Gemini Planet Imager

Operational Concept Definition Document

V4.1

Lawrence Livermore National Laboratory
National Research Council, Herzberg Institute of Astrophysics
National Science Foundation Center for Adaptive Optics
University of California, Los Angeles
Jet Propulsion Laboratory
University of California, Santa Cruz
American Museum of Natural History
Université de Montréal
University of California, Berkeley

Revision History

Date	Revision No.	Revised By	Description	Approval
June 30/04	0.0	Tessa Ford	Draft template for discussion	
Oct 15/04	0.1	James Graham	Initial Draft	
Feb 19/05	1.0	Bruce Macintosh	Release version for ExAOC CoDR	Bruce Macintosh
2007-05-08	2.0	James Graham Bruce Macintosh	PDR update. Updated observing sequences, added polarimetry, Io monitoring, and minor planet companion survey. Updated FoV requirement. Moved background contamination, operating wavelength, spectral resolution, polarization, and system Mueller matrix to Appendices Updated instrument sections to reflect current design Updated science summary from current PDR Vol. 1, Ch. 1. Reformatting, final requirements update, PDR mask sizes	Leslie Saddlemyer Bruce Macintosh
2008-4-25	2.1	James Graham	Reviewed contrast requirements & updated references	
2008-05-07	2.2 – 2.6	James Graham, Bruce Macintosh,		Leslie Saddlemyer
2008-05-07	3.0 – 3.1	Leslie Saddlemyer	Table of COR elements updated to latest.	Leslie Saddlemyer
2008-05-08	3.2	Leslie Saddlemyer	Fix Lyot table with blank	Leslie Saddlemyer
2008-11-26	3.3	Bruce Macintosh	Post-CDR updates: coronagraph table;	
2008-12-12	4.0	Leslie Saddlemyer	CCB approvals all received for V4.0, ready for submission to Gemini	Leslie Saddlemyer
2009-04-27	4.1	Leslie Saddlemyer	deltaCDR changes, mainly removal of unnecessary TBDs	Leslie Saddlemyer

Table of Contents

1 Purpose

This Operational Concepts Definition Document (OCDD) describes the operational concept model for the Gemini Planet Imager (GPI). This document summarizes the science cases that the instrument is supposed to address, relates these to the design requirements, and discusses the key functional and performance requirements that the instrument must meet. Key operational scenarios of the GPI are identified and discussed, especially in terms of the requirements the instrument places on other parts of the Gemini system. These scenarios are described in sufficient detail for technically and scientifically skilled, but non-expert, readers to understand.

2 References

[1] "GPI Functional and Performance Requirements Document". Leslie Saddlemyer, NRC-HIA…
[2] "Request For Proposal for Conceptual Design Studies for an Extreme Adaptive Optics Coronagraph (GPI) RFP No. N231802". AURA/Gemini Observatory, 16 January 2004.

Ackerman, A. S., & Marley, M. S. 2001, ApJ, 556, 872
Aime, C., Soummer, R., & Ferrari, A. 2002, A&A, 389, 334
Alcock, C., et al. 2000, ApJ, 542, 281
Allard, F., Hauschildt, P. H., Alexander, D. R., Tamanai, A., Schweitzer, A. 2001, ApJ, 556, 357
Alonso, R., et al. 2004, ApJL, 613, 153
Angel, J. R. P. 1994, *Nature*, 368, 203
Anic, A., Alibert, Y., and Benz, W. 2007, A&A, 477, 717
Ardila, D. et al. 2004, ApJ, 617L, L147
Asphaug, E. 1999, Nature, 402, 127
Augereau J. C., Lagrange AM, Mouillet D, Menard F. 1999. A&A, 350, L51
Armitage, P. J., Livio, M., Lubow, S. H., & Pringle, J. E. 2002, MNRAS, 334, 248
Backman, D. E. & Paresce, F. 1993 in *Protostars and Protoplanets III* Eds. Levy, E. H. & Lunine, J. I., 1253-1304, U. of Arizona Press, Tucson
Balick, B. & Frank, A. 2002, ARA&A, 40, 439
Benz, W., Slattery, W. L., & Cameron, A. G. W. 1986, *Icarus*, 66, 515
Benz, W., Slattery, W. L., & Cameron, A. G. W. 1988, *Icarus*, 74, 516
Bloemhof, E. E., Dekany, R. G., Troy, M., & Oppenheimer, B. R. 2001, ApJ, 558, L71
Bonnell, I. A., Bate, M. R., Clarke, C. J., & Pringle, J. E. 2001, MNRAS, 323, 785
Bonnell, I. A., & Bate, M. R. 2002, MNRAS, 336, 659
Boss, A.P. 2002, ApJ, 576, 462
Bouchy, F., Melo, C., Santos, N. C., Mayor, M., Queloz, D., Udry, S. A&A, 431, 1105
Burgasser, A. J., et al. 1999, ApJ, 522, L65
Burgasser, A. J., et al. 2000a, AJ, 120, 473
Burgasser, A. J., et al. 2000b, ApJ, 531, L57
Burgasser, A. J., et al. 2000c, AJ, 120, 1100
Burgasser, A. J. et al. 2002 ApJ, 564, 421
Burgasser, A. J., Burrows, A. & Kirkpatrick, J. D. 2006, ApJ, 639, 1095
Burrows, A. J., et al. 1997, ApJ, 491, 856

Burrows, A., Marley, M. S., Sharp, C. M. 2000, ApJ, 531, 438
Burrows, A., Hubbard, W. B., Lunine, J. I., & Liebert, J. 2001, Rev. Mod. Phys., 73, 719
Burrows, A., Sudarsky, David., Lunine, J. I. 2003 ApJ, 596, 587
Burrows, A., Sudarsky, D., & Hubeny, I. 2004, ApJ, 609, 407
Butler, P. et al. 2003, ApJ, 582, 455
Butler, P., Wright, J. T., Marcy, G. W., Fischer D. A., Vogt, S. S., Tinney, C. G., Jones H. R. A., Carter, B. D., Johnson, J. A. McCarthy, C., Penny, A. J. 2006, ApJ 646 505
Cameron, A. G. W. 1995, *Meteoritics*, 30, 133
Canup, R. M. 2005, *Science*, 307, 5709
Charbonneau, D., et al. 2005. ApJ *in press*, 20 June issue.
Chabrier, G., Brassard, P., Fontaine, G., & Saumon, D . 2000, ApJ, 543, 216
Chauvin, G., Lagrange, A.-M., Dumas, C., Zuckerman, B., Mouillet, D., Song, I., Beuzit, J.-L.& Lowrance, P. 2004, A&A, 425, L29
Chiang, E.I., Fischer, D., & Thommes, E. 2002, ApJL, 564, 105
Clampin, M., Ford, H. C., Illingworth, G., Petro, L., JPF Science Team 2001 BAAS 199 3302
Clampin, M., et al. 2003, AJ, 126, 385
Close, Laird M., Lenzen, Rainer, Guirado, Jose C., Nielsen, Eric L., Mamajek, Eric E., Brandner, Wolfgang; Hartung, Markus, Lidman, Chris, Biller, Beth. 2005, Nature, 433, 286
Cochran, William D. & Hatzes, A. P. 1993, ASPC, 36, 267
Cumming, A., Marcy, G. W., Bulter, P., & Vogt, S. S., 2008, PASP *in press—arXiv:0803.3357*
Delfosse, X., et al. 1997, A&A, 327, L25
Deming, D., Seager,S., L. J. Richardson & Harrington, J. 2005, *Nature*, in press
Dermott, S. F., Jayaraman, S., Xu, Y. L., Gustafson, B. Å. S., & Liou, J.-C. 1994, *Nature*, 369, 719
Duchêne, G., McCabe, C., Ghez, A. M., & Macintosh, B. A. 2004, ApJ, 606, 969
Endl, M., Kürster, M., Els, S. Hatzes, A. P., Cochran, W. D., Dennerl, K.. Döbereiner, S. 2002, A&A, 392, 671
Frank, A., Balick, B., Icke, V., Mellema, G. 1993 ApJ, 404L, 25
Fischer, D. A., Marcy, G. W., Butler, R. P., Vogt, S. S., Apps, K. 1999, PASP, 111, 50
Fischer, D. A., et al. 2003, ApJ 586 1394
Fitzgerald, M., Kalas. P., G. & Graham, J. R. 2007, ApJ *in press*
Gammie, C.F. 2001, ApJ, 553, 174
Gates, E. et al. 2004 ApJL, 612, 129
Give'on, A., Kasdin, N. J., Vanderbei, R. J., & Avitzour, Y. 2004, Proc. SPIE, 5490, 1438
Gold, T. 1975, Icarus, 25, 489
Goldreich, P., Lithwick, Y., & Sari, R. 2004, ARAA, 42, 549
Goldreich, P., & Sari, R. 2003, ApJ, 585, 1024
Goldreich, P. & Tremaine, S. 1980, ApJ, 241, 425
Goldreich, P., Lithwick, Y. & Sari, R. 2004, ARAA, 42, 549
Golimowski, D. A., et al. 2006, AJ, 131, 3109
Graham, J. R., Kalas, P., & Matthews, B. 2007, ApJ, 654, 595
Greaves, J. S., et al. 1998, ApJ, 506, L133
Grillmair, C. J., et al. 2007, ApJ, 658L, 115
Hansen, B. M. S. et al. 2004 ApJSS, 155, 551
Harrington, J. P. & Borkowski, K. J. 1994, AAS, 185, 9003
Hartkopf, W. I., McAlister, H. A., Mason, B. D. 2001, AJ, 122, 3480

Hartkopf, W. I., Mason, B. D., Wycoff, G. L., & McAlister, H. A. 2006, http://ad.usno.navy.mil/wds/int4.html
Heacox, W. D. 1999, ApJ, 526, 928
Hestroffer, D., Marchis, F., Fusco, T., Berthier, J., 2002, A&A, 394, 339
Hodgkin, S. T., Oppenheimer, B. R., Hambly, N. C., Jameson, R. F., Smartt, S. J., & Steele, I. A. 2000, *Nature*, 403, 57
Holland, W. S., et al. 1998, *Nature*, 392, 788
Holland, W. S., et al. 2003 ApJ, 582, 1141
Hayashi, C. 1981, Prog. Theor. Phys. Suppl., 70, 35
Hubickyj, O. Bodenheimer P. & Lissauer, J. J. 2004, Rev. Mex. A. & A. Conf. Series, 22, 83
Hubickyj, O., Bodenheimer, P., Lissauer, J. J. 2005, Icarus, 179, 415
Ida, S. & Lin, D. N. C. 2005, ApJ, 626, 1045
Jackson, A. A., & Zook, H. A. 1989, *Nature*, 337, 629
Jacquinot, P. & Roizen-Dossier, B. 1964, Progress in Optics, Vol. 3 (ed. E. Wolf)
Kasdin, N. J., Vanderbei, R. J., Spergel, D. N., & Littman, M. G. 2003, ApJ, 582, 1147
Kratter, K. M., & Matzner, C. D. 2006, MNRAS, 373, 1563
Krumholz, M. R., Klein, R. I., & McKee, C. F. 2007, ApJ, 656, 959
Johnson, B.M., & Gammie, C.F. 2003, ApJ, 597, 131
Jones, H. R. A., Butler, P. R., Marcy, G. W., Tinney, C. G., Penny, A. J., McCarthy, C., Carter, B. D. 2002 MNRAS, 337 1170
Jorissen, A., Mayor, M., Udry, S. 2001, A&A, 379 992
Jura, M. 2004, ApJ, 603, 729
Kalas, P., Graham, J. R., Clampin, M. 2005, *Nature*, 435, 1067
Kalas, P., Fitzgerald, M. P., & Graham, J. R. 2007, astro-ph/0704.0645
Kant, I., 1755 Allgemeine Naturgeschichte des Himmels, Konigsberg und Leipzig:Johann Friderich Petersen. (English translation:W. Hastie, Universal Natural History and Theories of the Heavens. In Kant's Cosmology. New York:Greenwood publishing (1968)Kastner, J. H., Zuckerman, B., Weintraub, D. A. & Forveille, T. 1997 *Science*, 277, 67
Kirkpatrick, J. D., et al. 1999, ApJ, 519, 802
Kirkpatrick, J. D., et al. 2000, AJ, 120, 447
Koerner, D. W., Sargent, A. I., & Ostroff, N. A. 2001, ApJ, 560, L181
Korycansky, D. G., Zahnle, K. J., & Mac Low, M.-M. 2000, *Icarus*, 146, 387
Korzennik, S. G., Brown, T. M., Fischer, D. A., Nisenson, P., Noyes, R. W. 2000, ApJL, 533, 147
Krist, J. E., Stapelfeldt, K. R., Ménard, F., Padgett, D. L., & Burrows, C. J. 2000, ApJ, 538, 793
Kroupa, P. 2001, MNRAS, 322, 231
Kuchner, M. J., & Holman, M. J. 2003 ApJ, 588, 1110
Landgraf, M., Liou, J.-C., Zook, H. A., Grün, E. 2002, AJ, 123, 2857
Laplace, P. S., 1795 Celestial Mechanics, Paris
Latham, D. W. 1998 Highlights of Astronomy Vol. 11A, as presented at Joint Discussion 14 of the XXIIIrd General Assembly of the IAU, 1997. Ed. Johannes Andersen. Kluwer Academic Publishers, p.419
Laughlin, G. L., Bodenheimer, P. & Adams, F. C. 2004, ApJL, 612, 73
Lecavelier Des Etangs, A., Scholl, H., Roques, F., Sicardy, B., & Vidal-Madjar, A. 1996, Icarus, 123, 168
Leggett, S. K., Ruiz, M.-T., & Bergeron, P. 1998, ApJ, 497, 294
Leinert, C., Richter, I., Pitz, E., & Planck, B. 1981, A&A, 103, 177L

Lin, D. N. C. & Ida, S. 1997, ApJ, 477, 781
Lineweaver, C. H., & Grether, D. 2002, AstroBio, 2, 325
Lineweaver, C. H., & Grether, D. 2003 ApJ, 598, 1350
Liou, J. -C. J. -C., Zook, H. A., 1999, AJ, 118, 580
Lissauer, J. J., & Safronov 1991, *Icarus*, 93, 288
Lynden-Bell, D. & Pringle, J. E 1974, MNRAS, 168, 603
Macintosh, Bruce A.; Becklin, E. E.; Kaisler, Denise; Konopacky, Quinn; Zuckerman, B. 2003, ApJ., 594, 538
Malbet, F., Yu, J. W., & Shao, M. 1995, PASP, 107, 386
Mamajek E. E., Lawson W. A., Feigelson E. D., 1999, ApJ, 516, L77
Marchis, F., de Pater, I., Davies, A.G., et al., 2002, *Icarus*, 160, 124
Marchis, F., Descamps, P., Hestroffer, D., Berthier, J., Vachier, F., et al. 2003, *Icarus*, 165, 112
Marchis, F., Descamps, P., Hestroffer, D., Berthier, J. 2005, *Nature*, 436, 822
Marcy, G. W., & Butler, R. P. 2000, PASP, 112, 137
Marcy, G. W., Butler, R. P., Fischer, D. A., Vogt, S. S. 2003, ASPC, 294, 1
Marcy, G. W. 2008 http://exoplanets.org
Marcy, G. W., et al. 2005, ApJ, 619, 570
Marley, M., et al. 1996, *Science*, 272 1919
Marley, M. Seager, S., Saumon, D., Lodders, K., Ackerman, A. S., Freedman, R. S., Fan, X. 2002, ApJ 568, 335
Marley, M. S., Fortney, J. J., Hubickyj, O., Bodenheimer, P., & Lissauer, J. J. 2007, ApJ, 655, 541
Matsuyama, I., Johnstone, D., & Murray, N. 2003, ApJ, 585L, 143
Mainzer, A. K., McLean, I. S., Sievers, J. L., & Young, E. T. 2004, ApJ, 604, 832
Marois, C., Doyon, R., Racine, R., & Nadeau, D. 2002, PASP, 112, 91
Marsh K. A., Silverstone M. D., Becklin E. E., Koerner D. W., Werner M. W., et al. 2002 ApJ, 573, 425
Mastrodemos, N. & Morris, M. 1999 ApJ, 523, 357
Marois, C. Macintosh, B., & Barman, T. 2007, ApJ, 654, L151
Marzari, F., & Vanzani, V. 1994, A&A, 283, 275
Marzari, F., & Weidenschilling, S.J. 2002, *Icarus*, 156, 57
Mayor, M. et al. 2003 *ESO Messenger*, 114, 20
Mayor, M. et al. 2004, A&A, 415, 391
McEwen, A.S., Keszthelyi, L., Spencer, J.R., et al., 1998, *Science*, 281, 87
McKee, C. F., & Tan, J. 2003, ApJ, 585, 850
McGovern, M R., Kirkpatrick, J. D., McLean, I. S., Burgasser, Adam J.. Prato, L., Lowrance, P. J. 2004, ApJ, 600, 1020
Merline, W.J., Weidenschilling, S. J., Durda, D. D., Margot, J.-L., Pravec, P., Storrs, A.D., Asteroids III, 289
Metchev, S. A., et al. astrop-ph/0412143
Mills, J. P., Gaiser, S. L., Diner, D. J., Watson, S. M. 1991 ApplOpt 30 3253
Monnier, J. D., Tuthill, P. G., & Danchi, W. C. 2002, ApJ, 567L, 137
Mouillet D, Lagrange A. M., Augereau J. C., Menard F. 2001. A&A, 372, L61
Murray, N., Hansen, B., Holman, M., & Tremaine, S. 1998 *Science*, 279, 69
Nakajima, T., Oppenheimer, B. R., Kulkarni, S. R., Golimowski, D. A., Matthews, K., Durrance, S. T. 1995 *Nature*, 378, 463

Neuhäuser, R., Guenther, E. W., Wuchterl, G. Mugrauer, M., Bedalov, A., & Hauschildt, P. H. 2005, A&A, 435, L13
Nidever, D. L., Marcy, G. W., Butler, P. R., Fischer, D. A., Vogt, S. S. 2002, ApJS, 141 503
Nisenson, P. & Papaliolios, C. 2001, ApJ, 548, L201
Oppenheimer, B. R., Kulkarni, S. R., Matthews, K., Nakajima, T. 1995 *Science*, 270, 1478
Oppenheimer, B. R., et al. 2001, ApJ, 550, 448
Oppenheimer, B. R., et al. 2001b, *Science*, 292, 698
Ozernoy, L. M., Gorkavyi, N. N., Mather, J. C., & Taidakova, T. A. 2000, ApJ, 537, L147
Papaloizou, J. C. B. & Terquem, C. 2001, MNRAS, 325, 221P
Pearl, J.C., Conrath, B.J., 1991, JGR, 96, 18921
Perrin, M. D., Sivaramakrishnan, A., Makidon, R. B., Oppenheimer, B. R., & Graham, J. R. 2003, ApJ, 596, 702
Pollack, J. B., et al. 1996, Icarus, 124, 62
Pravec, P. & Harris, A. W. 2007, Icarus, *in press*
Poyneer, L. A. & Macintosh, B. 2004, JOSA A, 21, 810
Quillen, A. C., & Thorndike, S. 2002, ApJ, 578, L149
Racine, R., Walker, G. A. H., Nadeau, D., Doyon, R. & Marois, C. 1999, PASP, 111, 587
Rages, K.A., Hammel, H.B., Friedson, A.J., 2004, *Icarus*, 172, 2, 548
Rasio, F. A. & Ford, E. B. 1996, *Science*, 274, 954
Reach, W. T. 1991, ApJ, 369, 529
Reach, W. T., et al. 1995, *Nature*, 374, 521
Reid, I. N. 1994, Ap&SS, 217, 57
Reid, I. N., et al. 2004. AJ, 128, 463
Richardson, L. J., et al. 2007, Nature, 445, 892
Roques, F., Scholl, H., Sicardy, B., & Smith, B. A. 1994, *Icarus*, 108, 37
Sano, T., Shoken, M.M., Umebayashi, T., & Nakano, T. 2000, ApJ, 543, 486
Sasselov, D. D., & Lecar, M. 2000, ApJ, 528, 995
Sato, B., Fischer, D., Henry, G., & Laughlin, G. et al. 2005, ApJ, 633, 465
Saumon, D., Marley, M. S., Lodders, K., & Freedman, R. S. 2003, IAUS, 211, 345
Schneider, G., et al. 1999, ApJ, 513, L127
Schneider, G., et al. 2004, American Astronomical Society Meeting Abstracts, Vol. 205
Shu, F. H., Adams, F. C., & Lizano, S. 1987, ARAA, 25, 23
Sivaramakrishnan, A., Koresko, C. D., Makidon, R. B., Berkefeld, T., & Kuchner, M. J. 2001, ApJ, 552, 397
Sivaramakrishnan, A., Lloyd, J. P., Hodge, P. E., and Macintosh, B. A. 2002, ApJ, 581, L59
Skrutskie, M. F., et al. 1997, The Impact of Large-Scale Near-IR Sky Surveys, ed. F. Garzon et al. (Dordrecht: Kluwer), p. 25
Slattery, W. L., Benz, W., & Cameron, A. G. W. 1992, *Icarus*, 99, 167
Smith, N., Campbell, R., Kuluhiwa, K., LeMignant, D., & Chaffee F. 2005 *private communication*
Soker, N. 1997 ApJSS, 112, 487
Song, I., Zuckerman, B., & Bessell, M. S. 2003, ApJ, 599, 342
Soummer, R., Aime, C., & Falloon, P. 2003, A&A, 397, 1161
Spangler, C., Sargent, A. I., Silverstone, M. D., Becklin, E. E., Zuckerman, B. 2001, ApJ, 555, 932.
Sparks, W. B. & Ford, H. C. 2002, ApJ, 578, 543
Stahler, S. W, Shu, F. H., & Taam, R. E. 1980, ApJ, 241, 637
Stepinski, T. F., Black, D. C. 2001, AA, 371, 250

Stern, A. S. 1992, ARA&A, 30, 185
Stevenson, D. J. 1987, Ann. Rev. Earth Plan. Sci., 15, 271
Strauss, M. A., et al. 1999, ApJ, 522, L61
Sudarsky, D., Burrows, A., & Pinto, P. 2000, ApJ, 538, 885
Tabachnik, S., & Tremaine, S. 2002 MNRAS, 335, 151
Takeuchi T. & Artymowicz P. 2001 ApJ, 557, 990
Torres, G., Konacki, M., Sasselov, D. and Jha, S. 2005, ApJ, 619, 558
Trauger, J., Backman, D., Brown, R. A., Burg, R., Burrows, A., Ftaclas, C., Kulkarni, S., Kirkpatrick, D., Lunine, J., Marcy, G., et al. 2000, BAAS 194 4907
Trilling, D. E., Lunine, J. I., & Benz, W. 2002, A&A, 394, 241
Tuthill, P. G., Monnier, J. D., Danchi, W. C., Wishnow, E. H., Haniff, C. A. 2000, PASP, 112, 555
Udry, S., Mayor, M., Naef, D., Pepe, F., Queloz, D., Santos, N. C., Burnet, M., Confino, B., Melo, C. 2000, A&A, 356, 590
Vogt, S. S., Marcy, G. W., Butler, R. P., Apps, K. 2000, ApJ, 536, 902
Wahhaj, Z., Koerner, D. W., Ressler, M. E., Werner, M. W., Backman, D. E., Sargent, A.I. 2003, ApJ, 584, L27
Weidenschilling, S. J. 1977, MNRAS, 180, 57
Weidenschilling, S. J. & Marzari, F. 1996, *Nature*, 384, 619
Weinberger, A.J., Becklin, E.E., Schneider, G., Smith, B. A., Lowrance P. J., et al. 1999, ApJ, 525, L53
Weinberger, A. J., et al. 2002, ApJ, 566, 409
Weinberger, A. J., Becklin, E. E., & Zuckerman, B. 2003 ApJ 584, L33
Wetherill, G. 1985, LPI, 16, 901
Willson, L. A. 2000, AARA, 38, 573
Wilner, D. J., Holman, M. J., Kuchner, M. J., & Ho, P. T. P. 2002, ApJ, 569, L115
Wu, Y. & Murray, N. 2003 ApJ 589 605
Wright, J. T., Marcy, G. W., Butler, R. P., & Vogt, S. S. 2004, ApJS, 152, 261
Zapatero-Osorio, M. R., et al. 2000, *Science*, 290, 103
Zucker, S. & Mazeh, T. 2001, ApJ, 562, 1038
Zuckerman B., Forveille T., & Kastner J. H. 1995 *Nature*, 373, 494
Zuckerman, B., 2001 ARAA, 39, 549
Zuckerman, B. and Song, I. 2004a ApJ 603 738
Zuckerman, B. and Song, I. 2004b AARA 42 685

3 Acronyms and Abbreviations

ADI	Angular Differential Imaging (speckle suppression)
ALTAIR	ALTitude-conjugate Adaptive optics for the InfraRed
AMNH	American Museum of Natural History
AO	Adaptive Optics
AU	Astronomical Unit (1.495 × 10^{11} m)
AURA	Association of Universities for Research in Astronomy
CfAO	Center for Adaptive Optics
CWS	Cold Working Surface

DHS	Data Handling System
EPICS	Experimental Physics and Industrial Control System
GPI	Extreme Adaptive Optics Coronagraph
FOV	Field of View
FPRD	Functional and Performance Requirements Document
GIS	Gemini Interlock System
HIA	Herzberg Institute of Astrophysics
ICS	Instrument Control System
IFU	Integral Field Unit
INO	Institut national d’optique
IRAS	Infra-Red Astronomy Satellite
ISO	Infra-Red Space Observatory
ISS	Instrument Support Structure
ITC	Integration Time Calculator
IWD	Inner Working Distance
JPL	Jet Propulsion Laboratory
LAO	Laboratory for Adaptive Optics
LLNL	Lawrence Livermore National Lab
mas	Milli-arcseconds (4.848×10^{-6} radians)
MTF	Modulation Transfer Function
MWI	Multi-Wavelength Imager
OCDD	Operational Concepts Definition Document (this document)
MCAO	Multi-Conjugate Adaptive Optics
MEMS	Micro Electro-Mechanical System
M_J	Mass of Jupiter (1.8988×10^{27} kg)
NICI	Near Infrared Coronagraphic Imager
OIWFS	On-Instrument Wavefront Sensor
OWD	Outer Working Distance
pc	Parsec (3.085×10^{16} m)
RMS	Root Mean Square
SDI	Spectral Differential Imaging (speckle suppression)
SCS	Gemini Telescope Secondary Control System
SNR	Signal to Noise Ratio
PSF	Point-Spread Function
SFWFS	Spatially Filtered Wave-Front Sensor
STScI	Space Telescope Science Institute
TPF-C	Terrestrial Planet Finder-Coronagraph
UCLA	University of California, Los Angeles
UdeM	Université de Montréal
VLT	Very Large Telescope
WFE	Wave Front Error
WFS	Wave Front Sensor

4 Introduction

The Gemini 8-m telescopes are among the first generation of telescopes designed to operate with adaptive optics (AO) systems and thereby deliver an unprecedented combination of large collecting area, low infrared background and high angular resolution. This has been demonstrated with Hokupa'a and ALTAIR on Gemini North, and soon will be extended with the commissioning of NICI on Gemini South. These AO systems deliver moderate levels of wavefront correction for $V \sim 12$ mag. and partial correction for stars as dim as $V \sim 15$ mag. The Gemini South Multi-Conjugate Adaptive Optics (MCAO) system is being designed to overcome the limited field of view of classical AO systems; MCAO will provide uniform, near diffraction-limited image quality at near-infrared wavelengths across a two arc minute diameter field-of-view. Multiple laser beacons will be deployed to provide broad sky coverage. By contrast, GPI delivers high precision and accuracy (< 5 nm RMS) AO correction towards a bright ($I \leq 9$ mag.), on-axis, natural guide star. This high fidelity correction will be used to feed a coronagraph, which suppresses diffraction, and in turn a science camera with the ability to cancel residual speckles in the final image. The result is a system capable of very high dynamic range ($> 10^7$) imaging from a few λ/D of the target star out to the edge of the control radius ($\lambda/2d$). At this contrast level a whole array of novel science avenues become accessible, including the direct imaging of exoplanets and circumstellar debris disks.

4.1. Technical Summary

A planet-detection AO system is not merely a conventional AO system with higher actuator count and Strehl ratio. GPI is fundamentally different from a classical AO system in two respects.

First, what matters is not the RMS wavefront error or the Strehl ratio, but the wavefront error over a specific range of spatial frequencies. A sinusoidal wavefront error with spatial frequency f in cycles/meter measured on the primary mirror will scatter light to a field angle in the PSF given by $\theta = \lambda f$ (Malbet et al 1995.) Consequently, by minimizing wavefront error over a range of spatial frequencies, we can create a PSF with a “dark hole” region where planets can be seen. Hence, GPI must be optimized to control the dominant mid-frequency ($0.25\ \mathrm{m}^{-1} < f < 3\ \mathrm{m}^{-1}$) sources of wavefront error, such as temporal bandwidth errors and wavefront sensor aliasing errors.

Second, the most detrimental form of wavefront errors to final image contrast are not random errors, such those induced by the atmosphere, but quasi-static errors. Examples of persistent wavefront errors include systematic wavefront calibration errors and uncorrected wavefront errors on optical surfaces. Such errors completely dominate the contrast-performance of current AO systems. To achieve our contrast goals, we must reduce these quasi-static wavefront errors to below 2-3 nm RMS over the mid spatial frequency range—compared to 50–100 nm for existing AO systems. We will achieve this goal with a combination of high-quality optics and a slow, accurate, and precise wavefront sensor, operating as close as possible to the science instrument in wavelength and location.

GPI uses two deformable mirrors in series, configured as a "woofer" and "tweeter", to provide fast, high stroke, and spatially dense wavefront control across the telescope pupil. Nyquist sampling of the wavefront on a spatial scale of 36 cm yields a dark hole diameter of 1.9 arc seconds at *H* (1.65 µm) and 2.5 arc seconds at *K* (2.2 µm). The wavefront is controlled in real time at up to 2.0 kHz by a spatially filtered visible-light (0.7–0.90 µm) Shack-Hartmann sensor. The precision infrared calibration sensor maintains the accuracy of the wavefront delivered to the science instrument at the 1-2 nm RMS level by feeding back offsets to the wavefront controller at a rate of ~ 1 Hz. The contrast performance of the design adopted for CDR is depicted in Figure 4-1, and an illustration of typical data is shown in Figure 4-2.

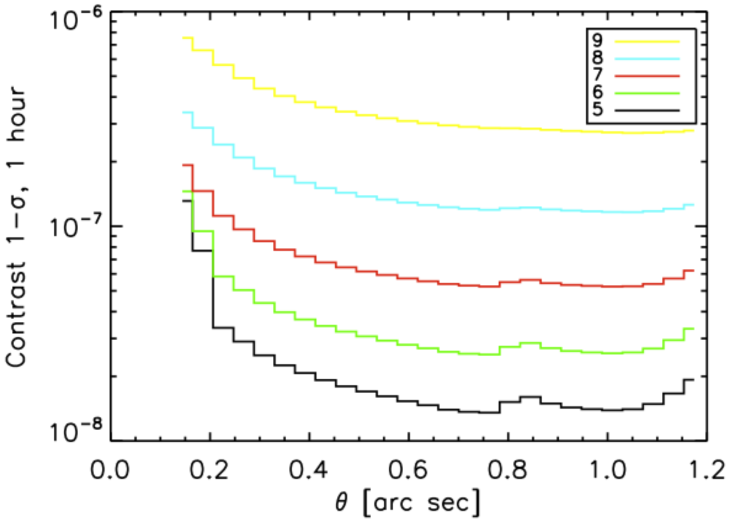

Figure 4-1: The final achieved 1-σ contrast in *H* for the final adopted CDR system including ADI and SDI speckle suppression. This calculation is for a 2.5% band-pass, although the coronagraph performance is nearly achromatic across the *H* band. The integration time is one hour. The system performance is a function of guide star *I*-band magnitude, which is shown in the legend at the upper right. The coronagraph architecture is an apodized pupil Lyot coronagraph. In each case, the atmosphere uses the standard Gemini recipe for Cerro Pachon and each layer has different wind speeds and directions. The total Fried parameter (at 500 nm) is 14.5 cm, which is appropriate for Cerro Pachon.

Light can be scattered by diffraction. Furthermore, wavefront errors and diffraction interact through the phenomenon of speckle pinning. High contrast can only be achieved by jointly controlling phase and amplitude. In our architecture, AO controls phase errors and a coronagraph controls the amplitude of the coherent part of the wavefront. Consequently, before the corrected wavefront is delivered to the science instrument the amplitude is controlled by an apodized-pupil Lyot coronagraph. This uses wavelength-specific combinations of pupil apodizers that mildly taper the transmission across the enterance aperture with focal-plane masks (FPM) that remove the central PSF core. These combine to channel the coherent part of the target starlight outside of a re-imaged Lyot pupil stop. Finally, the corrected and apodized wavefront is delivered to a science camera that provides simultaneous multiwavelength imaging (MWI), which is implemented as an integral field spectrometer (IFS). Speckles caused by phase errors at a pupil exhibit radial chromatic displacement in the focal plane whilst the location of genuine astrophysical sources remains fixed with wavelength. GPI uses this distinction to suppress the noise associated with residual speckles in post-detection image processing (ADI and SDI).

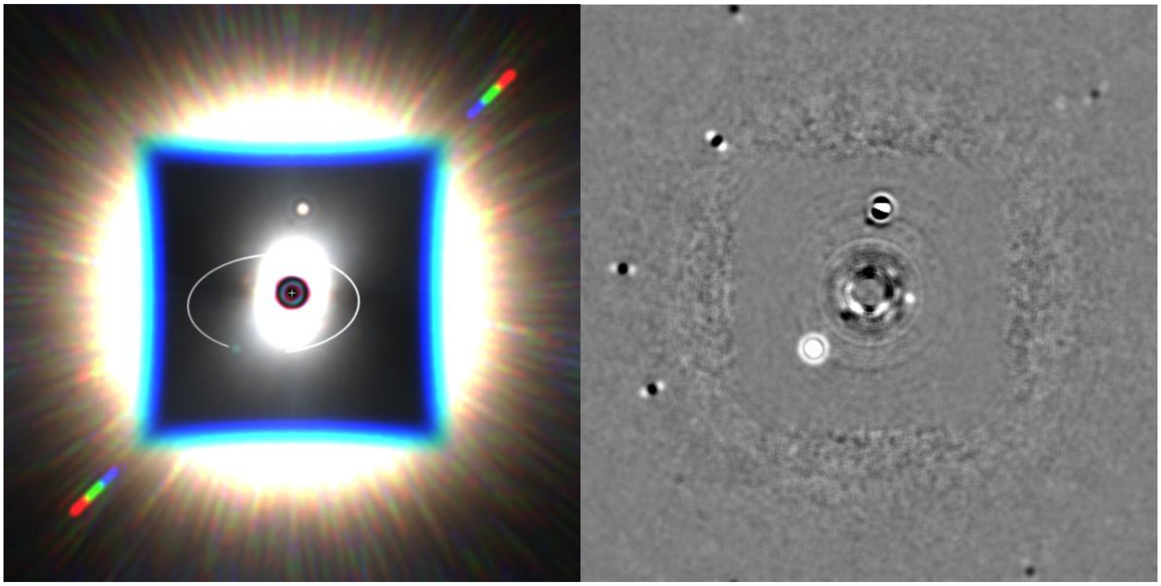

Figure 4-2: Simulated 1-hr GPI exposure of a G2V star at 20 pc. The star is 75-Myr old and hosts two planets: a 4 M_J planet at 11 AU (7 o'clock) and 1 M_J at 6 AU (3 o'clock). Background stars and dust artifacts are also visible. *Left:* Composite color-coded 1.48, 1.57, & 1.78 µm image. The elliptical trace shows the orbit of the 4 MJ planet. *Right:* Image after spectral differencing. This simulation includes both dynamic atmospheric errors and quasi-static GPI errors including Fresnel effects, Gemini M1 reflectivity variations, CAL system residuals, and dust on optical surfaces.

