsewhere in the surrounding nebula, which is additional evidence for UV emission from the very near vicinity of star D.

3.7 Conclusion

Infrared imaging arrays provide important spatial information not available from single element detectors scanned over an extended astronomical object. The reflection nebula NGC 7023 has complex spatial structure that lends itself well to the imaging observations presented and discussed here. NGC 7023 is a copious emitter of both continuum infrared radiation as well as the unidentified infrared dust emission features. The dust composition has been shown to vary widely across the face of the nebula as evidenced by the varying spectra at different locations, the line to continuum ratio, as well as the strength of the $3.29\mu\text{m}$ and $3.4\mu\text{m}$ emission features.

The most dramatic and striking feature in the images presented here are the bright filaments. While they are brighter than the surrounding, diffuse emission, they are at the same time redder than the ambient nebula. In addition, the filamentary interiors are significantly redder than their edges. We take these observations as evidence that the filaments are self shielding, *i.e.* that the envelopes of the filaments absorb and scatter radiation in such a fashion as to insulate their interior. There are two additional reasons for believing that the filaments are self shielding: 1) the width of the western filament in the $3.29\mu\text{m}$ emission feature is spatially more narrow than in the broadband frames, it is tucked within the broadband filamentary emission, and 2) the $I_c(3.29\mu\text{m})$ is extremely high inside filaments. The filaments show asymmetric profiles across their narrow dimension, suggestive of an expanding medium interacting with the ambient molecular cloud. There are several objects within the nebula that are in active phases of their lives (HD200775 has a well known stellar wind), and there is a bipolar outflow associated with NGC 7023, so it is not surprising to find evidence for a shock front. The $3.4\mu\text{m}$ emission feature is also stronger in the filaments than anywhere else in the nebula, the inferred $3.4\mu\text{m}$ emission feature reaches peak strengths of one half that of the $3.29\mu\text{m}$ feature in the brightest (and coldest) regions of the filaments.

We have examined the spatial distribution of the $3.29\mu\text{m}$ and the $3.4\mu\text{m}$ emission feature in NGC7023 and found both features to vary considerably over the central $1'$ of the nebula. Broadly

speaking, there are two components to the $3.29\mu\text{m}$ emission features; coherent spatial structures and diffuse, extended emission. Both features are strong and variable across the nebula; both the line to continuum ratio and the emission strength vary by more than an order of magnitude. Our observations are inconsistent with an intrinsic line to continuum ratio associated with the $3.29\mu\text{m}$ emission feature; the observed line to continuum ratio ranges from 6 to 70 over the face of the nebula. We have inferred the presence of the $3.4\mu\text{m}$ feature from our long wavelength CVF observation at $3.367\mu\text{m}$ and calculated the emission spectrum in the filaments, the location 30W20N as well as for the $1' \times 1'$ around the filaments. We have compared our $3.4\mu\text{m}$ feature data with a model from the literature that predicts the $3.29\mu\text{m}$ to $3.4\mu\text{m}$ emission feature strengths as a function of UV field and found disagreement with the model. We suggest that internal nebular extinction is the reason for the discrepancy.

Our broadband observations indicate, in agreement with other observers, constant infrared colors and temperatures in the 1500K range across the central $1'$ of the nebula; these observations cannot be explained by equilibrium dust emission. A simple Hubble law analysis shows that scattered light is not sufficient to account for this surface brightness. We conclude, in agreement with others, that there is a sizable contribution to the near infrared brightness of NGC 7023 due to transiently heated small grains.

The ratio of the far infrared luminosity to the $3.29\mu\text{m}$ emission feature luminosity in NGC 7023 appears to be different from that observed in IR bright galaxies. It is lower by a factor of 5 unless there is no additional $3.29\mu\text{m}$ emission outside the $1'$ area we have observed. If we have observed all the $3.29\mu\text{m}$ feature emission in the nebula, then the observed correlation in galaxies applies, in which case one of our suggested constraints on this ratio is sensible; that the FIR and $3.29\mu\text{m}$ emitting species do not have to be coextant as long as the FIR emitting dust is optically thick to the radiation exciting the $3.29\mu\text{m}$ feature.

HD200775, the central star of NGC7023, appears to be the dominant radiative excitation source in the nebula (including the filaments) but there is likely an additional source of mechanical energy input; perhaps stellar winds or the same mechanism that is driving the bipolar outflow is

responsible for the morphology. Of the seven stellar objects observed in the frame (excluding HD200775), two were found to be consistent with nearby late-type main sequence stars and the remainder were hypothesized to be protostellar objects due to variability and proximity to HD200775 which is the brightest star in a young stellar cluster. In particular, we observed 3.3 μ m and 3.4 μ m feature emission associated with star D (and none with HD200775). Star D also exhibits variability, infrared excess and colors that make it a highly likely young stellar object.

4. Charge Trapping in InSb Photovoltaic Infrared Detector Arrays

4.1 Introduction

Charge trapping in infrared detector arrays can compromise their utility and sensitivity. We have performed experiments on a particular manifestation of charge trapping, termed a latent image, that appears in low background indium antimonide astronomical detector arrays operated at cryogenic temperatures. Substantial evidence points to the detector surface as the site where the charge trapping associated with the latent image phenomenon occurs. We investigate the physics of surface states both experimentally and theoretically in order to gain an understanding of the surface contribution to the observed behavior, and present a model to explain the experimental results. Identification of these trapping centers should enable a reduction in their number density by appropriate changes in the material preparation and device fabrication. To this end, we present a plausible molecular candidate as being responsible for the latent images. Surface calculations, in the form of a solution to the Poisson equation at the surface of the InSb photodiode, are performed in order to verify some of the key features of the model.

4.2 Indium Antimonide Infrared Detector Arrays

Indium Antimonide (InSb) is a direct bandgap semiconductor that is conventionally used to detect infrared radiation. The bandgap of InSb is a function of temperature, and is 235meV at 10K, which corresponds to a cutoff wavelength of 5.28 μm ($E_{\text{gap}}=hc/\lambda$). InSb detectors are generally operated in the photovoltaic mode as photodiodes. The fundamental device structure of the detector is a p-n junction, formed by ion implanting p-type⁹ dopants such as beryllium or boron into an n-type InSb substrate, thereby creating a p-type semiconductor in contact with an n-type semiconductor. The physics of such planar devices is well understood and we refer the reader to any one of a plethora of well-written texts on the subject (Grove, 1967, Sze, 1981). For the astronomical detectors studied here, substrate and implant dopings are on the order of 10^{14} cm^{-3} and 10^{16} cm^{-3} respectively.

A typical detector geometry is shown in Figure 4.1. In order to maximize well depth¹⁰ and increase the detector resistance, the p-n junctions are operated under reverse bias conditions, which means that V_{bias} is negative relative to V_{detsub} in the figure. The IR photons are incident on the substrate, shown at the bottom of the diagram (termed the backside). Assuming that the photon energy is larger than the InSb bandgap, these photons are absorbed in the n-type substrate where they generate an electron-hole pair via photoionization of a bound electron into the conduction band. The photogenerated holes and electrons diffuse randomly through the substrate; holes that reach the edge of the depletion region are preferentially swept by the electric field in the depletion region into the p-type region and thereby counted as signal charge. The photogenerated electrons are repelled by the field at the edge of the depletion region and are either collected at the electrical contact to the substrate

⁹ The semiconductor type (n-type or p-type) refers to the flavor of the dominant charge carrier, *i.e.* electrons are referred to as the majority carriers in n-type material and holes are the majority carriers in p-type material.

¹⁰ Well depth is the total amount of charge that can be integrated in a pixel. The well depth is given by $Q=CV$ where C (for the configuration discussed here) is the detector, gate and stray capacitances at the maximum reverse bias V (see Figure 4.6). Thus, higher detector biases result in more charge capacity.

(V_{detsub} in Figure 4.1), or recombine in the bulk of the semiconductor. For astronomical purposes, InSb photovoltaic detectors are physically thinned to approximately $7\mu\text{m}$ (indicated on the left side of Figure 4.1) in order to increase their quantum efficiency¹¹. This assures that the thickness of the detectors is smaller than the minority carrier diffusion length¹². Were the detectors not thinned, the photogenerated holes would recombine in the bulk before being collected by the junction, resulting in a loss of signal charge, and hence a lower quantum efficiency. The minority carrier diffusion length in low doped, high purity n-type InSb ranges from at $\sim 25\mu\text{m}$ at 77K to $\sim 12\mu\text{m}$ at 20K.

The detector array is connected, via indium bumps, to a silicon “readout.” The readout is a CMOS integrated circuit (ROIC) that essentially performs three functions: (1) providing detector array stimuli (clocks and biases) necessary to operate the device, (2) interfacing the detectors electrically to the outside world via a chain of signal processing electronic circuit elements that typically contain source followers, amplifier/drivers, sample and hold, and clamp circuits, and (3) multiplexing circuitry to couple the $N \times M$ detectors (each detector is referred to as a pixel) onto a much smaller number of output lines. We will alternately refer to this composite device as a sensor chip assembly (SCA), a detector array, or simply an array.

The quantum efficiency and noise of an SCA ultimately limit the detection sensitivity. In addition to the photon noise inherent in the incident background, the dominant sources of noise in an InSb photodiode focal plane array stem from both the detector and the ROIC. The detector itself is a

¹¹ Quantum efficiency is the efficiency with which a detector converts photons into electrons, i.e. $QE = N_e / N_\gamma$, where N_e is the number of measured electrons, and N_γ is the number of incident photons.

¹² Diffusion length is the characteristic length over which the minority carrier density diffuses prior to recombining. It is given by $L = \sqrt{D\tau}$, where D and τ are the minority carrier diffusivity and lifetime (in this instance, holes in an n-type material). The diffusivity is given by $D = \mu kT / q$ where μ is the minority carrier mobility, k is the Boltzmann constant, T is the temperature and q is the electronic charge.

source of dark current¹³ shot noise, Johnson noise, photocurrent shot noise, and 1/f noise. Each of the various circuit elements that comprise the ROIC add noise to the signal as it propagates through the chain. SCAs are operated at cryogenic temperatures in order to reduce the detector dark current and associated noise. Typically doped InSb diodes perform poorly at room temperature due to large thermally generated diffusion currents in the bulk of the device. The reverse bias resistance of a photodiode (an often quoted figure of merit) is limited by generation-recombination (GR) mechanisms in the depletion region at 77K, which is a bulk material property, and by tunneling mechanisms at colder temperatures, which has both a surface and a bulk component.

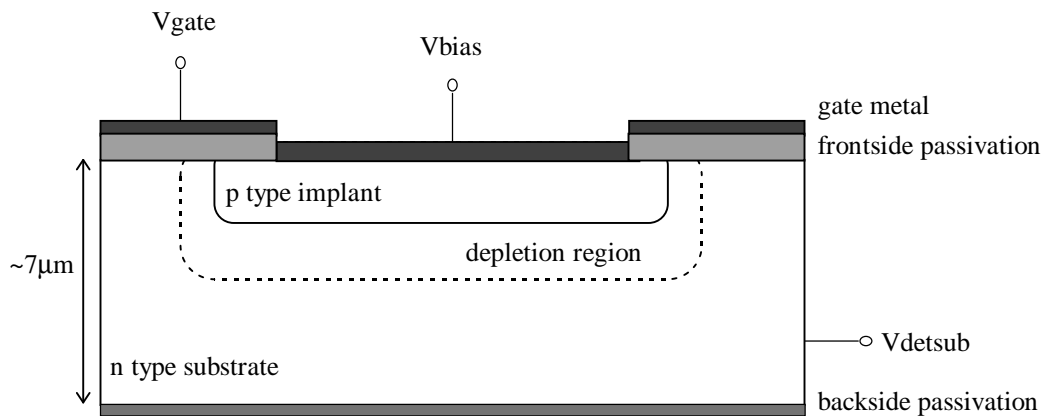


Figure 4.1 - Typical detector geometry with electrical connections indicated. The photons are incident from the bottom of the diagram (referred to as the backside). The detector is biased across V_{bias} and V_{detsub} while V_{gate} is used to control the potential of the surface.

¹³ Dark current is the detector (reverse bias) current in the absence of any photogenerated charge, *i.e.* in the dark.

4.3 Surface Passivation

There are a large number of localized energy states (also known as generation-recombination centers), both discrete and quasi-continuous, at the surface of a semiconductor due to the abrupt discontinuity of the crystal structure there. The surface atoms have anisotropic, dangling bonds that produce surface states that can lie within the forbidden energy gap. The existence of these states affects the statistical mechanics of the charge transport by permitting surface conduction. They can also lead to an increase in the recombination and/or generation rate at the surface and can trap mobile charge at the surface of the semiconductor. Overall, the presence of electrically active surface states can compromise the operational qualities of an imaging array. For these reasons, the surfaces of a semiconductor device are “passivated,” or chemically bonded to a material whose primary purpose is to ameliorate these effects.

To virtually eliminate surface effects in an imaging array requires reducing the surface/interface state density to an extremely low value ($<10^9 \text{ cm}^{-2} \text{ eV}^{-1}$) via passivation. However, present passivation techniques can only obtain densities on the order of $10^{10} \text{ cm}^{-2} \text{ eV}^{-1}$, which forces detector manufacturers to use a metal gate (typically aluminum) over the passivation to provide direct control of the surface potential (see Figure 4.1). By applying the proper voltage to this gate (termed the gate voltage, or V_{gate}), optimal performance can be achieved. Unfortunately, it is difficult to fabricate gated infrared detector arrays larger than 3–4mm square due to processing yield issues; for reference, a 256×256 array with 30 μm pixels is 7.7mm on a side. Thus, for production of large area staring arrays, the passivation alone must produce the proper surface conditions.

The purpose of a surface passivation is actually twofold: (1) to negate or minimize surface effects that can compromise the electrical properties of the detector, and (2) to protect or encapsulate the surface so that environmental effects (such as the adsorption of water) will be minimized. In addition, the passivant must also be thermally compatible with the detector material, exhibit excellent adhesion and be mechanically stable in order to permit indium bump bonding to the silicon ROIC. The

quality of the semiconductor/passivating layer interface and the dielectric properties of the passivating layer play crucial roles in determining the performance of InSb diode arrays. Unfortunately, as a result of both the processing chemistry and the material properties of the passivant, the passivation may also introduce undesirable charge that can affect the state of the surface as well as the electrical properties of the device.

Figure 4.2 illustrates the various components associated with a passivated semiconductor surface. The interface traps are dependent on the chemical composition of the interface and result from the InSb/passivant interface properties; the energies of these traps are found inside the forbidden energy gap of InSb, and exhibit a continuous density distribution with maxima near the band edges and a minimum near the middle of the gap (Kreutz *et al.*, 1977). In silicon, the fixed passivation charge is found within 30Å of the interface and cannot be changed over a wide range of surface potential (Sze, 1981). It is this constituent of the passivation that is utilized to adjust the semiconductor surface to the desired state¹⁴ for optimal device performance. The passivation trapped charge is associated with defects and/or impurities in the passivant, and can be charged and discharged by injecting carriers from the semiconductor or by direct creation of carriers inside the passivation. The mobile ionic charges move within the oxide under high temperature or high electric fields, neither of which are relevant for the operation of the devices under discussion here.

¹⁴ The charge state of a semiconductor surface falls into one of four (continuous) states: inverted, depleted, flatband and accumulated. An inverted surface is one where the carrier density at the surface is of the opposite type, and larger than the majority carrier density in the bulk of the material. A depleted surface is one where there are no free carriers; the surface charge, due to ionized impurities, is spatially fixed. In the flatband state, the surface has the same majority charge density state as that of the semiconductor bulk. An accumulated surface is one where the carrier density at the surface is of the same sign, and larger than the carrier density in the bulk of the material.