The IFS uses a single HAWAII-2RG 2048 × 2048 HgCdTe/CdZnTe molecular beam epitaxy detector to Nyquist sample, at 14 mas, the high-resolution direction of the apodized point spread function in the final science image at 1 µm and yields a square field of view of 2.8 × 2.8 arc seconds. The spectral resolution at *H* band, $R = \lambda/\delta\lambda$, is about 45 per two pixels, and the instantaneous band pass is 20%, e.g., 1.5–1.8 µm. This combination is suitable for speckle suppression, and establishing the effective temperatures and surface gravities of cool (T_{eff} < 1200 K) stars and warm planets (T_{eff} > 120 K). A comprehensive suite of broadband filters is available for operation between 0.96 µm and 2.4 µm. GPI combines high throughput (≈ 15%) with low background so that stellar halo photon shot noise is the dominant noise source. The primary potential source of background is the coronagraph apodizer mask, which has a 35% emissivity. A sub-array on HAWAII-2RG detector may be read out during the integration to yield unsaturated images near the core of the PSF and increase dynamic range.

5 Scientific Overview

This section contains a summary of the GPI science case, which is developed in full the PDR documentation, Volume 1.

5.2. Architecture of planetary systems

As of this writing, Marcy (2008) lists 228 extrasolar planets and 5% of targeted stars possess massive planets. Doppler surveys show that a diversity of exoplanet systems exists (Figure 5-1), but they leave several long-standing questions about planetary systems unanswered: How do planets form? Is the solar system typical? What is the abundance of solar-like systems? Doppler surveys also raise a host of new questions, including what produces the dynamical diversity in exoplanetary systems? Direct imaging can answer these questions by offering a fast alternative to Doppler surveys for searching the greatest stellocentric distances for planets. Characterizing the frequency and orbital geometries of planets beyond 3 AU will finally enable us to answer whether orbital configurations like our own planetary system are commonplace. Direct detection will also reveal the zone where planets may form by direct gravitational instability and uncover traces of planetary migration.

The abundance of circumstellar disks suggests that the frequency of planetary systems may be as high as 15 to 50%: a range that is defined by the occurrence of debris disks and protostellar disks, respectively. The low detection rate of planets may be a consequence of the biases inherent to search methods that detect orbital motion. So far, none of the Doppler surveys has the precision or lifetime necessary to detect Jupiter and thus do not yet constrain the frequency of solar system analogs. For a reliable detection, a significant fraction of an orbital period must elapse. For example, in the Keck radial velocity search, which began in 1996 July, only planets with $a < 4.9$ AU have completed one orbit (Butler et al. 2006). The median semimajor axis of known exoplanet orbits is 0.93 AU and only two planets have $a > 6$ AU (HD 24040 b at 6.85 AU, and HD 154345 b at 10.4 AU). The sample of exoplanets appears to be incomplete for $a > 3$ AU and the underlying distribution of planets in log(a) is at least flat (Stepinski & Black 2001) if not rising (Tabachnik & Tremaine 2002; Lineweaver & Grether 2003; Cumming et al. 2008). Thus, a direct-imaging search of outer solar system regions (4-40 AU), such as proposed here, would increase the total number of planets found relative to those in inner solar system orbits (0.4-4 AU). For a surface density law that meets the requirements of the minimum solar nebula ($\Sigma \sim r^{-3/2}$) such a search would approximately quadruple the total number of known planets. Our first goal is to assemble the first statistically significant sample of exoplanets that probes beyond the reach of indirect searches and quantifies the abundance of solar systems like our own.

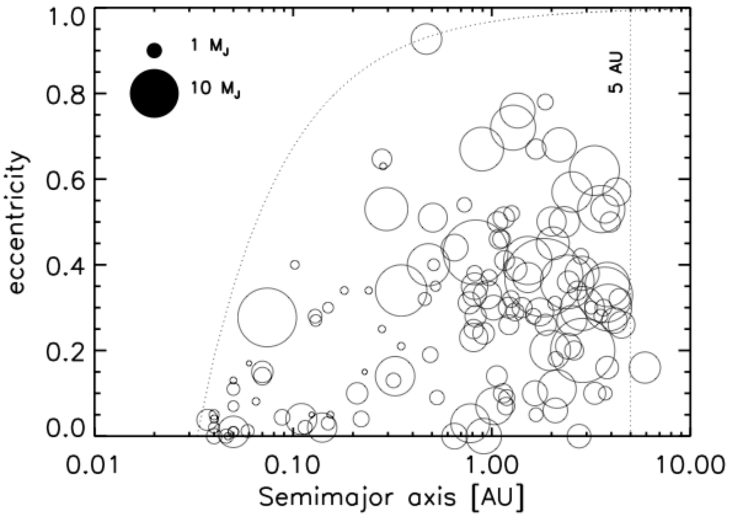

Figure 5-1: Eccentricity vs. semimajor axis distribution of Doppler-detected exoplanets. Solar system planets occupy a sliver of low eccentricity phase space ($0.006 < e < 0.21$) along the x-axis. Indirect planet searches, which rely on Kepler's third law, are restricted by the duration of the search to small semimajor axis separation. For a reliable detection, a significant fraction of one orbit must elapse (P_{Jupiter} = 11.9 yr, P_{Neptune} =163.7 yr.)

A second and fundamental motivation to image the outer regions of solar systems is to sample the regions where Jovian planets are thought to form and quantify the greatest distance out to which giant planets can form. The location of the region of interest depends on at least two competing factors: timescales for planet building and the availability of raw material. Dynamical and viscous time scales in the disk are shorter at small radii, while for typical surface density laws the amount of mass increases with radius, with a jump in the abundance of solid material beyond the "snow line" where ices condense. The change in the surface density of solid material occurs at 2.7 AU in the Hayashi model (Hayashi 1981). The location of this boundary depends on the disk structure (Sasselov & Lecar 2000) but for solar type stars, the zone of interest is beyond that which is readily probed by the Doppler method. The discovery of giant planets far beyond the snow line would tend to favor theories of planet formation by gravitational instability over solid core condensation and accretion. At larger orbital radii (> 30 AU), gas-cooling times become shorter than the Keplerian shearing time—a necessary condition for runaway gravitational instability (Gammie 2001; Johnson & Gammie 2003; Boss 2002)—while solid core growth by collisional coagulation of planetesimals proceeds prohibitively slowly (Goldreich, Lithwick, & Sari 2004).

A third reason to image the outer regions of extrasolar systems is to probe them for vestiges of planetary migration. Ninety percent of the Doppler sample consists of massive planets with $a < 3$ AU, suggesting that they migrated inwards to their present locations. A variety of mechanisms may drive orbital evolution; the tidal gravitational interaction between the planet and a viscous disk (Goldreich & Tremaine 1980), the gravitational interaction between two or more Jupiter mass planets (Rasio & Ford 1996), and the interaction between a planet and a planetesimal disk (Murray et al. 1998). It is energetically favorable for a Keplerian disk to evolve by transporting mass inward and angular momentum outward (Lynden-Bell & Pringle 1974). Consequently, inward planetary drift appears inevitable, and this is what is found in certain simulations (Trilling, et al. 2002; Armitage et al. 2002; Matsuyama, Johnstone, & Murray 2003). However, if planets form while the disk is being dispersed, or if multiple planets are present, outward migration can also occur. In a system consisting initially of two Jupiter-like planets, a dynamical instability may eject one planet while the other is left in a tight, eccentric orbit. The second planet is not always lost; the observed Doppler exoplanet eccentricity distribution can be reproduced if the 51 Pegasi systems are formed by planet-planet scattering events and the second planet typically remains bound in a wide ($a > 20$ AU), eccentric orbit (Rasio & Ford 1996; Marzari & Weidenschilling 2002). Divergent migration of pairs of Jupiter-mass planets within viscous disks leads to mutual resonance crossings and excitation of orbital eccentricities such that the resultant ellipticities are inversely correlated with planet masses (Chiang, Fischer, & Thommes 2002). Given decreasing disk viscosity with radius and the consequent reduction in planetary mobility with radius, we expect eccentricities to decrease with radius, perhaps sharply if the magneto-rotational instability is invoked (Sano et al. 2000). By contrast, excitation of eccentricity by disk-planet interactions requires no additional planet to explain the ellipticities of currently known solitary planets (Goldreich & Sari 2003). Clearly, observations of the incidence, mass, and eccentricity distributions of multiple planet systems would sharpen our nebulous ideas regarding how planetary orbits are sculpted.

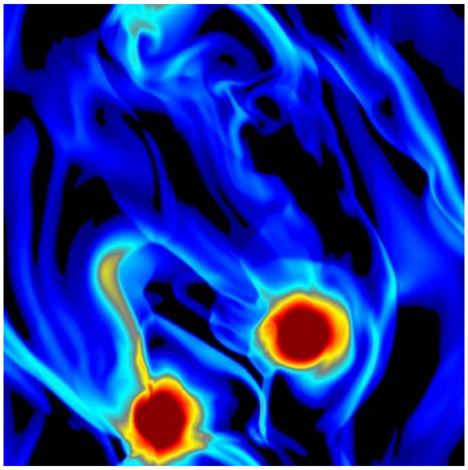

Figure 5-2: At ever-greater stellocentric distances, planet formation may proceed by gravitational instability of gas rather than by condensation of a seed rocky core. Direct imaging of the outer regions (10-30 AU) of solar systems may detect gas giants formed via this first channel. This image from Gammie (2001) displays two clumps of gas undergoing runaway gravitational collapse in a circumstellar disk where the cooling time is shorter than shearing time; colors trace disk surface density. Later, the two clumps collide and merge to form a self-gravitating protoplanet.

A number of other reasons motivate direct imaging of exoplanets. Imaging improves our knowledge of the physical properties of planets. In the absence of direct detection or an astrometric measurement of the motion of the primary the orbital inclination, *i*, and hence the mass of the planet is unknown to a factor of sin *i*. However, the photometric and spectroscopic properties of the exoplanet constrain its mass. Long-term astrometric monitoring exoplanets with GPI will also yield the Keplerian orbital elements. With the exception of a handful of transiting planets we have no information on the atmospheres of exoplanets. Detection of the light from planets opens their atmospheres to the study of temperatures, gravities and compositions. A program to search for planets with signatures of life will ultimately require spectroscopic analysis of planet light.

Imaging also provides a snapshot with the potential to reveal multiple planets, zodiacal dust and Kuiper belt structures, and brown dwarves or stellar companions. The Fourier decomposition that underlies Doppler and astrometric detection are subject to aliasing and beat phenomena, and suffer confusion when multiple planets are present. Therefore, indirect searches require the completion of multiple orbits to disentangle complex systems. The presence of stellar or sub-stellar companions and their potential dynamical influence on neighboring planets can be revealed immediately (cf. the case of HD 80606b studied by Wu & Murray (2003)). Furthermore, direct detection can access a wider variety of host stars than current Doppler techniques, including higher-mass A and early F stars with weak photospheric absorption lines, and pre-main-sequence stars whose chromospheric activity introduces kinematic jitter. Direct detection of planets orbiting young stars is practical because of the high luminosity of freshly assembled planets, and provides direct constraints on timescales for planet formation.

5.3. Planetary atmospheres

To understand why direct detection of luminous, young planets is feasible, we must consider their atmospheres. With the exception of rare transiting planets, e.g., HD 209458, we have no observations of the atmospheres of exoplanets, and we must be guided by observations of objects that are cooler (Jupiter & Saturn), hotter (T dwarfs), and by theory.

The discovery of Gliese 229B (Nakajima et al. 1995; Oppenheimer et al. 1995) and the 2MASS (Skrutskie 1997), Sloan (Strauss et al. 1999) and DENIS (Delfosse et al. 1997) surveys have launched a new era in stellar astronomy. The L and T dwarfs (Kirkpatrick et al. 1999, 2000; Burgasser et al. 1999, 2000a, 2000b, 2000c) are the first fundamental spectroscopic classes to be added to the stellar alphabet in nearly a century. The L dwarfs delineate the lower edge of the solar-metallicity main sequence, with effective temperature near 1700 K. More than 400 L dwarfs with T_{eff} = 2200–1300 K are now known. The coolest L dwarfs are brown dwarfs, objects with insufficient mass (< 80 M_J) to burn H on the main sequence (Burrows et al. 2001). About 60 T dwarfs have been cataloged, spanning the T_{eff} range from 1200 to 700 K: these are all brown dwarfs.

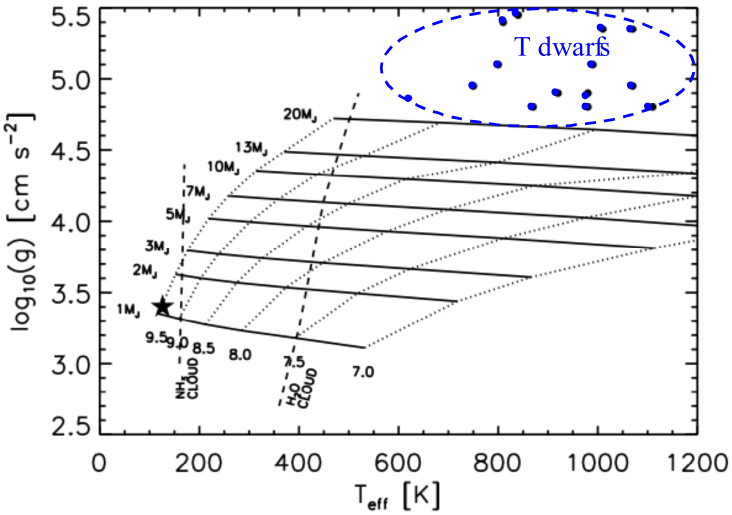

Figure 5-3: The predicted location of exoplanets in the surface gravity/T_{eff} diagram for a range of masses (1-20 M_J) and ages (0.01-3 Gyr) is shown (Burrows et al. 2003). Hot, young jupiters lie at the center of this figure. Solid curves are evolutionary tracks (labeled on the left with planet mass); dotted curves are isochrones (labeled at the bottom with log age). The condensation curves for H_2O and NH_3 are dashed lines. For objects to the left of these lines, the corresponding condensate will form in the atmosphere. Water ice clouds are expected to condense in the atmospheres of most of these objects, while NH_3 condenses only the lowest mass, oldest objects. Currently, the coolest known T dwarfs are in the upper right (Burgasser et al. 2006). Mature, solar-system planets (4.5 Gyr) are located at the lower left— the location of Jupiter is designated with a ★. Currently detectable objects occupy only a small fraction of the depicted phase space.

While there have been claims of the discovery of free-floating planetary mass (< 13 M_J) objects (Zapatero-Osorio et al. 2000), these are most likely brown dwarfs (McGovern et al. 2004). A wide gap in T_{eff} (150 K < T_{eff} < 800 K) exists between the currently known T dwarfs and cool, solar Jovian planets (see Figure 5-3). However, these objects must exist as the youthful progenitors of the known population of Doppler-detected exoplanets. Figure 5-4 shows theoretical spectra of a 5 M_J exoplanet as a function of age showing the distinctive peaks due to enhanced flux between the water vapor absorption bands (0.93, 1.1, 1.4, 1.8 & 6.5 µm). Thus, the ground-based near-IR *YJHK* bands, which are defined by the same H_2O opacity, are ideal bands in which to seek detection.

Detection and spectroscopy of the light from planets opens their atmospheres to the study of temperatures, gravities and compositions. These objects represent planetary *terra incognita*. For example, at T_{eff}'s below 400–500 K water condenses in planetary atmospheres. The appearance of water ice clouds constitutes a significant milestone along the path from the known T dwarfs to the giant planets. Associated with condensation is the depletion of the gas-phase abundance of that species above the cloud tops. Within 100 Myr, water clouds form in the atmosphere of an isolated 1 M_J object. The presence of clouds of any sort emphasizes the kinship of this transitional class with solar system planets, in which clouds play a prominent role. On Jupiter itself, water clouds are too deep below the ammonia cloud decks to have yet been unambiguously detected.

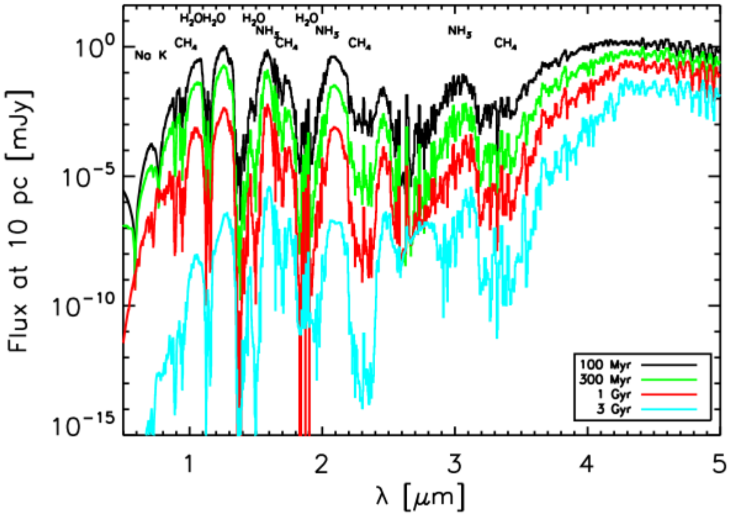

Figure 5-4: Spectra of a 5 M_J exoplanet as a function of age showing the distinctive peaks due to enhanced flux between the water vapor absorption bands (0.93, 1.1, 1.4, 1.8 & 6.5 µm) typical of brown dwarfs. Other general features are the broad hump at 4.5 µm, CH_4 features at 1.7, 2.2, & 3.3 µm and the NH_3 features at 1.5, 1.95, & 2.95 µm. The strengths of each of these features are functions of mass and age.

Field and cluster surveys will yield hundreds of exoplanet candidates for further study and follow-up. These objects will include false alarms that fall into several categories. Physically associated, low luminosity interlopers include brown dwarfs and white dwarfs. Even ancient (13 Gyr) white dwarfs are more luminous and are hotter than all but the youngest and most massive exoplanets. Although white dwarfs should be found in abundance in the field star survey, they will immediately distinguished from exoplanets by their luminosity and colors.

In addition to broadband filters, the GPI IFS can be used to synthesize narrow band-pass filters. For example, narrow band indices covering portions of the *H* band have been shown to correlate with spectral type over the entire L and T range (Mainzer *et al.* 2004). These include regions corresponding to “continuum” emission at 1.595 µm and molecular absorption at 1.495 µm (H_2O), 1.66 µm (CH_4), and 1.75 µm (H_2O). Features within the *J* band are particularly useful, for example the KI lines are sensitive to surface gravity, while FeH and H_2O bands are important for assigning spectral types (*e.g.* McGovern *et al.* 2004).

Old brown dwarfs and young planets can have similar effective temperatures. However, these two classes are distinguished by surface gravity. For example, Figure 5-3 shows that a 25 M_J brown dwarf at 5 Gyr has the same temperature as a 7 M_J planet at 300 Myr, but the gravity of these objects differs by 0.8 dex. The IFS shall provide sufficient spectral resolution and wavelength span to measure effective temperature and surface gravity to discriminate between these two classes of objects. Guided by the theory of exoplanet atmospheres, a precision of about 20 K and 0.1 dex in log(*g*) can be accomplished with spectral resolution, $\lambda/\delta\lambda \approx 45$, a wavelength span, $\Delta\lambda/\lambda \approx 20\%$, and moderate (10%) photometric accuracy.

5.4. Planets in open clusters and young associations

Targeted searches, in young clusters and stellar associations will yield core science regarding the formation and evolution of planetary systems. As products of a single star formation event, open clusters represent ideal targets samples with known ages, distances, and metallicities. Within a cluster, it is possible to conduct planet searches that eliminate uncertainties about different environmental factors. Combining all the open cluster targets provides a large sample of youthful stars (ages 90–660 Myr) that formed in regions of high stellar density. The young ages of open clusters enhance the planet detection threshold since young giant planets are more luminous, and the nearby distances ensure that separations comparable to the sizes of the solar system and circumstellar disks are resolvable. A plot of the ages and distances of the nearest open clusters is shown in Figure 5-5. The closest clusters are visible in the Northern Hemisphere, while the youngest clusters are located in the Southern Hemisphere (cf., Zuckerman & Song 2004b).

Ursa Majoris and Coma Berenices are sparsely populated, while the Hyades, Pleiades, and α Persei have substantial membership. Most clusters are out of the Galactic plane and the closest have high proper motions, allowing confirmation of true companions within a year. The total number of open cluster members is substantial if a limiting magnitude of $I < 8$ mag. (233 stars) can be achieved; this large sample is critical given the limited frequency of extrasolar planets detected by large scale radial velocity searches. For the closest clusters, both early-type and solar-type stars are feasible targets, however, only early-type stars are bright enough in the Pleiades and α Persei.

Until the late 1990s only two nearby (< 100 pc), coeval, comoving groups of stars were known: the rich Hyades and the sparse Ursa Majoris clusters. Both are hundreds of millions of years old. Then, beginning in the late 1990s, three more stellar groups—the TW Hydrae Association (Kastner et al. 1997), the Tucana/Horologium Association, and the beta Pictoris Moving Group (cf. Song et al. 2003)—were identified within about 100 pc of Earth. Recently, two more such groups have been identified in addition to the η Chamaeleontis cluster at 97 pc (Mamajek et al. 1999). Including all members of nearby young stellar groups plus young, nearby stars apparently not belonging to any known groups, about 300 young (8–50 Myr), nearby (< 100 pc) stars are known.

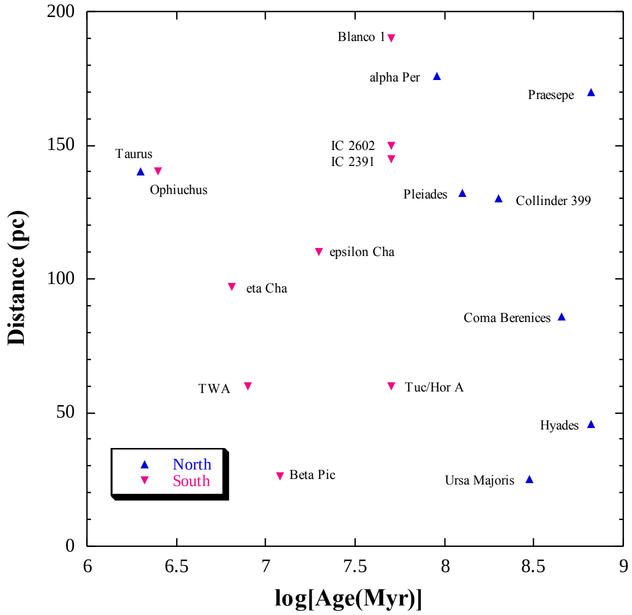

Figure 5-5: Nearby open clusters, associations, and star-forming regions.

These young, nearby stars are excellent targets for direct imaging detection of cooling giant planets because of their extreme youth and proximity to Earth. Current AO systems can detect massive planets in wide orbits (> 50 AU) around the later-type or closest members of these groups, GPI will be able to detect sub-Jovian planets to the inner working distance of its coronagraph—0.14 arc seconds, 4 AU at 30 pc. This will also enable imaging studies of the planetary debris disks and early evolution of planetary systems. The study of young associations in the solar neighborhood is an active field, and we will continue to assemble and critically review a catalog of young, nearby candidate stars for a direct imaging survey. Our input list will include one of the most thoroughly vetted samples of young stars, suitable for imaging planetary companions, that of the Spitzer FEPS (Formation and Evolution of Planetary Systems) Legacy Program. This target sample will represent the “first light” science for GPI, to be followed by larger scale surveys of field stars.

5.5. Adolescent stars

The young stars discussed in §5.4 are ideal targets for GPI because the planets that they host are luminous. However, given a Galactic star formation history that is relatively uniform, only about 1% of stars in the solar neighborhood can be young (< 50 Myr). Older stars are more numerous, and therefore offer the statistical riches necessary to characterize planet populations. Although the majority of young stars within 100 pc have been located, a much smaller proportion of “adolescent” stars have been identified.

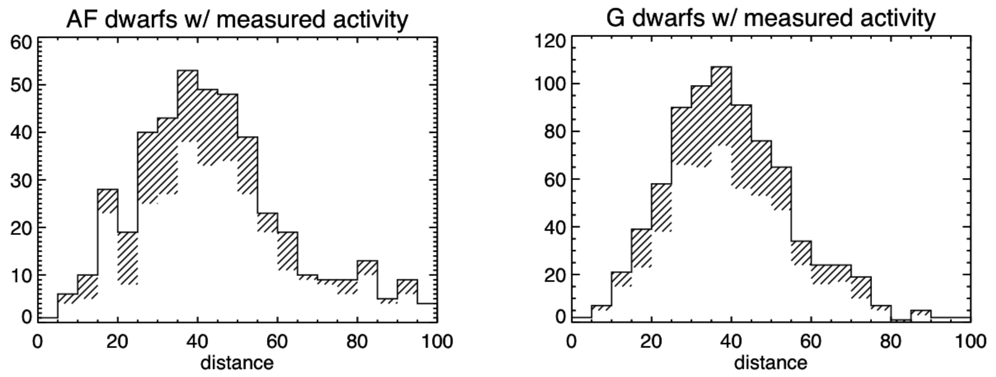

Figure 5-6: Histograms of candidate adolescent main sequence HIPPARCOS A, F, and G dwarf stars with *I* < 9 mag., d < 60 pc, and with measurements of chromospheric activity. Out of 1900 stars about 500 stars in the solar neighborhood show levels activity consistent with ages of less than 2 Gyr (shaded part of histogram).

We distinguish adolescent (0.1–2 Gyr) F, G, K, and M stars by their levels of chromospheric activity (Wright et al. 2004). In the HIPPARCOS catalog there are approximately 7500 G, K, and M main sequence stars within 60 pc and *I* < 9 mag. At least 1900 of these potential GPI targets have measured activity levels. Analysis is ongoing, but of the 1700 already considered roughly 500 show activity levels consistent with ages of less than 2 Gyr.

5.6. Scattered light imaging of debris disks

Debris disks are the extrasolar analogs of our Zodiacal dust disk (< 3 AU) and the dust complex generated in the Kuiper belt (40-50 AU; Landgraf et al. 2002). They are optically thin and gas-poor. Debris disks arise from the collisional erosion of larger solid objects, but may include a contribution from subliming icy bodies as they pass through periastron. Key motivations to study these disks include the following:

1. Unlike hot exoplanets, a handful of debris disks are detected and spatially resolved in scattered light by current high-contrast instrumentation, and these systems present an early observational test for GPI, and an opportunity to refine observing and data reduction procedures.
2. Dust optical depth and age are correlated (Spangler et al. 2001). Thus, stars with the dustiest debris disks are among the youngest stars in the solar neighborhood, and some of the best candidates for finding self-luminous planets.
3. Debris disk morphology gives several constraints on where to look for planetary bodies around a star, including: position angle, inclination of the system to the line of sight, and the radius of a central dust-depleted region approximating the outer radius of the planetary system (e.g., Roques et al. 1994).
4. Planet orbital parameters are further constrained by interpreting radial and azimuthal asymmetries in debris disks as the dynamical effects of planet-mass objects (Liou & Zook 1999). For example, the orbital eccentricity of a detected planet may be constrained by disk structure before the actual orbit is observed in multi-epoch data.
5. For older stars in the sample, where self-luminous objects are dim, the analysis of debris disks still characterizes the possible, unseen planetary system, perhaps leading to follow-up observations with future facilities such as Terrestrial Planet Finder. The debris disk analysis may include multi-epoch observations that show the rotation of disk features, which can then be linked to the mass and location of planetary objects (Ozernoy et al. 2000).

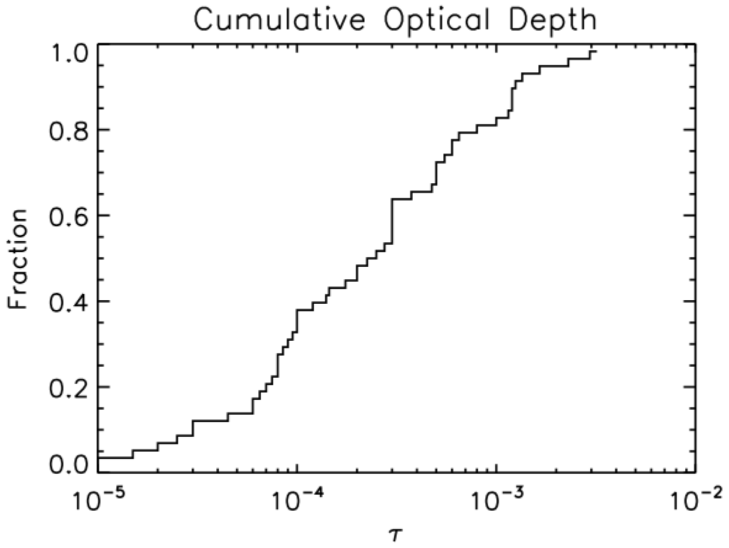

Figure 5-7: The cumulative optical depth distribution for debris disks from Zuckerman & Song's (2004a) comprehensive catalog of IRAS-detected Vega-like stars. Only a handful of disks at $\tau \sim 10^{-3}$ have been imaged in scattered light—and most of these are edge on systems like β Pictoris or AU Microscopii.

The relevance of high contrast imaging for debris disk studies for points (1) & (2) is illustrated by examining Figure 5-7 & Figure 5-8. Figure 5-7 shows the distribution of optical depths for the approximately sixty systems in the debris disk catalog of Zuckerman & Song (2004a). The cataloged debris disks spans three orders of magnitude in optical depth, yet only a handful of the highest optical depth systems, $\tau > 10^{-3}$, have been imaged. The scientific utility of higher contrast imaging is twofold.

First, better contrast performance will enlarge the sample of objects: an order of magnitude improvement would represent a significant accomplishment; two orders of magnitude would allow the entire sample to be observed. Greater contrast will not only enlarge the sample of systems that can be studied within the proposed classification scheme of Kuchner & Holman (2003), but also extend the range of ages and evolutionary states that can be observed (& Figure 5-8).

Second, increased sensitivity to low surface brightness increases the detectability of face-on disks. Beta Pic is one of the most intensely investigated stars in the sky (e.g., Zuckerman 2001), because of its prominent *edge-on* debris disk. Structure within the β Pic disk is resolvable at mid-IR wavelengths on the largest ground-based telescopes (Wahhaj et al. 2003, Weinberger et al. 2003). Disk structure can be probed closer to the star with mid-IR imaging than with scattered-light imaging, even with HST, because of the more favorable disk-to-star brightness ratio in the mid-IR. Augereau et al. (2001) model the grain properties and conclude that parent bodies for the dust are located between 20 and 150 AU with a peak at 110 AU. Weinberger et al. (2003) obtained spatially resolved spectroscopy that revealed evidence for amorphous and crystalline silicates within 20 AU of β Pic. However, because the β Pic disk is edge on, we have no information on the azimuthal structure needed to place the system within a morphological classification scheme.

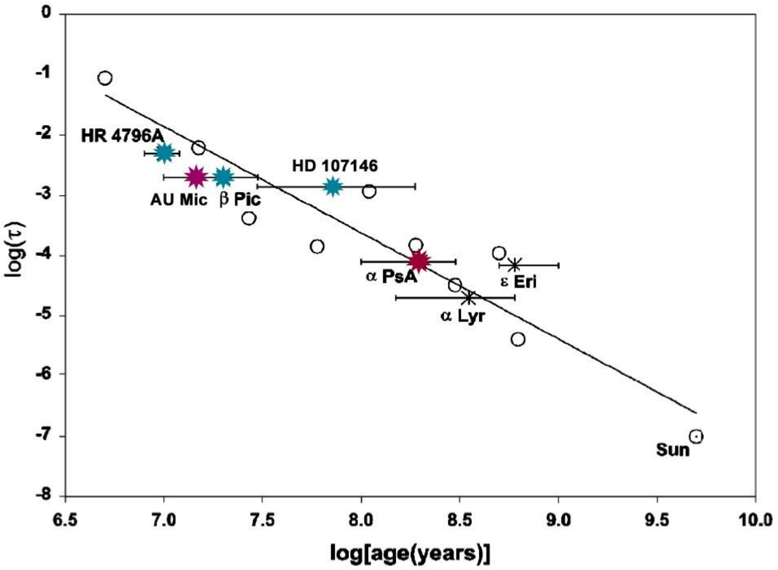

Figure 5-8: Trend of debris disk optical depth with age (adapted from Zuckerman 2001). Filled stars indicate systems already imaged by optical or near-IR coronagraphs. Open circles represent nearby clusters such as the Pleiades. The point for the Sun is based on the Kuiper belt dust mass inferred from models. Increasing contrast and corresponding sensitivity to smaller and smaller optical depths extends the age range accessible to scattered light imaging. The debris disks surrounding HD 107146 (Ardila et al. 2004) and Fomalhaut (α PsA) (Kalas et al. 2005) were first seen in scattered light in 2004 because of the sensitivity gains provided by the HST ACS coronagraph. Currently, only a handful of the highest optical depth, youngest and edge-on systems can be imaged with ground-based systems. Thus, current high contrast imaging facilities probe only optical depths of a few 10^{-4}.

The optical properties of dust grains in disks yield additional information regarding their sizes and composition, and hence their origin (e.g., Duchêne et al. 2004). For example, the *R* and *H* band colors of the AU Mic debris disk, together with the sub-mm SED shows a pronounced lack of small (< 0.3 µm) grains in the inner disk (Metchev et al. 2004; Fitzgerald, Kalas & Graham 2007). Small particles may be removed by coagulation into larger particles (grain growth), destruction by Poynting-Robertson and/or corpuscular drag, or radiation pressure blowout. However, the AU Mic disk becomes increasingly blue at larger radii, at least between 30–60 AU (Figure 5-9). The sub-micron grains scattering visible light at wide separations (where the collision and Poynting-Robertson time scales are longer than the age of the star) may be primordial, rather than being blown out from the inner disk. These grains may be a remnant from the proto-stellar cloud that never coagulated beyond an ISM grain size distribution (Krist et al. 2004). In addition to *YJHK* colors, which constrain the grain size, the scattering phase function and polarization measurements can be used to identify details of the grain size population.

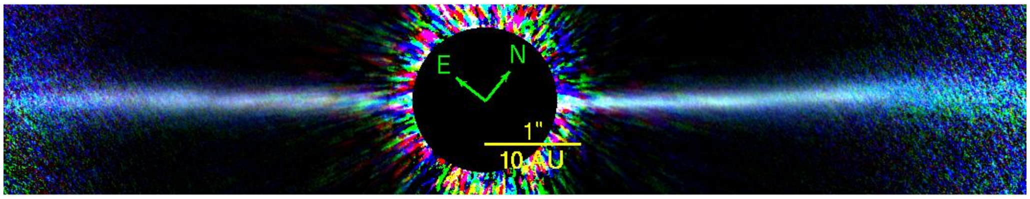

Figure 5-9: The AU Mic debris disk resolved by Keck-2 adaptive optics. The colors, blue, green and red correspond to *J*, *H*, and *K'* respectively (Fitzgerald, Kalas, & Graham 2007). In this edge-on disk information about the azimuthal structure is lost. In such a system, information about the scattering phase function, which relates directly to the grain size distribution, is lost.

5.7. Adjunct Science

Additional examples of GPI science are included in this section. The purpose is to demonstrate the broad astrophysical utility of high contrast imaging. One key design goal for our system is to have a broad scientific reach – with high performance operation down to I = 8 mag. (I = 9 mag. goal), high Strehl ratio, a stable and predictable PSF, and IFS and polarimetric capabilities, GPI will be capable of several additional scientific missions. Examples given here include solar system science, binary stars, and the study of mass-losing evolved stars.

5.7.1. Solar System Exploration

Over the last forty years, planetary science has been revolutionized by the development of spacecraft that permit *in situ* and high angular resolution observations. Recently, large, ground-based telescopes equipped with adaptive optics systems have become competitive with remote sensing platforms and bring new information about the nature, origin, and possible geological activity of the bodies orbiting around our Sun. Spacecraft have visited most of the major solar system bodies. After this preliminary phase of exploration, planetary science is now driven by a more precise study of these worlds. The environment of asteroids, moons and planets provides a unique laboratory to study complex physical processes. In fact, we are often surprised by phenomena that link geologic, atmospheric, and magnetospheric processes. Examples of these might include the interaction between volcanism on Io and the Jovian plasma torus, the formation of clouds in the dense atmosphere of Titan, and the geological changes on the surface of Europa. For these studies, high angular resolution observations spanning a large wavelength range, and good temporal sampling, can only be provided by AO on large ground-based telescopes. Moreover, the complex interplay of atmosphere and surface can only be explored and understand by long term synoptic monitoring over a periods that exceeds the typical lifetime of space missions. For instance, to record the variability of the prebiotic atmosphere of Titan and its interaction with the surface, including the methane precipitation that feeds the fluvial system, we need to follow seasonal effects over the Kronian year (29.46 yr). The solar system is a complex and chaotic system, and exceptional occurrences such as the impact of a comet can be predicted only months or even days before it happens. Consequently, a ground-based campaign is frequently the only practical response to many solar system events.

High contrast imaging with GPI has the potential to: a) Characterize the surface and atmospheric composition of Galilean satellites and Titan, and monitor the volcanic activity of Io, pinning down the highest temperature of the magma (McEwen et al. 1998; Marchis et al. 2002); b) Determine size, shape, surface morphology and composition of the 50 largest main-belt asteroids, search for multiplicity(Figure 5-10), and hence yield information about the bulk density and the formation of this remnant of the solar system formation (Merline, et al. 2002; Marchis et al. 2003); c) Monitor the atmospheric activity of Uranus and Neptune, focusing especially on the cloud formation and wind profile above the stratospheric haze near the southern pole of Uranus, which is now being exposed to sunlight (Rages et al. 2004). Study of Neptune's atmosphere yield information about the transport of energy and the source of its mysterious internal source of heat (Pearl & Conrath 1991).

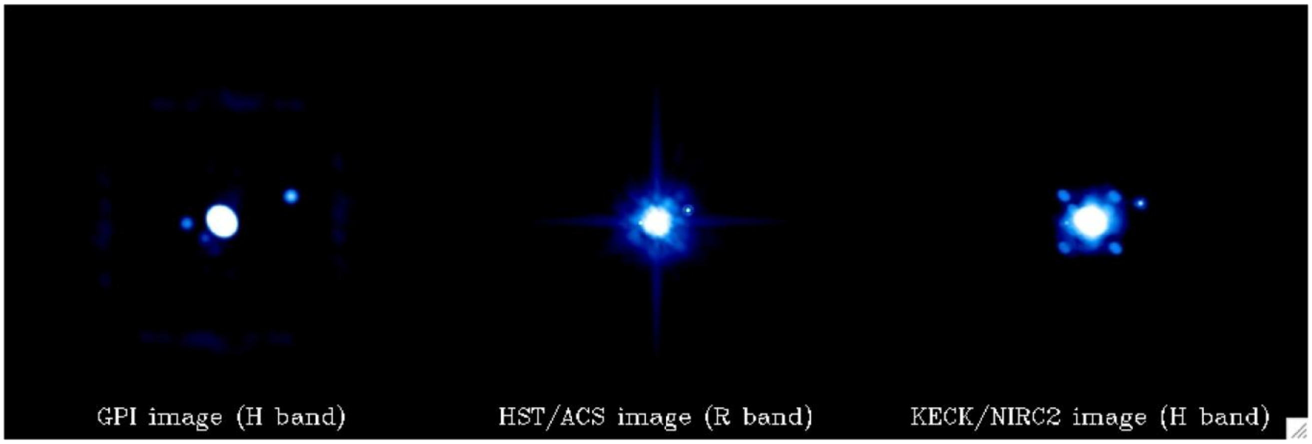

Figure 5-10: Simulated *I* = 9 mag. main belt asteroid with km sized moonlets. The companions are easily detectable with GPI, but only the biggest and most distant companion is visible to either (the now defunct) ACS or Keck-2/AO. There are about 10 asteroids bright enough to perform this experiment. At *I* = 10.5 mag. there are 50 main belt asteroids where Dactyl-sized (1-km) companions to 200-km parents are detectable.

Solar system science stresses non-coronagraphic imaging and performance at the shortest possible wavelengths with to achieve the highest possible angular resolution. Low-resolution spectroscopy in the near-IR is required for mineralogical studies of satellite and asteroid surfaces. The addition of diffraction limited operation in *Y* band (0.97–1.07 µm) will help to better characterize the surfaces through the detection of an absorption feature of pyroxene mineral centered at 1 µm. For example, combining spectroscopy and polarimetry, it will be possible to probe the atmosphere of Titan tracing the particle size and altitude of the haze layers (Figure 5-11).

Some solar system science also stresses limiting magnitude. For example, many asteroids have companions at small angular separations that are marginally detectable with Keck AO or HST. At these separations and contrasts, GPI could operate in a non-coronagraphic mode and open up the bandpass of the spatially-filtered wavefront sensor, allowing operation down to *I* = 10 to 11 mag. Strehl ratios at these magnitudes would be no better than current ALTAIR, but systematic and static errors would still be much reduced, significantly improving detectability of companions at small separations (Figure 5-10).

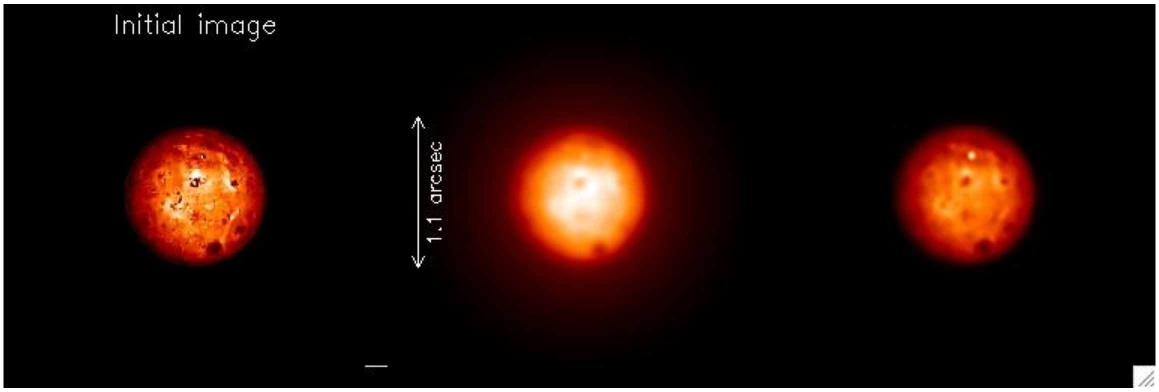

Figure 5-11: Simulated *H*-band GPI observations of Io. *Left:* a perfect image of Io as it would appear observed from the ground at an angular resolution of 10 mas (spatial resolution of ~ 30 km at opposition). Surface albedo features such as pateras, plume deposits and SO_2 frost regions are dominant on this image. *Center:* simulated Altair-NIRI observation. *Right:* simulated GPI image. The ALTAIR-NIRI image was derived using empirical PSFs. The GPI image represents operation in non-coronagraphic mode. The gain in contrast and angular resolution is evident. A faint artificial hot spot located close to the north pole of Io is detectable only by GPI. Detection of the thermal emission in *J* and *H* bands is essential to pin down the highest temperature component of the magma and relate directly to the internal structure of the Io (Marchis et al., 2002).

5.8. Stellar Astrophysics

5.8.1. Stellar Binaries

More than 8000 interferometric observations of stellar objects have been reported so far, of which 75% are of binary stars (Hartkopf et al., 2006). Figure 5-12 shows the distribution of separations measured in the Third Catalog of Interferometric Measurements of Binary Stars. The range of stellar separations to which GPI is sensitive, 0.1–1 arc seconds, is similar to that already probed by interferometric methods. However, interfometric techniques have limited dynamic range. When measuring visibility amplitudes, binaries with ratios more than 10:1 are challenging and 100:1 is about the limit, since the detectable flux ratio is roughly the fractional visibility calibration error. Some special techniques, e.g., non-redundant masking can achieve 200:1 (Tuthill, et al. 2000). Thus, GPI observations will complement on-going interferometric campaigns. The limited dynamic range of existing search methods represents one of the major forms of incompleteness in existing catalogs. Whereas interferometric binaries span only a few spectral types, a GPI-detected binary can span the entire MK sequence. An O5V primary with a M5V companion represents a contrast of 17.5 mag. or a factor of 10^{-7}, which is well within the reach of GPI. Dynamical masses have been measured for about a dozen stars with masses > 3 $M_{\odot}$ (Latham 1998). Thus, a long-term benefit of GPI observations will be improved calibration of the main sequence mass-luminosity relationship for A stars and earlier.

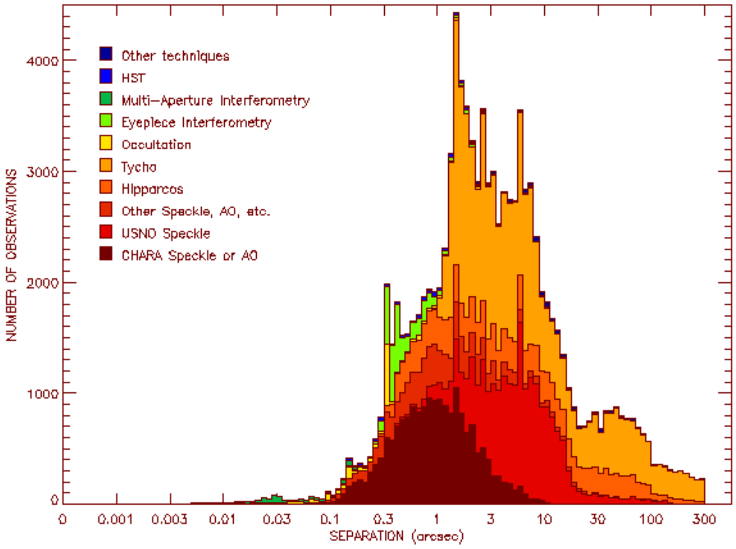

Figure 5-12: Distribution of binary separations from The Fourth Catalog of Interferometric Measurements of Binary Stars (Hartkopf, et al. 2006). GPI will discover stellar binaries with separations ≈ 0.1–1 arc seconds where existing AO and speckle measurements are limited by dynamic range. Whereas the typical dynamic range is at most a few magnitudes, GPI will reveal an M5V companion to an O5V primary ($\Delta m \approx 17.5$ mag.).

5.8.2. Multiplicity of O stars and high mass star formation

Although there is a standard model for the formation of solar type stars (Shu, Adams, & Lizano 1987), it is unclear to what extent this picture applies to massive stars. Unlike low mass stars, massive stars (type A and earlier) have no significant pre-main sequence phase; they have such short Kelvin-Helmholtz contraction time scales that they can still be accreting by the time that they reach the main sequence. Consequently, a number of theories have emerged to explain how massive stars form. These include competitive accretion (Bonnell, et al. 2001), the turbulent core model (McKee & Tan 2003), and, perhaps most radically, formation by stellar coalescence (Bonnell and Bate 2002). These models make radically different predictions regarding the binary companions of massive stars.

5.8.3. White dwarfs in binaries

Cool white dwarf (WD) stars with $T_{\mathrm{eff}} < 4000$ K are of great interest for several reasons. They are the end-stage remnants of main-sequence stars with masses less than about 8 $M_{\odot}$, and represent some of the oldest objects in the Galaxy. As such, they give direct information about star formation during the Galaxy's earliest epochs. Since WDs continue to cool and fade with time, the very coolest can place lower limits on the ages of various galactic components. In addition, recent microlensing searches have suggested that there may be a significant population of WDs in the Galactic halo (Alcock et al. 2000), offering a window into the early stages of the Galaxy and its formation. While the Sloan Digital Sky Survey has yielded a handful of free-floating ultra-cool WD, exhibiting the tell-tale strong collision-induced absorption from molecular hydrogen (Gates et al. 2004), the majority of the population is simply too faint to be detectable in this relatively shallow survey.

As a pointed survey on an 8-m telescope, GPI has the potential to find fainter and hence older WDs in binaries than the wide field Sloan Survey. WD companions to typical disk-population stars (5 Gyr) will be readily detectable. Even older WDs, for example those as old as the halo, (13 Gyr) will be detectable as companions to G, K and M stars.

5.8.4. Substellar companions

T dwarfs are substellar objects whose near-infrared spectra exhibit characteristic signatures of H_2O and CH_4 (Oppenheimer et al. 1995). They make up the coolest class of brown dwarfs currently known, with effective temperatures $T_{\mathrm{eff}} \approx 1300–1500$ K (Burgasser et al. 2002). Their atmospheric properties are therefore similar to class III and IV giant exoplanets (Sudarsky, Burrows, & Pinto 2000) but are more easily studied without the obscuration of a bright host star. Indeed, our understanding of exoplanet atmospheres has advanced in parallel with that of isolated brown dwarfs, since the latter differ only in the absence of an external radiating source. The observed properties of T dwarfs have served to constrain these models. The currently known nearby T dwarfs are also useful for studies of the substellar population in the solar neighborhood and the substellar initial mass function (Reid et al. 2004). Finally, since these objects do not significantly process their initial hydrogen supply to heavier metals, they may be used as a tracer population of the chemical history of the Galaxy, as long as temperature, gravity, and metallicity diagnostics can be disentangled.

Unlike lone brown dwarfs, those that are paired with nearby, luminous stars can have many shared characteristics inferred from the primary such as parallax, age, and metallicity. At present, the list of known companions is quite small at fourteen known L dwarf companions and four known T dwarf companions. Their separations range from 19 AU (for Gl 86) to 3600 AU (for Gl 584C). We hope to uncover brown dwarf companions that can help contribute to the scarcity of available statistics for brown dwarf parameters and help populate the "brown dwarf desert" for widely separated companions.

5.8.5. Evolved Stars and mass loss

The study of post main sequence stars represents a major area where significant progress awaits the deployment of high contrast, high angular resolution observational facilities at near infrared wavelengths.

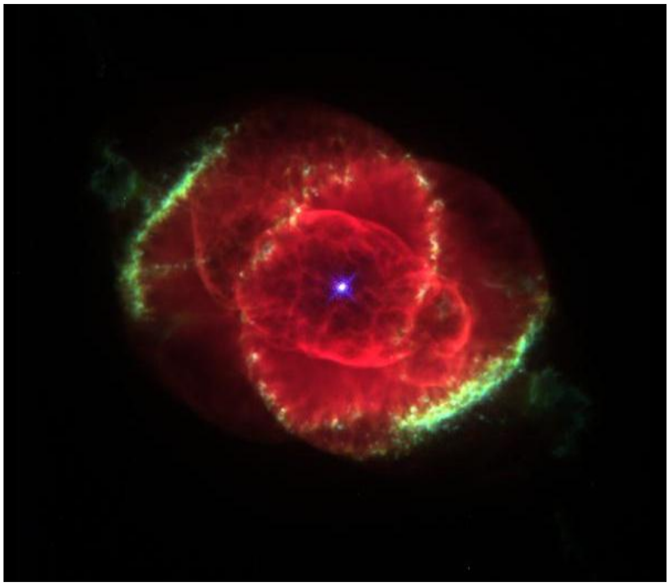

Figure 5-13: WFPC2/HST image of NGC 6543, nicknamed the “Cat's Eye Nebula” (Harrington & Borkowski 1994). Hubble fails to show the structures predicted by Frank et al. (1993), and instead reveals surprisingly intricate structures including concentric gas shells, jets of high-speed gas, and unusual shock-induced knots of gas.

As post-main sequence stars pass through the asymptotic giant branch (AGB) phase and the planetary nebula stage they experience episodes of prodigious mass loss. The initial outflow of the outer envelope of AGB stars is isotropic, however both the protoplanetary nebulae (pPNe) associated with post-AGB objects and the final planetary nebulae (PNe) are marked by axially symmetric morphologies (cf. Balick & Frank 2002: Figure 5-13). Historically, the morphologies of pPNe and PNe have yielded insights into the final stages of stellar evolution and mass loss.

AO and interferometry are beginning to shed light on the nature of mass loss and dust formation by delivering high contrast imaging of the circumstellar material itself. For example, Monnier et al. (2002) have used the techniques of speckle masking to reveal the dust plume and proper motions of newly formed dust in the colliding wind binary WR 140. Figure 5-14 shows an image of the circumstellar environment of RY Scuti. This massive contact binary may represent a rapid evolutionary phase in transition to a short-period WR+OB system. Conventional, natural guide star AO is able to reveal that a young nebula with unusual geometry surrounds it. With its higher dynamic range, stable PSF, and polarimetry, GPI will continue and expand this study.

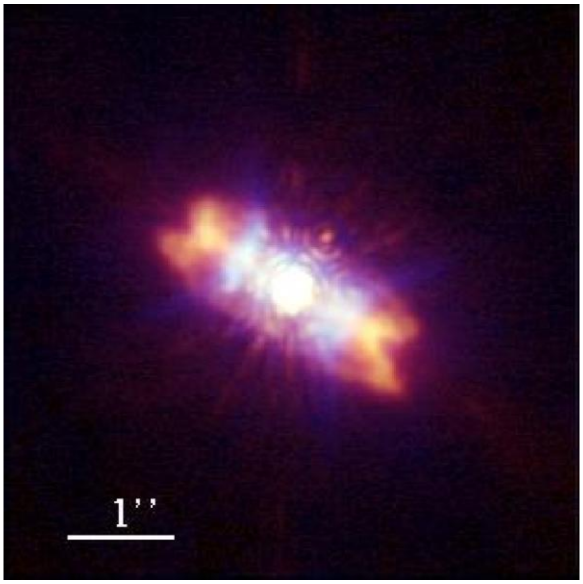

Figure 5-14: Keck/NIRC2-AO image of RY Scuti in the L_p filter (red) with HST Hα (blue). (Smith et al. 2005). RY Scuti is an eclipsing, symbiotic binary star of β Lyr type, which is believed to be the progenitor of colliding wind WR binaries.

6 Science Drivers & Requirements

In this section, we describe several primary scientific objectives of GPI with the purpose of identifying and quantifying the key requirements placed on the instrument. Where appropriate, the reduced science-reach implicit in scaling back from the requirement is stated.

The final instruments requirements are called out and labeled with the designation: **REQ-OCD-XXXX**, where **XXXX** is a unique number. These are requirements that can be inferred directly from a science goal, and can be stated unambiguously in terms of a verifiable instrument specification, e.g., a field of view (e.g., **REQ-OCD-0160**). These requirements are cited in other GPI documents such as the FPRD. A second type of requirement, e.g., the planet recovery rate is set by statistical goals within the science case, is given a different designation (e.g., **REQ-SCI-0010**) because it is not directly verifiable. However, the corresponding contrast requirement (**REQ-OCD-0011**), which is derived by making certain assumptions about the distributions of exoplanet properties—in this case exoplanet semimajor axis, mass and age—is verifiable.

6.1. Surveys for direct detection of planets

6.1.1. Field star survey

The most straightforward approach to quantifying the abundance of exosolar systems involves surveying a large sample (~ 1000) of main sequence stars in the solar neighborhood (< 50 pc). The purpose of this initial survey is to create a catalog for subsequent confirmation and characterization of atmospheric and orbital properties. The approach is analogous to that adopted by current Doppler surveys. The field star survey presents the opportunity to observe the state of mature (0.1–3 Gyr) planetary systems. While young targets drawn from associations and open clusters (§ 6.1.2) are likely to yield higher detection rates, their planetary systems are still evolving. Although this evolution is a vital topic for study, comparison with the results of Doppler surveys would be problematic. Main sequence, field F, G, K and M stars can also be sorted by age, using chromospheric activity, stellar rotation and kinematics. However, more stars need to be observed and given age estimates before this approach is fully viable.

The fundamental premise adopted when specifying requirements for the field star survey is that planetary systems are abundant, and that lower limits are set by the occurrences of debris disks (15%) and circumstellar disks (50%). It is possible, although extremely unlikely that Doppler surveys have already found all the planets in the solar neighborhood. The incompleteness of Doppler surveys is substantial. Of 1800 nearby solar-type stars being monitored in Doppler surveys, about 90 have detectable planets. Thus, at least 5% of stars possess planets, yet only 20% of the mass-period phase-space occupied by our solar system has been sampled. The true fraction of solar-type stars with planets may therefore be as large as 100%. For a first direct-imaging reconnaissance of the field population detection efficiency comparable to or higher than the Doppler rate of 5% is the minimum acceptable performance.

Limiting magnitude & detection rate: The degree of AO correction, and hence the achievable contrast, depends on the WFS SNR. Photon statistics dominate the error budget for the dimmest stars. On average, the planets orbiting dim stars will be found at smaller angular separations, where the achievable contrast is lower. Thus, the detection rate should be specified for a magnitude-limited sample, which spans a representative range of spectral types and distances. The target list for a field-star survey ($d < 50$ pc) is large (800 with $I < 6$ mag.; 3500 with $I < 8$ mag., see Table 6-1), promising a statistical exploration of exoplanet properties according to $M \sin i$, semimajor axis, and orbital eccentricity. This data set, for example, would be used to confront evolutionary scenarios such as the divergent migration picture of Chiang et al. (2002). Such a survey calls for a sample consisting of ≈ 240 exoplanets (five bins in each planetary property and 4-σ counting statistics per bin). Therefore, a planet detection rate of 7% (for $I < 8$ mag.) is required. If the pool of detections falls below about 135 planets then this exercise becomes statistically meaningless. We choose the higher rate of 7% as our fundamental science requirement.

For a bright star the wavefront can be measured at high temporal and spatial bandwidth, with the result that atmospheric wavefront error can be minimized. The requirements on temporal bandwidth are set by the characteristic coherence length, r_0, and the wind speed, while sub-aperture size, d, determines the largest angular scale on which wavefront errors can be corrected: $\lambda/2d$. Purely AO-based considerations to achieve high contrast drive the choice to small sub-apertures and high temporal bandwidths.

Table 6-1. Results of a GPI survey of the main sequence, field star population estimated for an idealized GPI architecture. Numbers for detected planets and AO hit rate in columns three and four are for 18 cm sub-apertures. Numbers in parentheses are for 12 cm sub-apertures. Please consult Critical Design Document Volume 1–Scientific Motivation for the performance of the actual CDR configuration.

Guide star magnitude (I)	No. of stars	No. of planets detected	AO hit rate (%)
< 2	4	2 (1)	50 (25)
< 4	66	13 (13)	20 (14)
< 6	833	87 (82)	10 (10)
< 8	3827	247 (208)	6.5 (5.4)

Table 6-1 lists the cumulative number of stars for a solar neighborhood main sequence field survey with declination range $-60^o < \delta < +0^o$. The third and fourth columns lists the number of planets and the planet detection rate in a Monte Carlo experiment that couples assumptions about the semimajor axis distribution ($dN/da \sim a^{-1/2}$) and mass ($dN/dM \sim M^{-1}$) with an idealized GPI system. This so-called straw man system has 44 (18 cm) sub-apertures across the Gemini primary mirror, 2.5 kHz bandwidth, and × 16 speckle noise suppression. The observation time is one hour. The AO system employs adaptive modal gain control and idealized Blackman pupil apodization in the coronagraph. The results for 68 (12 cm) sub-apertures are shown in parentheses. Throughout this section we use this same set of physical assumptions about GPI and astrophysical assumptions about the exoplanet population to translate between science requirements and instrument performance, e.g., from planet detection rate to contrast.

N.B. *This Monte Carlo simulation is based on an idealized realization GPI conceived to establish instrument capabilities and requirements. The results are used for consistency with previous versions of the OCDD. Please consult Critical Design Document Volume 1–Scientific Motivation for the performance of the actual CDR configuration of GPI.*

- **REQ-SCI-0010: Planet recovery ≥ 7% at *H* for *I* ≤ 8 mag field-star survey.**

Exposure time: The field star survey represents the most ambitious use of GPI because of the large size of the survey sample. High impact science results must accrue within the first year of operation. One scenario assumes completing one quarter of the survey (600 stars) within the first year of operation. Assuming 30 nights per semester (two nights per science team member) requires achieving desired detection rate (7%) in 60 minutes per target.

The exposure time requirements can be relaxed by improving the sensitivity to exoplanets or by refining target selection. Eliminating the older F, G, K and M stars from the sample, using chromospheric activity, stellar rotation, and kinematics, removes the old, cool and dim exoplanets to which we are insensitive. If we choose stars with ages < 3 Gyr, for example, the straw man GPI system would triple its exoplanet detection rate to about 22%. The current catalogs of < 3 Gyr F, G, K and M stars (e.g., Wright et al. 2004) are large enough (about 500 stars per hemisphere) to warrant scheduling for early investigation, but substantial work remains to be done to winnow the old stars from potential GPI target lists.

- **REQ-OCD-0090: Survey 2000 stars in 200 nights to REQ-OCD-0011 level**

Contrast: For point sources, contrast simply refers to the ratio of planet to star flux. However, the achievable star-planet contrast is necessarily a derived quantity that is set by the required exoplanet detection rate. A single number cannot specify detectable contrast. The achievable contrast is a function of angular separation from the star: GPI contrast jumps rapidly beyond the inner working distance (typically a few λ/D for conventional coronagraph architectures) then improves roughly in proportion to the power spectral density of uncorrected wavefront errors at spatial frequency, $f = \theta/\lambda$.

The broad range of contrast levels at which exoplanets may be detected is illustrated in Figure 6-1, which depicts the cumulative distribution for detections with the straw man GPI system (18 cm sub-apertures, 2.5 kHz bandwidth, I = 8 mag. guide star limit, × 16 speckle suppression, 1 hour observation). Planets are found with contrasts from 10^{-8} to 10^{-5}. The median star-planet contrast for warm Jovian planets in is a few 10^{-7} and a factor of five below this is required to accumulate a detection rate > 5%.

According to Figure 6-1, factors of a few in achieved contrast can be traded against observing time. However, if the delivered contrast is, say, 10^{-6} instead of 10^{-7}, then there will be insufficient planets for a full statistical exploration of exoplanet mass spectrum and orbital elements. If a contrast of only 10^{-6} were achieved, then GPI science would focus deriving the abundance of solar-like exoplanetary systems and on the properties of exoplanet atmospheres. Note that for bright stars, where speckle noise dominates, **REQ-OCD-0011** may be achieved only in a single, narrow, IFS spectral channel. For dimmer stars, where the photo halo shot noise dominates the entire *H* band may be used to achieve the required contrast.

- **REQ-OCD-0011: *H*-band point source contrast of $C \leq 3 \times 10^{-7}$ at 0.17 arc sec, dropping to 2×10^{-8} by 0.5 arc sec and maintained at this level to the edge of the dark hole 1.2 arc sec (1 hr 1-σ) for *I* ≤ 8 mag. May be achieved either in a single narrow band or a broad band, depending on brightness of the target star.**

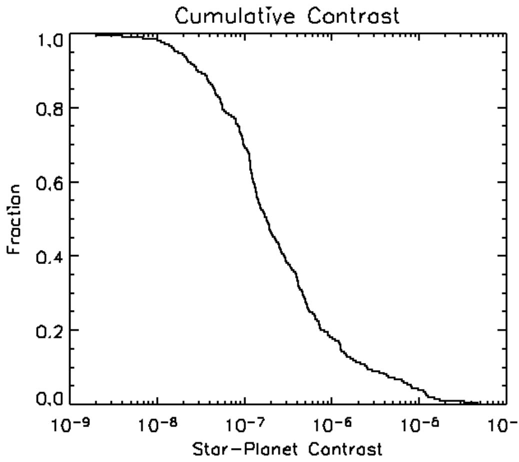

Figure 6-1. The cumulative contrast detection threshold distribution for exoplanet detections at *H* with the straw man GPI system (18 cm sub-apertures, 2.5 kHz bandwidth, *I* = 8 mag. guide star limit, × 16 speckle suppression, 1 hour). This system finds exoplanets in the field population (no age cut) at a rate of about 6%. Planets are found with contrasts from 10^{-8} to 10^{-6}. Contrast is a strong function of angular separation; it should be considered a requirement derived from the planet detection rate.

Bright star limit: The field star survey catalog includes a small number of bright stars ($I \sim 0$ mag.) This requires a specification that the wavefront sensor can lock without saturation on such bright targets.

- **REQ-OCD-0021: WFS bright magnitude $I \leq 0$ mag.**

False alarm rate: For a survey of a large sample, one might argue that the false alarm, η, rate should be a fraction of the expected planet detection rate. However, designing a system with a low rate of false alarm, ~ 10^{-4}, is difficult, because non-Gaussian fluctuations tend to dominate the wings of the probability distribution. Moreover, validating a high level of robustness is likely to be expensive. False positives are tolerable if observations can be repeated and the results of subsequent measurements are uncorrelated. Suppose that the survey is conducted by first observing all targets in the catalog and then re-observing candidates until they are discarded. On the null hypothesis, the time required to a complete a survey, including all re-observations, is a factor $\sum_{i=0}^{\infty} \eta^{i} = 1/(1-\eta)$ longer than that required to carry out the initial reconnaissance: a η = 50% false alarm rate doubles the time required to conduct the survey. For $\eta \ll 1$ we have $1/(1-\eta) \approx 1 + \eta$, so low false alarm rates are is in the realm of diminishing returns. Therefore, a false alarm that is not far out in the Gaussian wings, and comparable to the detection rate, say 3-5% (2.17–1.96 σ), is likely practical and verifiable.

Confusion with background stars and galaxies is discussed in the OCDD Appendix Star & Galaxy Confusion (§ 11).

- **REQ-OCD-0091: Statistical false alarm rate ≤ 5% per observation**

Operating wavelength: The entire near-infrared wavelength range from 1 to 5 µm is of interest for exoplanet atmospheres (see Figure 5-4). The bulk of radiation emerges between gaps in the H_2O and CH_4 opacity at 1.05, 1.3, 1.6, 2.1, & 3.0 µm, and effective temperature and surface gravity sensitive species are present throughout this wavelength range. The opacity gap at 1.6 μm is predicted to be especially deep and therefore this is the prime operating wavelength for GPI.

Single waveband operation is sufficient for planet detection. Additional wavelengths are required for confirmation and characterization. Moreover, multi-wavelength operation (a MWI implemented as a IFS) within a given band-pass is likely required for implementing various speckle suppression schemes. However, the science driver for wavelength flexibility is lack of empirical knowledge regarding opacity sources in exoplanet atmospheres—as Figure 5-3 emphasizes, the atmospheres of exoplanets are *terra incognita*. With the option to operate at *K* band we become sensitive to atmospheres that have stronger continuum opacity and are more blackbody like. This is also to be true at longer wavelengths.

Some exoplanet model atmospheres predict bright *L* and *M* band magnitudes. However, these colors are very sensitive to cloud properties (Saumon et al. 2003). For a ground-based instrument, the thermal background in the *L* and *M* bands from the atmosphere and the telescope is inescapable, and for this reason science drivers for these wavelengths and cannot be considered without due consideration of practical realities. Analysis of noise sources shows that there are two broad choices for instrument architecture: 1) an ambient temperature (< 280 K) AO system, coronagraph and MWI which delivers high order correction, control of diffraction and speckle suppression at *J, H* and K_p; 2) a cold (< 180 K), low background, conventional AO system and coronagraph. Delivered AO contrast only increases modestly with increasing wavelength, and the uncontrollable thermal background from the atmosphere and telescope renders speckle noise suppression irrelevant at 3–5 μm. Thus, a GPI system that delivers good AO correction and speckle suppression at *H* will always outperform a low emissivity AO system working at L_p.

The operating wavelength/system operating temperature trade is explored in the OCDD Appendix Thermal Conditions and Operating Wavelength (§ 12)

- **REQ-OCD-0040: Science operating wavelength: 1.1–2.35 μm.**

Speckle suppression and flat field accuracy: The requirement on spectral resolution is a derived quantity. High contrast is achieved through a combination of wavefront control (AO), control of diffraction (coronagraph) and speckle suppression. Speckle suppression may be implemented through a combination of ADI and SDI. An estimate based on the performance of the final CDR design suggests that at a factor of $\geq$ 5-10 speckle noise reduction is required. Speckles caused by phase errors at a pupil are readily identifiable by their chromatic behavior. Moderate bandwidth ($\Delta\lambda/\lambda \approx 20\%$) and spectral resolution ($\lambda/\delta\lambda \approx 20$–$60$) is required for implementation of multi-color speckle suppression (Sparks & Ford 2002). Multicolor speckle identification is a photometric procedure, and consequently places significant demands on the accuracy of flat fields. For example**,** Table 6-2 shows that with a flat field accuracy of 1% (per IFS pixel) speckle suppression is effectively limited to a factor of 16, and to a factor of 64 for 0.1% accuracy.

Table 6-2. Flat field accuracy required for speckle suppression.

Speckle suppression factor	Planet detection rate (%)	
	1% Flat field	0.1 % Flat field
1	2.31	2.45
4	2.91	4.11
16	3.10	5.87
64	3.19	6.38
256	3.19	6.52
∞	3.19	6.52

- **REQ-OCD-0030: Speckle noise suppression of uncorrected atmosphere and static aberrations to a level required to meet REQ-OCD-0011 using a combination of SDI and ADI (estimated to be × 5-10).**
- **REQ-OCDD-0130: Flat field accuracy 1.5×10^{-4} RMS per beam on spatial frequencies ~ $3\lambda/D$**
- **REQ-OCD-0150: 14±0.5 mas pixel sampling.**

Point source sensitivity: The need to detect faint exoplanet companions places requirements on throughput, instrumental background, and detector performance. The required sensitivity is derived from the detection rate. Although some exoplanets are expected to be as bright as $H \approx 15$ mag., to achieve a detection rate of 7% requires detection of planets with $H \approx 23$ mag. This implies a point-source sensitivity, in the absence of speckle and star-halo shot noise, comparable to existing Gemini AO instruments.

- **REQ-OCD-0070: $H = 23$ mag., ≥ 5-σ, 1 hour.**

Inner working distance: The angular search space required for the field star survey can be deduced from the distribution of semimajor axes (assuming $dN/da \sim a^{-1/2}$) and host star distances. Most planets are found at small semimajor axis separation. Therefore, there is a strong desire to minimize the “inner working distance”—the smallest angular separation where good contrast is delivered. However, the region within 3 AU is currently well explored by Doppler surveys and $a < 5$ AU will be charted by the time GPI is deployed. Thus, to maintain complementarity with indirect searches GPI must focus beyond this separation. The median distance of the $I < 8$ mag. solar neighborhood sample is 35 pc thus the inner working distance should be ≈ 0.14 arc seconds.

Current Doppler techniques are insensitive to the signature of exoplanets orbiting young (< 1 Gyr) stars because of photospheric kinematic jitter associated with chromospheric activity. However, if techniques are developed that can suppress this systematic error the most likely approach for resolving the inclination angle ambiguity for these exoplanets is not GPI but astrometry. The case for developing exotic coronagraph architectures, which deliver high contrast at separation of less than 3-4 λ/D is weak.

Halving the inner working distance would approximately double the number of detectable planets, and extend the reach of GPI into the semimajor axis domain of Doppler searches. This challenge should be reserved as a goal for the next generation of direct planet imagers, e.g., TPF-C.