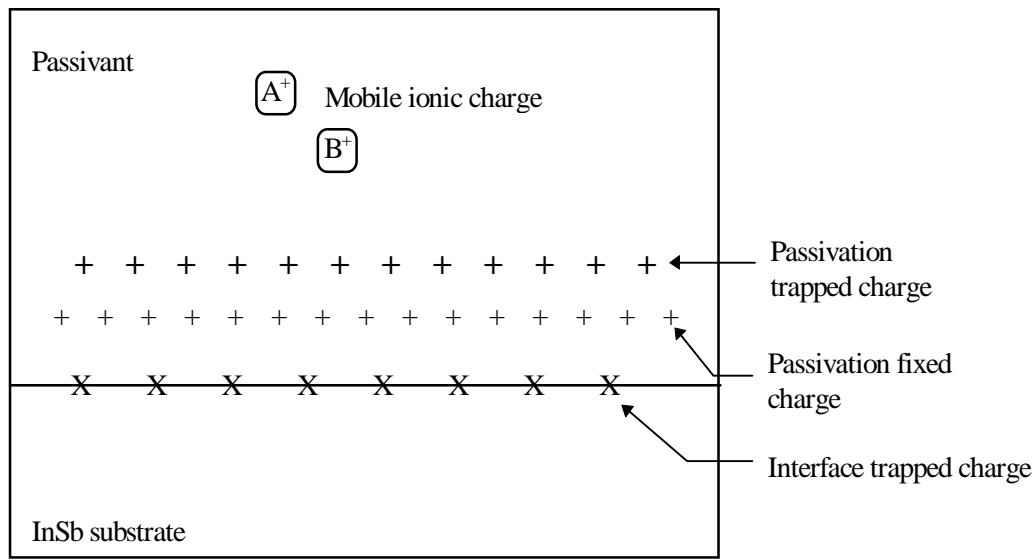


Figure 4.2 - Schematic of the InSb/passivant system (after Sze, 1982). Both the signs and the spatial locations of the various passivant charges are for illustrative purposes only, and do not represent the actual InSb passivation configuration.

The zincblende crystal structure of InSb results in two possible surface “faces,” an indium rich face (termed the A face), and an antimony rich face (termed the B face). Passivating both faces of InSb photodiodes is important because both surfaces are present and are well within a diffusion length of the p-n junction, and hence will have an effect on quantum efficiency, crosstalk, dark current and noise. The backside surface (upon which the photons are incident, see Figure 4.1) is typically the A face, and it is passivated to ensure that photo-generated holes are not lost to recombination at this surface.

A substantial flux Φ of photocarriers can be generated at an arbitrary distance x close to the backside according to

$$\Phi(x) = \Phi_0 e^{-\alpha x} \quad (4.1)$$

where Φ_0 is the incident photon flux and α is the absorption coefficient for InSb, approximated by the expression

$$\alpha(\lambda, T) = 2.6 \times 10^4 \left[\frac{hc}{\lambda} - E_g(T) \right]^{0.65} \cdot \text{cm}^{-1} \quad (4.2)$$

In this expression, h is the Planck constant, c is the speed of light, λ is the wavelength, E_g is the energy gap (in eV) and T is the temperature (Joe Rosbeck, Private communication, 1991). Figure 4.3 plots $e^{-\alpha x}$ as a function of the depth (x) into the substrate from the photo-incident side, from which we observe that 90% of the photocarriers generated by $2\mu\text{m}$ photons are generated within $1.6\mu\text{m}$ of the back surface. Thus, reducing the carrier recombination (or surface recombination velocity, SRV¹⁵) at this surface is crucial for high quantum efficiency. At longer wavelengths the backside plays a lesser role because the photocurrent is generated deeper into the substrate, *i.e.* long wavelength photons penetrate further into the InSb than do shorter wavelength photons. Only 22% of the photocarriers generated by $5\mu\text{m}$ photons are generated within $1.6\mu\text{m}$ of the back surface (see Figure 4.3). A passivation process that maintains the back surface in an accumulated state will help to accomplish the goal of reducing the SRV (for an n-type substrate). The resultant band-bending and attendant electric field will repel the diffusive flow of minority carriers from the recombination sites at the illuminated surface, thereby reducing the number of carriers that interact with the recombination/trapping centers, thereby reducing the backside surface recombination.

¹⁵ Surface recombination velocity is a measure of carrier recombination at localized surface defects. The surface recombination velocity is defined as $\text{SRV} = v_{\text{th}} \sigma N$, where v_{th} is the thermal velocity of the carriers, σ is the recombination cross section, and N is the areal density of the centers.

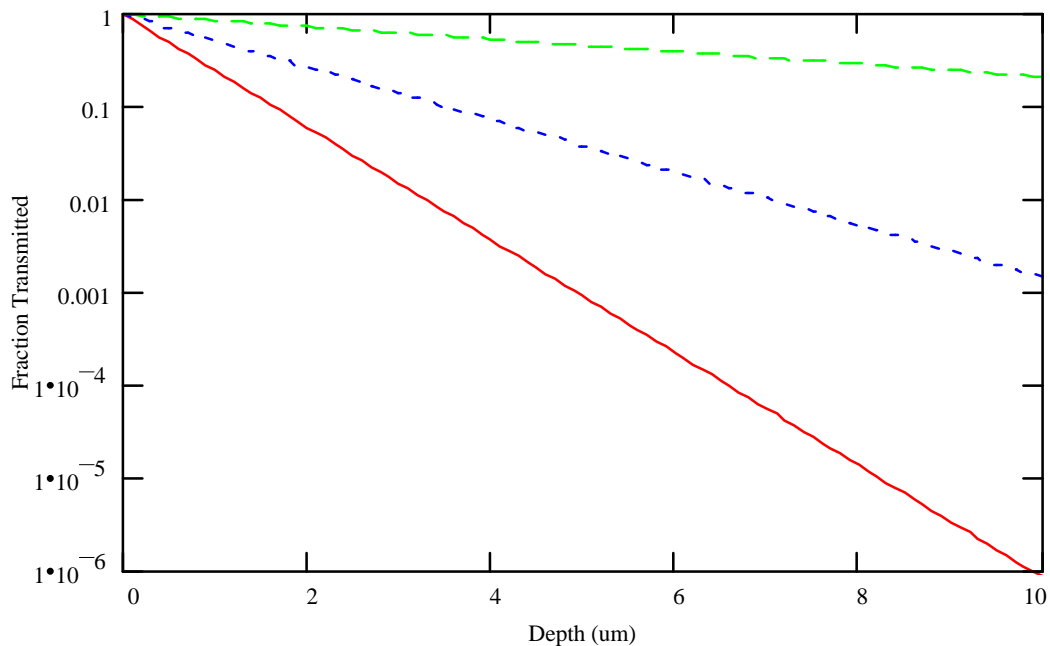


Figure 4.3 - The fraction of the incident photoflux transmitted into the bulk of the InSb as a function of depth from the photo-incident side at 10K according to equations (4.1) and (4.2). The solid line is the fraction transmitted at an incident wavelength of $2\mu\text{m}$, the dotted line is that at $3\mu\text{m}$, and the dashed line at $5\mu\text{m}$. Note that this plot does not include surface reflections.

Passivation of the front surface is more critical because surfaces of both type are present in planar devices. It is difficult to optimize because it must serve multiple purposes:

- 1) reduce SRV between the implant regions (requires an accumulated surface, similar to the backside passivation),
- 2) minimize charge trapping near the periphery of the p^+ diode implants (requires an accumulated surface, similar to the backside passivation),
- 3) minimize surface (tunneling) leakage currents near the periphery of the p^+ diode implants (requires a mildly depleted surface),
- 4) maximize the reverse bias well depth/charge capacity, and
- 5) reduce crosstalk between the diode implants (requires a depleted surface).

These requirements are not independent of one another. Reducing the SRV will also lead to a reduction in charge trapping if the trapped carriers diffuse or drift from the bulk to the trapping site (as opposed to being generated at the trapping site). Requirements 3 and 4 are also related in that tunneling breakdown mechanisms, which lead to large dark currents, place a limit on the maximum

reverse bias that can be applied to the detector. As will be discussed, the Si_3N_4 passivation is superior to SiO_x (where $1 < x < 2$) vis à vis most of the five parameters listed above.

Broadly speaking, there are two types of passivants: native oxides, which are typically grown directly on the semiconductor substrate, and deposited passivants. Both the SiO_x/InSb and the anodic oxide/ InSb interfaces are not abrupt transitions; the width of the transition region in both cases is on the order of 100\AA (Valcheva and Germanova, 1991). ESCA¹⁶ & Auger analysis show that both types of interfaces contain varying amounts of oxides of both indium and antimony, and in some cases, elemental In and/or Sb, as well as SiO_x (Baumgratz, 1988). Both oxides have been investigated in the literature, and we briefly summarize these results here.

Anodic oxides are grown on InSb crystals by passing a current through the semiconductor (the InSb is the anode) while it is suspended in an electrolyte with an oxygen bearing component. The anodic oxide grows as the electronic current flows through the oxide under an applied electric field (Nicollian and Brews, 1982). The resultant oxide consists predominantly of In_2O_3 , Sb_2O_3 , and Sb_2O_5 , with some reports of elemental Sb (Mattausch and Aspnes, 1981, Rastogi and Reddy, 1995). A wide variety of measurements have observed that anodized InSb surfaces tend to be n-type for both p-type and n-type substrates (Axt & Rogers, 1970). This suggests that these oxides trap positive charge, thereby producing an accumulated surface for an n-type substrate (which is desirable at the backside in order to repel the minority carriers as discussed previously) and a depleted or inverted surface in the case of a p-type substrate.

Deposited oxides have, in general, proven to be superior to native oxides for passivating InSb . This is particularly true for the front side, where the requirements of the passivation are more exacting than that of the back side. The literature is filled with inconsistent conclusions regarding how best to optimize this surface; for example, some authors claim that preservation of a native oxide between the InSb and the deposited SiO_x is crucial for low surface state densities (Langan and Viswanathan, 1979),

while others conclude that lower surface state densities are obtained by removing the native oxide prior to depositing the SiO₂ (Okamura and Minakata, 1985). There is also considerable evidence for charge trapping effects specifically associated with SiO_x and SiO₂ passivations across a wide variety of passivation processing methodologies, device structures, starting materials and measurement techniques.

Pagnia (1968) investigated light induced trapping in p-type InSb via current vs. time observations after illumination with photons of varying energies. He observed 3 separate decay modes (at 85K): 1) a fast, intrinsic response with a ~230meV excitation (the bandgap energy), 2) a slow relaxation (on the order of minutes) requiring ~600meV excitation, and 3) a weak, quasi-stable (frozen-in) response with a 2eV excitation. He interpreted his results as being consistent with a model whereby electrons are trapped in the native oxide after crossing the InSb/oxide barrier. The 0.6eV photons excite electrons from the valence band maximum (VBM) into the oxide traps while 2eV photons induce transitions to the conduction band of the oxide whereupon some of these electrons are trapped in the native oxide (at a higher energy than the 600meV traps).

Davis (1964) found two discrete surface states in InSb; one near the VBM and the other near the conduction band minimum (CBM) at 77K. Huff *et al.* (1966, 1968) found that the same states were always present over a wide range of crystal orientations and faces as well as in both n and p-type material. Furthermore, the surfaces in all cases were found to be p-type, and the degree to which the surface was p-type was directly related to the density of surface states near the VBM. This leads to the conclusion that these surface states are acceptor-like. Sewell and Anderson (1975) performed measurements on the decay of channel conductivity in thin-film InSb field effect transistors passivated with SiO_x. They noted that the measured conductivity decay was non-exponential, although we have fit their data with multiple time constants (*i.e.* it is exponential), and propose that their observations are consistent with electrons tunneling into interfacial traps located within ~20Å of the interface (see also

¹⁶ ESCA, or Electron Spectroscopy for Chemical Analysis, is a surface analysis technique involving the measurement of the energy distribution of electrons ejected from a sample under monoenergetic x-ray

discussion in section 4.7 and Figure 4.7. They hypothesized that the traps are associated with the stoichiometry of the insulator (rather than being impurity related) by varying the insulator preparation parameters, but their experiments were not conclusive in this regard. Barthe and Lile (1993) performed capacitance-voltage measurements on metal-insulator-semiconductor structures, in this case SiO_2 deposited on both n and p-type InSb. They observed that the p-type material exhibited larger hysteresis than n-type, that most of the hysteresis was associated with the negative voltage portion of the CV measurement and that the magnitude of the hysteresis increased with voltage swing. They concluded that their observations are consistent with hole traps in the oxide with energies close to the VBM.

bombardment.

4.4 Latent Images & The Physics of Trapped Charge

A latent image is an image that persists on the array after the illuminating source has been removed from the field of view. The brightness of a latent image is observed to decay with time, and its initial magnitude is proportional to the initial incident photon flux until a certain flux is reached, at which point the magnitude in the latent image remains constant (we call this saturation). Image latency presents substantial difficulties to both ground and space-based astronomy; consider slewing a telescope from a bright star to a faint galaxy and then observing the latent image of the bright star interfering with the galaxian image for as long as several minutes; or having a bright object cross the field of view while slewing to another faint galaxy and then observing a latent streak across the galaxian image; or worst of all, a very poor imaging duty cycle requiring many throwaway frames to remove the latent image from the frame. We have performed latent image experiments and analyses on both gate-controlled arrays (SiO_x surface passivation) as well as gateless arrays (Si_3N_4 surface passivation).

Charge trapped at the interface and/or within the passivant can manifest itself as image latency in two different ways, both of which may be occurring simultaneously. Throughout the remaining discussions, we hypothesize that photo-generated holes are trapped in the front side passivant as a result of illumination. In addition, we will be making frequent reference to two different depletion regions, the distinction between the two is critical to understanding the proposed model; the field-induced junction refers to the depletion region created at the surface of the n-type InSb by the applied gate bias, whereas the metallurgical junction refers to the p-n junction created by ion implantation of a p^+ region in an n-type substrate (*i.e.* the detector).

One plausible mechanism for explaining the latent image phenomenon is that the trapped holes are re-emitted from the passivant into the bulk InSb, whereupon they diffuse laterally to the junction and are collected as signal charge. If the InSb surface is in a mildly inverted or depleted state, then the release of trapped carriers may also be collected as signal charge via drift across a surface channel into

the p^+ implant, assuming that the electric field at the surface has a component parallel to the surface. Another possible collection mode is that the released charge travels via a conductive path in the interface region to the implant (p^+ region), where again they would be counted as signal charge. In both of these cases, it is possible that all the detrapped charge is not collected subsequent to removing the illuminating source. The collection efficiency would be determined by the surface conditions and recombination and scattering mechanisms in play. A simpler mechanism for collecting the trapped charge invokes the fact that the detector bias changes (debiases) as it integrates photocurrent. As the detectors integrate charge, the width of the metallurgical depletion region will diminish, exposing more of the surface to the bulk InSb until the detector reaches its zero bias width. This change amounts to approximately $0.5\mu\text{m}$ for the detector parameters of FPA84 at 150mV reverse bias, over all the temperatures used in our experiments. This area, once the SCA has been shuttered, will find itself within the depletion region after the detectors have been reset, thereby providing a simple path for the detrapped holes to enter the p^+ region and be counted as (latent) signal charge. The immediate perimeter of the junction also has access to the depletion region and still sees part of the detector bias. Further experiments to determine the trapped charge collection mechanism should be undertaken.

An alternative mechanism for producing the observed latency is more indirect. The positive trapped charge in the passivant pushes the InSb surface in the direction of accumulation (see Figure 4.4). If the initial charge state of the passivation (prior to illumination) is such that the InSb surface is either depleted or mildly inverted, the induced surface depletion region (field induced junction) will become narrower. At temperatures in the 77K range this will result in lower GR (dark) currents, while at lower temperatures there should be no effect as the GR currents decrease exponentially with decreasing temperature to a negligible fraction of the total reverse (dark) current, which is dominated by tunneling below approximately 60K. On the other hand, if the initial surface state is close to the flatband condition, the surface may become accumulated as holes are trapped in the surface, resulting in high tunnelling/field induced dark currents due to the metallurgical depletion region becoming pinched off at the surface. In this case, the total, integrated charge in the latent image can exceed that

of the trapped charge. The quantitative relationship between the extent of any trapped charge and the state of the InSb surface will be discussed in section 4.10. As the magnitude of the trapped charge decreases (“detraps”), the InSb surface will relax back to its original configuration and hence, the dark current will increase, decrease, or remain constant, depending on the initial configuration of the surface (and the operating temperature). Note that in all the cases discussed, the observed time constant should be proportional to the detrapping time.