A few Doppler exoplanets may lie at just at the threshold of detection at the natural IWD of GPI (Burrows et al. 2004). It is a goal to leave an option for a narrow field of view coronagraph mode to pursue detection of these targets.

- **See REQ-SCI-0010 and REQ-OCD-0011**

Field of view: The planet detection rate increases with field of view. For example, a few nearby (< 5 pc) stars may host planets in wide orbits with semimajor axes of 30–50 AU. Since such large angular separations lie well outside the likely AO control radius, $\lambda/2d$, of 1 to 1.5 arc seconds, these targets should be ceded to conventional AO systems such as NICI or the VLT planet finder. However, the statistics of planets in wide orbits for typical stars in the sample (median distance 35 pc) carries key information about the efficiency of various planet formation pathways and the mechanism of planetary migration. This reason requires that the field of view must extend at least to the edge of the control radius if not beyond. Inspection of the cumulative distribution for the straw man GPI system suggests that the point of diminishing returns is achieved at a field angle of 1.4 arc seconds, where 80% of the population has been detected. Thus, the required field of view is estimated to be at least 2.8 arc seconds × 2.8 arc seconds. Other science (in particular debris disk imaging) may benefit from slightly larger FOV.

- **REQ-OCD-0160: Field of view in direct imaging is ≥ 2.8 × 2.8 arc sec^2.**

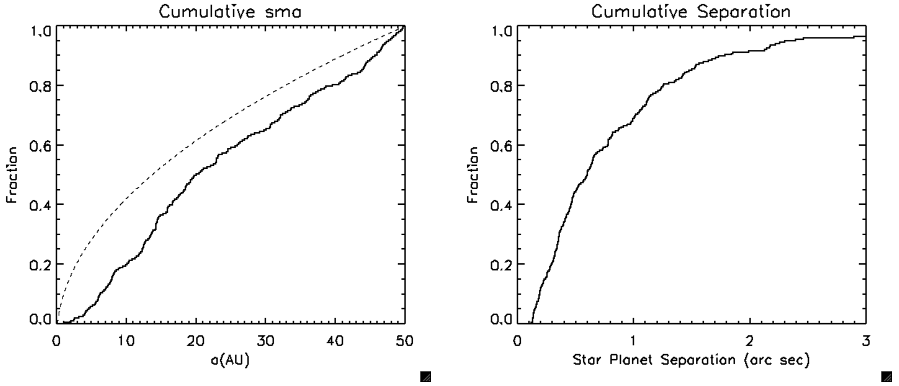

Figure 6-2. *Left:* Cumulative exoplanet semimajor axis distribution. The dashed line shows the underlying population and the solid line shows the semimajor axes of detected exoplanets in a field survey (*I* < 8 mag.). At 10 AU the sample is 50% complete, by 30 AU, the sample is nearly 90% complete. The planets are drawn from *H*-band detections with the straw man GPI using 18-cm sub-apertures. *Right:* Corresponding cumulative star-planet angular separations.

Astrometric accuracy: Astrometric measurements are required to confirm the common proper motion status of planetary companions. The discovery observation must have astrometric integrity so that follow-up observations can be planned without immediate re-observation to acquire first epoch data.

For field-star exoplanets at 50 pc, typical annual motion amounts to 28 mas (one axis). It is desirable that common proper motion status can be confirmed in an interval of less than one year so that additional follow-up observations can be scheduled in a timely fashion. This can be accomplished with 4 mas, 1-σ errors. The astrometric solution for the IFS must be stable or calibratable so that systematic errors do not exceed this threshold. GPI must either consistently locate the host star at the same point in the field of the science camera and maintain its orientation relative to the celestial reference frame, or provide an algorithm for measuring the position of the exoplanet candidate relative to the primary.

- **See REQ-OCD-0120, REQ-OCD-0121, & REQ-OCD-0122**

Zenith distance limit: A significant fraction of the celestial sphere must be accessible to accumulate a statistically significant sample targets and to enable efficient scheduling. We estimate that 6 sr or 48% percent of the sky must be The fraction of the sky observable at a site with latitude l and zenith limit, z, is $\cos(l)\sin(z)$. To cover this fraction of the sky requires a 30° zenith limit at Mauna Kea ($l = +19.6°$) and a 34^{o} degree zenith limit at Cerro Pachon ($l = -30.2°$).

- **See REQ-OCD-0100: and REQ-SCI-0110**

Photometric accuracy: The field star survey places weak constraints on photometric accuracy. If the age of the host star is known, then a photometric measurement with 10% accuracy is sufficient to estimate the mass of the exoplanet. Requirements that are more stringent are called out elsewhere, e.g., for speckle noise suppression.

- **REQ-OCD-0140: Absolute photometric calibration ≤ 10%**

6.1.2. Searches for exoplanets in young clusters and associations

Observations of young clusters afford an excellent opportunity for direct detection of Jovian-mass planets. The luminosity of a young exoplanet scales roughly as t^{-1}, so a 100 Myr-old planet is ten times more luminious than its mature Gyr-old counterpart. Consequently, the first targets for an initial reconnaissance of exoplanets will likely be drawn from catalogs of young, nearby stars. Table 6-3 shows an inventory of nearly 500 young stars with $I < 9$ mag.

The ages span 30–650 Myr . As described in the sample discussion, the target sample consists of four categories: (1) all field stars brighter than the system magnitude limit, (2) nearby field stars with spectroscopic, kinematic, and X-ray signatures of youth (< 100 Myr), (3) members of young (8–50 Myr) associations, and (4) members of young (90–650 Myr) nearby (25–165 pc) open clusters.

Table 6-3. Young, nearby groups, associations & clusters (*I* < 9 mag.)

	#	Age (Myr)	Median distance (pc)	Distance range (pc)
Young assoc.	101	8-50	44	10-105
U Maj	56	300	26	9-121
Hyades	169	660	47	30-67
Coma Ber.	35	400	~ 85	
Pleiades	48	125	~ 130	
α Per	48	90	~ 165	

The youngest stars in Table 6-3 are listed as "young associations", which includes members of β Pic, η Cham, Tucana/Horologium and TW Hydrae as well as young, nearby stars not apparently belonging to any known groups. Because of its youth, this group includes the most luminous exoplanets. Members of "young associations" also represent the most relevant sample for revealing the transition from debris-disk phase to mature planetary system. Since the contrast required to detect a 5 M_J exoplanet at 30 Myr could be only ≈ 1×10^{-5}, this group will be studied intensely by Gemini/NICI and the VLT planet finder. If the required GPI contrast is descoped from the level required for the field star survey, specified in § 6.1.1, then this group acquires even greater significance.

Zenith distance limit and accessible declination range: Figure 6-3 illustrates the well-known N/S asymmetry present in current catalogs of the youngest stars. There are three times as many targets accessible from Gemini S than Gemini N at a limiting magnitude of *I* < 8 mag. and a zenith distance limit of 40^{o}. Given that this sample of young stars and their solar systems will dominate the discovery phase of exoplanet imaging, it is a requirement that these targets be accessible to GPI.

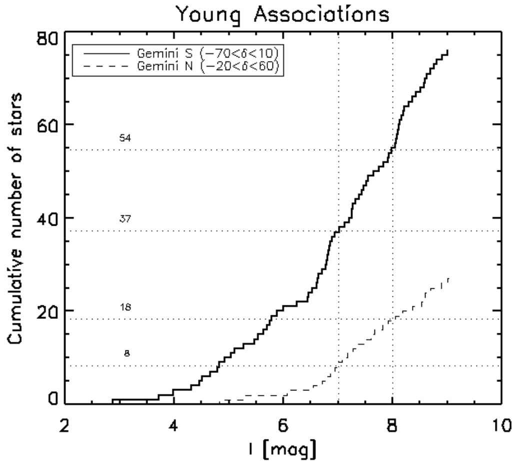

Figure 6-3. The cumulative distributions for young stars vs. magnitude assuming a zenith distance limit of 40°. This excludes open clusters but includes β Pic, η Cham, Tucana/Horlogium and TW Hydrae as well as young, nearby stars not apparently belonging to any known groups. There are three times as many targets in the southern hemisphere as in the north at *I* < 8 mag. and within a 40° zenith distance limit.

The Hyades, with an age of 660 Myr and a distance of 47 pc, represents a significant pool of young stars that place requirements on accessible declination range. See Table 6-4. The Hyades, with an initial mass of 1200 $M_{\odot}$, presumably contained a few O stars, and may have resembled the Orion Nebular cluster. The Hyades therefore presents an important setting in which to evaluate the process of planet formation in a region of massive star formation. For 100 Hyads with $I < 8$ mag. to be observable at Gemini S requires a zenith distance limit of 50^{o} given the declination of this group ($\delta \approx +15^{o}$). Alternatively, a two-telescope survey could cover the Hyades by moving to Gemini North.

- **REQ-OCD-0100: Zenith distance range ≤ 50^{o}**
- **REQ-SCI-0110: Working declination range is -70^{o} to $+15^{o}$**

Contrast: As in § 6.1.1, required contrast is a derived from the specified planet detection rate. There are currently about 170 known young (< 660 Myr), bright ($I < 8$ mag.) and nearby ($d < 60$ pc) targets that are observable from Gemini South with a zenith distance limit of 50°. The exoplanet detection rate for the young stars must be greater than that for field stars because the target list is shorter. A planet detection rate of 25% is required to measure the abundance of exoplanets hosted by young stars and it is a goal to achieve 40% to estimate the distributions of planetary masses and semimajor axes.

Figure 6-4 shows the cumulative distribution of contrast at which exoplanets are detected in a survey of young ($d < 660$ Myr) stars with the straw man GPI that achieves a detection rate of 36%. Although the younger exoplanets are brighter, the contrast requirements for the young star and for the field surveys are comparable. This is because the higher detection rate for young stars is achieved by probing the distribution exoplanets to lower masses, which are fainter. Nonetheless, significant progress could be achieved with a contrast that is a factor of three poorer than required for the field-star experiment, or about 10^{-7}. Figure 6-4 shows also shows the significant population of relatively bright exoplanets that are expected to be associated with the young star sample. At a contrast detection threshold of 10^{-6} at most a few dozen planets will be detectable: insufficient for a detailed statistical analysis, but sufficient to begin the study of exoplanet atmospheres.

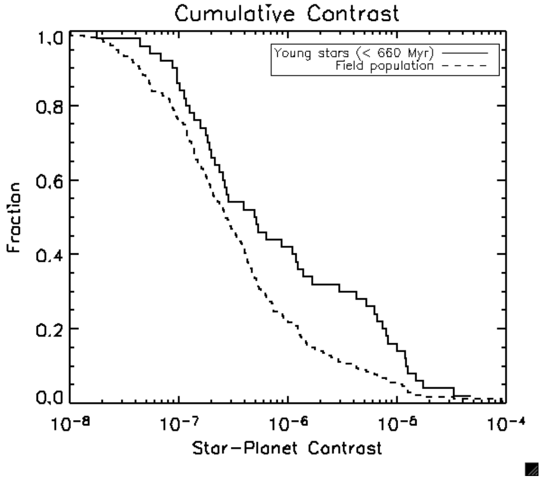

Figure 6-4: Cumulative fraction of the exoplanet population as a function of detection threshold. The field population of detected exoplanets (no age cut) shown in Figure 6-1 is compared here (dashed) with the exoplanet drawn from young associations and open clusters (age < 660 Myr). In both cases, the properties of the straw man GPI system have been assumed and all stars are within 50 pc.

It is a priority to observe this sample rapidly during the initial deployment of GPI. It is a requirement that the initial reconnaissance be completed in 20 nights of telescope time.

The required contrast is relaxed between zenith distances of 40° and 50° because of the relative abundance of Hyads.

- **See REQ-OCD-0011:**

Wavelength range: The young planets found in surveys of clusters and associations will be hotter than those found in the field, and consequently the contrast between parent star and planet is greater at shorter wavelengths. Figure 5-4 shows that 100–300 Myr-old planets are bright in J (1.25 µm) and Y (1.03 µm). In addition, the majority of clusters and associations are relatively distant, which argues for minimizing λ/D by working at the shortest practical wavelength. Operation in J band is therefore required.

- **See REQ-OCD-0040**

Spectral resolution: Exoplanets associated with young stars are expected for be warm (Figure 5-3). A 5 M_J object at 100 Myr has $T_{\rm eff} \approx 600$ K. This is above the condensation temperature for H_2O, so we can have confidence in our knowledge of the opacity sources and that CH_4 differencing is an appropriate method for speckle suppression. It is therefore required that multicolor speckle suppression based on three $\lambda/\delta\lambda \approx 30$ bands spanning the CH_4 bandhead at 1.59 µm (Marois et al. 2002). Analysis of the spectral resolution is further discussed in the Appendix on spectral resolution.

- **See REQ-OCD-0050**

Limiting magnitude: While Table 6-4 shows that significant numbers of target stars can be observed with a limiting magnitude of $I = 8$ mag., Figure 6-3 shows that a significant number of targets have $I =$ 8–9 mag. Since there are only modest numbers of these young targets, the performance of GPI should lock on stars as faint a $I = 9$ mag., and degrade in a graceful manner for $I \geq 8$ mag.

- **See REQ-OCD-0020**
- **REQ-OCD-0022: AO system will lock on I = 8–9 mag. stars**

Table 6-4. Number of targets in young associations and open clusters broken down by magnitude and declination. The Hyades dominates the N hemisphere. There are currently about 170 known young (< 600 Myr), bright (I < 8 mag.) and nearby (d < 60 pc) targets observable with a zenith distance limit of 50° from Gemini South.

	Young Assoc.	Ursa Maj.	Hyades	Coma Ber.	Total
North (δ > 0°)					
$I < 7$	8	31	91	17	146
$I < 8$	18	36	129	27	210
South (δ < 0°)					
$I < 7$	37	11	7	0	55
$I < 8$	54	11	7	0	72

Pixel sampling: Spatial pixel sampling is a derived requirement that depends on the shortest operating wavelength. Nyquist sampling at this wavelength is inferred from the derived requirement of astrometry and speckle suppression, both of which require good sampling at the cut-off spatial frequency of λ/D. This implies a pixel size of 14 mas, assuming a perfect pixel MTF.

- **See REQ-OCD-0150**

6.1.3. Confirmation and characterization of exoplanet candidates

Field and cluster surveys will yield hundreds of candidates for further study and follow-up. These objects will include false alarms that fall into several categories. Non-astrophysical false alarms will be eliminated by repeat observations, eliminating instrumental artifacts caused by statistical fluctuations. During the course of the survey, persistent artifacts, e.g., due to glints, ghosts, or uncorrectable phase or amplitude errors, will be catalogued. Candidates adjacent to known flaws will be assigned the lowest priority for follow-up.

Astrometric accuracy: The first class of astrophysical interlopers consists of foreground or background objects, which are projected along the line of sight. The first stage of winnowing will be done by measuring proper motion (see § 6.1.1). Physical association, however, is best confirmed by demonstrating orbital acceleration, and ultimately measurement of orbital elements. A minimum measurement necessary to demonstrate acceleration consists of three measurements, separated by an interval Δt. Performing error propagating, assuming a circular orbit and a solar mass primary, the required one-axis RMS positional error is $\sigma_1 = \sqrt{2/3}(\pi\Delta t)^2(da)^{-1}$ arc seconds, where d is the distance in pc and a is the semimajor axis in AU, and Δt is in years. For $\Delta t = 1$ yr, and a 3-σ detection at the 90-th percentile of the standard population ($I < 8$ mag., $dN/da \sim a^{-1/2}$, $d < 50$ pc) requires $\sigma_1 = 1.8$ mas. The astrometric solution for the IFS must be stable or calibratable so that systematic errors do not exceed this threshold.

It is expected that this astrometric accuracy will be achieved through use of a pupil-plane grid to generate artificial ghost images (Marois et al 2006, Sivaramakrishnan et al 2006).

- **REQ-OCD-0120: Astrometric accuracy ≤ 1.8 mas, 1σ, (1-axis)**
- **REQ-OCD-0121: Location of target star is known to << 1.8 mas, 1σ, (1-axis)**

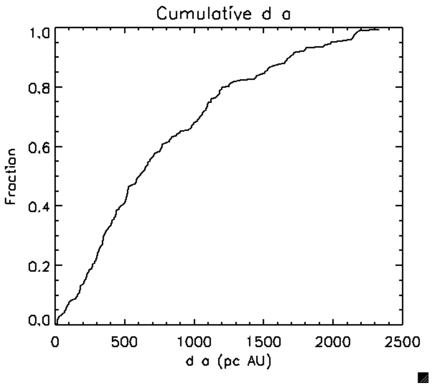

Figure 6-5. The cumulative distribution of the product of semimajor axis, *a*, and distance, *d*, in units of pc AU. A clear indication of physical association between star and candidate exoplanet is detection of acceleration due to Keplerian

motion. For a circular orbit and a solar mass primary, the acceleration is $4\pi^2/(da)$ arc seconds yr^{-2}. The standard population, I < 8 mag., $dN/da \sim a^{-1/2}$, d < 50 pc, has been assumed.

Instrument lifetime and long-term astrometric stability: An eccentricity precision of < 0.1 is adequate to distinguish planets of different origins, e.g., planets formed *in situ* ($\varepsilon \approx 0.0$) and planets scattered to large semimajor axis ($\varepsilon > 0.5$). Preliminary calculations show that long term monitoring rather than improved astrometric accuracy yields the most reliable way to estimate the orbital eccentricity, ε. For example, for a typical exoplanet (a = 30 AU, d = 30 pc), this level of precision can be achieved with $\sigma_1 = 1.7$ mas after an interval of about 10 years, with three to four one-hour observations per year. Although the final requirement on astrometric accuracy and precision for quantifying orbital elements is to be determined, it is evident that the instrument lifetime must be at least 10 years, and long term astrometric stability must be maintained or able to be calibrated over this period.

- **REQ-OCD-0122: 10-year instrument lifetime and astrometric stability**

Spectral resolution: Physically associated, low luminosity interlopers include brown dwarfs and white dwarfs. Ancient (13 Gyr) white dwarfs are more luminous and are hotter than all but the youngest and most massive exoplanets. Although white dwarfs should be found in abundance in the field star survey, they will immediately distinguished from exoplanets by their luminosity and *JHK* colors.

In addition to broadband filters, the IFS can be used to synthesize narrow band-pass filters. For example, narrow band indices ($\lambda/\delta\lambda \approx 50$, $\delta\lambda \approx 0.03\mu$m) covering portions of the *H* band have been shown to correlate with spectral type over the entire L and T range (Mainzer *et al.* 2004). These include regions corresponding to "continuum" emission at 1.595 µm and molecular absorption at 1.495 µm (H_2O), 1.66 µm (CH_4), and 1.75 µm (H_2O). Features within the *J* band are particularly useful, for example the KI lines are sensitive to surface gravity, while FeH and H_2O bands are important for assigning spectral types (*e.g.* McGovern *et al.* 2004).

Old brown and young planets can have similar effective temperatures. However, these two classes are distinguished by surface gravity. For example, referring back to Figure 5-3 shows that a 25 M_J brown dwarf at 5 Gyr has the same temperature as a 7 M_J planet at 300 Myr, but the gravity of these objects differs by 0.8 dex. The IFS is required to provide sufficient spectral resolution and wavelength span to measure effective temperature and surface gravity to discriminate between these two classes of objects. Guided by the theory of exoplanet atmospheres, Figure 6-6 suggests that a precision of about 20 K and 0.1 dex in log(g) can be accomplished with spectral resolution, $\lambda/\delta\lambda \approx 40$, a wavelength span, $\Delta\lambda/\lambda \approx$ 15% (including 1.5-1.7 microns at H band) and moderate (10%) photometric accuracy. Since sensitivity of GPI degrades (due to thermal backgrounds) beyond 2.2 microns, it is acceptable to divide K band into two sub-bands of lower spectral coverage.

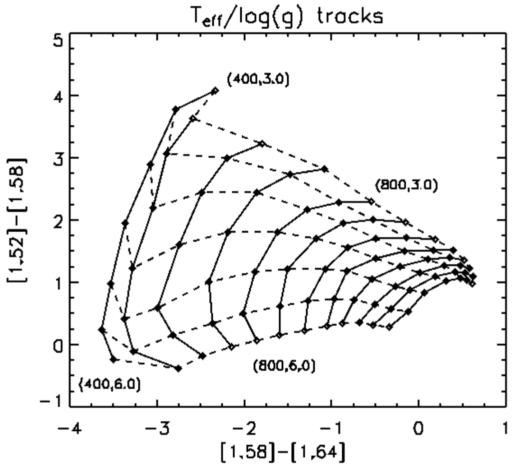

Figure 6-6. Color-color diagram defined from 2.5% band pass colors in *H* band at 1.52, 1.58 and 1.64 µm. Colors are expressed in magnitudes. Each track is at constant effective temperature (solid lines) ranging from 400 K (left) to 1500 K (right) in increments of 100 K. Each point on a given track refers to different surface gravity (dashed lines) ranging from log *g* = 3.0 (top) to 6.0 (bottom) in increment of 0.5 dex in CGS units. Color indices are derived from "COND" atmosphere models of Allard et al. (2001). Low-resolution ($\lambda/\Delta\lambda \approx 40$) spectroscopy within the *H* band provides a good diagnostic of effective temperature and surface gravity, especially at low (~ 500 K) temperature.

This requirement is refined to a minimum spectral resolution of $\lambda/\delta\lambda \geq 45$, based on a detailed analysis given in OCDD Appendix Spectral Resolution and Inferring Exoplanet Mass and Age (§ 13).

- **REQ-OCD-0050: H-band Spectral resolution $\lambda/\delta\lambda \geq 45$ over the majority of the field of view**
- **REQ-OCD-0051: Instantaneous wavelength coverage ≥ 15% for YJH bands.**

6.2. Debris disks

Debris disks are the extrasolar analogs of our Zodiacal dust and Kuiper Belt, but on account of their youth they are signposts of the transitional phase from protoplanetary disk to mature solar system. The handful of existing images at submillimeter and near-infrared wavelengths show substantial structure (e.g., Holland et al. 2003; Greaves et al. 1998; Schneider et al. 1999; Krist et al. 2000; Koerner et al. 2001; Wilner et al. 2002; Weinberger et al. 2002; Weinberger et al. 2003; Clampin et al. 2003). Specifically, β Pic, Vega, ε Eri, Fomalhaut, and HD 141569 all show distinct nonaxisymmetric structure. HR 4796 is orbited by a narrow dusty ring. Only the dust at TW Hya has failed to reveal any structure. Excepting perhaps HD 141569 (Clampin et al. 2003), the most likely explanation of the observed structures is gravitational perturbations by planets with semimajor axes comparable to the radius of the dusty rings and disks. For Vega, a 3 M_J planet is suggested (Wilner et al. 2002), while for ε Eri either a 0.2 M_J (Ozernoy et al. 2000) or a 0.1 M_J (Quillen & Thorndike 2002) planet has been proposed. Indeed, COBE found that the Earth is led and trailed in its orbit by clumps of dust particles (Reach et al. 1995).

The two categories of GPI investigation will be a survey mode and a case-study mode. The case-study investigations will likely comprise early observations for Gemini GPI because ground-based AO and

HST have already imaged a handful of debris disks. The fidelity of data for these targets will provide important early on-sky verification of the performance of GPI and represent an important opportunity to demonstrate superior relative to previous facilities. GPI observations will yield the first images at of these systems at near-infrared wavelengths, record polarization information, record structure closer to the central star then previously possible, and hold out the best chance of direct detection of planets embedded in the disk. Debris disks span a large range of surface brightness. As Figure 6-7 illustrates, the distribution of optical depths in the IRAS debris disk catalog of Zuckerman & Song (2004) covers three orders of magnitude, yet only a few of the highest optical depth systems, $\tau > 10^{-3}$, have been imaged.

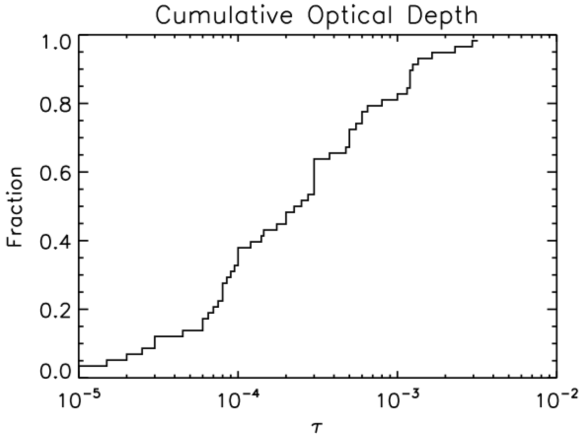

Figure 6-7. The cumulative optical depth distribution from Zuckerman and Song (2004), who list the most comprehensive catalog of IRAS-detected Vega-like stars.

Contrast: The goal of the survey is to image all IRAS-excess stars with fractional infrared luminosities $\tau \equiv L_{\rm dust} / L_{\rm star} \geq 3 \times 10^{-5}$. This corresponds to the majority of stars detected as Vega Phenomenon objects by IRAS or ISO, representing A–M main sequence stars with ages from ~ 10–600 Myr.

Defining contrast as $Q = \Omega_{dust} \tau / (\lambda / D)^2$ we can use the debris disk optical depth and solid angle subtended, Ω_{dust}, tabulated by Zuckerman and Song (2004) to estimate the ratio between the mean face-on surface brightness of the disk and the star in a diffraction-limited beam. The observed surface brightness will on average be greater than the face-on brightness, due to projection effects and radial and azimuthal structure, while the surface brightness at near-infrared wavelengths will be a factor of a few lower than this estimate because dust grains are not grey. Neglecting these refinements, Figure 6-8, shows that we require GPI to achieve $Q \approx 5 \times 10^{-8}$ (after post-processing with polarimetric or estimated-PSF subtraction) to reach a significant fraction (> 80%) of the debris disk sample.

- **REQ-SCI-0012: Detect $\tau \leq 3\text{x}10^{-5}$ debris disks for $I \leq 8$ mag.**
- **REQ-OCD-0013: Surface brightness contrast $Q \leq 5\text{x}10^{-8}$ per $(\lambda/D)^2$ beam for $I \leq 8$ mag.**

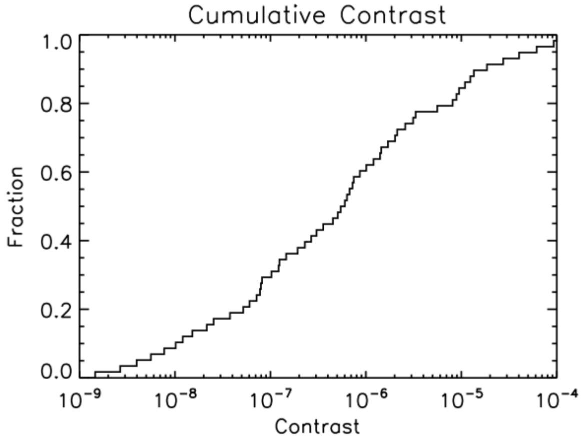

Figure 6-8. Cumulative debris disk contrast, *Q*, defined as the average face-on surface brightness of the disk compared to the brightness of the star in a diffraction-limited beam.

Limiting magnitude: All known debris disk stars with $Q \geq 5 \times 10^{-8}$ have $I < 7.9$ mag., this threshold must be achieved for a guide stars dimmer than $I = 8$ mag.

- **REQ-OCD-0020: WFS limiting magnitude $I \geq 8$ mag.**

Declination and zenith distance limit: A few famous debris disks are southern hemisphere objects, e.g., Fomalhaut, AU Mic and β Pic. Moreover, the majority of stars in young nearby associations such as the Beta Pictoris moving group and Tucanae-Horologium lie at southern declinations. These targets offer a coeval (~10 Myr) cohort by which we can compare disk evolution around stars with different masses. Although debris disk candidates exist both in the northern and southern hemisphere, Gemini South is the required location for this comparative study. Because the target lists includes only a modest number of stars, a zenith distance limit of 50° is required to cover 70% of the sample with $Q \geq 5 \times 10^{-8}$.

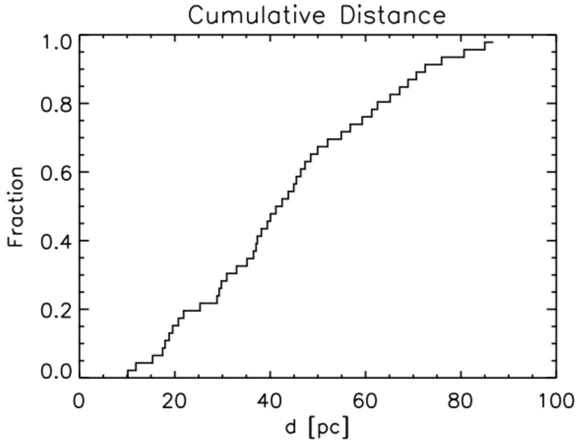

Figure 6-9. Cumulative distance distribution of Vega excess stars from Zuckerman & Song (2004) with $Q \geq 5 \times 10^{-8}$.

- **See REQ-OCD-0100 and REQ-SCI-0110**

Inner working distance: Many debris disks, including our Kuiper Belt, are truncated on the outside at 5-150 AU radii. Figure 6-9 shows the cumulative distance distribution of Vega excess stars from Zuckerman & Song (2004) with $Q \geq 5 \times 10^{-8}$. To reach the inner structure in at least half the sample the inner working distance is required to be 0.14 arc seconds.

- **REQ-OCD-0014: Inner working distance ≤ 0.14 arc sec**

Field of view: In addition to an outer truncation, inner debris disk holes are common, with radii between 5 and 100 AU. The radial sharpness and azimuthal structure of the inner debris disk boundary are critical for inferring the orbital parameters of an unseen planetary system. For a sample median distance of 42 pc in the $Q \geq 5 \times 10^{-8}$ sample the radial extent of the field of view should be at least 2.4 arc seconds radius. Assuming that this can oriented along a diagonal of a square field, the field is required to be 3.4 arc seconds × 3.4 arc seconds. The baseline GPI field of view does not fully support this; it is a goal to have a mode that allows the IFS field of view to be dithered with respect to the sky.

- **See REQ-OCD-0160 and REQ-OCD-0161**

Wavelength and polarization: The extended low surface brightness signal of debris disks means that *H* band is the best compromise between the higher thermal background of the *K* band, and the lower Strehl correction of *J*-band observations. A debris disk signal may be present, but indistinguishable from the scattered and diffracted light halo of the central star. However, scattered light from dust grains is polarized signal (≤ 30% for single scattering), whereas the stellar instrumental halo will not have a significant polarization component (the instrumental polarization signal can be measured). Dual channel polarimetry is robust against non-common path WFE, and therefore offers a significant advantage for suppressing speckle noise, enabling the detection of a debris disk signal, and characterizing the physical properties of the dust grains.

A detailed Mueller matrix analysis of a dual channel polarimeter is given in OCDD Appendix-A Mueller Matrix Analysis of Non-Idea Dual Channel Polarimeter (§14)

- **REQ-OCD-0060: Dual channel polarimetry mode**
- **REQ-OCD-0061: Precision of linear polarization ≤ 0.3%**
- **REQ-OCD-0031: Suppression of unpolarized speckles ≥ 100 (3-σ)**

PSF stability and repeatability: Achieving surface brightness contrast of $Q = 5 \times 10^{-8}$ places significant demands on knowledge of the PSF, especially for observations of debris disk targets whose geometry produce low polarization. If the PSF were stable, the first step for removing the uncorrected halo over a given few-hour observing block would be to observe PSF reference stars that will be subtracted from the science target data. This is a somewhat different problem than PSF subtraction for planet detection, in that what matters is not the detailed speckle pattern (dominated by static wavefront error effects), but the average PSF halo intensity (dominated by atmospheric and AO conditions). PSF variations are therefore inevitable, and what is required is not necessarily PSF stability, but sufficient knowledge of system performance to allow partial PSF reconstruction. The reconstruction of the average PSF halo intensity may be assembled from a combination of PSF star observations, AO telemetry and WFE information from the calibration system. Although disks can be detected through polarized light without additional reference PSF subtraction, quantitative polarimetry—a powerful tool for probing dust parameters (Graham et al. 2007) – requires photometry of the unpolarized component of the dust scattered light as well. This may not be possible for the dimmest disks, but reconstructing the PSF (from telemetry or reference stars) at 4% precision will allow quantitative polarimetry for disks with $Q > 1 \times 10^{-6}$ at the SNR=10 level.

- **REQ-OCD-0170: Goal: PSF reconstructability at 4% precision (RMS per pixel) over 3 hours.**

6.3. Adjunct Science

Adjunct science adds goals for instrument performance, not requirements.

6.3.1. Solar System Exploration

Goals for solar system exploration include:

- Clear field of view with no coronagraphic mask
- Operation in *Y* band (0.97–1.07 µm) for diffraction limited operation at the shortest possible wavelengths
- The ability of the wavefront sensor to lock on small extended sources such as Galilean moons, e.g., Io with diameter 1.2 arc seconds
- Moderate performance (exceeding ALTAIR image quality) for R =10–11 mag. targets
- Low spectral resolution ($R \approx 50$) for surface composition studies of icy bodies
- Spectopolarimetry to trace the particle size and altitude distribution of atmospheric hazes

7 Scientific Requirement Summary Matrix

	Exoplanet field star survey	**Young assoc. and cluster survey**	**Exoplanet properties**	**Debris & protostellar disks**	**Solar system**	**Final requirement**
Contrast vs. angular separation	Required: 7% planet recovery at *H* for $I \leq 8$ mag. C $\approx 3\times10^{-7}$ at 0.17 arc sec; improving to 2×10^{-8} at 0.5 arc sec (1 hr 1-σ) and maintained to edge of dark hole. Includes speckle suppression. May by narrow or broad band, depending on star brightness.	Required: 20% planet recovery. $C < 3\times10^{-7}$ between 0.17–0.6 arc sec.	Goal: 40% planet recovery rate.	Requirement: Detect $\tau \approx 3 \times 10^{-5}$ or Q $\approx 5 \times 10^{-8}$ (equivalent to HR 4796A/100 pole on). Requirement: Inner working distance 0.”14.		**REQ-SCI-0010:** Planet recovery $\geq$ 7% at *H* for $I \leq 8$ mag. field-star survey. **REQ-OCD-0011:** *H*-band point source contrast of $C \leq 3 \times 10^{-7}$ at 0.17 arc sec, $\leq 2 \times 10^{-8}$ at 0.5 arc sec (1 hr 1-σ) for $I \leq 8$ mag. to edge of dark hole. **REQ-SCI-0012:** Detect $\tau \leq 3 \times 10^{-5}$ debris disks for $I \leq 8$ mag. **REQ-OCD-0013:** Surface brightness contrast $Q \leq 5 \times 10^{-8}$ for $I \leq 8$ mag. **REQ-OCD-0014:** Inner working distance $\leq$ 0.14 arc sec.
WFS mag. limit & λ	Required: 700–900 nm, I = 0–8 mag.	Goal: Graceful performance degradation for $8 < I$/mag. < 9 stars for G stars in Hyades.		Required: Bright limit I = 0 mag. (selectable).	Goal: Lock AO on extended objects, e.g., Io diameter ~ 1.2 arc sec (selectable).	**REQ-OCD-0020:** WFS limiting magnitude $I \geq 8$ mag. **REQ-OCD-0021:** WFS bright magnitude $I \leq 0$ mag. **REQ-OCD-0022:** Graceful WFS degradation for I = 8–9 mag.
Speckle suppression	Required: speckle noise suppression using a combination of ADI & SDI.	Goal: Speckle noise suppression multi-band CH_4 differencing.		Required: polarimetry mode: speckle noise suppression $\geq \times$ 100		**REQ-OCD-0030:** Speckle noise suppression $\geq \times$ 5 using ADI/SDI. **REQ-OCD-0031:** Suppression of unpolarized speckles $\geq \times$ 100.
Science wavelength range	Required: 1.1-2.35 µm.	Goal: 0.96–2.4 µm sensitivity to young planet. Wavelength in selectable $YJHK_s$ bands.		Required: polarimeter works at *H* & *K*. Goal: constrain grain size distribution $0.96 \leq \lambda/\mu\text{m} \leq 2.4$.	Goal: 0.8 µm for best angular resolution at moderate Strehl (selectable).	**REQ-OCD-0040:** Science operating wavelength: 1.1–2.35 µm.
Spectral resolution	Required: multi-color speckle rejection, $R \geq 30$ or 5 or more “colors” in *H*. Instantaneous band pass $\geq$ 15%.	Goal: instantaneous band pass $\geq$ 20%, but not at expense of FOV.	Required: $R \approx$ 40-50 in *JHK* to measure T_{eff} to $\leq$ 100 K and log*g* to $\leq$ 0.1 dex.	Goal: $R \approx 50$ for ice and SiO bands.	$R \approx 50$ for ice mineralogy at 2 µm with instantaneous band pass $\geq$ 20%.	**REQ-OCD-0050:** Spectral resolution $\geq$ 45 in *H* band. **REQ-OCD-0051:** Instantaneous wavelength coverage $\geq$ 15% in YJH band.
Polarimetry	Required: Distinguish exoplanets & zodiacal blobs: sensitive to 0.3% linear polarization.			Required: Dual-channel polarimetry for immunity to non-common path WFE.	Atmospheric haze particle size distribution.	**REQ-OCD-0060:** Dual channel polarimetry mode **REQ-OCD-0061:** Precision of linear polarization $\leq$ 0.3%
Sensitivity	Required: detection of faint exoplanets, $H \leq 23$ mag. when not speckle noise limited.		Goal: *H*-band throughput $\eta >$ 25%, top of atmosphere to photo-electrons.			**REQ-OCD-0070:** H = 23 mag., $\geq$ 5-σ, 1 hour.
Detector	Required: detector noise increases residual speckle & photon shot noise by $\leq$ 20% at $R \approx$ 50.	Required: fast, low read noise permits ISS stationary-mode observations without significant image blur.	Goal: detector noise increases residual speckle & photon shot noise by < 20% at $R \approx$ 50.			**REQ-OCD-0080:** Detector noise is $\leq$ 20% of residual photon shot noise and residual speckle noise at $R \approx$ 50.

Operability, reliability	Required: Nominal operation in best 50% seeing. Survey 2000 stars in 200 nights or 90% open shutter time. Statistical false alarm rate ≤ 5%.	Requirement: acquire, configure AO and observe up to 20 targets per night		Accumulate deep exposures	Capture rare events	**REQ-OCD-0090:** Survey 2000 stars in 200 nights to **REQ-OCD-0011** level. **REQ-OCD-0091:** Statistical false alarm rate ≤ 5%
Zenith distance limit	Required: ≤ 40^{o} for reach sufficient numbers of field stars	Goal: ≤ 50^{o}, graceful degradation > 45^{o}. to reach Hyades ($\delta \approx +15^{o}$)				**REQ-OCD-0100:** Zenith distance range ≤ 50^{o}
Accessible Dec. range	Required: Consistent with ZD limit.	Required: young assoc. and nearby open clusters, $-70^{o} \leq \delta \leq +15^{o}$).		Required: cohort of debris disk stars in β Pic & Tuc-Hor moving group $-65^{o} \leq \delta \leq 0^{o}$).	Goal: $-25^{o} < \delta < 25^{o}$	**REQ-SCI-0110:** Working declination range is -70^{o} – +15^{o}
Astrometric accuracy	Required: 4 mas 1σ (1-axis) for common proper motion. Star on FPA is set or measured to << 4 mas.	Required: orbital acceleration of companions ≤ 1.8 mas, 1σ, (1-axis).	Required: 10 yr term instrument lifetime and astrometric stability for measurement of orbital elements.			**REQ-OCD-0120:** Astrometric accuracy ≤ 1.8 mas, 1σ, (1-axis). **REQ-OCD-0121:** location of target star is known to << 1.8 mas, 1σ, (1-axis). **REQ-OCD-0122:** 10-yr instrument lifetime & astrometric stability
Flat fielding accuracy	Required: ≤ 0.1% RMS/pixel or 1.5×10^{-4}/ beam on spatial frequencies ~ λ/D for ADI/ SDI (speckle suppression).			Sensitivity to surface brightness		**REQ-OCDD-0130:** Flat field accuracy ≤ 1.5 × 10^{-4} RMS per beam on spatial frequencies ~ λ/D.
Photometric accuracy	Required: absolute ≤ 10%		Required: relative narrow band indices to ≤ 5% to measure T_{eff} and log$g$			**REQ-OCD-0140:** Absolute photometric calibration ≤ 10% **REQ-OCD-0141:** Relative photometric accuracy ≤ 5%.
Pixel sampling	Required: 1.5-2 × Nyquist @ H for speckle suppression. Nyquist @ 1.1 μm					**REQ-OCD-0150:** 14±0.5 mas pixel sampling.
Field of view	Required: FOV 2.8 × 2.8 arc sec^{2} instantaneous viewable.			Required: FOV 2."8 × 2."8 set by outer extent of debris disks. Full FOV available in polarimetry mode	Unobscured high contrast, high angular resolution imaging	**REQ-OCD-0160:** Field of view in direct imaging is ≥ 2.8 × 2.8 arc sec^{2} **REQ-OCD-0161:** Polarimetry field of view is square ≥ 2.8 × 2.8 arc sec^{2}
PSF stability				PSF stability or knowledge for PSF subtraction at 4% precision over 3 hours		**REQ-OCD-170:** PSF stability at 4% precision over ≥ 3 hours

8 Instrument Overview

8.1. Basic parameters

The GPI system consists of five key subsystems:

- The AO system is responsible for fast measurement of the instantaneous wavefront, and for providing wavefront control via deformable mirror(s), as well as for opto-mechanical interface with the telescope.
- The coronagraph provides control of diffraction.
- The calibration subsystem provides precise and accurate measurements of the time-averaged wavefront at the science wavelength, so that the final image is not dominated by persistent speckles caused by quasi-static wavefront errors.
- The science instrument produces the final scientific image or data cube, including simultaneous multiple wavelength channels to suppress residual speckle noise.
- The upper-level software coordinates communication between subsystems and the observatory software

Some aspects of the system functionality are not localized within one sub-system, e.g., components of the polarimeter are distributed between the science instrument and the coronagraph.

Figure 8-1 shows a block diagram of the system.

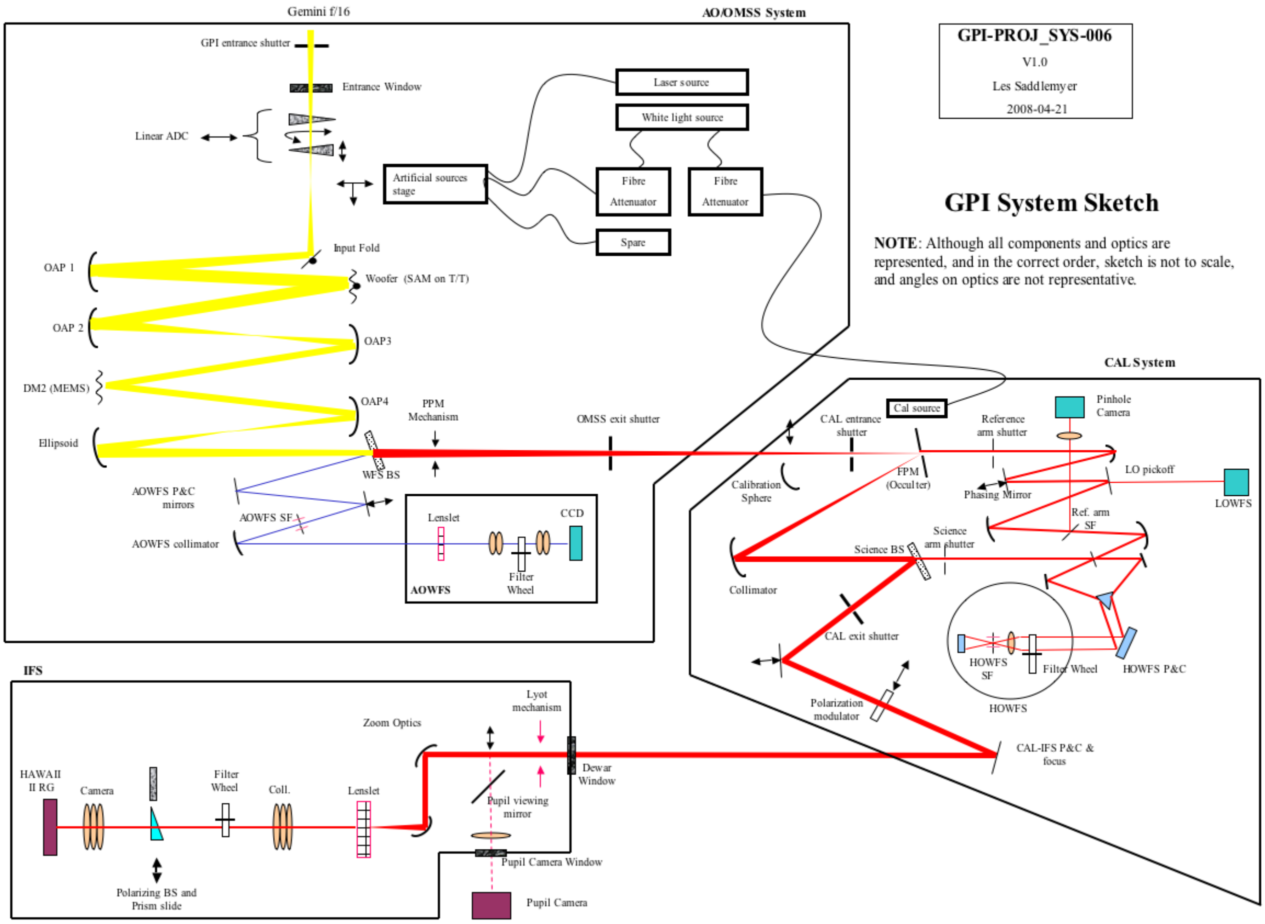

Figure 8-1. Schematic of the GPI system. GPI consists of four major subsystems: AO (fast wavefront sensor and wavefront control); coronagraph (focal plane mask and pupil apodization); integral field spectrograph (science instrument); and the high accuracy wavefront calibration system.

8.2. Subsystem Overview (Functional description of subsystems)

8.2.1. Adaptive Optics system

The heart of GPI is its adaptive optics system, which is responsible for producing the best wavefront quality yet achieved by a ground-based telescope. The AO system employs a MEMS deformable mirror to provide fine wavefront control. The MEMS DM has approximately 43×43 illuminated actuators on the Gemini South telescope, corresponding to sub-aperture size $d \approx 18$ cm, and operates at 1.5 kHz. The MEMS has only ≈ 3 µm stroke, so a conventional pizeoelectric deformable mirror provides low-order "woofer" correction. The piezo DM will also provide fast tip/tilt correction.

The primary wavefront sensor for GPI is a spatially filtered Shack-Hartmann wavefront sensor (SFWFS, Poyneer and Macintosh 2004.) The SFWFS measures the instantaneous wavefront slopes over a bandpass of 0.7–0.9 µm. An adjustable spatial filter in the input focal plane of the WFS prevents aliasing errors. The typical size of the (square) spatial filter is λ_{wfs}/d = 1” on a side, but it can be opened wider to allow initial AO lock or in poor seeing.

The wavefront is reconstructed using a Fourier-domain algorithm (Poyneer et al 2002.) Individual Fourier modes will have their own independent control loop gains in an optimal modal controller similar to ALTAIR but using Fourier modes (spatial frequencies) rather than Zernike modes (Poyneer & Veran 2004 submitted.) This will automatically optimize the control loop parameters for each mode in response to system telemetry. A possible upgrade could use complex Fourier mode information to predict the motion of independent atmospheric layers a millisecond into the future; such “predictive control” could enhance dim-star performance by ~1 magnitude.

The AO bench will also be responsible for interface to the telescope and field steering. Steering mirrors will keep the AO and telescope pupils aligned during operation, using WFS intensity information. Field steering will be accomplished by using a pointing and centering pair in the WFS leg.

- Wavefront controlled by a conventional DM and a 64 × 64 MEMS device
- Spatially-filtered Shack Hartmann wavefront sensor with selectable bandwidth, gain and sub-aperture size
- Fast (1.5 kHz) wavefront reconstructor

8.2.2. Coronagraph

The Coronagraph sub-unit of the GPI has the primary purpose of optically suppressing the diffraction component of the central star’s PSF within the required search area. The coronagraph is receives a high-fidelity wavefront produced by the AO system in tandem with the calibration subsystem.

Removal of starlight is achieved with a set of optical masks. The primary mask set uses the Apodized Pupil Lyot Coronagraph, combining a mild input pupil apodization (optimized for a particular waveband) with a conventional Lyot coronagraph. This provides good performance to quite small IWD (~2.5 λ/D). The exact optimization of the mask is a function of the desired IWD, contrast, and wavelength range; two sets will likely be provided for each wavelength.

From the AO system, the optical beam forms a pupil image at the first coronagraphic mask where greyscale apodizer masks are located. The first focus past this initial pupil plane is used to place a focal plane mask (FPM), in baseline a hard-edged mask that completely transmits or completely reflects each part of the focal plane. This mask operates in reflection – the off-axis beam is reflected to the science instrument while the inner part of the PSF is dumped into the calibration system, where it is used for pointing control and as a reference for the remaining calibration system leg. The reflected light is diffracted and reorganized in the subsequent pupil plane. At the following pupil plane a second set of pupil stops are provided, each being matched to the particular diffraction pattern due to the combination of first pupil plane stop and focal plane mask. This pupil plane is cryogenic and located inside the science instrument, serving also as the instrument cold stop. This architecture provides flexibility for future upgrades with other coronagraph technologies (e.g. band-limited coronagraphs, achromatic shaped pupils, or even some classes of pupil reformatters) should they prove practical for ground-based AO.

- Two selectable pupil plane optics, including apodizers in the input pupil and one or more hard Lyot masks in the second
- Focal plane masks, including transmission of spatially filtered beam for the calibration subsystem.
- All masks will be on wheel mechanisms. The mechanisms permit selection of the mask. There is no provision, however, for independent rotation of the masks with respect to GPI. All current apodizer masks are rotationally symmetric, but the Lyot masks – which block the telescope spiders –are not. As a result, GPI is constrained to a limited set of orientations with respect to the Gemini telescope pupil (nominally 0 and 45 degrees.)

Although the APLC is an excellent high-performance and high-throughput coronagraph, it is inherently chromatic. The APLC coronagraph operates by carefully matching the pupil-plane response of its FPM to the apodizer mask. Since the FPM's effective size in units of λ/D is obviously a function of wavelength, this tuning operates less well as λ changes. Each APLC apodizer/FPM combination must therefore be optimized a specific broadband bandpass.

The current APLC mask set has excellent achromatic performance but somewhat limited contrast in the region of 3-5 λ/D. It is a design goal to come up with additional narrowband apodizers that enhance performance at small working angles.

8.2.3. Science camera

The proposed MWI science instrument is implemented as an integral field spectrograph with low spectral resolution ($R \approx 45$) but fine spatial sampling and excellent wavefront properties. The primary driver for the spectroscopy is to suppress speckles within the PSF. The instrument samples a two dimensional field covering at least the two arcsecond primary null region, out to a slightly larger field (2.8 arcsec). It is an instrument design goal to be able to steer this field by an additional +-1 arcsec relative to the coronagraph optical axis for imaging of specific targets.

A single fixed lenslet array is used to sample the focal plane and feed the spectrographic optics. A fixed lenslet along with a fixed dispersion element (likely a prism) provides a very stable data product that will allow us to create a robust data reduction pipeline. Figure 8-2 shows the layout of the spectra on the detector. A complete broad-band spectrum (*Y, J, H, K1 or K2*) will be included in every exposure. The GPI data pipeline will reconstruct this into a spectral cube, yielding essentially 18 narrow band images simultaneously over the field of view. This will require five filters in a selectable mechanism. Only one plate scale will be provided, but it will be fine enough (0.014 arcsec) to sample the AO corrected PSF and perform speckle suppression.

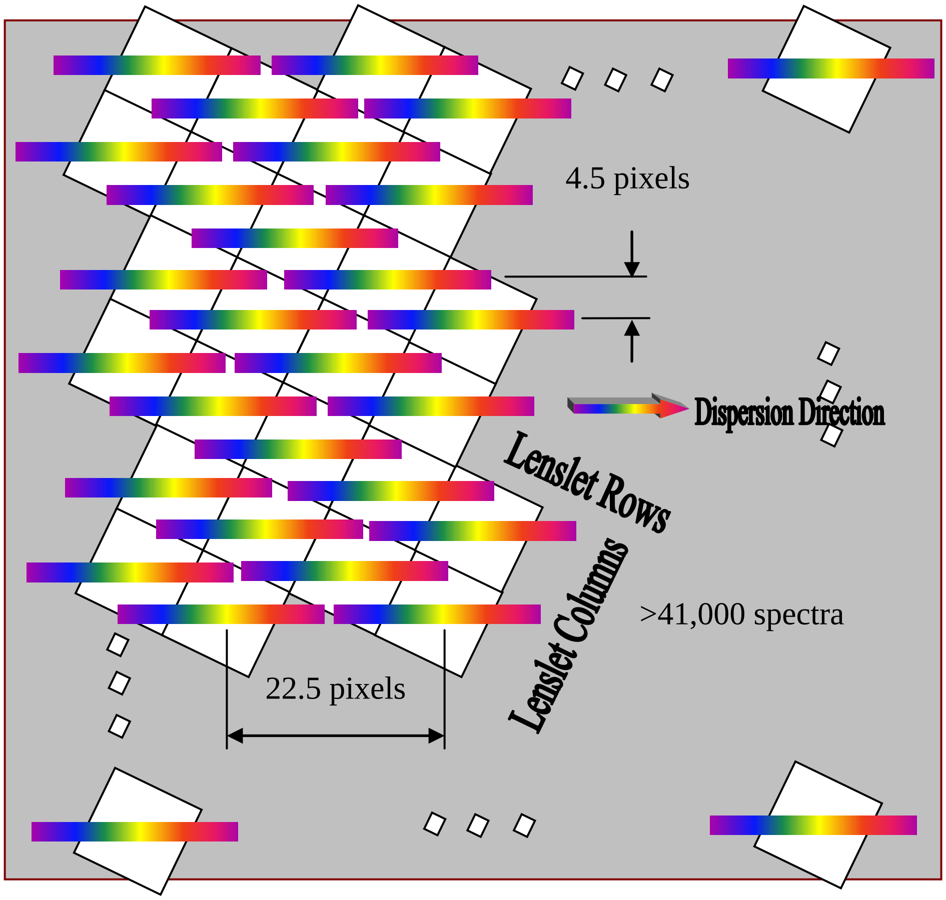

Figure 8-2 Schematic of the layout of spectra on the detector. The lenslet array is rotated 18.33 degrees relative to the dispersion axis so the spectra are interleaved between each other. Each spectrum is separated from neighboring lenslets by 22.5 pixels along the dispersion direction and 4.5 pixels perpendicular to the dispersion. The filter bandpasses and the dispersion of the prism are set so that spectra don't overlap.

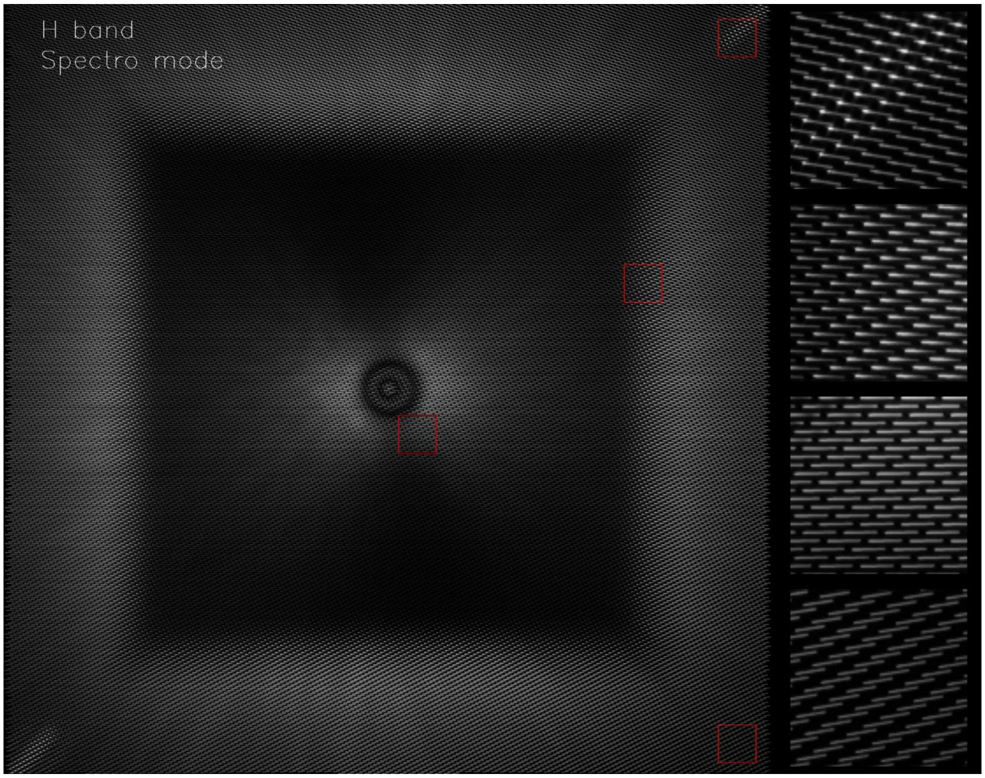

Figure 8-3 Simulated image for H band, spectral mode. Red squares indicate the locations of zoomed-in regions at right.

The spectral dispersive element is a BaF2/S-FTM16 prism. The dispersion of this prism varies slightly across the science bands (and, to a lesser extent, the field of view), resulting in lower spectral resolution at Y band and higher at K band. Regardless of dispersion, each individual spectrum is limited to 16-18 pixels in length to avoid crosstalk (each spectrum center is 21.5 pixels along its length from its nearest neighbor). As a result, the Y and J bands are slightly extended and the K band split into two subfilters, K1 and K2. The bandpasses, expected spectral resolutions and number of spectral channels for each filter are given in Table 8-1.

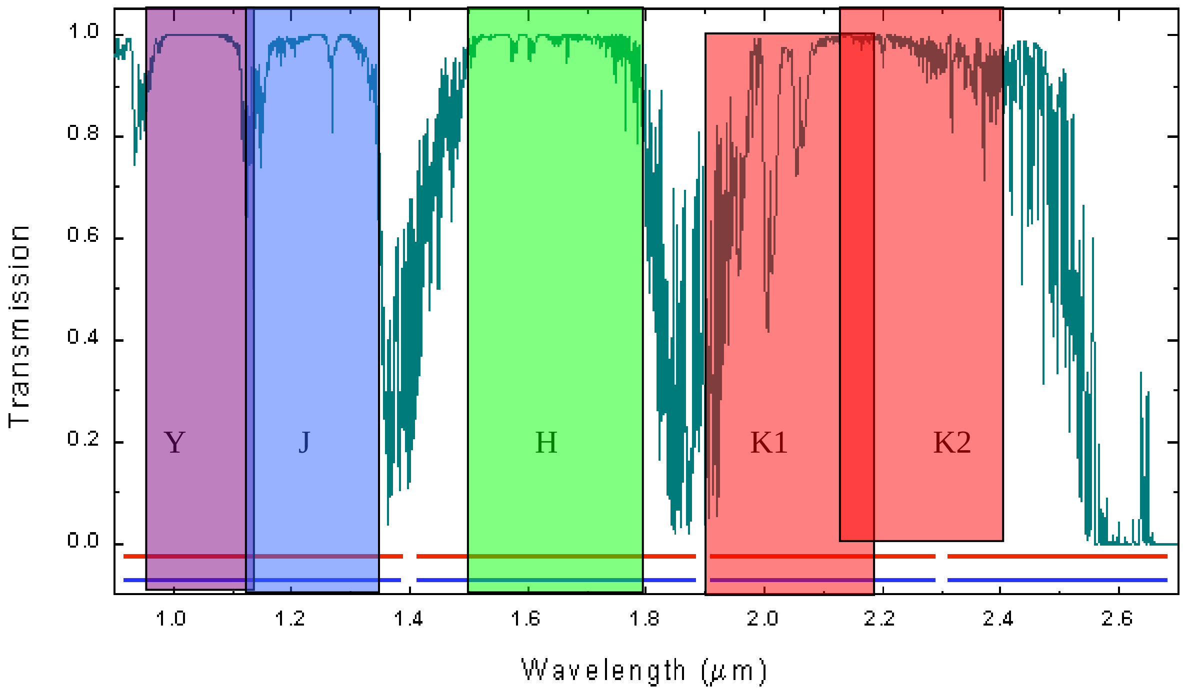

Figure 8-4 -Atmosphere transmission with approximately 1.6 mm of precipitable water vapour overlayed with the GPI IFS filter complement.

Filter Name	½ power Wavelengths	%bandpass	Spectral Resolution	# spectral pixels
Y	0.95-1.14μm	18%	34-36	12-13
J	1.12-1.35	19%	35-39	13-15
H	1.50-1.80	18%	44-49	16-18
K1	1.9-2.19	14%	62-70	18-20
K2	2.13-2.4	12%	75-83	18-20

Table 8-1 -Filter half power points, required spectral resolutions and number of pixels along the spectral axis.

2.1.1 Spectrograph Optics

Multiple cryogenic pupil stops will be provided in a selectable mechanism. Polarization measurements will be made using an achromatic half-wave plate in the coronagraph section accompanied by a cryogenic Wollaston prism near the internal pupil. When used, the polarization components will produce two images on the lenslet array allowing simultaneous measurements of two polarization states through the same optics.

The science detector will be a Hawaii-2RG device from Rockwell Scientific. This detector has a detector layer of HgCdTe grown on CdZnTe and provides 2048x2048 pixels with a pitch of 18 μm and the ability to read out small windows for guiding during long exposures. It is sensitive from 0.85 to 2.5 microns and the instrument will provide a minimum operating wavelength range from 0.96 to 2.4 µm.

- Wavelength range: 0.96–2.4 µm
- Spatial sampling: 14 × 14 mas (Nyquist at 1.1 µm)
- Instantaneous field-of-view: 2.8 x 2.8 arc $\sec^2$
- Selectable focal plane occulting spots
- Broad-band blocking filters: *Y, J, H, K*1, *& K*2
- $R \approx 45$ dispersing element
- Selectable cold stops and pupil masks (to match coronagraph pupil)
- Polarization modulator—achromatic wave plate (external)
- Polarization analyzer—achromatic Wollaston polarizing beam splitter
- Pupil viewing mode (fixed 1.5-1.7 micron InGaAs camera, >200 pixels across pupil)
- Detector: 2048 × 2048 Rockwell HAWAII-2RG HgCdTe/CdZnTe MBE, 18 µm pixels with on-chip guide box
- Low noise, high speed data acquisition system

8.2.4. Dual Channel Polarimeter

A two-channel polarimeter consists of a polarizing prism or analyzer that splits the incoming beam into two orthogonal polarization states. When polarization of light can be is described in the intensity domain (ignoring interference) it is useful to adopt the Stokes parameters representation. The Stokes parameters are contained in a 4 × 1 column vector, (*I*, *Q*, *U*, *V*), called the “Stokes vector”. Each component has units of specific intensity, which is the amount of radiant energy passing normally through an elemental area in a direction confined to a differential solid angle element per unit time per unit frequency or wavelength interval. The Stokes representation is advantageous as all elements possess the same units and are real and measurable quantities.

Conventional detectors that rely on the photoelectric effect to create electron-hole pairs, respond equally to *I*, *Q*, *U* and *V*, so measurement of the four Stokes parameters requires a polarizing beam splitter and a means of rotating the reference plane. Polarizing beam splitters, fabricated from birefingent crystals formed into prisms, introduce an angular or linear displacement between orthogonal polarization states. This effect is used to deliver two simultaneous images of the field of view to the detector in orthogonal polarization states. Half of the original field of view must be masked to prevent overlap of the two images. These two images have associated wavefront errors and point spread functions designated PSF+ and PSF-. Ideally, the polarizing beam splitter, is located as close to the detector as practical, to minimize non-common path aberrations and hence differences between PSF+ and PSF-. Rotation of the reference plane is achieved by mechanical rotation of a half-wave retarder. A possible upgrade would use a $LiNbO_3$ or BBO, electro-optic Pockels cells that operates at infrared wavelengths—such devices may be available in the near future with clear apertures > 10 mm.

A single measurement with a dual channel polarimeter, which records both outputs of the analyzer, can measure, for example, +*Q* and -*Q*. The difference between the two channels, 2*Q*, is insensitive to aberrations common to both channels, e.g., uncorrected atmospheric or telescope wavefront errors. Other Stokes parameters are measured by rotating the reference plane. Stokes *U* is measured by using the retarder to rotate the reference plane by 45 degrees.

Rotation of the reference plane makes the measurement robust against non-common path wavefront errors and relaxes the tolerance on the similarity of PSF+ and PSF-. A second measurement of *Q* is obtained by rotating the reference plane by 90 degrees so that the two orthogonal beams now encode -*Q* and +*Q*, i.e., +*Q* and -*Q* have been swapped between PSF+ and PSF-. Thus, by rotating the reference plane, we can measure each polarization state first through one beam path, then through the other. Both polarization states are measured through both beam paths, and any static speckle pattern will vanish in the subtraction. Figure 8-5 depicts an illustration of speckle subtraction in differential polarimetry.

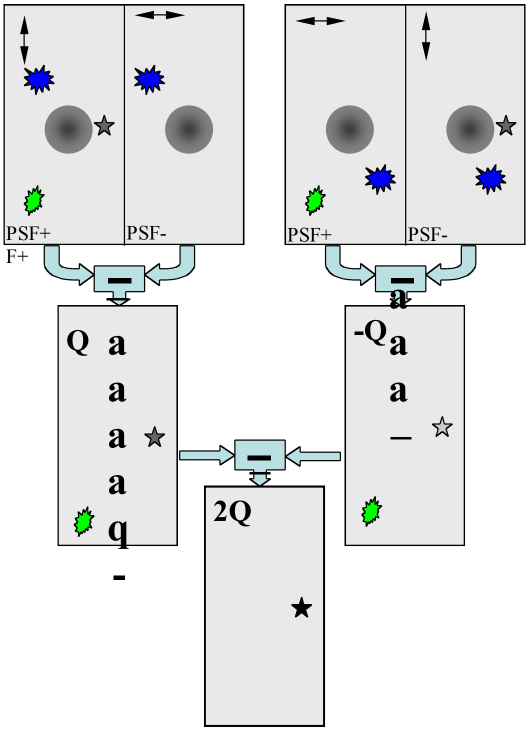

Figure 8-5: An illustration of speckle subtraction in dual-channel differential polarimetry. Each raw image (top row) consists of two half-field sub-images (PSF+ and PSF-), in perpendicular polarizations as indicated by the two-headed arrows; the waveplate has been rotated by 45 degrees between the left hand and right hand pair. In this conceptualization, there is an unpolarized central source (grey) surrounded by an atmospheric speckle (blue spot at 11 o'clock on the left and 5 o'clock on the right) and a non-common path speckle (green spot at 7 o'clock) and a polarized astrophysical source (grey five-pointed star at 3 o'clock). Single subtractions (second row) reveals the polarized astrophysical source, remove the unpolarized halo, cancel common-path speckles but leave non-common-path speckles (green) intact. Non-common path speckles have the same sign in both single-difference images, while the astrophysical polarized signal has opposite signs in the Q and –Q images, due to the waveplate rotation. Therefore, by taking the double difference (third row) we can cancel out the non-common-path speckles and leave only astrophysically polarized light.

The dual channel polarimeter is a distributed system, and includes the following components:

- Achromatic, retarder for *J*, *H*, & *K'* on a precision rotation stage upstream of the science camera;
- A Wollaston polarizing beam splitter for *J*, *H*, & *K'* providing a ~10 pixel beam deviation between e- and o-rays, located in collimated space within the science camera, resulting in two extremely low-resolution spectra per lenslet corresponding to the e- and o-rays.

A detailed Mueller matrix analysis of a dual channel polarimeter is given in OCDD Appendix-A Mueller Matrix Analysis of Non-Idea Dual Channel Polarimeter (§14). This appendix also discussed approaches to calibration and reduction of data in polarimetry mode.

Computation of the system Mueller matrix, which describes the systems properties in polarized light is given in OCDD Appendix-System Mueller Matrix (§ 15).

8.2.5. Calibration system

An image from a ground-based telescope is dominated by atmospheric seeing. Even with an extreme AO system and coronagraph, the residual starlight is many orders of magnitude brighter than a planet approximately 10^{-5} of the star. The residual speckle pattern of a short exposure image after the AO system and coronagraph is due primarily to phase errors in the wavefront not corrected by the AO system. After integrating thousands of short exposures, the image is much smoother. If the optical system is free of aberrations, spatial structure in a ~1hr long exposure image is ~10^{-7} limited by averaging of the atmospheric speckle pattern. But persistent wavefront errors, in either amplitude or phase, will produce persistent speckles that are much brighter than the planet light. The calibration system described here works by measuring the wavefront of the starlight, not suppressed by the coronagraph/AO system at the science wavelength and through as many of the science optics as possible. This information is time-averaged and fed back to the main AO system to adjust the control offsets of the SFWFS so that the AO system is producing the desired wavefront on the science camera. The calibration system may also use that information to generate a synthetic PSF. This PSF is subtracted from the long exposure (science) image, thereby removing image artifacts caused by static wavefront errors in the optics/AO paths, as well as the residual atmospheric speckle. The net result is ultimately limited only by detected photon statistics

This calibration scheme is implemented as a phase-shifting Mach-Zehnder interferometer that is tightly integrated with both the coronagraph and the science camera (Figure 8-6). It uses light that would otherwise be lost at the occulter for a reference, and is combined with a fraction of the light that is intended for the science camera. Because the “reference arm” of the interferometer uses only light from the star, the resulting wavefront and speckle removal image is unbiased by planet light, which is incoherent with respect to the reference starlight. Because it uses light that is in the same spectral band as the science camera, chromatic errors are mitigated. Likewise, integrating it around the coronagraph/science camera reduces effects due to non-common path optics. Finally, the ability of the calibration unit to measure both residual phase and amplitude errors post-coronagraph provides a powerful tool for the description (and potential correction) of errors internal to the whole instrument that have no current or planned means of compensation. The calibration system operates in concert with a star simulator, during periods off-the-sky, and is the difference between the wave front as seen by the science camera and that measured by the calibration unit. During a science observation, the wave front measured by the calibration unit is used to: 1) apply slow updates to the AO control offsets and; 2) estimate the wave front (both amplitude and phase) as seen by the science camera. This estimate is formed from instantaneous measurement of the amplitude and phase with an interferometer and the knowledge of the wave front difference between the calibration wave front sensor and the science camera (previously measured in the off-the-sky calibration). Because this wave front is measured with a reference beam from the parent star alone, it produces a pure measurement of the wave front error attributable only to the starlight.

In post processing, the phase and amplitude information gathered by the calibration unit as well as the wave front error between the calibration and science instruments are used to form an estimate of the residual speckles in the final science image that are due only to the star light.