The mechanism for populating the traps is most likely twofold: 1) photo-generated minority carriers diffusing from the bulk to the surface and being trapped there, and 2) direct excitation of the traps via photons that penetrate to the front surface. The latter mechanism would be more apparent at longer wavelengths and high incident fluxes (see Figure 4.3; at 10K less than 0.01% of the incident photons with $\lambda=2\mu\text{m}$ reach the front surface of a $10\mu\text{m}$ thick detector, compared with ~30% of $5\mu\text{m}$ incident photons). Thus we would expect the latent image strength to be larger at longer wavelengths for the same incident flux level. This experiment should be undertaken in the future to determine the degree to which the trapped charge is generated in the bulk as compared to the front surface, as well as to detect potential trapping contributions to the latent image from the backside.

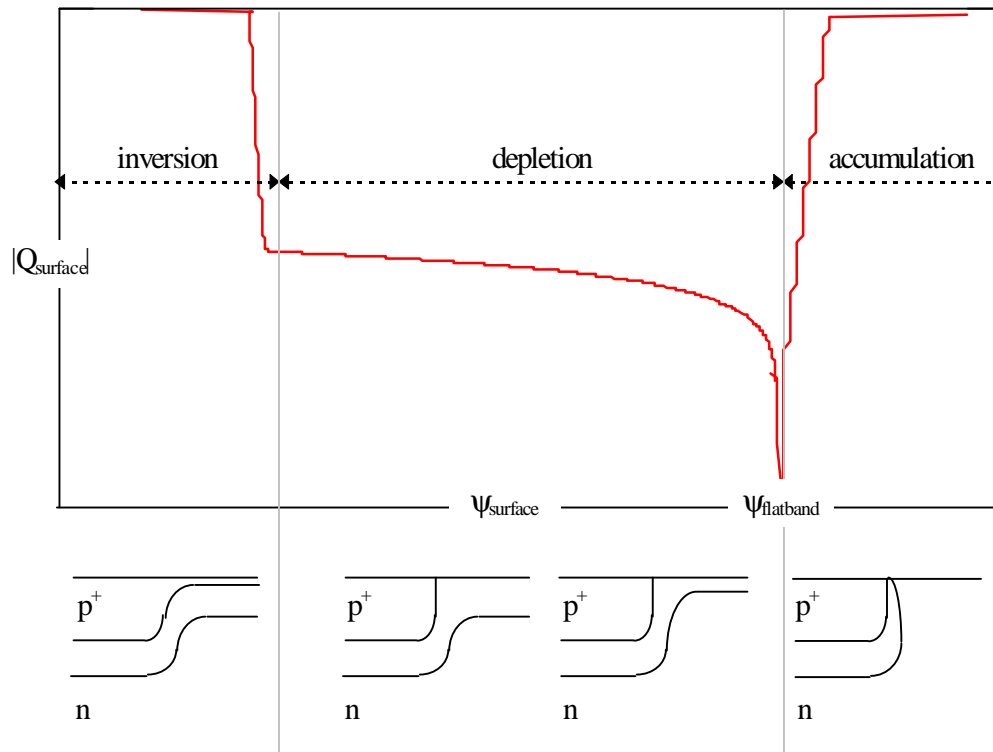


Figure 4.4 - Charge induced in the InSb surface as a function of surface potential. Note that the sign of the surface charge in the accumulated state is actually negative, and has been displayed with inverted sign here for compactness. The point where the surface charge goes to zero is the flatband condition ($\Psi_{\text{surface}}=0$). The pictures at the bottom of the graph show the state of both the metallurgical depletion region as well as the induced depletion region at the surface.

4.5 Experimental Procedure

The basic latent image experiment consisted of illuminating an array for 0.5 seconds to 15.5 minutes with a defocused miniature pinlight bulb and then inserting a 4.2K shutter (internal to the dewar) into the beam to block the light source, whereupon successive 5 second integration time images were recorded. The pinlight was powered by a DC voltage supply, and was located behind a pinhole. A cold, infrared broadband filter ($\lambda_0=2.158\mu\text{m}$ $\Delta\lambda=0.385\mu\text{m}$) was in the beam during photoexposure so that the only photons incident on the detector array were at $2.2\mu\text{m}$. Two different sampling techniques were utilized in acquiring the data, depending on the camera system in place at the time of the experiment. The array was reset between successive integrations, *i.e.* the detector bias was returned to its initial, pre-integration voltage level. Figure 4.5 shows a typical latent image decay series obtained by this method (after the photon source was removed). Latent image decays were recorded for FPA84 and ASM244 at varying incident photon flux levels at approximately constant fluence (fluence = incident photon flux \times integration time, *i.e.* total integrated signal/pixel) until the signal level of the *initial* latent image reached a constant value (which we term “saturation”). A latent image decay series was then recorded at several operating temperatures ranging between 6–26K to study the decay time constant(s’) dependence on temperature.

Figure 4.5 - Typical latent image decay grayscale. The vertical bar frame is composed of 16 distinct rectangular subframes, each separated in time by 5 seconds so that the total time from the top to the bottom of the image is 80 seconds. The top image was acquired approximately 1 second after the photon source was removed from the field of view.

The detector arrays used in this investigation are all InSb detector arrays fabricated at Santa Barbara Research Center (Goleta, CA) and they are listed in Table 4.11, along with their primary characteristics. Two types of front-surface passivations were employed in the devices used for the investigations discussed here: SiO_x and Si_3N_4 . The backside passivation for every array studied was an anodic oxide. Each array's unit cell (pixel) configuration was a source follower per detector (see Figure 4.6). Reverse bias is directly applied to the detectors by supplying a reset voltage to the diode's cathode (labeled V_{bias} in Figure 4.1 and Figure 4.6) and integrating the photocharge (as well as dark current) on the capacitances of both the detector and the source follower MOSFET gate and stray capacitances. As this charge integrates, the detector debiases (V_{bias} becomes more positive) until the diode is forward biased such that the dark current exactly balances the photocurrent. This point is known as the open-circuit voltage. All the 58×62 arrays and one of the 256×256 (FPA84) had SiO_x front surface passivations with gate controlled surfaces; the remaining 256×256 arrays were passivated with Si_3N_4 and had no gates. The metal gate and oxide passivation layer overlap the p^+ implant by approximately $2\mu\text{m}$, except for arrays ASM243 and ASM244, which were fabricated with a variety of gate configurations: different gate overlaps as well as varying gaps between the gate and implant region. The nominal operating gate voltage in all cases was chosen to both minimize the dark current and as to maintain reasonable uniformity of response over the entire SCA. When we refer to the surface being in a particular state or condition, *e.g.* inversion, we are referring to the state of the n-type substrate under the gate metal at the front surface, unless otherwise specified. Additionally, when we refer to a voltage level (*e.g.* V_{gate}), that voltage is with respect to the detector substrate (V_{detsub} in Figure 4.1 and Figure 4.6).

Array Designation	Size	Pixel Size (μm)	Doping (N_D-N_A)	Gate	Front Passivant	Thickness (μm)
ASM244	58×62	22/ 60	3E+14	yes	SiO _x	7
SCA02	58×62	60	1.7–2.3E+14	yes	SiO _x	~10
FPA84	256×256	30	3E+14	yes	SiO _x	6.2–7.8
FPA41	256×256	30	2–4E14	no	Si ₃ N ₄	10
FPA48	256×256	30	2–4E14	no	Si ₃ N ₄	10

Table 4.11 - Summary of key arrays used to study latent images. The size column refers to the number of pixels as rows×columns, doping is the net carrier concentration of the InSb substrate in cm^{-3} , front passivant is the type of front surface passivation material, and thickness is the measured thickness of the InSb detector material. ASM243 & 244 were experimental, composite arrays with both square and octagonal pixels of two different sizes ($76\mu\text{m}$ pitch), with regions of gapped gate electrodes. The $60\mu\text{m}$ pixel size is in boldface type because it is the pixel geometry used in our investigation of latent images on ASM244.

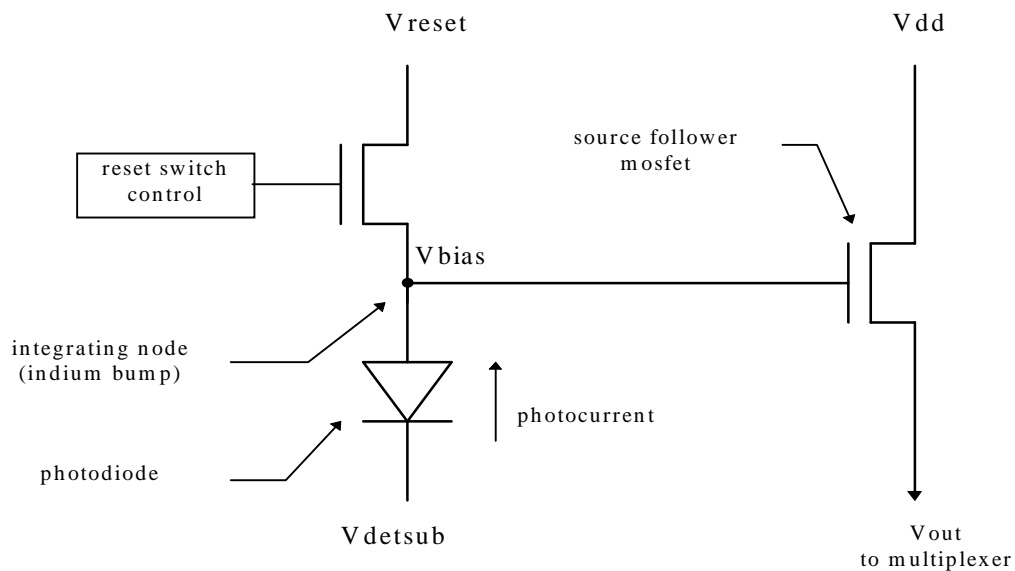


Figure 4.6 - A schematic of the circuitry associated with a single detector (photodiode). The direction of the photocurrent is indicated. The MOSFET directly above the detector is a switch that resets the detector to its maximum reverse bias at the start of the integration time (also see Figure 4.1 for a physical representation of the detector). The reset switch is open (non-conducting) while integrating photocharge on the capacitances of the detector and the source follower gate.

Both of the gated arrays (SiO_x passivation) studied had measurable latent images at detector temperatures in the range 6–26K for signal fluxes as low as $1500 \text{ e}^- \text{ sec}^{-1} \text{ pixel}^{-1}$, while neither of the

two gateless arrays (Si_3N_4 passivation) examined has shown latent images. The arrays that did not exhibit image latency were measured at 6K; FPA48 exhibited no latent images above a level of $50e^-$ in both 5 and 10 second integrations after prolonged exposures to a $2 \times 10^6 e^- \text{ sec}^{-1} \text{ pixel}^{-1}$ signal flux and FPA41 exhibited no latent images above a level of $200e^-$ in both 5 and 10 second integrations after prolonged exposures to a $3 \times 10^7 e^- \text{ sec}^{-1} \text{ pixel}^{-1}$ signal flux. We interpret this as evidence for a large reduction of front surface state charge trapping in the region of the gate oxide through the use of Si_3N_4 passivation. Our ability to change the characteristics, and in most cases, remove the latent image via cycling the gate voltage and/or forward biasing the detectors on SiO_x passivated arrays also supports this conclusion. Cycling the gate bias required setting the detectors to the unbiased condition (zero bias results when V_{bias} is set to the same potential as V_{detsub} , see Figure 4.6), followed by setting the gate voltage such that the surface was strongly accumulated and then stepping downwards in voltage until the proper gate bias was attained (mildly depleted surface condition). For the case of ASM244, this amounted to the following sequence for V_{gate} (after zero biasing the detectors): +2, +1.5, +1.0, +0.5, 0.0, -0.5, -1.0, -1.3, and variations on this general theme. Gate voltage cycling occurred rapidly (typically <2 sec) and was adjustable with a delay between steps. This methodology yielded varying results; the latent image was successfully removed, but resulted in highly non-uniform dark currents across the array for many minutes before equilibrium was attained. Additionally, in some cases after cycling V_{gate} and subsequently exposing the array to the external photon source, the latent image was much worse than it was prior to the gate cycling. We were not able to discern any correlation as to when this occurred, though we note that hysteresis is commonly associated with charge trapping, *i.e.* the history of a device determines its behavior. Forward biasing the detectors was more uniformly successful in removing the latent image, requiring approximately 300mV forward bias to completely remove the latent image for ASM244. Unfortunately, forward biasing also resulted in many minutes worth of detector instability over the entire array (*i.e.* every pixel exhibited this instability, not exclusively those exhibiting the latent image).

ASM244, the first array used in our series of image latency experiments, was studied at temperatures of 8K, 10K, 12K, and 14K. This array contained a variety of different pixel geometries (see Table 4.11). The geometry used for the latent image studies in this thesis were $60\mu\text{m}$ square implants on a $76\mu\text{m}$ pitch with a $2.5\mu\text{m}$ gate overlap. The upper limit to the dark current at 10K and a gate voltage of -1.35V was $0.21\text{ e}^-/\text{sec}$ for the $60\mu\text{m}$ pixels. The noise level was 230e^- at a temperature of 9.4K using correlated triple sampling (Fowler mode sampling was not yet implemented at the time these experiments were performed). The noise level determines the sensitivity with which we can measure the signal in a latent image, *i.e.* we cannot measure a latent image that is below the measured noise.

Latent images on the detector array FPA84 were studied at temperatures of 6K, 10K, 13K and 26K. The detectors were operated at an *applied* bias of 600mV, which corresponds to a bias across the metallurgical p-n junction of approximately 150mV (in addition to the built-in potential of the junction) at the start of the integration time (Forrest *et al.*, 1993). The operating gate voltage was -2.3V relative to the detector substrate; this condition resulted in near flatband operation (a depleted/slightly inverted surface, see discussion in section 4.10 and Figure 4.15). The low temperature ($T \leq 27\text{K}$) dark current measured for all the arrays studied here was less than $1\text{e}^-/\text{sec}$. The lowest readout noise attained was 50e^- in the correlated triple sampling mode at temperatures between $\sim 24\text{--}50\text{K}$ and below 16K. For multiply sampled read noise using Fowler sampling (Fowler & Gatley, 1991) we are able to achieve 30e^- at 6K and 20e^- at 26K, both at 8 sample pairs. No further improvement with increasing number of sample pairs was seen.

4.6 Trap Statistics Under Illumination

We begin by investigating the possibility that the trap occupation statistics are attributable to Shockley-Read-Hall (SRH) trapping centers (Hall, 1952 and Shockley & Read, 1952) near the SiO_x/InSb interface. The traps are expected to be shallow levels near to the valence band maximum or conduction band minimum, *i.e.* the band edges. Because we collect minority carriers (holes generated in the n-type substrate) at the integrating node, and the sign of the latent image indicates positive signal, we hypothesize that the traps are hole traps *i.e.* the trap energy level is close to the valence band maximum.