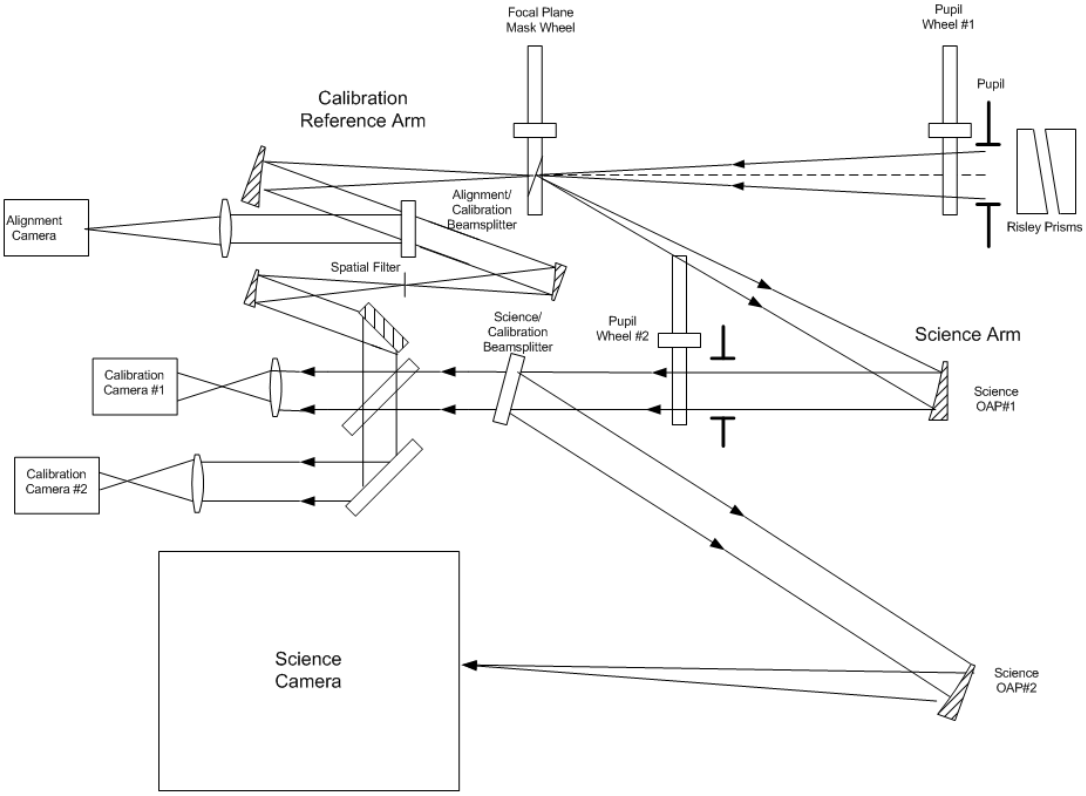

Figure 8-6. A simplified figure of the calibration system, a phase-shifting Mach-Zehnder Interferometer. Reference and Science beams are split at the focal plane; the reference beams passes through another stage of spatial filtering before it is combined with a sampled version of the science light. This system will measure the amplitude and phase of the science beam post-coronagraph, and will provided off-sets to the active wave front sensor to remove the effects of long-term speckles during an observation. A goal of the calibration system is to use the same amplitude and phase information to post-processing to improve the ultimate contrast of the system.

In summary, the wavefront calibration system provides precise and accurate wavefront sensing to measure and correct non-common path errors in the wavefront delivered to the science camera.

The baseline cameras in the CAL system are commercial InGaAs systems. This will limit their sensitivity to <1.7 microns – replicas of the science H band filters will further window this to 1.5-1.7 microns. Chromatic offsets in the aberrations (primarily pointing and focus) will be added in software by the CAL system when observing at other wavelengths.

- Pupil-plane diversity for internal calibration between high contrast AO WFS and Science Camera.
- Samples science light with a broadband 20% beam splitter.

8.2.6. Calibration sources

8.3. Operating Modes

8.3.1. Instrument configuration

The primary decision that the astronomer has to make when using GPI is the hardware configuration of coronagraph focal plane stop and pupil mask and the observing wavelength and desired contrast/inner working distance. The following table lists some possible configurations; see the appropriate sections for lists of the elements in each individual wheel.

In operation, the user would select a configuration indicating whether they want direct (for adjunct science), coronagraphic, or unblocked coronagraphic (for determining primary star position and Strehl ratio) mode, and which wavelength they desire; the instrument will automatically configure the various wheels. (Lower-level commands will be available for user-defined modes.). Examples include:

Table 8-2: Standard configurations

Config name	Filter	Apodizer PPM	FPM	Lyot (estimate)	Notes
Y_coron	Y	APOD_Y_56	FPM_Y_56	Lyot_3	
J_coron	J	APOD_J_56	FPM_J_56	Lyot_3	
H_coron	H	APOD_H_56	FPM_H_56	Lyot_4	
K1_coron	K1	APOD_K1_56	FPM_K1_56	Lyot 4	
K2_coron	K2	APOD_K2_51	FPM_K1_56	Lyot 4	
H_SIWA	H	APOD_H_S	FPM_H_4	Lyot_7	Could use Y or J FPM
H_LIWA	H	APOD_H_L	FPM_K1_56	Lyot 3	
H_star	H	H_star	FPM_H_56	Lyot_4	
Y direct	Y	none	mirror	open	
J direct	J	none	mirror	open	
H direct	H	none	mirror	open	
K1 direct	K1	none	mirror	open	
K2 direct	K2	none	mirror	open	
Y unbl	Y	APOD_Y_56	FPM_Y_56 + offset	Lyot_3	Could use mirror
J unbl	J	APOD_J_56	FPM_J_56+	Lyot_3	Could use

Config name	Filter	Apodizer PPM	FPM	Lyot (estimate)	Notes
			offset		mirror
H unbl	H	APOD_H_56	FPM_H_56+ offset	Lyot_4	Could use mirror
K1 unbl	K1	APOD_K1_56	FPM_K1_56+ offset	Lyot 4	Could use mirror
K2 unbl	K2	APOD_K2_51	FPM_K1_56+ offset	Lyot 4	Could use mirror
HS unbl	H	APOD_H_S	FPM_H_4+ offset	Lyot_7	Could use mirror
HL unbl	H	APOD_H_L	FPM_K1_56+ offset	Lyot 3	Could use mirror
Hstar unbl	H	H_star	FPM_H_56+ offset	Lyot_4	Could use mirror

The unblocked “check the star” mode will likely be supported by moving the FPM slightly so the star is unblocked – probably the equivalent of 5 or so lambda/D.

8.3.2. AO configuration

The operation and coordination of the wavefront sensors, wavefront controllers, and calibration system is autonomous. Optimum operation is established by prior calibration, and analysis of the WFS errors in real time. When the astronomer prepares an observation, the information includes B and V magnitudes, spectral type (preferred if available) or V-R color. This information, combined with estimates of r_0 and τ_0 from prior, internal, WFS telemetry, is used to pre-configure the AO system, including loop gains, bandwidths, and to select the sub-aperture size for the fast WFS. A menu of performance metrics will be established for the astronomer to select from at the time of planning the observation, including best Strehl performance, maximum contrast (between IWD and OWD), and best PSF stability. The astronomer must also choose between using the hardware ADC or software dispersion compensation, which accounts for the difference between the visible light sensor leg and the IR science light.

8.3.3. Coronagraph configuration

Coronagraph configuration involves selection of 3 different optical masks: Pupil Plane 1, Focal Plane and Pupil Plane 2. In the normal mode, the choice of masks are set by desired inner working distance (IWD) and wavelength (the masks are highly chromatic), with typically two masks – one with smaller IWD but lower contrast, one with larger IWD and larger contrast – available per waveband. Given the choice of desired IWD, the system will automatically select the correct mask for the current waveband.

In the baseline coronagraph, the second pupil (inside the science instrument) does not need to be undersized relative to the telescope pupil and hence does not need to be reconfigured, though smaller masks may be made available for specialized observations or to e.g. block out regions of extra WFE near the edge of the pupil. Rather than supporting arbitrary orientations, the Lyot masks are pre-designed for fixed orientations of GPI relative to the Gemini pupil.

The choice of masks is dictated by the science goals of the observer. A full set of possible combinations will be provided to achieve all of the required science goals. **Error! Reference source not found.** gives a representative list.

Table 8-3: Apodizer positions

Position	PPM name	Description
1	CLEAR	Oversized open
2	CLEAR GPI	Gemini pupil with secondary (no spiders)
3	APOD_Y_56	Grey 5.6 λ/D for Y
4	APOD_J_56	Grey 5.6 λ/D for J
5	APOD_H_56	Grey 5.6 λ/D for H
6	APOD_K1_56	Grey 5.6 λ/D for K1
7	APOD_K2_51	Grey 5.08 λ/D for K2
8	APOD_H_S	Small IWA narrowband mode - Optimized for 3-5 l/D at H
9	APOD_H_L	Matched to K FPM for characterization
10	H_Star	200/50 petal H-band Princeton/Cady APLC starshape

Table 8-4: Focal plane mask positions

FPMs		Diameter mas	Diameter microns	Notes
1	Mirror			Could be omitted
2	Pinhole	5.0	12.3	Alignment fixture. JPL responsibility
3	FPM_Y_56	156.2	383.3	5.6 λ/D for Y
4	FPM_J_56	184.7	453.3	5.6 λ/D for J
5	FPM_H_56	246.7	605.5	5.6 λ/D for H
6	FPM_K1_56	306.3	751.9	5.6 λ/D for K1, 5.08 for K2
7	FPM_H_4	176.2	432.5	could be replaced with 3 or 4
others				

Note that all sizes in λ/D are defined at the band centers as per the wavelength table. Diameters here are pending final verification by Soummer; see the COR SSDD for final versions.

Table 8-5: Lyot mask combinations. Standard inner/outer diameters are to be determined by Remi Soummer based on final flexure numbers and Fresnel simulations in 2009. Spider widths are the full width of the spider vanes as a fraction of the pupil diameter.

Lyots		OD/ID	spider width	Orientation	Notes
1	Open	10			Full GN pupil.
2	Blank				
3	Lyot_2_0	standard	2%	0	
4	Lyot_3_0	standard	3%	0	
5	Lyot_3_45	standard	3%	45	
6	Lyot_4_0	standard	4%	0	
7	Lyot_4_45	standard	4%	45	
8	Lyot_7_0	standard	7%	0	
9	Lyot_10_0	undersized	10%	0	
10					

The Gemini Planet Imager's Lyot masks do not support arbitrary rotation of the instrument with respect to the telescope secondary spiders. Instead, some subset of the masks are rotated by 45 degrees with respect to the instrument orientation.

8.3.4. Science camera configuration

The IFS science instrument has few moving parts, to simplify operation and increase stability. The basic parameters are operating wavelength (set by filters) and whether the polarimetry mode is in use.

The science camera incorporates the following mechanisms:

Cold pupil mask. (See above.)

Filter wheels. Blocking filters are used to isolate diffraction orders and to control the background reaching the detector array. These will closely approximate the standard *Y*, *J*, *H* filters, plus two custom filters (*K*1 and *K*2) to span the two-micron window

Name	Bandpass (1/2 power points) [μm]	Central λ [μm]	Bandpass [%]	Spectral res.	# spectral elements
Y	0.95 – 1.14	1.04	18	~ 60	19
J	1.12 – 1.35	1.23	19	~ 50	15
H	1.50 – 1.80	1.64	18	~ 45	16
K1	1.90 – 2.19	2.04	14	~ 45	17
K2	2.13 – 2.40	2.26	12	~ 45	16

Polarization analyzer. A polarizing beam splitter, a Wollaston prism, introduces an angular deflection between the ordinary and extraordinary rays producing two images with orthogonal polarization states on the focal plane array. Simultaneously, a “un-prism” deploys that partially cancels the spectral dispersion of the main prism. Together with the external modulator (the baseline is rotating half wave plate) the analyzer permits measurement of the Stokes parameters

The spectrograph’s pupil-viewing camera always operates at a fixed wavelength (*H* band).

8.3.5. Data Acquisition System

The data acquisition system supports a selection of specific modes of array readout at the device level. The purpose of this subsystem is to enable observations, which are limited in execution, cadence, or level of performance only by shot noise of the signal or background or the intrinsic read noise of the detector.

These functions include:

Pixel reset Modes

- Pixel reset— each pixel is independently addressed and individually reset.
- Ripple reset—each row of pixels is independently addressed and reset simultaneously
- Global reset—all pixels are reset at the same time

Detector Sampling modes

- Fast Readout – single frame (reset/read) for high background
- Single CDS Readout-difference of two full frames (reset/non-destructive read/integrate/read)
- IR Fowler Sampling
- IR Sample-up-the-ramp - fixed time intervals through readout

Region of Interest Selection

- Select a single subarray

Frame Accumulation Modes

- Co-addition - accumulation of a sequence of frames with one output image
  - Used at high signal/background to improve efficiency by restricting the number of frames passed through the data pipeline to storage
- Multiple co-addition - accumulation of a sequence of frames sorted in multiple output images, such as phase diversity channels or chop/nod positions

System Calibration Modes

- The system provides capabilities to support the calibration modes:
  - Optical phase diversity sequences (e.g, for image sharpening)
  - Dark frames
  - Bias frames

- Flat Fields
- Bad pixel masks

8.3.6. Science detector configuration

The on-chip integration time is set by two considerations. First, the integration time must be long enough that speckle noise or photon halo shot noise must dominate over detector read noise and dark current. Second, the on-chip integration time must be sufficiently short that trailing of images at the OWD does not significantly degrade detectability of faint companions. For a tolerable Strehl degradation is 0.8, then the maximum image motion during an exposure is 1.05 λ/D.

In an integration time, t, saturation occurs in the brightest pixel for a $H = 7.2 + 2.5 \log(t/\mathrm{s})$ star on the HAWAII-2RG array given a full well depth 150,000 e- at VRESET of 0.75V This assumes a 20% system efficiency, IFS spectral resolution of $R \approx 40$, and Nyquist sampling at 1 µm. Care must be taken to prevent saturation, since this can cause ghost images in subsequent readouts.

For bright stars, the H2RG will support subarray readout modes allowing very short exposures. Although H2RGs support on-chip guiding, this mode will not be implemented in the firmware build provided by Teledyne to GPI.

.

9 Setup and configuration requirements

This section will be detailed further in PDR as the instrument architecture is frozen and with cooperation from Gemini instrument scientists.

9.1. Daytime calibrations

9.1.1. AO calibration

The daytime calibration procedure is required to set the control point of the calibration unit during night time observations. The first step in this process is to minimize the residual wave front errors as seen by the science camera. In this mode, the calibration system serves as an intermediary by first, interpreting pupil images of the science camera under different diversity condition, second, reconstructing the science camera wave front and then provide offsets to the active wave front sensor, and finally, once the previous step has converged to a satisfactory result, measures the wave front that will be used as the target wave front during a science observation. The details follow.

The stimulus is inserted into the optical path of the instrument, and the science camera is requested to pupil mode. The AO loops are closed. The alignment camera determines if the PSF is centered on the occulting mask, offsets are given to the AO system until this is achieved.

The calibration system must first calibrate itself. The calibration camera filter wheel is closed and a dark exposure taken. The filter wheel is then set at a band pass filter to match the science camera. The science arm is shuttered and a measurement of the reference arm is made. After closing the reference arm shutter and opening the science arm shutter, another exposure is made, this time of the pupil in the science arm. From these images, the dynamic range of a calibration measurement is calculated, and the exposure time is adjusted to maximize it. An initial measurement using the phase-shifting interferometer is made, and from this measurement a few parameters are calculated, they are: 1) the step size of the phase-shift and 2) the system contrast. The step size is adjusted to the appropriate value and the zero point of the phase-shifting mirror is adjusted to ensure the interferometer is working on the white light fringe (maximum of contrast).

Trail pupil images are taken with the science camera. Images are checked to insure no saturation and that a large dynamic range in these images is achieved. This may involve some iteration on the stimulus brightness and/or the science camera exposure time. Once these images are deemed adequate, they are passed to the calibration system. The calibration system then request some diversity introduced into the system. (Currently, the two leading candidates are a series of translations of the occulting mask in the coronagraph image plane or different values of focus applied to the AO system DM.) From a series of images with different states of system diversity, the calibration system determines the wave front as seen by the science camera. Offsets are given to the AO DM to minimize the residual wave front errors, and this process is iterated until the residual wave front is determined to be satisfactory.

At this point, calibration system makes a measurement. This then becomes the nominal wave front set point for science observations.

9.1.2. Science instrument calibration

The IFS spectra are staggered on the science detector in a relatively complex pattern and spectra are allowed to have a few percent of overlap with each other. This means that they must be "extracted" from the detector and reassembled into the final data cube like returning straws to a strawbox.

To make this a tractable problem, we have designed the IFS such that there are no moving parts between the lenslet array and the detector. This makes the layout of the spectra on the detector stable, and the dominant motion is due to variations in diffraction of the pupil due to changes in the image plane illumination. Thus, the calibration and data reduction process focuses on mapping the stable spectral layout by determining the fraction of a given lenslet's light that falls on a particular pixel. This can be accomplished by flood-illuminating the entire spectrograph field, much like a traditional flat field. Since the spectra partially overlap, only a subset of the lenslets can be illuminated at any particular time in order to measure the wings of the PSF. This is accomplished by passing a slit behind the lenslet array in the focal plane formed by the lenslets. In this way, the PSF of each lenslet can be accurately calibrated. In the GPI IFS, if every third column is allowed through, then no significant overlap will occur. This will be achieved with a custom slit mask consisting of approximately 70 slits each 100 microns wide and 29 mm long, spaced by 300 microns from each other. By stepping the mask by 100 microns, a different set of columns is allowed through and, in general, three step positions are needed to map the full set of PSFs. If the slit separation (300 microns) proves too small for easy manufacture, then fewer slits will be cut and more positions required for calibration.

Since there are no moving components to change the calibration, it is expected that scans for the four filters can be accomplished very infrequently (months apart) and will require only a few minutes to accomplish, even if ten or more step positions are required.

9.2. Setup Prior to Observation

9.2.1. Science Observation Sequence

9.3. Nighttime Calibration

9.3.1. AO calibration

Before observations are ready to begin, the process of daytime calibration can be verified. If the system is stable, this should require only a cursory examination of the science camera wave front and the set point of the calibration system. This procedure is identical to the daytime calibration procedure detailed previously. Final contrast is verified with observations of the internal reference source.

9.3.2. Science instrument calibration

No additional nighttime calibration is expected beyond standard dark and flat-field measurements.

9.4. Quicklook software and data reduction pipeline

See discussion in Chapter 10 and appendices of CDR Volume 2.

10 Observing Scenarios

10.1. Typical Observing Sequences

Here we here describe several typical observing sequences to establish the series of events that occur when executing a science program. These sequences form the basis for the detailed descriptions of observing sequences provided in Vol. 3 of the CDR documentation (Software). The current purpose is to answer questions such as:

- Does the instrument functionality of hardware, software and its interfaces, as conceived, support the complete set of actions necessary to achieve calibrated science data?

This exercise emphasizes operations at the system level, and illustrates how the AO, coronagraph, science camera and the calibration system interact with one another and with the observatory. This will help highlight deficiencies, e.g.,

- Missing hardware & software components
- Missing interface components
- Conflicts for resources

The level of detail should be sufficient to take the first steps towards maximizing the science delivered by making the most efficient use of telescope time.

The goal here is *not* to exhaustively enumerate all possible science cases or observing scenarios for GPI, nor to restrict or prevent other operational possibilities or sequences of actions beyond those listed here.

In the following sequences, optional steps are denoted by [square brackets]. These denote different variations on the same observing sequence, such as could be selected by an observer based on their particular science goals.

10.2. Target Acquisition Sequences

10.2.1. Goal

This sub-scenario describes the processes through which a new target is acquired by the AO system and initially delivered to the science camera, and optionally coronagraph and calibration system. The same acquisition process is a common element to many of the observing scenarios, so it is called out here separately to minimize repetition in the following.

The first acquisition sequence describes the acquisition process up through obtaining AO-corrected, unocculted images. At the end of this sequence, the instrument is ready to obtain unocculted science observations of both stellar or solar system targets.

The second acquisition sequence continues on to perform coronagraphic acquisition. At the end of this sequence, the instrument is ready to obtain occulted science observations of stellar targets.

10.2.2. Sequence: Target Acquisition

The Gemini ISS is operated in stationary pupil mode for all observations, such that the instrument remains fixed orientation with respect to the telescope pupil. In this mode the field of view, the projection of the pupil onto the sky, and the Stokes (polarization) reference frame appear to rotate with respect to the sky, with a rotation rate that depends on the instantaneous pointing of the telescope. Note that due to the nature of the Stokes parameters, the coordinate transformation for the Stokes frames rotates as twice the rate of the image and pupil (for the Mueller matrix describing polarization see OCDD Appendix Polarization).

Observation of the previous science target terminates, including completion of

- Science data readout
- Assembly and formatting of FITS keywords describing circumstances of observation
- Distillation of AO system telemetry to parameters relevant to image quality
- Science calibration data, e.g., PSF reconstruction information from wavefront calibration system

AO and calibration system end current observation and prepare for a telescope slew

- Open AO loops
    - Flatten both DMs
    - Set TT to zero
    - Insert Calibration sphere for Cal system
    - Preset & adjust WFS pointing & centering pair
- Close AO enclosure entrance shutter

- Mitigate potential for detector hysteresis[1] during acquisition of targets which may saturate the detector, e.g.,
    - Array to idle/continuous reset mode
    - Shutter closed/ND filter inserted.

Whenever possible, aspects of the previous observation should complete in parallel with preparation for the next target. For example, buffering of AO telemetry to disk will occur during the telescope slew.

Telescope

- Coordinates of next target are loaded. This includes loading non-sidereal rates for solar system targets
- Slew to next target begins
- Rotation of ISS to chosen position angle for next target begins[2]

The following preparation steps should occur in parallel with the slew:

- Prepare science instrument for target acquisition
    - Select IFS Filter for desired observing wavelength
        - When slewing filter wheel and other mechanisms, do not let unfiltered starlight fall on the detector, to avoid saturating it[3]
    - Set exposure time and other readout parameters for acquisition process, based on expected target magnitude at chosen wavelength
    - [Optional: Insert/remove polarimetry modulator (wave plate) and analyzer (birefringent prism) assemblies if the observing mode for next target differs from previous one. In polarimetry mode both the modulator and analyzer are need. In non-polarimetry neither are needed. If in polarimetry mode, set wave plate rotator mechanism angle to zero.]
- Prepare coronagraph for target acquisition
    - Apodizer slews into position appropriate for observing wavelength
    - Select Lyot stop to match ISS orientation and observing wavelength
    - Occultor moves to unocculted position
- Prepare Cal system for target acquisition
    - Select Cal system filter
    - Set exposure time and other readout parameters

Prepare AO system for next target based on target properties and observing conditions. Knowledge of target properties and quality of atmospheric seeing are used to speed the convergence of the AO control system.

- AO system preconfigures
    - Wavefront control system
        - Estimate control loop parameters given current atmospheric conditions (background, seeing, correlation time, wind profile), target brightness,

---

[1] This step is intended to mitigate potential deficiencies of the science detector, such as quantum efficiency hysteresis or elevated dark current associated with charge trapped in lattice defects.

[2] It is expected that blocks of targets and astronomical calibration sources will be observed at the same ISS orientation with respect to the telescope pupil.

[3] For example, do not permit accidental alignment of open slots in the IFS filter wheels during slewing to a new filter combination.

        color/spectral type, and target elevation. Atmospheric seeing information is estimated from previous instrument telemetry; values do not need to be extremely accurate since this is merely a starting point for the control optimization.
    - AO hardware configuration
        - Fast WFS filter wheel selects mode to prevent saturation
            - If required, neutral density filter based on star magnitude and color
        - Spatial filter aperture is set to open
        - ADC is slewed to configuration for target acquisition

Telescope

- Telescope slew ends
- Target is acquired and delivered to the GPI science aperture

Cal system calibration [Note: Targets that require unocculted observations, e.g., solar system objects, will not require the LOWFS calibration step because it receives no light from the target. Therefore, this step is skipped for unocculted targets. In the situation where the occultor is not used and no light is delivered to the Cal system it is necessary to preserve the Cal system information from the previous target and apply it to the current target.]

- Cal system acquire calibration sphere wavefront data
- Retract calibration sphere assembly
- Configure for target brightness and science observing wavelength using new calibration sphere wavefront data. [In the case where the new calibration sphere data is not acquired preserve information from previous target.]

AO

- Apply pointing model
    - Input fold steering mirror & pointing and centering pair slew to position for target acquisition
    - Ingest differential refraction information from TCS
- AO prepares for closed loop operation, e.g.,
    - ADC begins tracking
- AO captures the target
    - Check for light on WFS

The telescope delivers the target, assumed to be good enough so the star is present in the AO WFS field of view, or approximately 3” × 3” per sub-aperture. AO WFS measures counts per sub-aperture and verifies that it is seeing light from the target. (For stellar targets, catalog positions will almost always be accurate enough for this to work. Solar system targets will have more positional uncertainty, so it should be possible to disable the automatic verification process for such targets.) If the target is not located within the expected AO WFS counts, a search algorithm is initiated to find the target, or an alarm rings to invoke supervision of the support astronomer. If the AO WFS counts are near the expected value, acquisition proceeds, but a warning is issued to the support astronomer.

- Begin wavefront measurement and control
    - Close AO loops

    - Begin off-loading tip-tilt and low-order Zernike modes to telescope
    - Measure and control beam locations and initiate tracking input fold of telescope tracking pupil
        - Check coronagraph/IFS tracking reading
- Wait for AO loop optimization to update
    - Wait AO system to settle (~1 s)
    - Compare, e.g., AO WFS, TT, etc. sensor output with expected signal levels
    - If HOWFS signal factor of 2 different than predicted, adjust AO frame rate
        - Abort if below expected threshold for continuation and invoke supervision of the support astronomer
- Close AO WFS spatial filter iris
- Wait for full (~ 10 s) AO loop optimization
    - Compare AO performance with criteria for continuing observation, including information from wavefront Cal system

Science camera

- Establish satisfactory target acquisition based on science camera data. Confirms that the target is centered in the field of view.
    - Configure for unsaturated image, setting exposure time as necessary depending on chosen filter and mode. In polarimetry mode, because of non-dispersed operation, the per-pixel intensity will be higher than in regular IFS mode, requiring shorter integration times for unsaturated imaging.
    - Acquire unsaturated image with SNR ~ 100 in center (use science detector RG subarray mode or ND filters). This image may use either the full array or a smaller RG mode subarray, depending on target brightness and expected precision of coordinates.
    - Data is immediately passed to real-time pipeline for acquisition quality verification
    - Pipeline invokes image quality verification subroutine(s), with various routines available depending on observing mode, e.g.,
        - Confirm centering. The algorithm which does this shall be capable of accepting both spectroscopy mode or polarimetry mode data
        - For stellar targets, automatically estimate Strehl ratio and compare to desired/minimum Strehl criteria
        - [Optional: For stellar targets, examine unsaturated image to determine whether the target is a close binary. If the current observing program wishes to avoid binaries, then choose to abandon target.]
    - Reacquire target if unsatisfactory based on results of verification routine(s)
    - Abandon target if repeated attempts fail, or support astronomer intervention
- Configure science detector RG subarray for astrometric image stabilization

The initial target acquisition process is now complete. The instrument is ready to acquire unocculted science data, or may go on to perform the coronagraphic acquisition to acquire occulted data.

10.2.3. Sequence: Coronagraphic Acquisition

This sequence begins with the AO system locked on the science target and the coronagraph in the unocculted position (i.e. at the end of the Target Acquisition process, or following the end of unocculted science observations).

Coronagraph

- Occultor moves into position

Cal system

- Calibration system now receives light in its reference arm
- Cal system shutters open
    - Close loop on Cal pinhole tracking if necessary
    - Cal LOWFS begins reporting focus and tip tilt to SCC
    - Phasing mirrors adjust to white-light fringe
    - Step down cal system spatial filter
    - Cal HOWFS begins measurements

Cal system begins measuring instantaneous wavefronts and time-averaging them (probably ~<1 minute based on looking at the time statistics of the wavefront).

- Cal system reports when first iteration is complete and triggers science observations
- Cal system begins accumulating PSF reconstruction data

AO reports ready, following Cal system first iteration.

The coronagraphic acquisition is now complete. The instrument is ready to acquire occulted science observations.

10.3. Field star survey

Nominal system configuration: IFS with apodized pupil Lyot coronagraph. This would use the H band filter and the H1 coronagraph configuration, tuned for best performance around the bright 1.59 µm CH_4 feature in planetary spectra.

10.3.1. Science goal

This program consists of a survey of main sequence stars in the solar neighborhood. The scientific goal is to accumulate statistical information regarding the abundance of planets. The survey includes initial visits to targets, but also includes repeat visits to complete the full range of azimuthal coverage, or for confirmation and follow-up observations that will constrain common proper motion, Keplerian orbital elements and atmospheric properties ($\log(g)$, T_{eff}) of targets with candidate planets. A typical target catalog comprises 1000 stars, which will be observed for approximately 1 hour each, requiring 100 nights.

10.3.2. Target selection

With a large pool of potential candidates it is possible to optimize the use of telescope time by careful construction of nightly target lists which minimizing observing overheads (telescope slews, instrument calibration) and make the best use of current observing conditions.

Before observing begins, a target selection algorithm chooses from the master catalog a subset of stars to construct nightly target sequences given the LST at midnight. The master catalog includes target flags and information that quantifies observing priority based on factors which include youth and distance, spectral type of the parent star, the presence of a debris disk, or previous observations with GPI indicating candidate planets. The details of the prioritization algorithm are to be explored by the science team in later phases. The selection process may also consider atmospheric conditions such as wind direction or the presence/absence of clouds in certain parts of the sky.

Information in the master catalog includes, as a minimum,

- Target position
    - RA, DEC, proper motions, epoch and equinox
        - ICRS data are strongly preferred
        - The target positional information must error include information to permit estimation of the uncertainty of the position at the current epoch
        - Information to quantify the AO configuration
    - Brightness (*R*), colors (*V*-*R*), spectral type
        - Information to asses target priority, e.g.,
    - Parallax
    - Candidate companion from previous observation
    - Age, Doppler companion, debris disk
    - Binary, background star or galaxy, etc.

A subset of this information is retrieved and used for target acquisition and instrument configuration. Sequences of stars are generated with the same instrument configuration and a similar range of elevation angles, so that they may serve as mutual PSF references. Development of this advanced scheduling tool is a project goal (not a deliverable), possibly for the large-scale survey phase.

Selection of subsets from the master catalog may use information on predicted seeing conditions (e.g., r_0, τ_0 and wind direction from meteorological forecast) to restrict the zenith distance and brightness.

10.3.3. Science data calibration resources

Execution of the survey assumes that the following valid science data calibration data sets are in hand, and include:

- IFS science detector calibration, e.g.,
    - Darks frames
    - Linearity correction
    - Hot and cold pixel masks
- IFS calibration, e.g.,
    - Flat field response (wavelength and hence occultor and apodizer specific)
    - IFS cube reconstructor (spectroscopic mode; wavelength specific)
    - Photometric zeropoints (wavelength specific)
    - Astrometry
- System calibration
    - PSF reconstruction information from Cal system

10.3.4. Sequence: Field Star Survey observation

Perform Target Acquistion and Coronagraph Acquisition Sequences, as described above.

IFS:

- Prepare for first science exposure: Based on target magnitude and colors, compute optimum readout mode
    - Set frame time
        - Individual frame time is determined by the requirement not to trail a point source at the edge of the field of view
        - An image-motion induced Strehl degradation of 0.8 corresponds to 1.05 λ/D
    - Select readout mode based on frame time
        - Choose CDS, MCDS, or Fowler
        - Select number of samples to optimizes SNR/minimize overhead given the frame time
    - Select number of coadds
    - Select total number of coadded frames to get desired total exposure
- Acquire science frames
    - Acquire one science frame and write to disk
    - Put distilled AO system telemetry results (r_0, RMS wavefront error as a function of spatial frequency), and Cal system PSF reconstruction data into FITS headers or extensions
    - Hand off data to pipeline
    - Continue to loop, acquiring more science frames to required total
- Simultaneously with the above, real-time pipeline processes data.
    - Pipeline ingests science data in real time
    - Pipeline already has necessary instrument calibration data
        - Apply detector calibration information, e.g., dark frames, detector linearity
        - Apply instrument calibration, e.g., flat fields, photometric zero points, IFS spectral cube reconstruction
    - Science frames are reduced as they arrive. Based on reduced data,
        - Update status of target in master catalog
        - Update quicklook display
        - If data fail to meet quality, alert support astronomer with information regarding reasons for failure, e.g.,
            - High rms wavefront error
            - Variable counts in AO WFS Cal system
            - Convergence failure in AO control loop
- [After all occulted science sequence is complete optionally obtain direct, unocculted science frames. Unocculted data may be taken either prior or after the occulted observations, at the discretion of the observing program.
    - Cal system saves current calibration state
    - Cal system shutters and opens calibration loop, continuing to use saved calibration.
    - Coronagraph occultor moves to unocculted position
    - Science camera reconfigures frame time and readout mode for unsaturated, unocculted observations
    - Repeat “Acquire science frames” and “Real-time pipeline processes data” steps from above, until sufficient unocculted data has been taken.]
- If observations are desired in multiple filters, switch to next filter.
    - AO system continues running
    - Cal system opens loops.
    - IFS selects next filter and Lyot stop

  - [IFS inserts/removes polarimetry optics if necessary]
  - Cal system selects filter
  - Coronagraph selects apodizer to match chosen filter
  - Coronagraph selects occultor to match chosen filter
  - Continue with *Target Acquisition Process* from above, starting at "Science camera establishes satisfactory target acquisition" step.

- Finished with obtaining all science data at this position.

Terminate observations and move to next target, following the *Target Acquisition Process* above.

- The next target may be either an entirely different star, or the same star just observed but at a different ISS rotation angle (if the observing program requires obtaining multiple rotation angles back to back).

10.4. Polarization survey observation sequence

Nominal system configuration: IFS in polarimetry mode, with apodized pupil Lyot coronagraph. This would use the H2 or K2 configuration, tuned for best broad-band performance over the whole band.

10.4.1. Science goals

Here are two example science programs for polarimetric observations. The details of the observation sequence will be similar between these and other polarimetric programs, with the main difference being the number of targets and

- **Debris disk survey.** In this scenario, we seek to observe a large sample of nearby main-sequence stars to image previously unseen debris disks in linearly polarized light. The scientific goal is to obtain statistical information regarding the prevalence of such disks, their composition and properties, and the presence of any substructures or features at lower surface brightness levels than have been observable with previous instruments. Targets will be selected on the basis of spectral information indicating dust excess. Many targets will be observed per night. Initial visits to targets will typically attempt disk detection in a single bandpass, initially *H*, which as the best sensitivity to scattered light. For objects with successful detections, return visits will obtain observations in additional colors in order to constrain dust grain properties. A typical target catalog comprises 50 stars, which will be observed for 3 hours each, requiring 20 nights.

- **Targeted follow-up disk observations**. In this scenario, we seek to obtain detailed observations of a small number of targets, which are known to possess resolved disks, either from previous GPI observations or from other instruments (NICI, HST, Keck, etc.). Such programs will typically target no more than a few targets per night, and will obtain data in multiple wavelengths. It may be useful to conduct observations at multiple angles of the instrument rotator, in order to reduce PSF artifacts by roll-differencing, and/or to obtain additional area coverage by repositioning the diagonal corners of the field of view. It may be desirable to be able to dither the star relative to the FOV, or to have a mode in which circular polarization is measured.

10.4.2. Target selection

For the debris disk survey program, the target list will be drawn from nearby stars which have indications for circumstellar dust based on infrared excesses from e.g. Spitzer observations, or based on knowledge of disks previously resolved with other instruments.

Nightly target lists are constructed to optimize the use of observing time. Before observing begins, a target selection algorithm chooses from the master catalog a subset of stars to construct nightly target sequences given the LST at midnight. The master catalog includes target flags and information that quantifies observing priority based on factors which include youth and distance, spectral type of the parent star, strength of infrared excess indicating the presence of a debris disk, or previous observations with GPI indicating candidate planets. The details of the prioritization algorithm are to be explored by the science team in later phases.

Information in the master catalog includes, as a minimum,

- Target position
    - RA, DEC, proper motions, epoch and equinox,
        - ICRS data are strongly preferred
        - The target positional information must include information to permit estimation of the uncertainty of the position at the current epoch
- Information to quantify the AO configuration
    - Brightness (*R*), colors (*V*-*R*), spectral type
- Information to asses target priority, e.g.,
    - Parallax
    - Candidate companion from previous GPI observation
    - Age, Doppler companion, debris disk
    - Infrared excess fraction indicating presence of debris disk.
    - Binary, background star or galaxy, etc.

A subset of this information is retrieved and used for target acquisition and instrument configuration. Sequences of stars are generated with the same instrument configuration and a similar range of elevation angles, so that they may serve as mutual PSF references. Development of this advanced scheduling tool is a project goal (not a deliverable), possibly for the large-scale survey phase. It is expected that multiple different methods will be used to calibrate the PSF, including the use of data from the calibration subsystem and the observation of PSF stars.

Selection of subsets from the master catalog may use information on predicted seeing conditions (e.g., r_0, τ_0 and wind direction from meteorological forecast) to restrict the zenith distance and brightness.

10.4.3. Science data calibration resources

In order to maximize the efficiency of night operations, our goal is to conduct as much of the polarimetric calibration as possible during daylight or twilight hours. This requires the instrumental polarization to be stable on timescales of days to weeks (or longer), and that mechanism repeatability is sufficient so that calibrations may be re-used. Typical calibration operations involve observations of sources of known polarization (or unpolarized sources), such as lamps, lamps seen through polarizers, and the twilight sky.

The polarimetry mode is a significantly different mode of operation for the IFS, and thus will require substantial amounts of dedicated code in the data reduction pipeline. In particular, it will require its own reconstruction matrices for transforming raw detector frames into dual-channel ‘polarization pairs’. These matrices will be obtained through the same process as those used to obtain the regular IFS reconstructors. The goal is for these matrices to be stable over long timescales. Special data reduction pipeline modules will be needed for extracting reduced Stokes vector images from the polarization pairs (see OCDD-Appendix: Polarization Operation Calibration & Reduction).

Execution of the survey assumes the presence of all the same calibration data as described in the field star survey, with the addition of

- Instrument calibration
    - IFS polarization mode reconstructor
    - Photometric zeropoints for polarization mode (distinct from regular IFS zeropoints)
    - Modulator (wave plate) rotational zeropoint
    - Instrumental polarization calibration (i.e. the measurement matrix derived from the first row of the system Mueller matrix)

10.4.4. **Sequence: Configuration Changes to Enter Polarization Mode**

To place the instrument from regular IFS mode into polarimetry mode, the following steps must be taken.

Complete any ongoing science observations and archive data.

AO system may continue to operate, but cal system and science camera must quit current observations.

- LOWFS cal system terminates operation
- IFS Hawaii-2 RG goes to safe mode, to mitigate potential for detector hysteresis, if light scatters off of the moving polarimetry optics during insertion.
    - Array to idle/continuous reset mode
    - Shutter closed/ND filter inserted

Insert polarimetry optics modulator & analyzer

- Wave plate insertion mechanism places the wave plate into the beam.
- IFS Wollaston mechanism inserts Wollaston prism into the beam.

The motion of these components must be sufficiently deterministic to achieve the required tolerances here. The baseline is for analyzer and modulator motion to be deterministic. The wave plate rotation mechanism must have sufficient repeatability in rotation angle to achieve the desired tolerances in polarization position angle (tolerances are specificed in OCDD-Appendix: Polarization Operation Calibration & Reduction).

LOWFS cal system loop, if opened, may now be re-closed and science observations obtained, as described in the Observing Sequence, below.

For debris disk observations, accurate astrometric references are only necessary to follow up the detection of a point source companion, and the “french fry slicer” astrometric grid should be removed from the beam.

To exit polarimetry mode, reverse these steps.

10.4.5. Sequence: Polarimetry observation

Key differences from the Field Star Survey are highlighted in green.

Perform Target Acquistion Sequence, as described above.

IFS:

- Prepare for first science exposure: Based on target magnitude and colors, compute optimum readout mode
    - Set frame time. Individual frame time is determined by the requirement that field rotation does not distort or smear the polarization signal. There are two contributions to this: Direct geometric rotation of the image, and the changing orientation of the image relative to the polarizer, which changes the set of Stokes parameters measured in each frame. Accurately assessing the limits set by these two processes, and their relative importance, is still in progress.
    - Select readout mode based on frame time
        - Choose CDS, MCDS, or Fowler
        - Select number of samples to optimizes SNR/minimize overhead given the frame time
    - Select number of coadds
    - Select total number of coadded frames to get desired total exposure
    - Rotate wave plate to zero position
- Acquire science frames in polarimetry mode: For each angle in the wave plate settings list, perform the following steps. A possible wave plate settings list could be {0º, 45º, 22.5º, 67.5º}. Another possibility is 0º-360º in steps of 22.5º. The list will have at least three positions.
    - Acquire one science frame and write to disk
    - Put distilled HOWFS telemetry results (r_0, RMS wavefront error as a function of spatial frequency), and wavefront Cal system PSF reconstruction data into FITS headers or extensions
    - Hand off data to pipeline.
    - In parallel with data output to disk, rotate the wave plate to the next angle.
    - Continue to loop until all wave plate angles have been obtained.
    - Continue acquiring multiple sets of wave plate iterations, acquiring more science frames up to required total
- Simultaneously with the above, real-time pipeline processes data.
    - Pipeline ingests science data in real time
    - Pipeline already has necessary instrument calibration data
        - Detector calibration information, e.g, dark frames, detector linearity
        - Instrument calibration, e.g., flat fields, photometric zero points, IFS spectral cube reconstruction
        - Polarimetry mode reconstructor, system polarization response (measurement matrix derived from the first row of the system Mueller matrix)
    - Science frames are reduced as they arrive. Based on reduced data,
        - Update status of target in master catalog
        - Update quicklook display

        - If data fail to meet quality, alert support astronomer with information regarding reasons for failure, e.g.,
            - High rms wavefront error
            - Variable counts in AO WWFS or Cal system
            - Convergence failure in AO control loops

[After all occulted science frames are obtained, optionally obtain unocculted science frames.

    - Cal system saves current calibration state
    - Cal system shutters and opens calibration loop, continuing to use saved calibration.
    - Coronagraph occultor moves to unocculted position
    - Science camera reconfigures frame time and readout mode for unsaturated, unocculted observations
    - Repeat “Acquire science frames” and “Real-time pipeline processes data” steps from above, until sufficient unocculted data has been taken.]

- If observations are desired in multiple filters, switch to next filter.
    - AO system continues running
    - Cal system opens loops.
    - IFS selects next filter and Lyot stop
    - [IFS inserts/removes polarimetry optics if necessary]
    - Cal system selects filter
    - Coronagraph selects apodizer to match chosen filter
    - Coronagraph selects occultor to match chosen filter
    - Continue with target acquisition process from above, starting at “Science camera establishes satisfactory target acquisition” step.
- Finished with obtaining all science data at this position.

Terminate observations and move to next target.

- The “next” target may be either an entirely different star, or the same star just observed but at a different ISS rotation angle (if obtaining multiple rotation angles back to back.

10.4.6. Open questions in polarimetry mode:

- Required total exposure depths. Calculate based on instrument throughput, PSF, and expected target disk surface brightness
- Maximum exposure time per frame, given constraints on image-rotation-induced smearing of polarization.
- Detailed calibration plan. Instrumental polarization, wave plate position angle zeropoint, photometric zeropoint in polarimetry mode, etc.
- Desired set of wave plate angles? Assume least-squares data reduction. Nominal set : {0, 22.5, 45, 67.5}. At least three positions.

10.5. Binary Star Astronomy

To be added in early CDR, and will incorporate current activity to estimate reduction in contrast due to a bright neighbor.

10.6. *Volcanic activity of Io*

10.6.1. Science goals

This program consists of monitoring the volcanic activity of Io, a Galilean satellite of Jupiter, searching for the thermal signature of high temperature magma detectable between 1 and 2.5 µm. Io will be observed regularly at *K* band (2 µm). The detection of a high temperature hot spot will activate additional observations at shorter wavelengths (*Y*, *J*, & *H*) to measure the properties of the magma (temperature and flow rate).

10.6.2. Target selection

Only one target, Io, is selected for this program. Access to an ephemeris program (Horizons at JPL[4] or Skybot at IMCCE[5]) is necessary to calculate the celestial coordinates and proper motion of the satellite. The ephemeris is also to ensure that the satellite is observable: Io could be transiting, occulting or in the shadow of Jupiter. Scattered light from Jupiter can also a problem that will limit the AO correction if the satellite is too close to the limb of the planet. Io's angular diameter is 1.2" at its opposition.

10.6.3. Science data calibration resources

Execution of this program assumes the same calibration data are in hand as the field star survey program.

10.6.4. Sequence: Observations of Io

Acquire the target for unocculted observations, following the *Target Acquisition Sequence* above.

IFS:

- Prepare for first science exposure: Based on target magnitude (mv=~5-6) and colors (G0 type), compute optimum readout mode.
  - Set frame time ~1 second. Individual frame time is determined by saturation limits, plus the requirement for image rotation not to blur features on the moon beyond some threshold.
    - Choose CDS, MCDS, or Fowler
    - Select number of samples to optimizes SNR/minimize overhead given the frame time
  - Select number of coadds
  - Select total number of coadded frames to get desired total exposure
- Acquire science frames
  - Acquire one science frame and write to disk
  - Put distilled AO system telemetry results (r_0, RMS wavefront error as a function of spatial frequency), and wavefront Cal system PSF reconstruction data into FITS headers or extensions
  - Hand off data to pipeline
  - Continue to loop, acquiring more science frames to required total
- Simultaneously with the above, real-time pipeline processes data.

---

[4] http://ssd.jpl.nasa.gov/?horizons

[5] http://www.imcce.fr/page.php?nav=webservices/skybot/ws_skybot.php?method=skybot

    - Pipeline ingests science data in real time
    - Pipeline already has necessary instrument calibration data
        - Detector calibration information, e.g, dark frames, detector linearity
        - Instrument calibration, e.g., flat fields, photometric zero points, IFS spectral cube reconstruction
    - Science frames are reduced as they arrive. Based on reduced data,
        - Update quick-look display
        - If data fail to meet quality, alert support astronomer with information regarding reasons for failure, e.g.,
            - High RMS wavefront error
            - Variable counts in AO WFS or Cal system
            - Convergence failure in AO control loop
- If hot spot is detected in K band (with visual inspection by operator) then additional filters are desirable. Switch to next filter (*Y*, *J*, *H*) and obtain additional observations.
    - AO system continues running
    - Cal system opens loops.
    - IFS selects next filter and Lyot stop
    - Cal system selects filter
    - Continue with *Target Acquisition Process* from above, starting at "Science camera establishes satisfactory target acquisition" step.
- Finished with obtaining all science data at this position.

Observations complete. Terminate observations and move to next target.

10.7. Surface characterization and search for satellites of main-belt asteroids

10.7.1. Science goals

This program consist of mapping the surface of about 35 main-belt asteroids using the low resolution IFS capabilities (*Y*, *J*, *H*, & *K* filters) and searching for companions by direct imaging.

10.7.2. Target selection

Around 35 targets are selected for this program. An access to an ephemeris program (Horizons at JPL[6] or Skybot at IMCCE[7]) is mandatory to be able to calculate their celestial coordinates and their proper motion. The observations should be performed when the asteroid is the closest to Earth, in order to maximize the spatial resolution. At the beginning of each night, the ephemeris for each target should be calculated. A target is observable if: 1) its magnitude in *I* band is at least 10; 2) its elongation is larger than 140°. To improve the mapping accuracy of asteroid surface features, each target should be observed six times with various Central Meridian longitudes (separated by approximately 60°±10°). Based on their rotational rate, we can estimate if a target deserves to be re-observed (avoiding similar longitudes as have already been observed). These main-belt asteroids are relatively small: 1 Ceres, the largest, has an angular diameter of 0.7" at opposition.

---

[6] http://ssd.jpl.nasa.gov/?horizons

[7] http://www.imcce.fr/page.php?nav=webservices/skybot/ws_skybot.php?method=skybot

For solar system observations, accurate astrometric references are not necessary and the "french fry slicer" astrometric grid can be removed from the beam.

10.7.3. Science data calibration resources

This program requires the same calibration data resources, as the field star survey, above.

10.7.4. Observing Sequence: Asteroid observations

Acquire the target for unocculted observations, following the *Target Acquisition Sequence* above.

- Prepare for first science exposure: Based on target magnitude and colors (always G0-G2 type), compute optimum readout mode.
  - Set frame time. Individual frame time is determined by saturation limits, plus the requirement for image rotation not to blur features on the moon beyond some threshold.
    - Choose CDS, MCDS, or Fowler
    - Select number of samples to optimizes SNR/minimize overhead given the frame time
  - Select number of coadds
  - Select total number of coadded frames to get desired total exposure
- Acquire science frames
  - Acquire one science frame and write to disk
  - Put distilled AO telemetry results (r_0, RMS wavefront error as a function of spatial frequency), and wavefront Cal system PSF reconstruction data into FITS headers or extensions
  - Hand off data to pipeline
  - Continue to loop, acquiring more science frames to required total
- Simultaneously with the above, real-time pipeline processes data.
  - Pipeline ingests science data in real time
  - Pipeline already has necessary instrument calibration data
    - Detector calibration information, e.g, dark frames, detector linearity
    - Instrument calibration, e.g., flat fields, photometric zero points, IFS spectral cube reconstruction
  - Science frames are reduced as they arrive. Based on reduced data,
    - Update status of target in master catalog
    - Update quicklook display
  - If data fail to meet quality, alert support astronomer with information regarding reasons for failure, e.g.,
    - High rms wavefront error
    - Variable counts in AO WFS
    - Convergence failure in AO control loops
- Switch to new filters. Observe each target in all four available broadband filters *Y, J, H, K*.
  - AO system continues running
  - Cal system opens loops.
  - IFS selects next filter and Lyot stop
  - Cal system selects filter
  - Continue with *Target Acquisition Process* from above, starting at "Science camera establishes satisfactory target acquisition" step.

- Finished with obtaining all science data at this position. Terminate observations and move to next target.

11 OCDD Appendix-Star & Galaxy Confusion

James R. Graham (UCB)

v 1.2 March, 30, 2007

11.8. Confusion by stars and galaxies

In this section we estimate the false alarm rate due to physically unassociated objects and establish protocols that would reduce the amount of time needed to winnow sources masquerading as planets.

All GPI stars that are targets for exoplanet searches will be members of the solar neighborhood, and consequently will have substantial proper motion—the median proper motion of stars within 50 pc is 170 mas/yr (Figure 11-1). Therefore, measurement of proper motion will afford a definitive test to discriminate between physically associated companions and background sources. Establishment of common proper motion requires observations at least two epochs, with multiple measurements of course being preferable. Assuming rapid pipeline processing of observations and timely identification of candidates, it should be possible to schedule follow-up observations during the same observing season with the potential for measuring this amplitude of motion. The requirement that 75% of false positives be eliminated by a multi-epoch observation sequence with a cadence of four to six weeks translates to an astrometric accuracy of 2 mas (two axis).

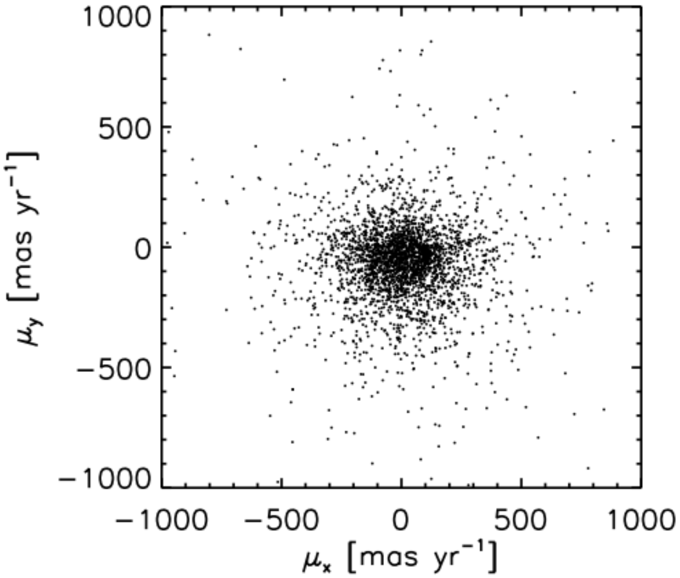

Figure 11-1: Proper motions for stars in the solar neighborhood (*d* < 50 pc). The median proper motion is 170 mas/yr and 75% of stars have a proper motion greater than 105 mas/yr.

The median exoplanet brightness detected by GPI is approximately $H = 22$ mag. At this flux level background galaxies are likely to be a source of interlopers. Galaxy counts have traditionally been a key tool for exploring cosmic evolution—any credible model of galaxy formation must be consistent with differential galaxies counts. Therefore, there have been significant efforts to make galaxy surveys from the near-UV to far-IR to the faintest possible magnitudes. Fortunately for GPI, surveys in near-IR have received special attention, because they can yield a distant galaxy's rest-frame optical flux at intermediate redshifts, which can in turn be related to fundamental properties such a galaxy's total mass in stars. The small size of near-IR detectors and the high sky at these wavelengths has made surveys a time-consuming undertaking. Only in the last few years, with the advent of larger-format detectors, has it become practical to survey deeply larger areas of the sky in the near IR to cosmologically significant flux levels. Figure 11-2 shows a recent compilation, including the results from the NICMOS survey of the Hubble Deep Field-South (Metcalfe et al. 2006).

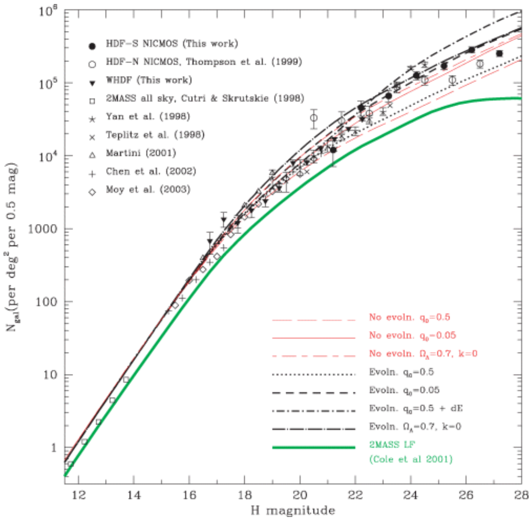

Figure 11-2: Differential galaxy number counts in half-magnitude intervals as a function of *H*-band magnitude from Metcalfe et al. (2006). Also shown are the predictions of the various models—low- and high-q_0 non-evolving model, together with a non-evolving Λ model, and four evolutionary models, low- and high-q_0, high-q_0 with an added dwarf galaxy component and a spatially flat Λ model. Error bars are calculated from the number of galaxies in each magnitude bin using Poisson counting statistics.

As a typical example of the distribution of detected exoplanet fluxes consider the results of the simulated 2 Gyr solar neighborhood survey depicted in the histogram presented in Figure 11-3. A total of 134 exoplanets are recorded in a survey of 504 stars. The median exoplanet brightness is $H = 22.1$ mag. Also plotted is the mean number of galaxies per half magnitude bin derived from the deep surveys summarized in Figure 11-2. The mean number of galaxies was computed from the observed surface density by multiplying the observed counts per solid angle by the instantaneous field and the number of stars surveyed. Here we have assumed that the radius of the dark hole with 18-cm sub-apertures gives the useful field of view. At $H \leq 22$ mag. exoplanets outnumber galaxies by almost five to one. However, in the range 22 mag. $\leq H \leq$ 25 mag. the number of exoplanets and galaxies are likely to be comparable. Figure 11-3 also shows predicted star counts for a low latitude field in the anticenter direction ($l = 180°$, $b = 10°$). In this example there are 11 stars brighter than $H = 22$ mag. and a total of 17 stars above the detection threshold. Except at low Galactic latitude and at the brightest fluxes ($H \leq 17$ mag.) are stars likely to be significant interlopers.

A false alarm rate of 20% at $H = 22$ mag. is not particularly onerous in terms of additional observing time required. However, beyond this flux limit, galaxies will become an increasing large fraction of the detections.

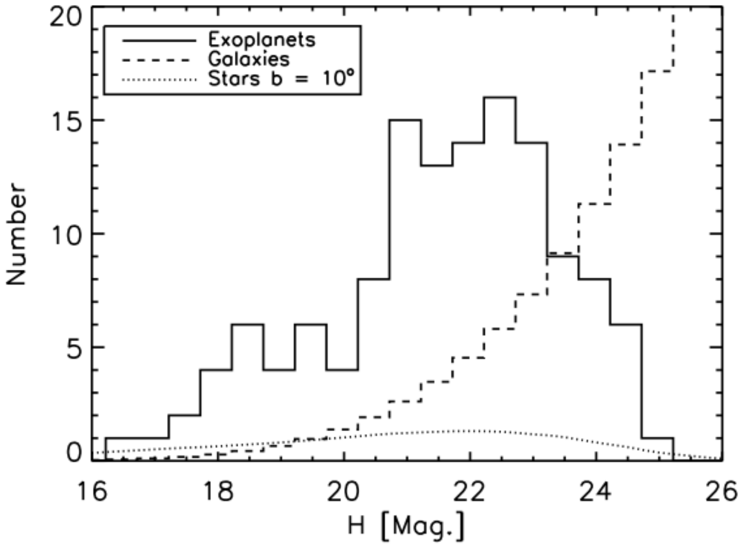

Figure 11-3: Histogram of the *H*-band magnitudes for the 134 exoplanets found in the 2-Gyr field star survey. The median exoplanet brightness is *H* = 22.1 mag. The dashed line represents the differential number counts from Figure 11-2. The average number of galaxies detected per magnitude bin is found by multiplying the galaxy counts by the field of view of GPI dark hole (1.9 arc seconds diameter at *H*) and by the number of stars in the survey (504). The differential number counts plotted here represent the upper envelope of the data in Figure 11-2. The dotted line represents the number of Galactic stars in a low latitude field in the anticenter direction. There are a total of 17 stars above the detection threshold in this example.

Figure 11-4 suggests that although galaxies will be significant contaminant, the angular size of faint, high redshift galaxies, can be used to distinguish them from exoplanets. A convenient measure of galaxy size is the effective radius, r_e, which is the radius interior to which half of the total light of the system is emitted. The median effective radius of faint galaxies in the NICMOS HDF-S is 250 mas, while galaxies above the first quartile of the size distribution have $r_e \geq 100$ mas. These galaxies will be readily resolved by GPI. The size distribution below $r_e = 100$ mas is poorly known. Some faint galaxies may have sizes appreciably smaller than this, and QSOs will be unresolved. Most of the work on establishing the statistics of galaxies sizes has been done with NICMOS/NIC-3, which has a pixel scale of 200 mas, explaining our ignorance on smaller scales. Nonetheless, it is evident that a substantial majority of extragalactic false alarms will be excluded immediately on the basis of size.

Spectral lines in distant galaxies may also be useful to identify them. Prominent emission lines fall in the *H* and *K* bands over a range of redshifts. For example Hα 6563Å will be detectable at z = 1.3–1.7 and 2.0–2.5, and [OIII] 5007 Å at z = 2.0–2.6 and 3.0–3.6.

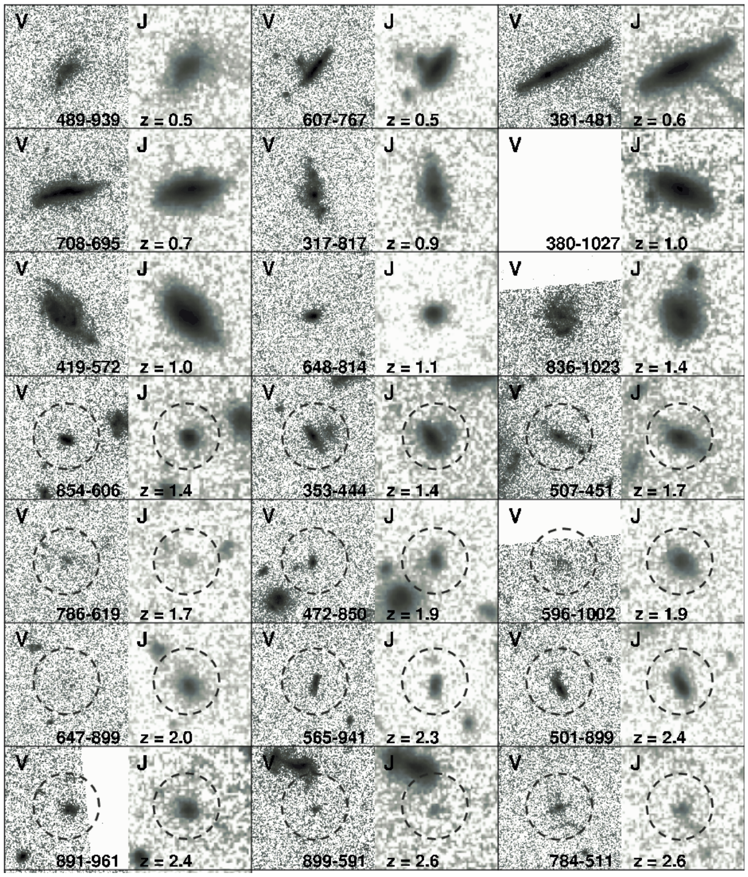

Figure 11-4: Faint, high redshift galaxies as they appear on STIS (*V*) and NICMOS (F110W) images of the HDF-S NICMOS field. Each square is 4.5 arc seconds on a side; *z* is photometric redshift (Tamm & Tenjes 2006). The median effective radius of faint high redshift galaxies measured by NICMOS in the Hubble Deep Field is 0.25 arc second (Zirm et al. 2007). At a redshift of 2.5 one arc second corresponds to 8 kpc.

11.9. References

Metcalfe, N., Shanks, T., Weilbacher, P. M., McCracken, H. J., Fong, R. & Thompson, D. MNRAS 2006, 370, 1257

Tamm, A., & Tenjes, P. 2006, A&A, 449, 67.

Zirm, A., et al. 2007 ApJ 656, 66

12 OCDD Appendix-Thermal Conditions & Operating Wavelength Range

James R. Graham (UCB)
v 1.3 February 7, 2005

12.10. Monte Carlo simulations and choice of observing wavelength

The choice of observing wavelength provides an important illustration of the Monte Carlo approach. The performance of a GPI is a strong function of wavelength. For example, the contrast between exoplanet and its host star depends on wavelength (see Volume 1). Angular resolution and diffraction effects depend on wavelength. Thus, for given field-angle, θ, the spatial frequencies of the wave front errors which contribute to speckle noise at that location scale as λ^{-1}. Since polishing errors and atmospheric wave front errors exhibit power spectral densities that decline steeply with spatial frequency, observations at longer wavelength may be favored. Similarly, the AO control radius scales as λ. Ground based observatories also suffer from photon shot noise due to the inescapable thermal background arising from the atmosphere and telescope.

Photon shot noise associated with the thermal background from telescope, AO and coronagraph can be ignored at *H* (1.65 µm). This is not true at *K* (2.2 µm) or longer wavelengths, e.g., *L'* (3.6 µm) or *M* (5.4 µm). Longer wavelengths may be important because they are closer to the blackbody peak for cool planets, and consequently the predicted planet fluxes are less sensitive to the details of radiative transfer and opacity sources in the exoplanet atmospheres. Observations at these wavelengths may also be valuable on account of their diagnostic use, both for planets and debris disks.

Consider the performance of an GPI system with a limiting magnitude of $I = 7$ mag., no speckle suppression, a telescope at 273 K and an “ambient” temperature AO system and coronagraph at 300 K. The coronagraph mask is assumed to have a 50% emissivity. An exposure time of 1 hour is assumed and 2164 stars are observed yielding 51 planets or a 2.4 % detection rate. The noise budget for this system is shown in Figure 12-1. In this example speckle noise dominates. Evidently, without speckle noise suppression, speckle noise dominates.

In a simple modification to the previous experiment, we suppose that we now that the IFS suppresses speckle noise by a factor of 10. Now 82 planets are detected corresponding to a 3.8% success rate, representing a 61% improvement compared to no speckle suppression. Inspection of Figure 12-2 shows that (for the median case) star halo photon short noise, followed by coronagraph thermal emission are now dominant in the noise budget.

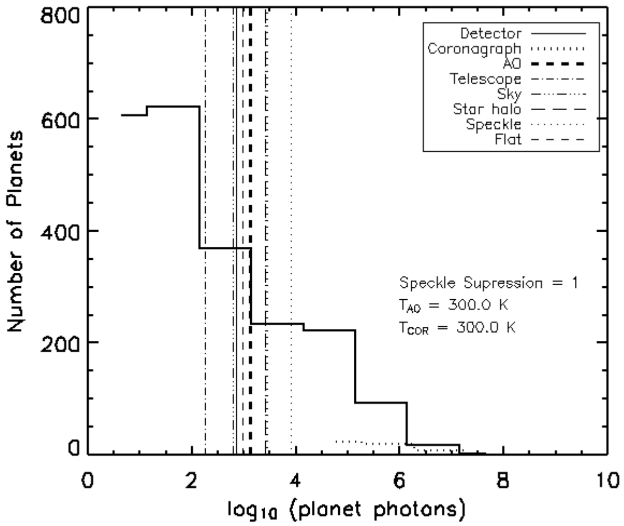

Figure 12-1: The noise budget for GPI operating at K_s. The histogram shows the photon signal accumulated from the planets in the population described Volume 1. The vertical lines indicate the RMS noise in equivalent photons from: detector read noise and dark current; thermal emission from the coronagraph, AO system, telescope and sky; the star halo photon shot noise; speckle noise; and gain noise associated with flat field errors. The star halo photon noise, the speckle noise and the gain noise all depend on the magnitude of the target star and hence only the median values are indicated. In this example speckle noise dominates. The faint dotted histogram shows the 5-σ detected planets.

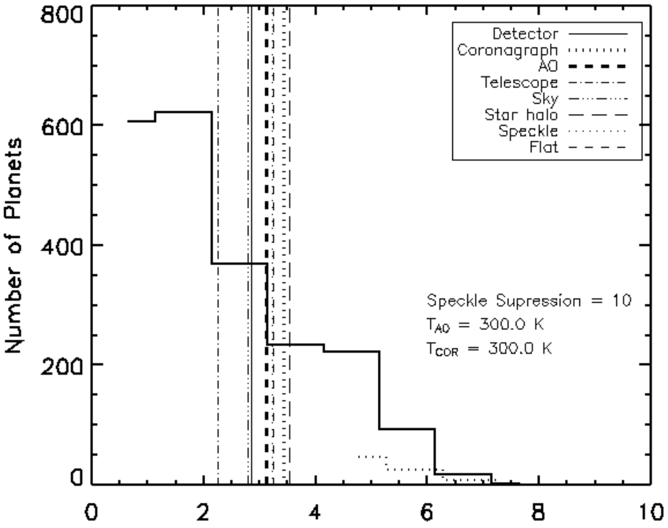

Figure 12-2: The noise budget as in Figure 12-1, except that the IFS suppresses speckle noise by a factor of 10. The star halo photon shot noise and the thermal emission from the coronagraph are now dominant. Note that the vertical lines only indicate the median value of each noise source.

A factor of 10 times speckle noise suppression represents a modest goal, and, in principle, an IFS should be able to deliver substantially more suppression than this. However, for additional speckle noise suppression to be valuable the thermal shot noise from the coronagraph must be controlled. Star halo photon noise dominates and coronagraph thermal noise is second. Consequently, cooling the instrument can increase the planet detection rate (Figure 12-3). A coronagraph temperature of 240 K and AO system of 260 K yields 100 planets or a 4.6% detection rate. There is no significant grain achieved by cooling below 250 K at K_s with 10 times speckle noise suppression. The same ExAOC system operating at *H* can achieve planet detection rates approaching 10% for the field survey target list, and therefore the shorter wavelength is preferred.

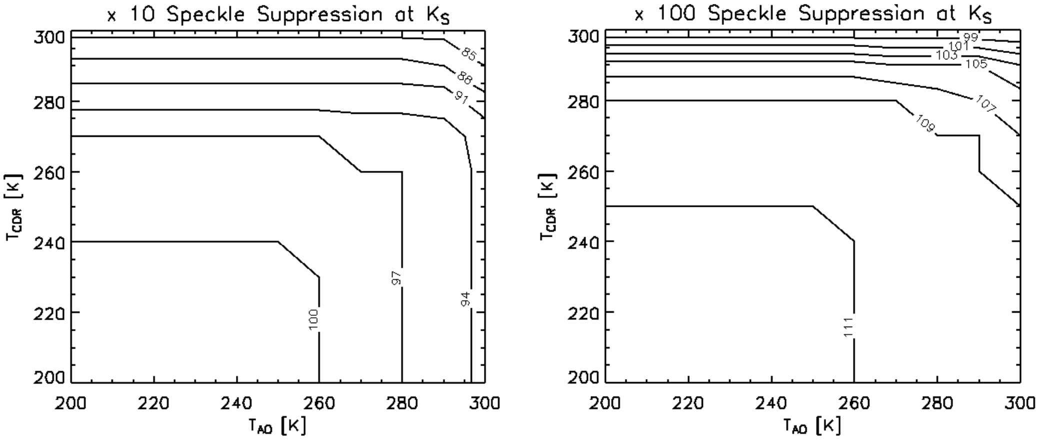

Figure 12-3: The effect of controlling the AO and coronagraph temperatures on planet detection for speckle noise suppression factors of 10 (left) and 100 (right). Contours are labeled with the number of planets detected. Since the star halo photon shot noise is a major component in the error budget (see Figure 12-2) reducing the instrument temperature is only modestly effective in boosting the planet detection rate. An additional ≈ 10% performance (as quantified by the planet detection rate) can be eked out by a factor of 100 speckle suppression on an instrument operating at 270 K. The 100-planet detection contour corresponds to a planet detection rate of 4.6% for this experiment.

Now consider operation at *L'*. In all respects this experiment is identical to that represented in Figure 12-1, except that the science observing wavelength has been changed from K_s to *L'*. The noise budget in Figure 12-4 shows that thermal photon shot noise from the instrument dominates. Consequently, the planet detection rate is only 1.9%, or a factor of nearly two and a half times worse than the best performing K_s system. Speckle noise suppression is irrelevant at *L'*. The only way to improve the performance is to cool the instrument and thereby eliminate the thermal photon shot noise. Figure 12-5 shows that the performance of a warm AO system and coronagraph at *L'* is comparable to the K_s system operating at 300 K with no speckle noise suppression. However, cooling the *L'* instrument to below 180 K makes it competitive with the best K_s system operating at 273 K with times 10 speckle noise suppression. Speckle noise suppression is never effective at *L'*, not even for a cooled system, because thermal photon noise from the telescope and atmosphere will dominate.