The fractional number of hole-filled traps, f_p , is governed by the differential equation:

$$\frac{df_p}{dt} = e_n(1-f_p) - e_p f_p + pc_p(1-f_p) - nc_n f_p \quad (4.3)$$

where e_p , e_n , c_p , c_n , n and p are the trap thermal emission coefficients for holes and electrons, the trap thermal capture coefficients for holes and electrons, and the free electron and hole densities, respectively. The coefficients are defined so that the hole capture and emission rates are given by $pc_p n_T(1-f_p)$ and $e_p n_T f_p$, respectively, where n_T is the trap volume density. The volume density of hole-filled traps is then given by $f_p n_T$. The solution to equation (4.3) is:

$$f_p(t) = A \cdot e^{-t(e_n + e_p + pc_p + nc_n)} + \frac{(e_n + pc_p)}{e_n + e_p + pc_p + nc_n} \quad (4.4)$$

where A is a constant. For the case where the array is illuminated by a source of photons at time $t > 0$, we expect $p = p_0 + p_\gamma$, the sum of the equilibrium (un-illuminated) and photo-generated hole densities. The boundary condition for this case is

$$f_p(t \rightarrow 0) \rightarrow \frac{(e_n + p_o c_p)}{e_n + e_p + p_o c_p + n_o c_n} \quad (4.5)$$

where n_o , and p_o are the free electron and hole densities in thermal equilibrium, respectively. The assumption that the trap energy is close to the valence band maximum implies $e_p \gg e_n$ in the neighborhood between the gate oxide and the underlying InSb. Assuming that the surface trapping volume for the n-type substrate is in either depletion or mild inversion (consistent with the nominal gate bias), we can neglect terms in n and, from boundary condition (4.5), solve for A:

$$A = \frac{-\frac{e_p}{c_p} p_\gamma}{\left(p_o + \frac{e_p}{c_p}\right) \left(p + \frac{e_p}{c_p}\right)} \quad (4.6)$$

$$\approx -\frac{p_\gamma}{p_\gamma + \frac{e_p}{c_p}}$$

for $p_o \ll e_n/c_n$ and $p_o \ll p_\gamma$ ($p_o \sim 10^{10} \text{ cm}^{-3}$ at 10K!!). The volume at the surface where the trapping occurs is, at this point, unknown, but hypothesized to be exceedingly small. This is because the surface, under optimal biasing conditions, is in either a mildly inverted or a depleted state. In these cases, the thickness perpendicular to the surface where the inversion layer exists will be on the order of tens of angstroms (Grove and Fitzgerald, 1966), thereby enhancing the photo-hole density in the bulk (p_γ in equation(4.6)) by many orders of magnitude. Substituting equation (4.6) into equation (4.4) along with the aforementioned approximation yields:

$$f_p(t) = \left[\frac{p_\gamma}{p_\gamma + \frac{e_p}{c_p}} \right] \left\{ 1 - \exp \left[-\frac{t}{\tau} \left(1 + \frac{p_\gamma c_p}{e_p} \right) \right] \right\} \quad (4.7)$$

where $\tau \equiv e_p^{-1}$, the hole trap decay time constant. Equation (4.7) gives both the flux (p_γ) and the fluence ($t p_\gamma$) dependence of the filled trap population under illumination. If we assume that Fermi-Dirac statistics govern the steady state trap occupation probabilities, then

$$f_p = 1 - f_n = \frac{1}{1 + e^{F-E_T/kT}} \quad (4.8)$$

where E_T is the trap energy and F is the Fermi level (or the quasi-Fermi level in the non-equilibrium case) and we have assumed that the trap energy level is non-degenerate. Taking the limit $t \rightarrow \infty$ in equation (4.7) for the steady state case, and comparing this with equation (4.8) yields:

$$\frac{e_p}{c_p} = p_\gamma e^{F-E_T/kT} \quad (4.9)$$

In the absence of illumination, once the array has been shuttered, we expect that the deviation in the carrier density from the thermal equilibrium value is small once there is no source of photo-generated holes, *i.e.* $p \sim p_0$. We can replace p_γ with p_0 for the thermal equilibrium case which, along with

$$p_0 = N_V e^{E_V - F/kT} \quad (4.10)$$

gives the well known result:

$$\frac{e_p}{c_p} = N_V e^{\frac{E_V - E_T}{kT}} \quad (4.11)$$

where N_V is the effective density of states in the valence band and is given by:

$$N_V = 2 \left[\frac{2\pi m k T}{h^2} \right]^{3/2} \quad (4.12)$$

where m is the effective mass of the valence band. Equation (4.10) and (4.11) are strictly valid only in thermal equilibrium (Hall, 1952 and Shockley & Read, 1952, Sah, 1967). For small deviations from equilibrium, we assume that these relations are still valid, and that the electric fields are not large enough to make the emission and capture coefficients field dependent. Using equation (4.11), we find:

$$\ln(\tau T^2) = \ln \left[\frac{T^2}{c_p N_V} \right] + \frac{E_T - E_V}{kT} \quad (4.13)$$

where τ and T are given in seconds and Kelvins, respectively.

Since $N_V \propto T^{1.5}$ and $c_p \propto T^{0.5}$ (see equation (4.14)), there is no temperature dependence in the first term on the right hand side of equation (4.13) and plotting $\ln[\tau T^2]$ versus $1/kT$ will provide a direct measure of the trap's energy level relative to the valence band edge (McNutt & Meyer, 1981). Note that this method assumes that the trap cross sections are independent of temperature. In addition, the intercept provides the value of c_p , with which we can approximate the trap cross section from

$$c_p \sim \sigma_T v_{th} \quad (4.14)$$

where σ_T is the trap cross section and v_{th} is the hole thermal velocity (Grove, 1967).

After the source of photo-generated holes (the light bulb) has been removed from the field of view, the traps will emit the captured holes en route to thermal equilibrium on some characteristic timescale τ . We expect emission transitions to dominate over capture transitions in this case, and that the approximations made previously are still applicable (specifically, $n_o, c_n n_o, p_o, c_p p_o, e_n \ll e_p$). Assuming that we collect all the emitted carriers, the signal we measure in a particular latent image, t_o seconds after the source has been removed (*i.e.* after the array has been shuttered) is given by

$$\begin{aligned} S(t_o + \Delta t) &= N_T \int_{t_o}^{t_o + \Delta t} \left| \frac{df_p}{dt} \right| dt \\ &= N_T f_p e^{-t_o/\tau} (1 - e^{-\Delta t/\tau}) \end{aligned} \tag{4.15}$$

where N_T is the total number of traps, Δt is the integration time and f_p is the fraction of hole-filled traps when the array was shuttered.

4.7 Derived Trap Parameters

The trap's emission time constant is determined by fitting the latent image data points from a given decay to equation (4.15). The data reduction methodology employed was to select the region on the array where the latent image persisted (corresponding to the location where the signal from the illuminated, defocused spot was a maximum, typically 4×4 or 5×5 pixels), and compute the mean signal level from these pixels. An unilluminated "background" area, at least 25 pixels away was selected, and the mean signal in this area was subtracted from the mean signal in the latent image area to remove any DC offset from the decay. Both the signal and background pixels for each of the frames in the decay remained the same in a given measurement.

In all cases, we find that the latent image decay data are better fit by an exponential decay with two time constants (as compared with a single time constant), implying that there is more than one trapping state involved. Figure 4.7 shows a typical latent image decay for FPA84 with both one and two time constant fits overlaid, illustrating the unambiguous presence of 2 time constants. The experiments performed on ASM244 were not optimal; latent image decay data was acquired for only 50 seconds after shuttering the SCA, compared with FPA84, where we measured the image latency for 100 seconds. This resulted in very poor measurements for the long time constant in ASM244. Furthermore, the data for ASM244 exhibited excess noise, making determination of the time constant(s) subject to large errors. In Figure 4.8, we compare the latent image decay for ASM244 with that of FPA84 (both at temperatures of 10K), and conclude that the decays are virtually identical during the first 50 seconds after shuttering the SCA. Importantly, the behavior of the extracted time constants as a function of temperature is the same in ASM244 and FPA84, leading to the clear conclusion that we are observing the same phenomenon, and that the traps are the same in both arrays. The error in the long time constant also results in an inaccurate total number of traps per pixel as inferred from the exponential fit of the decay extrapolated to time zero (when the shutter was closed). We have

determined the total number of traps per pixel for ASM244 in an independent experiment, which will be discussed in section 4.9.

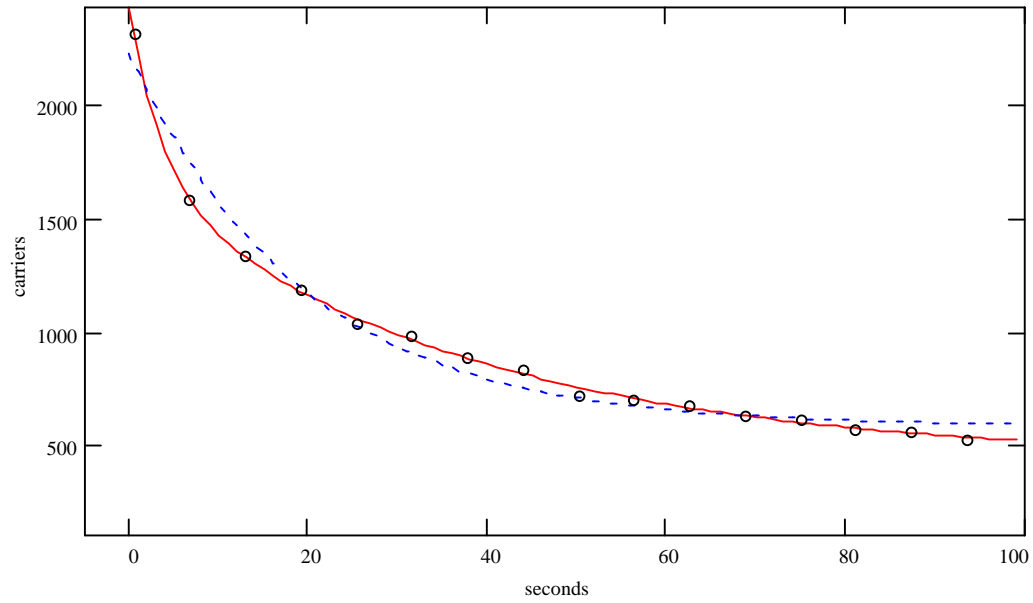


Figure 4.7 - A latent image decay for FPA84 at 6K after exposure to $2.2\mu\text{m}$ photons. The circles are the measured data after removal of the light source. Each is a five second integration on five second centers after shuttering the array. The solid line is a two time constant fit following the methodology described in the text ($\tau_1 \sim 4\text{sec}$, $\tau_2 \sim 37\text{sec}$), and the dashed line represents a single time constant fit to the data ($\tau \sim 19\text{sec}$).

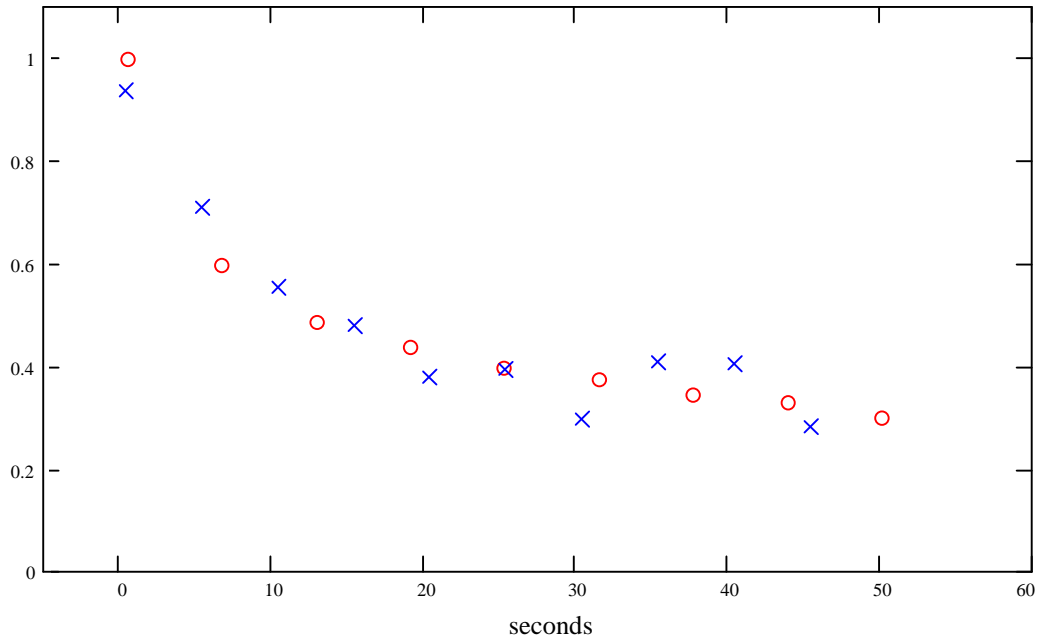


Figure 4.8 - A comparison of a latent image decay for FPA84 (circles) with ASM244 (crosses) at 10K, showing that the decay time constants during the first 50 seconds are the same. The data have been normalized to show the similarities of the time constants. The illustrated decay was after exposure to $2.2\mu\text{m}$ photons

There is some evidence for a very long time constant in several of the latent image decays we have observed in both FPA84 and ASM244. This evidence derives from the presence of a non-zero asymptotic signal level required to fit our data. For the case of FPA84, this level varied monotonically from roughly 100 e^- at 26K to 300 e^- at 6K. Our data do not permit us to extract this time constant, because we only acquired latent image decays for 100 seconds and the magnitude of this time constant is easily $>300\text{sec}$. Further experimentation should be undertaken to investigate the properties of this additional time constant, paying close attention to its behavior at higher temperatures.

The presence of two time constants indicates two possibilities; that there are two entirely different trapping species/defects, or that a single trap species with multiple energy levels is responsible for the latent images. We assume a single trapping species with two energy levels in the following arguments. We constrain the ratio of the two trap populations to be given by a Boltzmann factor, so that the relative populations will depend on the energy difference between the two traps. We expect that this difference will be on the order of 1meV , the thermal energy at 12K. The implicit assumption

in using the Boltzmann factor is that the trapped holes have redistributed themselves between the two energy levels rapidly enough for thermal equilibrium to prevail. We believe this to be the case based on several experimental observations. First, both the magnitude and the decay time constants of the latent image were experimentally repeatable. The latent image vs flux experiment was also repeatable (see section 4.9), lending additional support to the assumption that the trapping states have equilibrated during the exposure to illumination. Finally, we note that the technique described above converged to a stable solution rapidly, requiring no more than 3 iterations (see description in next paragraph).

The methodology described leaves five free parameters in the fit, *i.e.* the energy difference between the two traps, the two time constants, the total number of traps and the asymptotic signal level (indicative of either an electronic offset in the data, or a very long time constant trap with $\tau > 300$ sec). We vary the total number of traps to minimize the rms error in the fit, and perform this routine on all the decay data to obtain the time constants as a function of temperature (for each of the 4 temperatures used in the experiment). From this we can extract the trap energies and cross sections by using equation (4.13). The trap energy difference derived in this way is compared to that assumed in initially fitting the time constants and we iterate using this new value for the trap energy difference until we obtain convergence to better than 0.5 meV. The final values for the trap parameters derived for FPA84 in this way are given in Table 4.12. We find trap energies of 2.0 ± 0.7 and 2.4 ± 0.2 meV *below* the valence band edge for the fast and slow components, respectively. The fast trap's time constant has a large (percentage) error associated with it due to its rapid decay and because of the long integrations required to obtain reasonable signal to noise ratios in the latent images. Figure 4.9 shows the final plot corresponding to equation (4.13) from which both the fast and slow trap energies were extracted. The fit to the theory is seen to be excellent over more than a factor of four in thermal energy (6–26K), even though the errors in the time constant are large for the fast trap. When we perform the same analysis for the time constants measured on ASM244, we obtain trap energies of 1.1 ± 2.2 and 2.2 ± 0.8 meV below the valence band edge for the fast and slow traps, respectively. These values are consistent with those derived for FPA84. Although the long time constant is known to be in error for this SCA, as was

discussed previously, the functional form, and hence, the slope as measured by equation (4.13) with respect to temperature appears to be correct. We have not extracted the trap cross sections for ASM244 because these will certainly be inaccurate due to the error in the time constants.

Temperature (K)	Number of Measurements	N_{fast} (per pixel)	τ_{fast} (sec)	N_{slow} (per pixel)	τ_{slow} (sec)
6.2	8	1700	3.4 ± 1.1	10200	42.6 ± 6.3
10.2	3	1400	4.2 ± 1.6	7700	54.0 ± 5.7
13.2	2	900	3.2 ± 1.7	8400	45.4 ± 5.3
25.8	2	1400	3.1 ± 1.6	1900	22.8 ± 3.4

Table 4.12 - Trap parameters for FPA84 from fitting equation (4.15) to the latent image decays. The fitting scheme was iterative, as described in the text. The error in the short time constant is considerably larger than that of the long time constant. The column labelled number of measurements is the number of decays for which time constants were determined and averaged to yield the fast and slow time constants listed in this table.

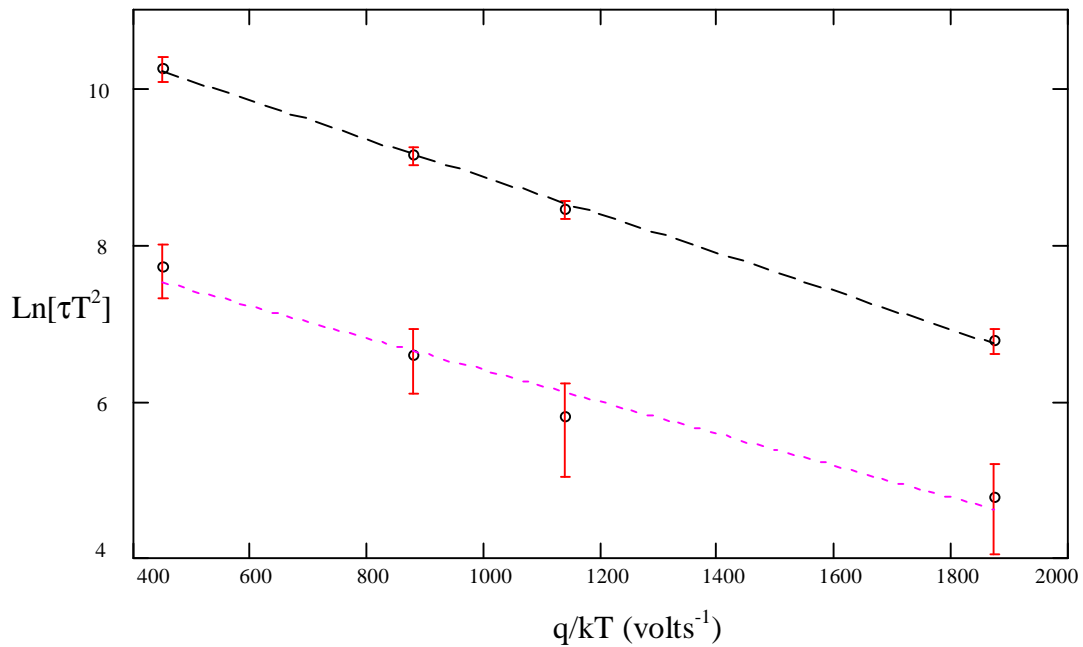


Figure 4.9 - Plot of equation 11 from which trap parameters were extracted for FPA84. The top dotted line is the slow trap and the bottom is the fast trap. The slopes yield energies of 2.0 ± 0.7 and 2.4 ± 0.2

meV below the valence band edge for the fast and slow traps, respectively. Note the large error bars on the fast trap data.

Our results are consistent with those of Barthe and Lile (1993), who have also found evidence for hole traps in SiO₂ passivated InSb just below the valence band edge based on C-V measurements of metal-oxide-semiconductor structures at 77K. Their conclusion is based on two experimental observations: first that the majority of the hysteresis seen occurs in the negative bias region of the CV curve (in which case holes will be attracted to the surface), and second that p-type material exhibits substantially greater hysteresis than does n-type InSb.

We find cross sections of $2 \times 10^{-25} \text{ cm}^2$ and $1 \times 10^{-26} \text{ cm}^2$ for the fast and slow trap components, respectively using equations (4.13) and (4.14) and calculating the hole thermal velocity to extract the cross section. These cross sections are considerably smaller than atomic cross sections, and are, at best, only marginally consistent with previous determinations of cross sections for repulsive trapping centers (Lax, 1960). The inferred trap cross sections and energies are consistent with trapping centers located within the passivation. In this case, hole tunneling from the InSb surface into the passivation through the interface is the trapping/detrapping mechanism. We will show that a tunneling mechanism can account for both the small cross sections as well as the small energies.

4.8 Trapping Mechanism

A commonly found native oxide of InSb, namely In_2O_3 , may be the hole trap responsible for the latent images observed in the arrays investigated here. Figure 4.10 illustrates the proposed energetics for an n-type InSb substrate in contact with the SiO_x passivant with In_2O_3 “impurities” close to the surface. In the figure, the InSb surface is in the flatband condition and the surface traps are within the first 50\AA of the interface (not to scale). The energy levels of In_2O_3 are indicated on the diagram (Geib *et al.*, 1984, Wager *et al.*, 1983, Okamura and Minakata, 1985, Wilmsen, 1976), and we see that the conduction band edge of the In_2O_3 is situated approximately in the middle of the InSb bandgap. We note that these energy levels are only accurate to $\pm 200\text{meV}$ (the temperatures at which they were measured was not provided), so that the precise energy locations are indeterminate as far as the accuracy we require to explain the traps observed here. Nonetheless, the energetics suggest that In_2O_3 may very well be the species responsible for the latent image phenomenon, and that further experimentation should be undertaken to unambiguously identify the nature of these traps. Additionally, the locations of these various energy levels will be shifted by the gate bias. The applied gate bias will raise the In_2O_3 conduction band edge in such a manner as to bring it further from the valence band edge of the InSb. Furthermore, under normal operating conditions, the surface is in the depleted/mildly inverted state, which will raise the InSb valence band maximum by the amount of the surface potential. Both of these effects will be further discussed in section 4.10, where we will estimate the magnitude of the field across the oxide as well as the band bending of the InSb surface based on the solution to the Poisson equation at the surface in conjunction with the operating parameters of the SCA.

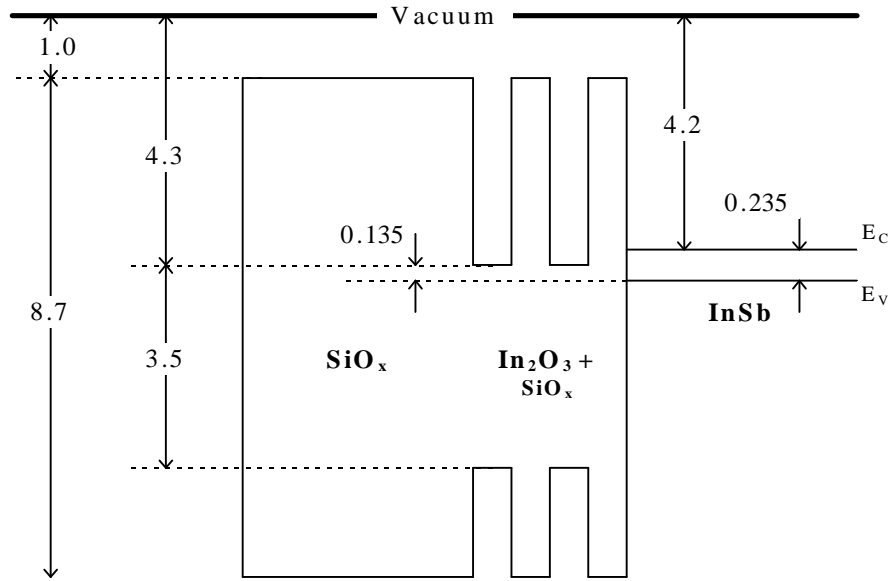


Figure 4.10 - Energy band diagram for the gated area of the bulk InSb in thermal equilibrium (not to scale). All energies are in electron volts (eV) and are relative to the vacuum. E_C and E_V are the conduction and valence band energies respectively. The location of the trap within the oxide is much closer to the InSb surface than is shown here. The bandgap of In_2O_3 is indicated (3.5eV). The electric field across the oxide due to the applied gate bias will lower the conduction band of the In_2O_3 relative to the InSb valence band.

Semi-classical quantum theory, in the form of the WKB approximation, can be used to provide an estimate for the probability of a hole tunneling from the InSb VBM into a bound state in the oxide. Approximating the barrier as a rectangular (i.e. the barrier height is independent of distance into the oxide), we calculate the tunneling transmission as a function of depth into the oxide. The transmission probability through a potential barrier is given by (Landau and Lifschitz, 1977):

$$T = \left[1 + \frac{\psi_o \sinh \left(d \sqrt{\frac{2m^*(\psi_o - E)}{\hbar^2}} \right)}{4E(\psi_o - E)} \right]^{-1} \quad (4.16)$$

where ψ_o is the barrier height, E is the energy of the hole, m^* is the effective hole mass, d is distance into the barrier and \hbar is the Planck constant divided by 2π . In the context of the present model, we see from Figure 4.10 that the tunneling barrier is on the order of $\sim 3\text{eV}$ for the holes in the InSb to tunnel

into the In_2O_3 . We use the drift velocity of the holes to calculate their kinetic energy at the surface, with a hole mobility of $10^4 \text{ cm}^2 \text{ volt}^{-1} \text{ s}^{-1}$ and an electric field strength of $2000 \text{ Volt cm}^{-1}$ in the surface field induced junction (see Table 4.14, and the discussion in section 4.10 for an explanation of how this field was calculated). The results of this calculation are presented in Figure 4.11, from which we see that the tunneling transmission probability is small, and falls rapidly within a very short distance into the oxide. This implies that the traps must be close to the interface, and also accounts for the anomalously small effective cross sections we have inferred from the time constant measurements presented in section 4.7.

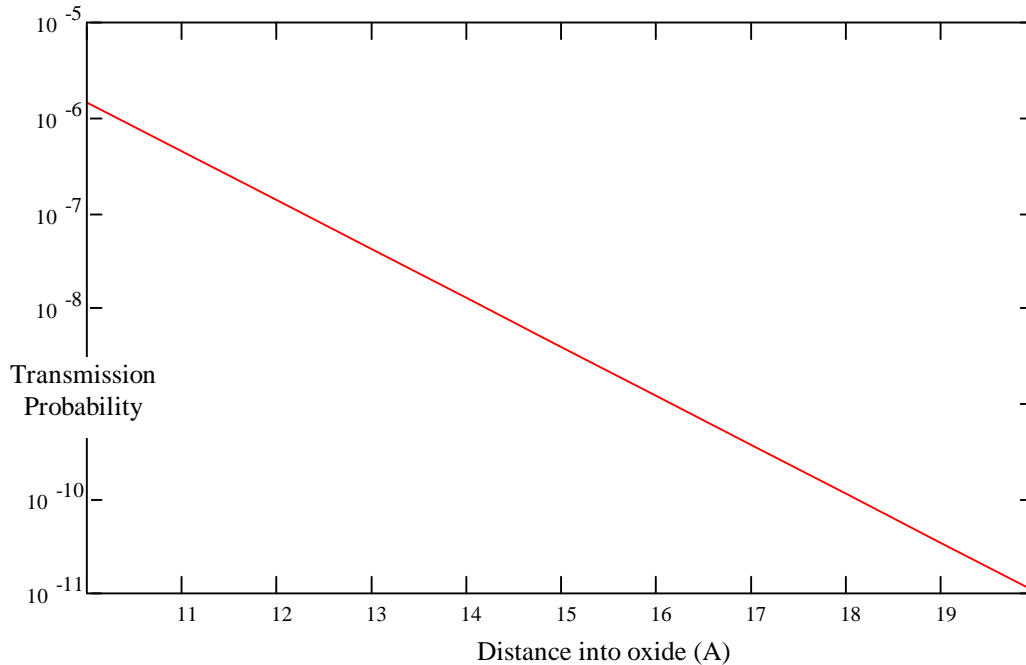


Figure 4.11 - WKB tunneling transmission factor for a 3.4eV barrier. The parameters used in the calculation are given in the text.

Sewell and Anderson (1975) derive the trap cross section as a function of the (tunneling) distance into a passivant using the same classical approximation:

$$\sigma(x) = \sigma e^{-2K_0 x} \quad (4.17)$$

where $\sigma(x)$ is the trap cross section at the distance x into passivant. K_0 is the wave number (called the wave function decay constant by Heiman and Warfield, 1965), and is given by

$$K_o = \sqrt{\frac{2m^*}{\hbar^2}(\psi_o - E_x)} \quad (4.18)$$

where E_x is the hole's kinetic energy in the direction perpendicular to the interface. Using equations (4.17) and (4.18) along with the same parameters used in plotting equation (4.16) we find that $\sigma(15\text{\AA})=2\times 10^{-8} \sigma$ and $\sigma(25\text{\AA})=1\times 10^{-13} \sigma$. If we assume that the traps are atomic (or molecular) in nature, rather than being defects in the oxide, then we require that the cross sections be consistent with atomic cross sections. This requirement results in traps located 20\AA from the interface (in agreement with the observations and calculations of Sewell and Anderson (1975)), and "true" trapping cross sections of $4\times 10^{-15} \text{ cm}^2$ and $2\times 10^{-16} \text{ cm}^2$ for the fast and slow traps, respectively. This model explains the low values previously inferred for the trap cross section, and suggests that the holes are trapped to a depth in the oxide on the order of 20\AA .

4.9 Latency Dependence on Operating Conditions

4.9.1 Latency Dependence on Gate Bias

From an operational point of view, a detector array's dark current is one of its most critical figures of merit. There are many details that affect the dark current of a photovoltaic detector: material and processing factors, temperature of operation, as well as detector and gate bias. For astronomical use, the detector bias is also chosen to maximize the well-depth, so that the only remaining "knobs" available to minimize the dark current are temperature and gate bias. In order to study the dependence of dark current on gate voltage, we have measured the dark current as a function of V_{gate} at 12K in 100 second integrations with the array staring at a 4.2K dark slide. The results are shown in Figure 4.12, where we have plotted the average dark current in the same region of the array where we performed the latent image experiments. There is a range of gate bias where the dark current is minimized, which corresponds to the surface being between the flatband and mildly inverted states (indicated on the plot). While the accumulated side of the dark current plot is clear, the inverted side is subject to discussion. We drew the inverted surface state line at -2.5V because that is the gate voltage at which we also observed significant cross talk between the pixels over more than half of the array, indicating the presence of an inversion layer at the surface which electrically connects neighboring pixels together.

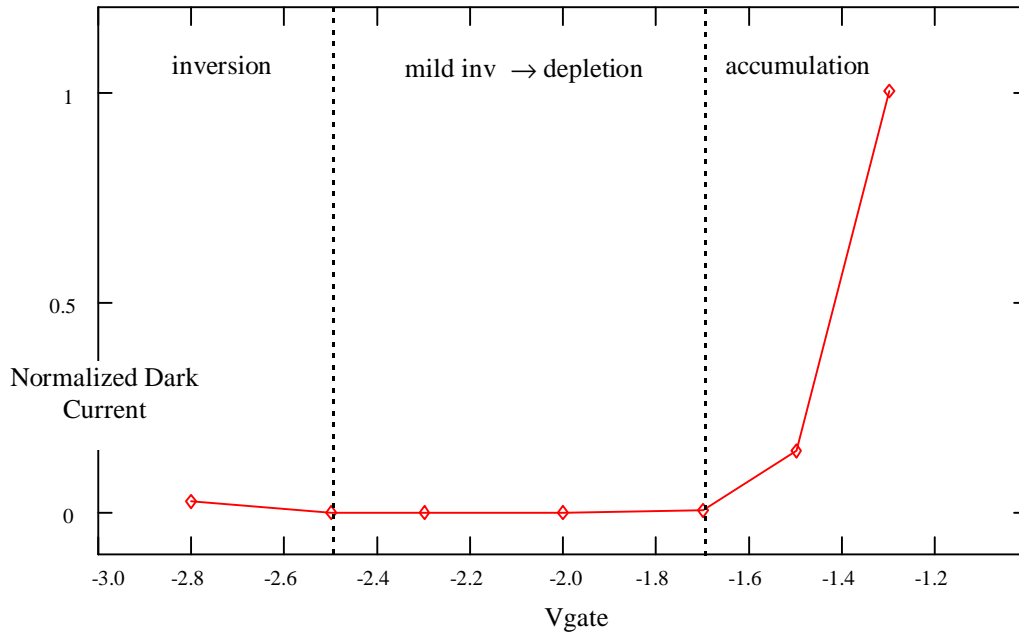


Figure 4.12 - Dark current vs. gate voltage for FPA84 at 12K. The dark current has been normalized to its maximum value at $V_{\text{gate}} = -1.3\text{V}$, where it measured to be $150\text{ e}^-/\text{sec}$. The measurement was performed by subtracting a 1 second integration from a 101 second integration. The increase in the dark current from $V_{\text{gate}} = -2.5\text{V} \rightarrow -2.8\text{V}$ is not clearly visible on the plot, but its magnitude increases from near zero to $15\text{ e}^-/\text{sec}$.

Figure 4.13 shows the initial strength of a latent image as a function of gate bias (after the light source was removed). The vertical axis has been normalized to the maximum value observed at $V_{\text{gate}} = -1.3\text{V}$, which was 7700 e^- . The fact that the accumulated side of the figure exhibits stronger image latency is consistent with our theory that the band-bending that occurs when the surface is accumulated pinches off the depletion region, resulting in higher, surface driven dark current. The strength of the latent images varies with gate voltage in the same manner as the dark current; the dark current vs. V_{gate} ; Figure 4.12 is remarkably similar to that of Figure 4.13,. Both sets of curves are flat in the region where the surface is depleted and rise in both inversion and accumulation. The rise on the inversion side in Figure 4.13 occurs 200mV earlier (closer to $V_{\text{gate}}=0$) than it does in the case of the dark current experiment. We hypothesize the behavior in the region of $-2.5\text{V} < V_{\text{gate}} < -2.3\text{V}$ is due to a mildly inverted surface exposing more of the surface traps, in the sense that there is a conductive path from the traps to the junction through the mildly inverted surface.

The similarity between the two plots strongly suggests that both phenomena are surface driven, though the physical mechanisms at play are likely to be different, as has been previously discussed. Comparing the accumulated regions of Figure 4.13 and Figure 4.12, we note that magnitude of the dark current at $V_{\text{gate}} = -1.3\text{V}$ was $150\text{ e}^-/\text{sec}$, which is an order of magnitude smaller than the current associated with the first latent image ($1540\text{ e}^-/\text{sec}$) when the gate was biased to produce an accumulated surface. This supports our theory that the traps are increasing the dark current through pinch-off (see Figure 4.4 and discussion in 4.10).

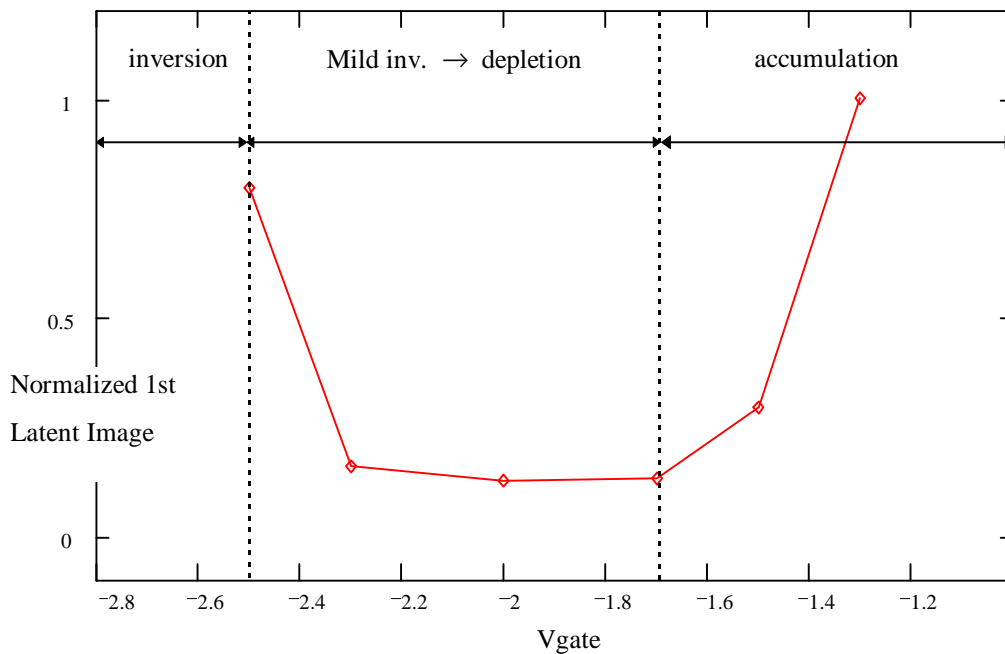


Figure 4.13 - Magnitude of the saturated latent image (as measured by the strength of the first latent image in a decay) vs. V_{gate} . The data is for FPA84 at 12K, with an integration time of 5 seconds. The vertical axis has been normalized to the maximum value observed at $V_{\text{gate}} = -1.3\text{V}$, which was 7700 e^- . The regions corresponding to the InSb surface are indicated at the top of the plot.

4.9.2 Latency Dependence on Photoflux

Additional experiments were performed to investigate the latent image decay parameters as a function of incident photon flux. Latent image decays were recorded for FPA84 and ASM244 at varying incident photon flux levels at approximately constant fluence until the signal level of the *initial*

latent image reached a constant value (which we term “saturation”). This saturation level provides the total number of traps per pixel. For the case of FPA84, the latent image decays were performed at the same flux level that saturated the latent image at 6.2K. The flux range used in this experiment ranged from $10^3 \text{ sec}^{-1} \text{ pixel}^{-1}$ to more than $10^6 \text{ sec}^{-1} \text{ pixel}^{-1}$.

We obtain the total number of traps (N_T) directly from equation (4.15) when the initial latent image is saturated, i.e. when the signal level in the first latent image no longer changes with increasing flux, by taking $f_p=1$. The total numbers of traps inferred for FPA84 and ASM244 are listed in Table 4.13, along with calculated areal trap densities. We note that these values varied by ~60% over a given SCA. In order to convert this to an areal density, we need to know the area under the passivation where the trapping occurs. If we take the total area under the gate (excluding the metallurgical junction) as the active trapping area for FPA84, we obtain $4.4 \times 10^9 \text{ cm}^{-2}$, as a lower limit to the areal trap density since it is likely that the charge trapping is not uniform across the area under the gate, but occurs within the first few microns of the depletion region around the immediate perimeter of the metallurgical junction as discussed in section 4.4. For the reasons presented there, we take a $2\mu\text{m}$ width around the perimeter of the junction to be the active trapping area. These calculated densities are listed in Table 4.13 for comparison with the areal densities computed for the total area under the gate. For the sake of completeness, we include ASM244 in the table even though our inability to accurately determine the time constants for ASM244 makes our determination of the total number of traps for this array somewhat uncertain. Table 4.13 shows that the trap densities are comparable for both arrays if the $2\mu\text{m}$ perimeter is used for the active trapping area. The lattice constant, a , of InSb is $a \sim 6.5 \text{ \AA}$ (at room temperature), which implies an areal density of $2 \times 10^{14} \text{ cm}^{-2}$ ($1/a^2$) for the lattice sites, from which we conclude that the InSb surface is well passivated in that there is roughly one latent image trap per 10^5 surface atoms.

SCA	Flux ($\times 10^6 \text{ sec}^{-1} \text{ pix}^{-1}$)	N_T pixel	Trap Areal Density ($\times 10^8 \text{ cm}^{-2}$)	
			Total area under gate	$2\mu\text{m}$ perimeter
FPA84	0.1	1.2×10^4	44	120
ASM244	0.5–5	$0.9\text{--}1.5 \times 10^4$	5–8	35–58
SCA02	0.7–3	0.8×10^4	4.4	31
FPA48*	2	< 50	< 0.2	< 0.5
FPA41*	30	< 200	< 0.7	< 2

Table 4.13 - Total number of traps inferred from saturated latent image data in conjunction with the fluxes that generated the saturation level for 5 SCAs. The two arrays listed with asterisks are gateless devices, while the remaining three have gates.

Using equation (4.7) in equation (4.15) we can plot the signal in the initial latent image against the incident photon flux and fit the photo-generated hole density to this data. This enables us to indirectly estimate the trapping volume by using the previously determined values for the trap areal density along with the illumination fluence. We obtain a starting estimate for the values of p_γ in the bulk InSb from the $p_\gamma^{\text{bulk}} = i_\gamma / q A_{\text{pix}} v_{\text{th}}$ where i_γ is the photocurrent, A_{pix} is the pixel area and q is the electronic charge. Typical values obtained for the case when the initial latent image saturates are $p_\gamma^{\text{bulk}} \sim 10^4 \text{ cm}^{-3}$. For an intrinsic detector quantum efficiency of $\sim 80\%$, there will be a maximum of $0.2 \times p_\gamma^{\text{bulk}}$ carriers available to be trapped. If all the uncollected holes reach the edge of the surface depletion region on the bulk side, they will be swept across to the interface between the InSb and the oxide. This effectively enhances the hole density at the surface relative to that in the bulk due to the extremely small volume at the surface into which the holes are “squashed.” Figure 4.14 shows the experimental data for the varying flux experiment, with the model fit overlaid. The minimum incident photoflux which experimentally caused the latent images to saturate was in the range $3 \times 10^4 \text{ sec}^{-1} \text{ pixel}^{-1}$, corresponding to fluences of greater than $6 \times 10^5 \text{ pixel}^{-1}$. The trap parameters derived from the time constant vs temperature experiment were used as inputs to the fit, so that there was one free parameter

in the fit, namely the multiplicative factor which converts the flux to carrier density at the trapping site. The model fit requires hole densities in the range $p_{\gamma} \sim 4 \times 10^{17} - 4 \times 10^{19} \text{ cm}^{-3}$. If we assume that the trapping is occurring in the immediate perimeter of the implant we can solve for the thickness required by this area to generate the hole density required by the model fit. This methodology yields 1–50Å, which is consistent with the depth of an inversion layer at the interface determined by using Gauss's law as well as the depth implied by the tunnelling calculations.

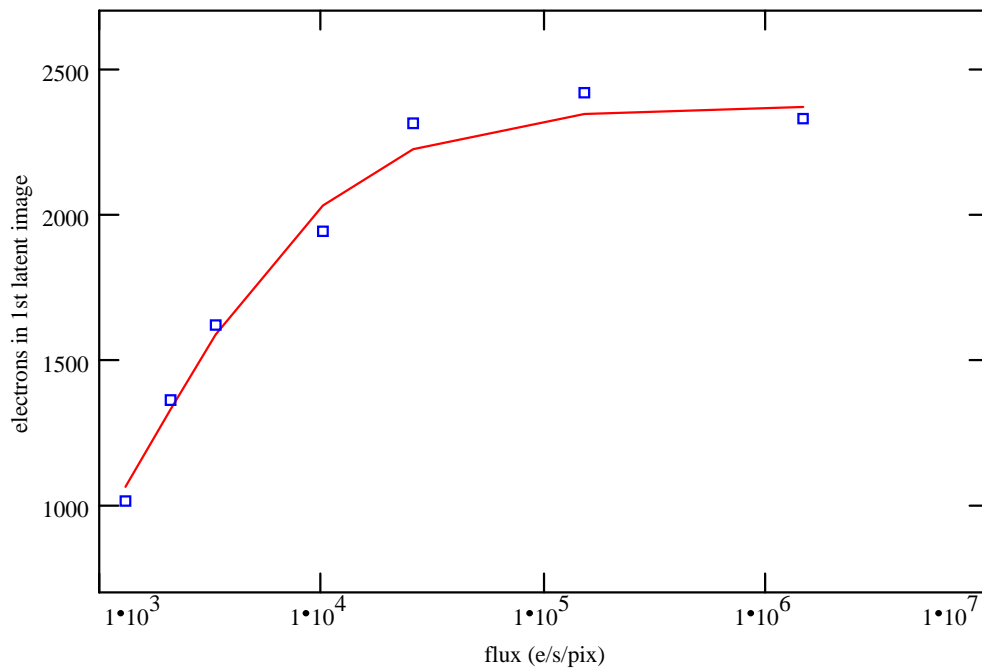


Figure 4.14 - Signal (e^-) in the first latent image, after the light source was removed, as a function of incident flux for FPA84 at 6.2K. The solid line is the best fit obtained from equations (4.7) and (4.15).

A saturated latent image decay series was recorded at various gate voltages (with the SCA at a temperature of 11.2K) in order to investigate its dependence on gate bias and surface potential. The results of this experiment are shown in Figure 4.13. These data strongly suggest that the effective number of traps increases when the surface is either accumulated or inverted, and that the effective number of traps are at a minimum when the surface is in the depleted state. This also lends credence to the notion that the active trapping area (when the SCA is biased at its normal operating point) is in the immediate perimeter of the metallurgical depletion region because it is only when the surface is

inverted that the entire surface under the gate can communicate with the p^+ implant. On the inversion side, we hypothesize that the induced conduction channel enables more of the surface traps to communicate with the implant, hence the increase in the magnitude of the latent image.

The fact that we observe latent images when the surface is biased in accumulation cannot be explained by the simple model whereby the signal in the latent image is due exclusively to detrapped photo-charge for two reasons: the electric field associated with an accumulated surface will repel the holes from the surface, and also because any holes tunneling into an accumulated surface from the passivant will recombine before they have the opportunity to diffuse or drift to the implant. In this case, the model associated with Figure 4.4 is the likely mechanism for the latent image, the additional charge trapped in the surface further pinches off the junction, resulting in higher dark currents which manifests itself as image latency. We would anticipate the latent image decay for an accumulated surface to be different from that observed in depletion or inversion. First because the total integrated signal in the latent image may well exceed the quantity of trapped charge, and second because we expect that the relaxation time constant of the surface in the accumulated state will be considerably faster than that in depletion. There is evidence that this is the case from observations of the time it took for FPA84 to equilibrate after gate bias changes; the SCA took less than a minute to equilibrate after V_{gate} was changed while the surface was in the accumulated state compared with ~10 minutes in both the inverted and depleted states. Equilibration in this case was defined to be the array output remaining stable over successive acquired frames.

4.10 Surface Calculations

We have solved Poisson's equation in one dimension for the surface region of the n-type bulk under the gated passivation, over the range of gate voltages used in the experiments in order to determine: 1) the surface potential, to study correlations of surface potential with the strength of the latent image as well as other operational parameters, 2) the magnitude and sign of any induced surface charge, and 3) the fixed oxide/surface state charge density. This calculation will also allow us to determine, based on the magnitude of the saturated trapped charge from the latent image experiments, the range over which the InSb surface state can swing as a result of a given exposure to illumination. We also use these calculations to verify some of the conclusions reached in the preceding analysis and discussion.

The formalism used follows that of Grove and Fitzgerald (1966, see also Nicollian and Brews, 1982), and we refer the reader to this well-written and seminal work for the mathematical details of the derivation. The surface charge density induced in the surface of the InSb is given by:

$$Q_s(u_s) = 2qn_i L_D \left[e^{u_B - u_s} - e^{u_B} + e^{u_s - u_B + v_j} - e^{-u_B - v_j} + 2u_s \sinh u_B \right]^{1/2} \quad (4.19)$$

where n_i is the intrinsic carrier density, L_D is the intrinsic Debye length¹⁷, u_B and u_s are the unitless potential ($u=q\phi/kT$, where ϕ is the potential) in the bulk InSb and at the surface respectively, and v_j is the unitless detector bias voltage ($=qV_{\text{bias}}/kT$). We see, from equation (4.19), that before solving for any induced surface charge, we must know the potential at the surface of the InSb (u_s). This potential will depend on the applied gate bias; in the absence of interface charge, the potential difference between the gate metal and the InSb surface is:

¹⁷ The Debye length is the characteristic screening length for a small charges in a material, and is given by $L_D = \sqrt{\epsilon_s kT / 2q^2 n_i}$, where ϵ_s is the semiconductor dielectric constant, and n_i is the intrinsic carrier density.

$$V_{\text{gate}} - \phi_s = V_{\text{ox}} + \phi_{\text{ms}} \quad (4.20)$$

where V_{gate} is the voltage applied to the gate, ϕ_s is the potential at the InSb surface, V_{ox} is the voltage drop across the passivation, and ϕ_{ms} is the metal-semiconductor work function. The flatband condition occurs, by definition, when the surface potential of the InSb is zero ($\phi_s=0$). Thus, the gate voltage required to bring the surface into flatband will be given by $V_{\text{gate}} \equiv V_{\text{fb}} = V_{\text{ox}} + \phi_{\text{ms}}$, from which it is clear that the oxide's dielectric properties play a large role in determining the flatband voltage, and hence, the state of the surface for both gated and un-gated arrays. Invoking charge neutrality according to $Q_{\text{surface}} + Q_{\text{ox}} + Q_{\text{gate}} = 0$ along with $Q_{\text{gate}} = V_{\text{ox}} C_{\text{ox}}$ enables us to rewrite equation (4.20), solving for the charge induced in the InSb surface:

$$Q_{\text{surface}} \equiv Q_s(\phi_s) = C_{\text{ox}} [\phi_s - V_{\text{gate}} + \phi_{\text{ms}}] - Q_{\text{ox}} \quad (4.21)$$

where C_{oxide} is the passivant capacitance per unit area. Any trapped interface charge will add another term to the right hand side ($Q_{\text{ox}} = Q_{\text{fix}} \rightarrow Q_{\text{fix}} + Q_{\text{it}}$). Equations (4.19) and (4.21) simultaneously determine the surface potential for a given gate voltage after the passivant's properties have been determined. The dielectric constant can be measured via ellipsometry, and the flatband determined from capacitance-voltage measurements of MOS structures. The flatband voltage can also be inferred from SCA level measurements, as will be discussed shortly. The value used for the InSb-Al work function difference was $\phi_{\text{ms}} = -0.1\text{V}$ (Rosbeck, private communication, 1993). C-V measurements yielded an estimate for the value of the interface trapped charge, approximately one-tenth that of the fixed oxide charge, *i.e.* $\sim 2 \times 10^{-9}$ Coul/cm² (Rosbeck, private communication). Once all these parameters have been determined, the oxide fixed charge (plus any interface trapped charge) can easily be calculated. We also note that if the interface and/or oxide fixed charge is a function of temperature, then we would expect that the threshold voltage (the surface potential that causes the surface to invert) will be a function of temperature as well.

The flatband voltage for FPA84 was determined from dark current versus V_{gate} measurements. The mathematical formalism used to calculate the flatband voltage derives from Grove and Fitzgerald

(1966). We first assume that the large increase in dark current at the more negative gate voltage (see Figure 4.12) is due to inversion of the n-type substrate (the “turn-on,” or “threshold voltage”), and then calculate the expected flatband voltage. We compare this voltage to the value expected from the dark current vs. V_{gate} measurement, assuming that the increase in dark current on the more positive side of the gate voltage occurs at the transition between depletion and accumulation, *i.e.* flatband. We repeated the calculation in the reverse; by assuming that the increase in dark current at the more positive gate voltage occurred at flatband and calculate the inversion voltage. Both methods yielded consistent results, in agreement with the measurements to better than 250mV. We find additional confirmation for our choice of flat-band and inversion potentials from Figure 4.13, from which we observe that the magnitude of the trapped charge increases at the same gate biases we have interpreted as being flatband and within 200mV of that which we have interpreted as inversion from the dark current measurements. The mechanism that this conclusion is based on has been discussed previously, and is consistent with these observations. Figure 4.15 shows the theoretical results for the total induced surface charge per unit area as a function of unitless surface potential ($U=qV/kT$) for the n-type substrate at a temperature of 12K at our nominal operating gate voltage of $-2.3V$. It also shows the surface charge as a function of gate voltage (in dimensionless units). The regions of accumulation, depletion, and weak and strong inversion are indicated. The intersection of these two curves is the solution for the surface state and the total fixed oxide charge ($Q_{\text{ox}}=Q_{\text{fix}}+Q_{\text{it}}$). Note that in Figure 4.15, the sign of the induced charge for the accumulated case is actually negative, its absolute value has been plotted for the sake of compactness, and thus the solution is the intersection at $U=-265$ (the more negative of the two curve intersections). Using, $Q_{\text{surface}}=4\times 10^{10} \text{ e}^- \text{ cm}^{-2}$ from the solution illustrated in Figure 4.15, the fixed oxide charge density was found (equation (4.21)) to be $+7.1\times 10^{-8} \text{ Coul/cm}^2$, or $4.4\times 10^{11} \text{ cm}^{-2}$. This value falls well within the range for the fixed charge typically measured in oxide passivations (Bloom and Nemirovsky, 1993), and is ~ 100 times larger than the value we have measured for the number of traps per square centimeter associated with the latent images. Depending on the initial charge state of the latent image traps, they may or may not be included in the fixed passivation

charge, and they represent a small subset of the total oxide charge if they are included. We find additional confirmation that the substrate surface is inverted when $V_{\text{gate}} \sim -2.5\text{V}$ from the observation of strong signal blooming of high dark current pixels into neighboring pixels indicating that the individual diodes (pixels) have been shorted together via the conduction channel created by the inverted hole population under the gate metal.

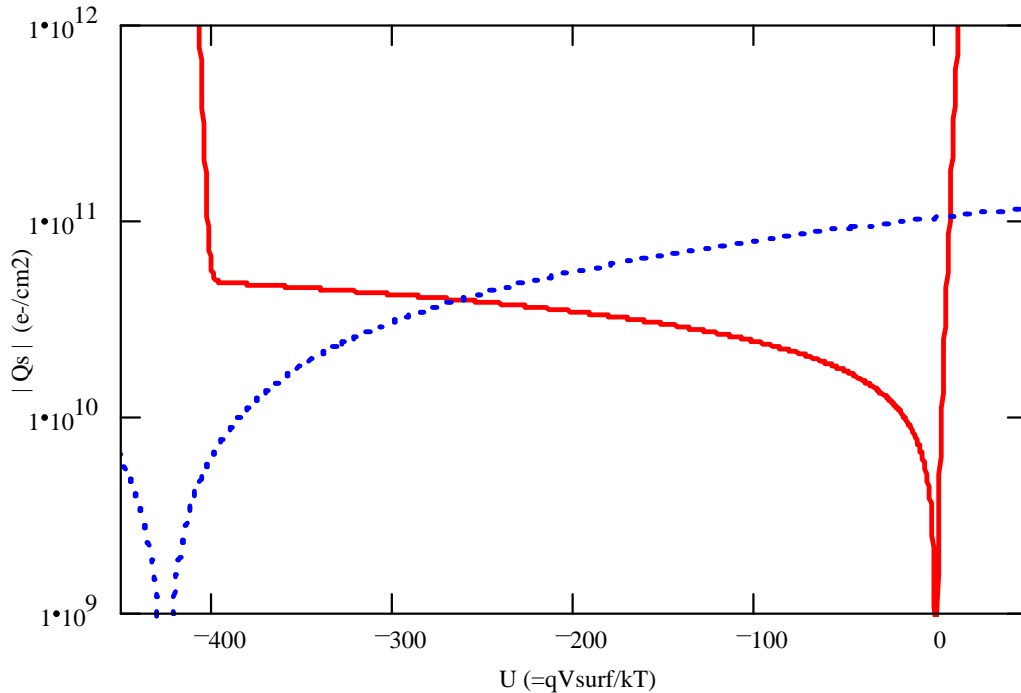


Figure 4.15 - Surface solution for n-type substrate assuming $T=12\text{K}$, $V_{\text{gate}} = -2.3 \rightarrow V_{\text{surf}} = -265\text{mV}$. The solid curve is the solution to Poisson's equation (equation (4.19)); the flatband condition occurs where $q\phi_{\text{surf}}/kT=0$, depletion and inversion of the surface occur to the left of this point and accumulation to the right. Note that the sign of the induced charge for the accumulated case is actually negative, its absolute value has been plotted here. The dotted curve represents the solution to the voltage drop equation (equation (4.21)). The intersection of the two curves represents the surface solution, i.e. under what condition the surface exists for a given gate bias.

Figure 4.15 enables us to graphically determine the change in the surface state expected for a given trapped charge in the oxide. We use the previously estimated areal density of the traps associated with the saturated latent images in the range $4\text{--}7 \times 10^9 \text{ cm}^{-2}$. Moving along the solid curve in the figure by this quantity of charge amounts to a negligible change in the surface state when the surface is inverted or mildly inverted, but can significantly change the operating point when the surface

starts close to flatband. For example, we can see by using the graph in Figure 4.15, that if the surface starts at $Q_s = -1.1 \times 10^{-9}$ coul/cm² ($\phi_s = -10$ mV, or $V_{\text{gate}} = -2.04$ V corresponding to $u_s = -9$) and then traps 15,000 holes in the surface (i.e. 1.1×10^{-9} coul/cm²), the surface will go into the flatband state. Table 4.14 summarizes the results of the surface calculations, both with and without the additional trapped charge. Note that in all cases, the surface potential is more positive when there is positive charge trapped in the passivant. The surface potential, Φ_s , gives the amount of band-bending that occurs in the InSb from the bulk to the surface. The fact that it is larger than the forbidden gap when the surface is inverted indicates that band-to-band tunneling may be responsible for the large dark currents observed when the surface is inverted. It is also interesting that very few holes are required to invert the surface; for an inversion channel 100Å deep by the entire area under the gate, a mere 1000 holes per pixel will invert the surface.

The calculated field strength in the field induced junction is reasonable; for comparison, the magnitude of the electric field in the metallurgical junction for a reverse bias of 150mV is ~ 3000 Volt cm⁻¹, 1.5–2 times larger than that calculated for the field induced junction at its nominal operating point in depletion. While the electric field strength in the passivation is large, it is well below the breakdown field. The magnitude of the passivation field enables us to estimate how far the In₂O₃ conduction band will be raised relative to the InSb, which is relevant to the trap energies we have inferred. We estimate this value by multiplying the electric field by the distance into the passivation where the trapping is presumed to occur: $\Delta V = 2 \times 10^5$ Volt cm⁻¹ \times 20Å = 40mV. The electric field was approximated by dividing the voltage drop across the oxide (~ 2 V) by the oxide thickness (1000Å), and was verified by more rigorous calculations (see Table 4.14, and the discussion in section 4.10 for an explanation of how this field was calculated). The total band bending at the surface, determined by Φ_s , amounts to approximately 270mV at our nominal operating gate bias of -2.3 V. When we adjust the band edges in Figure 4.10 per the preceding offsets, we find that the conduction band edge of the In₂O₃ is ~ 100 meV below the InSb valence band edge. Given the uncertainties in the energies associated with Figure 4.10, the variable band-bending that occurs at the surface, as well as the trap energies determined in this

work, the agreement between the derived trap energies ($\sim 2\text{meV}$) and that of the conduction band edge of In_2O_3 is very good ($\sim 100\text{meV}$). We conclude that In_2O_3 is a strong candidate for the traps responsible for the latent images, and that further experimentation should be undertaken to determine the validity of this hypothesis.

Owing to the small derivative ($dQ_s/d\Phi_s$, see Figure 4.15) in the depleted state, a given quantity of trapped charge will have a larger effect on the surface potential than it will for any other surface condition, resulting in a lower electric field across the field induced junction. Once the surface is inverted, the table shows that field induced junction has reached its maximum width, at which point any additional gate bias will be accommodated by an exponential increase in the inversion layer charge which effectively screens the field induced junction from the effect of the additional gate charge. The change in the inversion layer charge (Q_{sp}) at $V_{\text{gate}} = -2.5\text{V}$ and -1.9V are particularly instructive. The inversion charge in the case where we have included surface traps is three orders of magnitude smaller in the case where no trapping occurs. The presence of the trapped charge in this case effectively delays the onset of inversion (moves it to more negative gate voltages), maintaining the surface in the depleted state. On the flatband side of the table, we observe that the onset of accumulation (*i.e.* $\Phi_s > 0$) occurs at more negative gate voltages. From this observation, we conclude that holes trapped in a surface that is initially in or close to the flatband condition will become accumulated and hence lead to an increase in dark current via the pinch-off mechanism discussed earlier.

V_{gate}	without trapped charge						with trapped charge					
	-1.9	-2.1	-2.3	-2.5	-2.7		-1.9	-2.1	-2.3	-2.5	-2.7	
state	~flatband	depleted	depleted	inverted	inverted		~flatband	depleted	depleted	inverted	inverted	
Φ_s (mV)	-10	-126	-272	-413	-415		+1	-106	-250	-406	-415	
E_{fij} (V/cm)	190	1250	2000	2500	2510		30	1120	1900	2480	2510	
W_{fij} (μm)	0.52	1.01	1.36	1.65	1.65		0.42	0.95	1.32	1.64	1.65	
Q_{fij} ($10^9/\text{cm}^2$)	15.7	30.2	40.9	49.5	49.6		12.7	28.5	39.5	49.1	49.6	
Q_{sn} ($10^9/\text{cm}^2$)	8.49	3.10	1.03	0.35	0.34		15.7	3.58	1.22	0.36	0.34	
Q_{sp} ($10^9/\text{cm}^2$)	0	0	0	4.88	51.6		0	0	0	.006	44.8	
Q_s ($10^9/\text{cm}^2$)	7.18	27.1	39.9	54.1	101		2.97	24.8	38.3	48.8	94.3	
E_{ox} (10^5 V/cm)	1.79	1.87	1.93	1.99	2.18		1.80	1.89	1.95	1.99	2.18	

Table 4.14 - Summary of surface state with and without passivation trapped charge. The trapped charge ($6.8 \times 10^9 \text{ cm}^{-2}$, the saturated value) was added to the passivation fixed charge for the purposes of calculation after the fixed charge was determined. All calculations were performed for $T=12\text{K}$, $\Phi_{ms}=-0.1\text{V}$, and $Q_{it}=0.1 \times Q_{ss}$. The rows are as follows: state refers to the state of the InSb surface: accumulated, flatband, depleted or inverted, Φ_s is the surface potential, E_{fij} , W_{fij} and Q_{fij} (fij=field induced junction) are the average electric field strength, width and fixed charge, respectively, in the field induced depletion region, Q_s , Q_{sn} and Q_{sp} are the total, electron, and hole areal charge density, respectively, induced in the InSb surface, and E_{ox} is the average electric field strength in the oxide.

4.11 Conclusion

We have shown that the latent image phenomenon in astronomical InSb photodiode infrared focal plane arrays passivated with SiO_x can be explained by charge trapping at the front surface of the detector array. The type of front surface passivant plays a crucial role in the magnitude of this effect; SiO_x exhibits significantly more image latency than does Si₃N₄. The variation in image latency with gate bias is very similar to that exhibited by the detector dark current; when the detector's front surface (outside the metallurgical junction) is accumulated or inverted, the magnitude of the image latency has been experimentally shown to increase substantially.

Behaviorally, the latent images are consistent with a modified Shockley, Read, Hall trapping model (Hall 1952, Shockley and Read, 1952). The temporal decay of a latent image exhibits at least two decay time constants at temperatures below 30K. The variation of the decay time constant with temperature results in trap energies which are approximately 2meV below the InSb valence band maximum. Trapping cross sections extracted from the same temperature vs. time constant data are in the range 10^{-25} – 10^{-26} cm². These trap parameters, extracted from the modified SRH model, are consistent with hole traps in the SiO_x that lie within approximately 50Å of the SiO_x/InSb interface. Quantum mechanical tunneling is the proposed mechanism for populating and depopulating the traps; it accounts nicely for the small cross sections we inferred from experiment. We propose the molecular species In₂O₃ as a plausible candidate for the trapping species responsible for the latent images in SiO_x passivated InSb. While we did not observe latent images in either of the two Si₃N₄ passivated arrays which we investigated, the proposed mechanism and trapping species may operate in these devices as well.

5. Bibliography

Abramowitz, M., Stegun, I.A., eds., *Handbook of Mathematical Functions*, 1965, Dover Publications, Inc., New York.

Allamandola, L.J., Tielens, A.G.G.M., Barker, J.R., *ApJ*, **290**, L25, 1985.

———. *ApJSuppl*, **71**, 733, 1989.

———. in *Polycyclic Aromatic Hydrocarbons and Astrophysics*, eds. A. Léger, L. d'Hendecourt, N. Boccarda, Reidel, Dordrecht, p.255, 1987.

Allen, C.W., *Astrophysical Quantities*, 3rd Ed., Athelone Press, London, 1976.

Altamore, A., Baratta, G.B., Cassatella, A., Grasdalen, G.L., Persi, P., Viotti, R., *AA*, **90**, 290, 1980.

Anderson, W.W., Hoffman, H.J., *J. Applied Physics*, **53**, 9130, 1982.

Barthe, W. and Lile, D., *Thin Solid Films*, **229**, 54, 1993.

Baumgratz, B., "ESCA Studies of InSb Natural Oxides." Master's Thesis, University of California, Santa Barbara, 1988.

Black, J.H., Dalgarno, A., *ApJ*, **203**, 132, 1976.

Bloom, I., Nemirovsky, Y., *IEEE Transactions on Electron Devices*, **40**, 2, 309, 1993.

Bregman, J.D., Rank, D., Temi, P., Hudgins, D., Kay, L., *ApJ*, **411**, 494, 1993.

Buss, R.H., Cohen, M., Tielens, A.G.G.M., Werner, M., Bregman, J., Witteborn, F.C., Rank, D.M., Sanford, S.A., *ApJ*, **365**, L23, 1990.

Casey, S., *ApJ*, **371**, 183, 1991.

Chen, H., Garnett, J.D., Solomon, S.L., Forrest, W.J., Pipher, J.P., Helmbock, A., **IRAC TM92-3004**, June, 1992.

Chen, H., Garnett, J.D., Solomon, S.L., Libonate, G.S., Pipher, J.P., Forrest, W.J., **IRAC TM93-3005**, April, 1993.

Clayton, G.C., Mathis, J.S., *ApJ*, **327**, 911, 1988.

Cohen, M., Allamandola, L.J., Tielens, A.G.G.M., Bregman, J., Simpson, J., Witterborn, F., Wooden, D., Rank, D., *ApJ*, **302**, 1986

Davis, J.L., *Surf. Sci.*, **2**, 33-39, 1964.

deMuizon, M., d'Hendecourt, L.B., Geballe, T.R., in *Polycyclic Aromatic Hydrocarbons and Astrophysics*, eds. A. Léger, L. d'Hendecourt, N. Boccara, Reidel, Dordrecht, p.295, 1987.

Dennefeld, M., Désert, F.X., *AA*, **227**, 379, 1990.

Draine, B.T., Anderson, N., *ApJ*, **292**, 494, 1985.

Draine, B.T., Lee, H.M., *ApJ*, **285**, 89, 1984.

———. *Proc 22nd ESLAB Symposium: IR Spectroscopy in Astronomy*, ed. B.H. Kaldeich, Noordwijk, Neth., ESA Pub Div, 1989.

Duley, W.W., in *Polycyclic Aromatic Hydrocarbons and Astrophysics*, eds. A. Léger, L. d'Hendecourt, N. Boccara, Reidel, Dordrecht, p.373, 1987.

Duley, W.W., Williams, D.A., *MNRAS*, **196**, 269, 1981.

Duley, W.W., Williams, D.A., Moorhouse, A., *MNRAS*, **253**, 505, 1991.

Finkenzeller, U., Mundt, R., *AA*, **55**, 109, 1984.

Forrest, W.J., Chen, H., Garnett, J.D., Solomon, S.L., Pipher, J.P., *SPIE Proceedings: Infrared Detectors and Instrumentation*, **1946**, 18, 1993.

Fowler, A.M., and Gatley, I., *SPIE Proceedings, Infrared Sensors: Detectors, Electronics and Signal Processing*, **1541**, 127, 1991.

———. *Ap.J.*, **353**, L33, 1990.

———. *SPIE Proceedings, Infrared Sensors: Detectors, Electronics and Signal Processing*, **1541**, 127, 1991.

Fuente, A., Martin-Pintado, J., Cernicharo, J., Brouillet, N., Duvert, G., *AA*, **260**, 341, 1992.

Garnett, J.D., Chen, H., Solomon, S.L., Forrest, W.J., Pipher, J.P., **IRAC TM91-3001**, April, 1991.

Garnett, J.D., Forrest, W.J., in *SPIE Proceedings, Infrared Detectors and Instrumentation*, **1946**, 395, 1993.

Gatley, I., Hasegawa, T., Suzuki, H., Garden, R., Brand, P., Lightfoot, J., Glencross, W., Okuda, H., Nagata, T., *ApJ*, **318**, L73, 1987.

Geballe, T.R., in *Stardust to Planetesimals*, ASP Conference Series, eds. Y.J. Pendleton and A.G.G.M. Tielens, **122**, 1997.

Geballe, T.R., Lacy, J.H., Persson, S.E., McGregor, P.J., Soifer, B.T., *ApJ*, **292**, 500, 1985.

Geballe, T.R., Tielens, A.G.G.M., Allamandola, L.J., Moorhouse, A., Brand, P., *ApJ*, **341**, 278, 1989.

Geballe, T.R., Tielens, A.G.G.M., Kwok, S., Hrivnak, B.J., *ApJ*, **387**, L89, 1992.

- Geib, K., Goodnick, D., Lin, D.Y., Gann, R.G., Wilmsen, C.W., *J.Vac.Sci.Technology*, **82**, 516, 1984.
- Grady, C.A., Imhoff, C.L., and Bjorkman, K.S., in *A Decade of UV Astronomy with IUE, Proc. Celebratory Symposium*, ed. E.J. Rolfe (ESA SP-281), **2**, p.109, 1988.
- Grove, A.S., Fitzgerald, *Solid State Electronics*, **9**, 783, 1966.
- Grove, A.S., *Physics and Technology of Semiconductor Devices*, John Wiley & Sons, New York, 1967.
- Habing, H.J., *Bull. Astron. Inst. Netherlands*, **19**, 421, 1968.
- Hall, R.N., *Phys.Rev.*, **87**, 387, 1952.
- Hartmann, L., Kenyon, S.J., Calvet, N., *ApJ*, **407**, 219, 1993.
- Heiman, F.P., Warfield, G., *IEEE Trans. Electron Devices*, **ED-12**, 167, 1965.
- Helou, G., Smith, E.A., *IPAC Interoffice Memorandum*, 240-87-167, 1987.
- Herbig, G.S., *ApJ Suppl.*, **4**, 337, 1960.
- Hillenbrand, L.A., Strom, S.E., Vrba, F.J., Keene, J., *ApJ*, **397**, 613, 1992.
- Huff, H., Kawaji, S., Gatos, S.C., *Surf. Sci.*, **5**, 399, 1966.
- . *Surf. Sci.*, **10**, 232, 1968.
- Jackson, J.D., *Classical Electrodynamics* (2nd ed.), John Wiley & Sons, New York, 1975.
- Jenkins, E., Savage, B., Spitzer, L., *ApJ*, **301**, 355, 1986.
- Joblin, C., Ph.D. Thesis, Université Paris, 7, 1992.
- Joblin, C., Tielens, A.G.G.M., Allamandola, L.J., Geballe, T.R., *ApJ*, **458**, 610, 1996.
- Jourdain de Muizon, M., d'Hendecourt, L.B., Geballe, T.R., *A&A*, **227**, 526, 1990.
- Kittel, C., *Introduction to Solid State Physics* (6th ed.), John Wiley & Sons, New York, 1986.
- Korneef, J., *AA*, 128, **84**, 1983.
- Lada, C.J., Adams, F.C., *ApJ*, **393**, 278, 1992.
- Landau, L., Lifschitz, *Quantum Mechanics - Non-Relativistic Theory*, Pergamon Press Inc., New York, 1977.
- Langan, J.D., Viswanathan, C.R., *J. Vac. Sci. Technol.*, **16**, 1474, 1979.
- Lax, M., *Phys.Rev.*, **119**, 1502, 1960.

- Léger, A., d'Hendecourt, L., in *Polycyclic Aromatic Hydrocarbons and Astrophysics*, eds. A. Léger, L. d'Hendecourt, N. Boccara, Reidel, Dordrecht, p.232, 1987.
- Léger, A., Puget, J.L., *A&A*, **137**, L5, 1984.
- Lemaire, J.L., Field, D., Gerin, M., Leach, S., Pineau des Forêts, G., Rostas, F., Rouan, D., *A&A*, **308**, 895, 1996.
- Margalit, S., Shappir, J., *Solid State Electronics*, **19**, 789, 1976.
- Martin, P.G., Whittet, D.C.B., *ApJ*, **357**, 113, 1990
- Mathis, J.S., Rumpl, W., Nordsieck, K.H., *ApJ*, **217**, 425, 1977
- Mathis, J.S., Mezger, P.G., Panagia, N., *AA*, **128**, 212, 1983.
- McNutt, M.J., Meyer, W.E., *J. Electrochemical Soc.*, **128**, 892, 1981.
- Milkey, R.W., Dyck, H.M., *ApJ*, **181**, 833, 1973.
- Mizutani, K., Suto, H., Maihara, T., *ApJ*, **346**, 675, 1989.
- Moorwood, A.F.M., *AA*, **166**, 4, 1986.
- Mouri, H., Kawara, K., Taniguchi, Y., Nishida, M., *ApJ*, **356**, L39, 1990.
- Nicollian, E.H., Brews, J.R., *MOS (Metal Oxide Semiconductor) Physics and Technology*, John Wiley & Sons, New York, 1982.
- Okamura, M., Minakata, M., *J. Appl. Phys.*, **57**, 2060, 1985.
- Osterbröck, D.E., *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei*, University Science Books, Mill Valley, CA, 1989.
- Piché, F., *A Near-Infrared Study of the Star Forming Region NGC2264*, Ph.D Dissertation, Univ. Washington, Seattle, Wa, 1992.
- Sah, C.T., *Proc. IEEE*, **55**, 654, 1967.
- Sah, C.T., Noyce, R.N., Shockley, W., *Proc. IRE*, **45**, 1228, 1957.
- Sakata, A., Wada, S., Tanabe, T., Onaka, T., *ApJ*, **287**, L51, 1984.
- Satyapal, S., "Probing the Dust Enshrouded Nuclei of the Starburst Galaxy Systems M82 and Arp 299: High Spatial Resolution Near-Infrared Fabry-Perot Imaging Observations." Ph.D. Dissertation, Univ. Rochester, Rochester, NY, 1995.
- S. Satyapal, D. M. Watson, J.L.Pipher, W. J. Forrest, D. Coppenbarger, S. N. Raines, S. Libonate, F. Piche, M. A. Greenhouse, H. A. Smith, K. L. Thompson, J. Fischer, C. E. Woodward and T. Hodge. *ApJ*, **448**, 623, (1995).

- Schutte, W.A., Tielens, A.G.G.M., Allamandola, L.J., *ApJ*, **415**, 1993.
- Seab, C., in *Interstellar Processes*, D.Reidel, Dordrecht, 492, 1987.
- Sellgren, K., Allamandola, L., Bregman, J., Werner, M., Wooden, D., *ApJ*, , **299**, 416 1985.
- Sellgren, K., *AJ*, , **88**, 985, 1983.
- . *ApJ*, , **277**, 623, 1984.
- . *ApJ*, , **305**, 399 1986.
- . *ApJ*, **245**, 138, 1981.
- Sellgren, K., Tokunaga, A.T., Nakada, Y., *ApJ*, **349**, 120, 1990.
- Sellgren, K., Werner, M.W., Allamandola, L.J., *ApJ Suppl.*, , **102**, 369, 1996.
- Sellgren, K., Werner, M.W., Dinerstein, H.L., *ApJ*, , **271**, L13, 1983.
- ., *ApJ*, , **400**, 238, 1992.
- Sewell, H., Anderson, J.C., *Sol. State Elect.*, **18**, 641, 1975.
- Shockley, W., Read, W.T., *Phys.Rev.*, **87**, 835, 1952.
- Siebenmorgen, R., Krugel, E., *AA*, **259**, 614, 1992.
- Simon, M., Felli, M., Cassar, L., Fischer, J, Massi, M., *ApJ*, **266**, 623, 1983.
- Skinner, S.L., Brown, A., Linsky, J.L., *ApJ*, **357**, L39, 1990.
- Spitzer, L., *Physical Processes in the Interstellar Medium*, John Wiley & Sons, New York, 1978.
- Strom, S.E., Strom, K.M., Grasdalen, G.L., *ARAA*, **13**, 187, 1975.
- Strom, S.E., Strom, K.M., Yost, J., Carrasco, L., Grasdalen, G.L., *ApJ.*, **173**, 353, 1972.
- Sze, S.M., *Physics of Semiconductor Devices* (2nd Ed.), John Wiley & Sons, New York, 1981.
- Tauber, J.A., Tielens, A.G.G.M., Meixner, M., Goldsmith, P.F., *ApJ*, **422**, 136, 1994.
- Tielens, A.G.G.M., Allamandola, L.J., in *Interstellar Processes*, eds, D.J. Hollenbach and H.A. Thronson, Reidel, Dordrecht, p.397, 1987.
- Tokunaga, A.T., *IRTF Photometry Manual*, 1986.
- Tokunaga, A.T., Sellgren, K., Smith, R.G., Nagata, T., Sakata, A., Nakada, Y., *ApJ*, **380**, 452, 1991.
- Valcheva, E.P., Germanova, K.G., *Superlattice and Microstructures*, **10**, 319, 1991.

- Wager, J.F., Geib, K.M., Wilmsen, C.M., Kazmerski, J., *J.Vac.Sci.Technology*, **81**, 778, 1983.
- Watt, G.D., Burton, W.B., Choe, S.U., Liszt, H.S., *AA*, **163**, 194, 1986a.
- Watt, G.D., Mundy, L.G., Schwartz, R.D., *ApJ (Letters)*, **303**, L61, 1986b.
- Whitcomb, S.E., Gatley, I., Hildebrand, R.H., Keene, J., Sellgren, K., Werner, M., *ApJ*, **246**, 416, 1981.
- Whittet, D.C.B., *Dust in the Galactic Environment*, Institute of Physics Publishing, Pennsylvania, 1992.
- Wilking, B.A., *PASP*, **101**, 229, 1989.
- Willner, S., Soifer, B., Russell, R., Joyce, R., Gillett, F., *ApJ Letters*, **217**, L121, 1977.
- Wilmsen, C.W., *J.Vac.Sci.Technol.*, **13**, 64, 1976.
- Witt, A., Cotrell, M.J., *ApJ*, **235**, 899, 1980.
- Witt, A., Schild, R., *ApJ Suppl*, **62**, 839, 1986.
- Zinnecker, H., McCaughren, M.J., Wilking, B.A., in *Protostars and Planets III*, ed E.H. Levy, J.I. Levine, and M.S. Matthews, University of Arizona Press, Arizona, 1992.

6. Appendix A - Optical Distortion

Optical distortion results in an image when the magnification (*i.e.* platescale) is a function of the distance from the optic axis. Introducing a stop into a lens system usually introduces distortion unless the stop is coincident with the optical center (e.g. for a single lens, if the stop is placed at one of the lens surfaces). Optical distortion comes in two flavors: “pin-cushion”, or positive distortion, where the lateral magnification increases with distance from the optic axis, and “barrel”, or negative distortion, where the lateral magnification decreases with off-axis distance. In general, positive optical elements exhibit positive distortion, and negative elements exhibit negative distortion.

Distortion has two adverse effects which need to be corrected in our images; the first is to correct the position of every pixel on the array relative to the optic axis, and the second is to correct for the fact that pixels further from the optic axis see a smaller solid angle than do pixels closer to the axis. This latter effect explains, in part, why our flat field images exhibit smaller signal with increasing distance from the optic axis (*i.e.* the platescale decreases with off-axis distance). This effect is also due to the fact that pixels further from the optic axis see a smaller solid angle due to the geometrical projection of the Lyot stop. Dividing a source frame by a flat field (both uncorrected for distortion) conserves the surface brightness of the source. In order to obtain the flux from a given source, the surface brightness must then be multiplied by the area of the pixel. In this way, a star near the corner of the array (having larger counts/pixel) has a smaller effective area in steradians so that the total flux from the star is the same at any position on the array. We have found a reasonable approximation for the distortion to be:

$$\delta = a \cdot \rho^3 \tag{A1}$$

where δ is the distortion (in pixels), ρ is the distance from the optic axis, and a is a constant that measures the amount of distortion in the image. The position of a pixel after correcting for (positive) distortion will be given by:

$$\begin{aligned}\rho' &= \rho - \delta \\ &= \rho \cdot (1 - a\rho^2)\end{aligned}\tag{A2}$$

One can see that after applying this correction, the area of a single pixel will be less than that of the “distorted” pixel. Calculating the area of the undistorted pixel relative to the distorted pixel provides the second correction, and is given by the Jacobian of the distortion transformation shown in equation (A2):

$$J = 1 - 4a\rho^2 + 3(a\rho^2)^2\tag{A3}$$

Thus, to correct a given flat-fielded pixel one simply multiplies the signal in that pixel by the value $|J|$. The reimaging optics used for these observations is referred to as a doublet, and consists of a single, positive silicon element with closely spaced, negative CaF₂ element. This achromatic doublet produced a negligible amount of pincushion distortion at the corners of focal plane.

7. Appendix B - Detailed Calculation of Inferred Spectrum

8. Appendix C - Detailed Calculation of Surface Potential