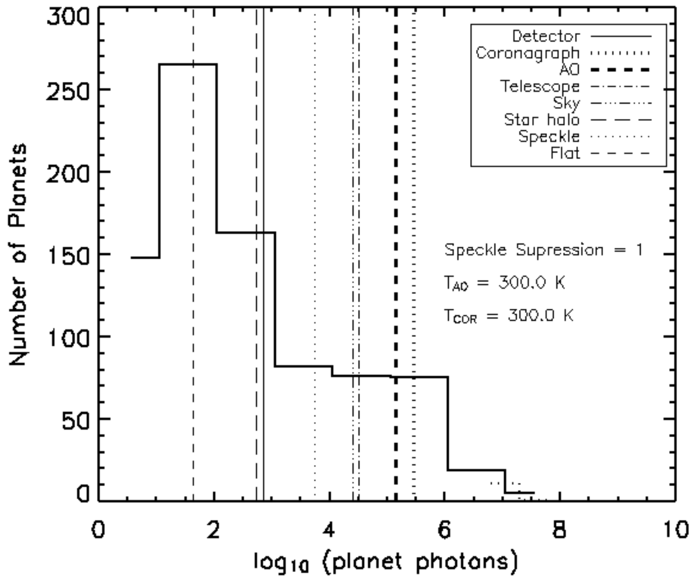

Figure 12-4: The noise budget for GPI operating at *L'*. The experiment is the same as that depicted in Figure 12-1, except that the observing wavelength has been changed from K_s to *L'*. The thermal shot noise from the coronagraph and the AO system dominate.

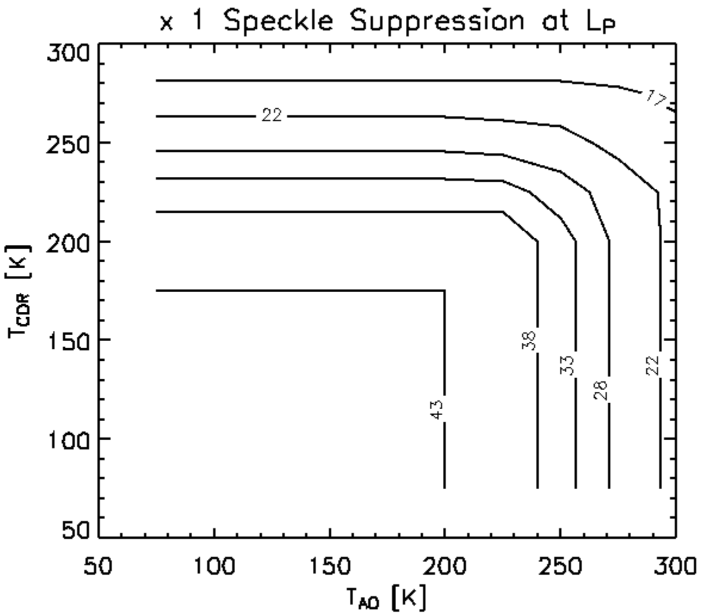

Figure 12-5: Planet detection contours for an ExAOC system working at *L'*. Since thermal photon shot noise from the instrument dominates (Figure 12-4), cooling has a profound effect on the planet detection rate. Cooling to temperatures significantly below 270 K are needed, and full performance is only achieved below ≈ 180 K. The planet detection rate on the highest contour, 43 detected planets, corresponds to a planet detection rate of 4.9%.

The conclusion from this exercise is that both K_s and *L'* are scientifically valid options. Broadly there are two choices for GPI: 1) A "warm" system ($\leq$ 280 K) operating at $\lambda \leq 2.4$ µm and incorporating substantial speckle noise suppression ($\geq$ 10); 2) a "cold" instrument ($\leq$ 180 K) operating at *L'* with no speckle suppression. The fact that speckle noise never dominates at *L'* means that an “extreme” AO performance is not required at the longer wavelengths. The benefits of equipping a “warm” system with a science camera that operates at wavelengths longer than 2.4 µm are minor, since it may be outperformed at these wavelengths by a lower performance AO system that is optimized for low background operation, e.g., one that uses a deformable secondary or a cryogenic deformable mirror.

13 OCCD Appendix-Spectral Resolution and Inferring Exoplanet Mass & Age

James R. Graham (UCB)

v 1.4 November 20, 2006

13.1. Introduction

The science driving spectral resolution to higher values is the desire to record the information present in the strength and shape of various atomic molecular features, e.g., NaI, KI, CH_4, NH_3, and H_2O. This must be achieved without compromising the core goal of detecting planets. Ultimately, follow-up instruments will use high-resolution spectroscopy to measure effective temperature, gravity, chemical composition and chemical abundances.

13.2. Model Atmospheres

Burrows, Sudarsky, & Lunine (2003) have calculated a grid of evolutionary models for cooling planets with masses between 1 and 15 M_J that represent isolated objects with effective temperatures from 130 K to 800 K. The models are coupled to model atmospheres and therefore predict the emergent spectra as a function of age and mass, providing a useful context for examining the optimal spectral resolution for GPI. At young ages the formation history is likely still imprinted on internal entropy content of the planet. Therefore, predictions for the resultant radius and effective temperature earlier than 100 Myr may be unreliable (Marley et al. 2006). As the hosts of the young planets are located in relatively distant star forming regions, e.g., Taurus (140 pc), these stars are too faint to be primary targets for GPI.

Previous trade studies of the spectral resolution requirements for GPI have focused on the need for speckle suppression, which calls for $R = \lambda/\Delta\lambda > 40$. We have also shown that the gravity and effective temperature estimates for L and T type atmospheres can be constrained by considering two or three colors indices within the *H* or *K* bands. However, as model atmosphere calculations imply that the emergent spectra of young planets exhibit a wealth of information, further study is called for. We have analyzed the Burrows et al. spectra with the goal of constructing an objective method to discover whether or not the planet age and mass can be deduced from narrow-band spectral indices. We consider the ideal case where we know *a priori* the planet mass and age. Investigating this purely theoretical relation between the observable properties (colors) and the fundamental planet properties (age and mass) as a function of spectral resolution and signal to noise ratio is straightforward. Practical application of this transformation between color indices and mass and age assumes that a set of templates for calibrating exoplanet atmospheres exists. These in turn need advances in calibrating the ages of parent stars and testing and refining the fidelity of spectroscopic indicators of effective temperature and gravity. Nonetheless, to the extent that state-of-the-art models capture the relevant physics and chemistry of planetary atmospheres, we can establish lower limits on spectral resolution and signal to noise requirements.

13.3. Science Requirements

Planet masses span roughly an order of magnitude, from 1 M_J to 10 M_J, therefore relatively good precision is called for. For example to make a reliable mass function with ten bins requires an rms error in $\Delta M/M < 10\%$. Note that in computing the mass we will not assume that the age of the planet is known, i.e., we will marginalize over age as a nuisance parameter. The dynamic range in age is larger, and therefore a larger fractional error may be acceptable. Some young planets in solar neighborhood associations are as young as 12 Myr, e.g., the β Pic moving group, and we are sensitive to some planets that are as old as a Gyr. Thus 20% rms error in $\Delta t/t$ is acceptable for this preliminary reconnaissance.

13.4. A Regression Analysis

Consider the matrix **C**, which represents an arbitrary linear combination of the *n* colors corresponding to a vector of *m* model planets with masses M_{i}. The number of colors depends on the spectral resolution across the *Y*, *J*, *H* and *K* windows. Suppose that the relationship between the observed quantity **C** and the model properties, **M**, can be expressed as

$$\mathbf{Cd} = \mathbf{M}.$$

To infer **d** we can treat this as a least squares problem, by multiplying both sides by $\mathbf{C}^T$ and the pseudo-inverse[8] of **C**,

$$\mathbf{d} = \left(\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T\mathbf{M}.$$

If we assume that the planet mass (or age) can be found from a linear combination of the colors then the form of **C** that describes terms up to, say, the quadratic is

$$\mathbf{C} = \begin{pmatrix} 1 & c_{0,0} & c_{0,1} & \cdots & c_{0,n-1} & c_{0,0}^2 & c_{0,1}^2 & \cdots & c_{0,n-1}^2 \\ 1 & c_{1,0} & c_{1,1} & \cdots & c_{1,n-1} & c_{1,0}^2 & c_{1,1}^2 & \cdots & c_{1,n-1}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & c_{m-1,0} & c_{m-1,1} & \cdots & c_{m-1,n-1} & c_{m-1,0}^2 & c_{m-1,1}^2 & \cdots & c_{m-1,n-1}^2 \end{pmatrix}.$$

13.5. Color Indices

We measure colors by convolving the high-resolution Burrows et al. spectra to a given resolution and defining non-overlapping band-passes that lie within the terrestrial atmospheric windows (see Figure 13-1). The resultant fluxes are used to construct colors within each atmospheric window[9]. Thus, each set of colors represents an instantaneous measurement with GPI. In the current example (Figure 13-1), there are six independent spectral channels in the *H* band, so this window contributes five colors. We do not assume that the *H* and *K* fluxes can be cross-calibrated so that the color indices within the *H* and *K* windows are constructed separately.

---

[8] In practice singular value decomposition should be used in preference to the Moore-Penrose pseudo-inverse.

[9] Colors are defined in the conventional astronomical sense as logarithms of flux ratios.

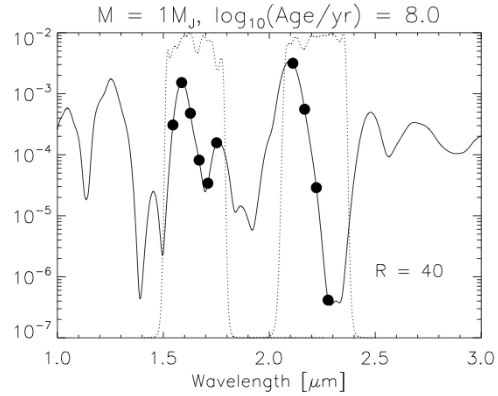

Figure 13-1: The spectrum of a 100 Myr, 1 Jupiter mass planet from Burrows et al. (2003) smoothed to a spectral resolution of 40. The dots represent sampling of the smoothed spectra in non-overlapping spectral channels. In this example 10 independent flux measurements are made. In the current analysis for this example these are combined to form to five colors in *H* and three colors in *K*. The dotted lines represent the transmission of the NIRI *H* and *K* filters, showing the extent of the atmospheric windows.

13.6. Line Spread Function

The details of the instrumental line spread function, at fixed *R* and signal-to-noise ratio (*SNR*), determine the ability of the instrument to discriminate between different models. Therefore, we have performed these simulations using a Gaussian line spread function and more realistic sinc^2 profile. If two instruments with the same FWHM are compared (see Figure 13-2), but one has a Gaussian and the other has a sinc^2 response, the instrument with the Gaussian line spread function has greater ability to discriminate between models. The $\Delta\lambda^{-2}$ Lorentzian envelope of the sinc^2 profile throws sufficient energy into the line wings so that the effective spectral resolution of such an instrument is inferior to one with a Gaussian line spread function of equal FWHM.

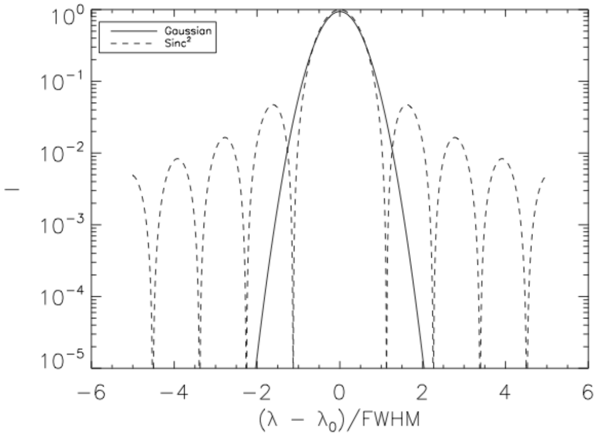

Figure 13-2: Comparison of Gaussian and sinc^2 line spread functions with equal FWHM. In the line wings the $\Delta\lambda^{-2}$ Lorentzian envelope of the sinc^2 profile throws sufficient energy to make the effective resolution of the Gaussian instrumental profile a better discriminant between models. Each curve is normalized to the same total energy.

13.7. Results

We have constructed the array of colors for a range of spectral resolutions and signal-to-noise ratios. The utility of the extracted colors are evaluated by examining the residuals in mass and age inferred from application of the regression matrix **d**. We have investigated using both linear and quadratic terms in **C**. Using a simple linear relation between colors and the planet properties is inadequate to capture the information present in the synthetic spectra; a regression that includes quadratic terms relating the colors to the planet properties give substantially reduced residuals. Comparison of Gaussian and sinc^2 line spread functions with equal FWHM shows differences in the ability to discriminate between models and confirms that the details of the profile must be taken into consideration when choosing the spectral resolution.

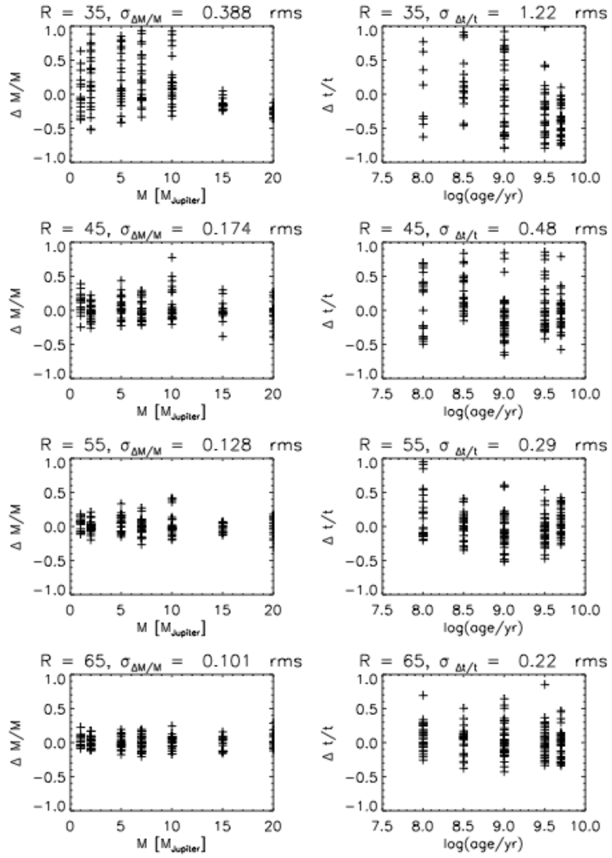

Figure 13-3: Results of the quadratic regression analysis showing the residual between planet properties based on colors in *H* and *K* bands. Fractional residuals for planet mass are shown in the right column, fractional residuals for age are on the left. The title of each plot lists the spectral resolution and the rms residual. All these results are for *SNR* = 10 per color. The ability to discriminate both mass and age increases monotonically with increasing spectral resolution. The line-spread function is a sinc^2.

Figure 13-3 shows the residuals in planet mass ($\Delta M/M$) and planet age ($\Delta t/t$) derived from colors for quadratic regressions for spectral resolutions of R = 35, 45, 55, & 65. The signal to noise ratio in Figure 13-3 is 10; an example with *SNR* = 20 is shown in Figure 13-4. It is evident from inspection of Figure 13-3 that the reliability of the measured planet properties increases with spectral resolution.

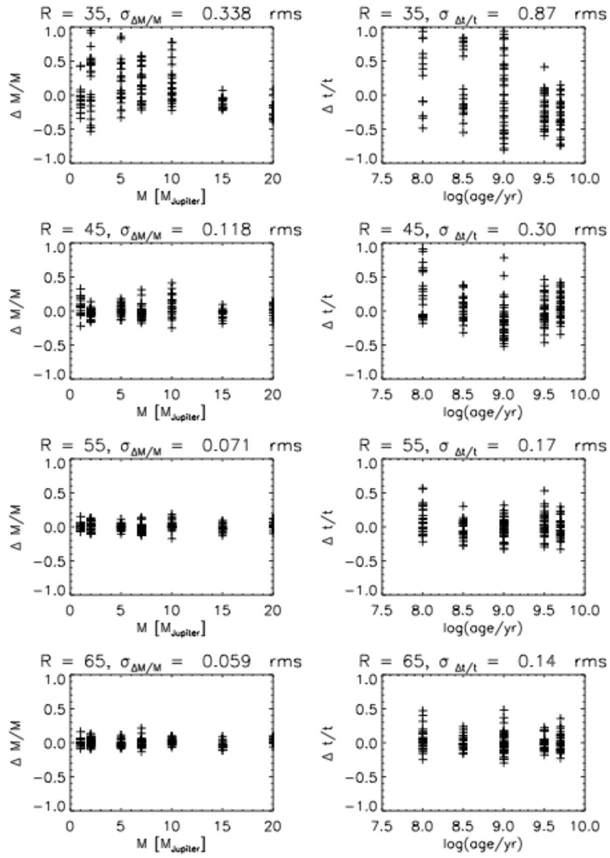

Figure 13-4: Same as Figure 13-3, except for *SNR* = 20. The ability to discriminate both mass and age increases with *SNR*.

Comparison of Figure 13-3 and Figure 13-4 shows the effect of increasing the *SNR*, and that spectral resolution can be traded off against the signal-to-noise ratio. The combined effects are best seen in Figure 13-5, which shows the variation of error in mass and age as a function of spectral resolution. This figure shows a steep improvement in the ability to discriminate between models as *R* increases from 30 to 60 followed by a slower improvement. The effects of improved *SNR* are also evident.

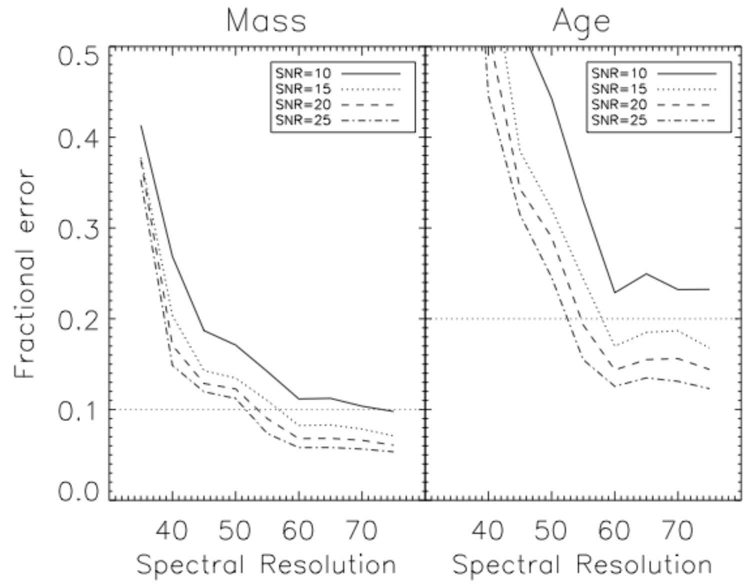

Figure 13-5: Variation of fractional rms error in mass (left) and age (right) vs. spectral resolution. These plots summarize the information presented in Figure 13-3 and Figure 13-4. Four curves are shown in each panel for *SNR* $\in\{10, 15, 20, 25\}$. The horizontal dotted lines at 10% in mass and 20% in age represent the requirements.

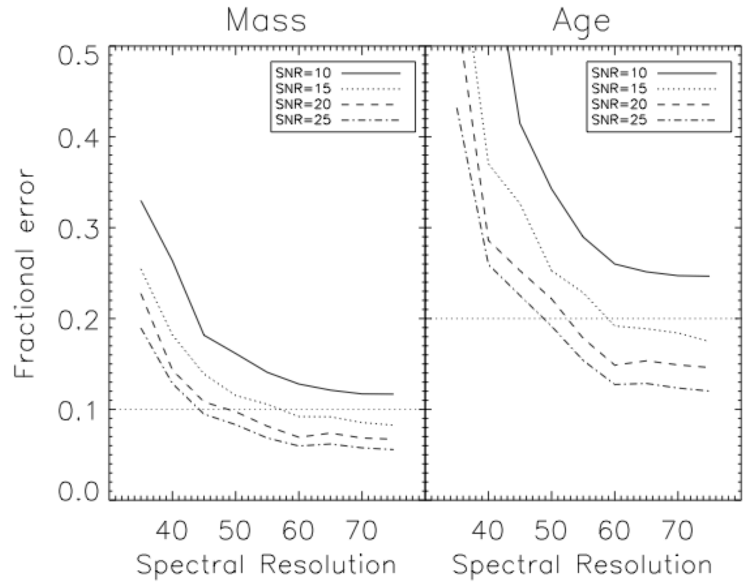

Figure 13-6: Same as Figure 13-5, except for colors in the *H* and *Kp* filters.

13.7.1. Conclusions

Inspection of Figure 13-3 and Figure 13-4 shows that $R \leq 40$ is inadequate to take advantage of the information that is likely present in the spectra of young exoplanets. Doubling the SNR at this resolution does not improve the situation significantly. Systematic errors in the integral field spectrograph, e.g., flat field errors, will limit the precision with which colors can be measured to a few percent. Demanding performance of *SNR* > 20-25 would place unrealistic demands on the spectrograph. The alternative is to adopt a higher spectral resolution. The steepness of the *R* vs. error plot means that even a modest increase of spectral resolution returns significantly more information. Assuming that colors can be measured to a precision of 4–5% (the lower two curves in Figure 13-5) a spectral resolution of 45–50 meets our science goals and is sufficient for an initial survey of the atmospheres of exoplanets. Comparison of Figure 13-5 and Figure 13-6, which show the difference between the *H*/*K* and *H*/*K*p filters, imply that additional investigation of small changes in the filter band passes needs to be investigated.

13.7.2. Further Work

This work highlights several key steps for required for GPI science and several additional avenues of investigation. These include:

- Include realistic errors terms in the SNR estimate;
- Investigate the impact of small changes in the filter band passes;
- Design experiments to validate exoplanet atmosphere predictions and establish spectral templates;
- Develop techniques to improve age estimates for parent stars;
- Exploit of multi-planet systems where siblings presumably have similar ages;
- Determine if composition variations in exoplanet atmospheres be constrained.

13.7.3. References

"Beyond the T Dwarfs: Theoretical Spectra, Colors, and Detectability of the Coolest Brown Dwarfs", Burrows, A., Sudarsky, D., and Lunine, J. I. 2003, ApJ, 596, 587

"On the Luminosity of Young Jupiters", Marley, M. S., J.J. Fortney, J. J., Hubickyj, O., Bodenheimer, P., Lissauer, J. J. astro-ph/0609739

14 OCDD Apendix-A Mueller Matrix Analysis of Non-Ideal Dual Channel Polarimeter

James R. Graham (UCB)

v1.8 16 January 2007

14.1. Introduction

Sensitivity to the polarization state of incident radiation measured by GPI is required for two reasons. First polarization capability is required to aid the detection of low surface brightness emission. Second, the ability to measure the degree of polarization is required to measure the scattering properties of circumstellar dust grains.

Consideration of differential phase retardance between the s- and p- components is also crucial to general operation of GPI. Stress induced birefringence can induce polarization-dependent phase errors that cannot be sensed or corrected. Ideal dielectric layers on will introduce angle-dependent phase retardation too. If these elements are present in a converging or diverging beam they can also introduce polarization state dependent aberrations. Examples include any optics with anti-reflection coatings or dichroic beam splitters.

14.2. Polarization Requirements

Speckle noise dominates the detection of sources in the vicinity of the bright natural guide stars that are GPI's science targets. As speckles are a diffraction phenomenon, the intensity and location of a speckle changes with wavelength. For point sources, we plan on using the variation of the location with wavelength to identify speckles that arise from pupil plane phase errors. In addition, for sources with T dwarf spectra the deficit of flux in the 1.6 µm $2\nu_3$ CH_4 band provides an additional discriminant between exoplanets and speckles.

For extended sources, such as debris disks, differential imaging between adjacent wavelengths cannot be used to discriminate between true sources and artifacts. As dust scatters the light of the parent star, to first order, the spectrum of star and the disk are the same. The primary difference between the starlight and the disk is caused by the fact that dust grains selectively scatter one polarization state selectively over the other. Small dust grains preferentially scatter light with the **E** vector oriented perpendicular to the scattering plane (the plane that contains the source, the scatter and the observer). To the extent that GPI does not introduce instrumental polarization, or that the instrumental polarization can be characterized, this intrinsic difference between the starlight and the disk light can be exploited to separate the two signals.

An example of the polarized signal from a debris disk is shown in Figure 14-1. In this case the degree of polarization reaches 40% at 5 arc sec (50 AU). In general the degree of polarization increases with distance from the star, as the fraction of scattering events with scattering angle close to $\pi/2$ (where polarization is maximum) increases.

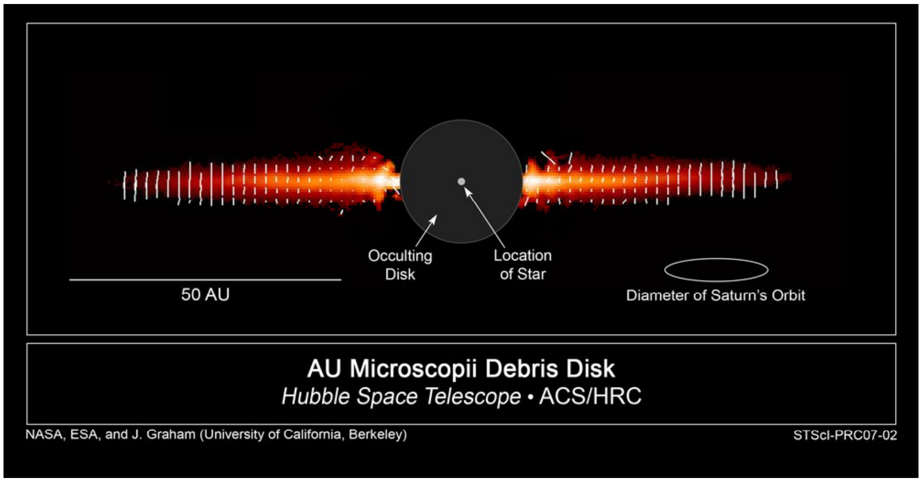

Figure 14-1: The AU Mic debris disk detected in polarized light at 606 nm (Graham et al. 2007 ApJ 654, 595). The orientation and degree of linear polarization is indicated by the white tick marks. The degree of polarization rises from a few percent close to the star up to about 40%.

The polarization requirements are

- Speckle suppression
    - ×100 of unpolarized light (detection of AU Mic/50 at an inclination of 60°)

- Stokes parameters calibration to characterize dust grain properties
    - Measure degree of linear polarization to $\leq 1\%$
    - Measure PA to $\leq 1°$
    - Full Stokes parameters

14.2.1. Not discussed

Wavefront errors of the polarizing components are not addressed here. Specifically, the polarization modulator, e.g., a mechanical wave plate, can modulate the speckle pattern as it rotates.

14.3. Measurement of the Stokes Vector

Several techniques are known for the measurement of components of the Stokes vector. Under the assumption that a single detector is available, all four components of the Stokes vector can only be measured quasi-simultaneously. Such a method must necessarily rely on introduction of rotation between a polarizer and a retarder that introduces a differential phase delay between the s- and p-polarization states. A classical system for the measurement of the Stokes vector comprises a quarter-wave plate and a linear polarizer placed in series with a detector. The orientation of the fast axis of the wave plate is rotated relative to the axis of the linear polarizer by means of a rotation stage. The method is only quasi-simultaneous, because a minimum of four orientations is required to measure the four components of the Stokes column vector $(I, Q, U, V)^T$. True simultaneous measurement of I, Q, U, and V can be achieved using four detectors that are placed to measure the reflections of an incident beam.

The current purpose is to investigate methods for deriving the tolerances required to achieve accurate Stokes polarimetry when a polarimeter consists of a rotating wave plate and a fixed linear polarizer. The Stokes vector, represented $\mathbf{S} = (I, Q, U, V)^T$, specifies the total intensity I and three other parameters that indicate both the ellipse of polarization and the degree of polarization.

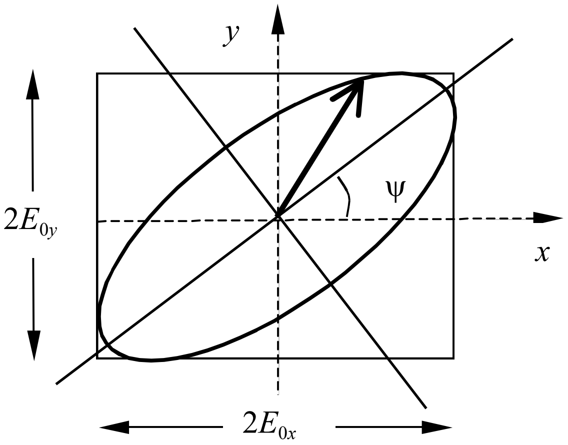

Figure 14-2: The polarization ellipse showing the projection of the tip of the electric field vector onto the (x, y) plane. The heavy line terminated with an arrow represents the instantaneous electric vector. The time dependent x- and y-components of the electric field are $E_x(z, t) = E_{0x}\cos(\omega t - kz + \delta_x)$ and $E_y(z, t) = E_{0y}\cos(\omega t - kz + \delta_y)$ where $\omega = 2\pi f$ is the angular frequency, $k = 2\pi/\lambda$ is the wavenumber, $\delta = \delta_y - \delta_x$ is an arbitrary phase. The position angle of the polarization ellipse is $\tan 2\psi = 2E_{0x}E_{0y}\cos\delta \Big/ \left(E_{0x}^2 - E_{0y}^2\right)$.

Referring to Figure 14-2, the Stokes vector is defined in terms of the amplitude of the electric field and the phase delay, δ, between the x- and the y-components

$$\mathbf{S} = \begin{pmatrix} I \\ Q \\ U \\ V \end{pmatrix} = \begin{pmatrix} E_{0x}^2 + E_{0y}^2 \\ E_{0x}^2 - E_{0y}^2 \\ 2E_{0x}E_{0y}\cos\delta \\ 2E_{0x}E_{0y}\sin\delta \end{pmatrix}.$$

For linearly polarized light $\delta = 0$, and V, which measures circular polarization is zero. For linearly polarized radiation with $E_{0y} = 0$ (horizontal polarization) the Stokes vector is $\mathbf{S}_{HP} = I_0\begin{pmatrix}1 & 1 & 0 & 0\end{pmatrix}^T$. For vertically polarized light with $E_{0x} = 0$ the Stokes vector is $\mathbf{S}_{HP} = I_0\begin{pmatrix}1 & -1 & 0 & 0\end{pmatrix}^T$. Conventional astronomical polarization notation frequently rotates this convention by 90° counterclockwise, so that linearly polarized light oriented along the NS direction corresponds to $Q \geq 0$.

In general the Stokes vector describes either partially or completely polarized radiation, thus $I^2 \geq Q^2 + U^2 + V^2$. The equals sign applies for fully polarized light. The degree of polarization is defined as

$$p = \sqrt{Q^2 + U^2 + V^2} \Big/ I,$$

and the position angle with respect to the *x*-axis is given by $\tan 2\psi = U/Q$. The most convenient source of polarization of known orientation is the twilight sky. The details are given in the Appendix.

14.4. Mueller & Measurement Matrices

The idealized system under consideration consists of a wave plate in conjunction with a linear polarizer at a fixed position angle. The input and output Stokes vectors are $\mathbf{S^i}$ and $\mathbf{S}^\mathrm{o}$, respectively, and θ is the position angle of the fast axis of the wave plate. Wave plates have the property that along the *x*-axis (the fast axis) the *x*-component of the electric field experiences a phase shift of $\phi/2$, and the component of the electric field along the *y*-axis experience a phase shift of $-\phi/2$, giving rise to a total phase shift ϕ between the orthogonal electric field components. Therefore the Mueller matrix describing a wave plate with its fast axis oriented along the *x*-axis is

$$\mathbf{M}_{WP}(\phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\phi & -\sin\phi \\ 0 & 0 & \sin\phi & \cos\phi \end{pmatrix}.$$

Common types of wave plates include half wave retarders, $\phi = \pi$, or quarter wave retarders, $\phi = \pi/2$, often known as half wave and quarter-wave plates, respectively. Polarizing elements are defined with respect to the *x*-axis. If the wave plate is rotated counterclockwise through an angle θ, the Mueller matrix for the rotated component is $\mathbf{M}(\theta) = \mathbf{M}_{ROT}(-\theta)\mathbf{M}_{WP}\mathbf{M}_{ROT}(\theta)$, where $\mathbf{M}_{ROT}$ is the Mueller matrix for a rotation

$$\mathbf{M}_{ROT}(\theta) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta & 0 \\ 0 & -\sin 2\theta & \cos 2\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

A linear polarizer changes the amplitude of E_{0x} and E_{0y}, either by absorption or by splitting according to polarization state. The absorption coefficients in the amplitude are $0 \le p_x \le 1$ and $0 \le p_y \le 1$. The Mueller matrix for a linear polarizer is therefore

$$\mathbf{M}_{LP} = \frac{1}{2}\begin{pmatrix} p_x^2 + p_y^2 & p_x^2 - p_y^2 & 0 & 0 \\ p_x^2 - p_y^2 & p_x^2 + p_y^2 & 0 & 0 \\ 0 & 0 & 2p_x p_y & 0 \\ 0 & 0 & 0 & 2p_x p_y \end{pmatrix}.$$

Intensity rather than amplitude transmission factors are frequently used when characterizing a linear polarizer. The transmittance of a polarizer for linearly polarized light whose angle is chosen such as to maximize the transmitted intensity is designated k_1; the minimum transmission as a function of the angle of a linearly polarized beam is k_2. As k_1 and k_2 define the intensity transmission, it is evident that $k_1 = p_x^2$ and $k_2 = p_y^2$, if $p_x > p_y$. The transmission of unpolarized light by two parallel analyzers is $(\mathbf{M}_{LP})^2_{0,0} = (k_1^2 + k_2^2)/2$ and two crossed analyzers is $(\mathbf{M}_{LP}\mathbf{M}_{ROT}(\pi/2)\mathbf{M}_{LP})_{1,1} = k_1 k_2$, and the ratio $k = k_2/k_1$ is called the extinction ratio.

14.5. A Single-Channel Polarimeter

For a perfect linear polarizer which transmits E_{0x} perfectly and attenuates E_{0y} perfectly $p_x = 1$ and $p_y = 0$. The system Mueller matrix for our simple polarizer consisting of a rotating wave plate and an ideal polarizer is $\mathbf{M}_{SYS}(\theta,\delta) = \mathbf{M}_{LP}\mathbf{M}_{ROT}(-\theta)\mathbf{M}_{WP}(\phi)\mathbf{M}_{ROT}(\theta)$. Using the top row of the system Mueller matrix the observed intensity is

$$\begin{aligned} 2I^O(\theta_j,\phi) = {} & I^i + \\ & \left(\cos^2 2\theta_j + \sin^2 2\theta_j \cos\phi\right)Q^i + \\ & \left[\cos 2\theta_j \sin 2\theta(1-\cos\phi)\right]U^i + \\ & \sin 2\theta_j \sin\phi V^i \end{aligned},$$

where the output intensity, $I^o(\theta_j,\phi)$ is measured at angle $j = 1, 2, 3, \ldots N$. Since there are four real components of the Stokes vector, a minimum of four values of $I^o(\theta_j,\phi)$ must be measured at four different angles. A system using a half-wave plate ($\phi = \pi$) is insensitive to V^i , and only three measurements are needed. The column vector of measurements are related by a measurement matrix, **M**’, to the input Stokes vector

$$\begin{pmatrix} I_1^o \\ I_2^o \\ I_3^o \\ \vdots \\ I_N^o \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & \cos^2 2\theta_1 + \sin^2 2\theta_1 \cos\phi & \cos 2\theta_1 \sin 2\theta_1(1-\cos\phi) & \sin 2\theta_1 \sin\phi \\ 1 & \cos^2 2\theta_2 + \sin^2 2\theta_2 \cos\phi & \cos 2\theta_2 \sin 2\theta_2(1-\cos\phi) & \sin 2\theta_2 \sin\phi \\ 1 & \cos^2 2\theta_3 + \sin^2 2\theta_3 \cos\phi & \cos 2\theta_3 \sin 2\theta_3(1-\cos\phi) & \sin 2\theta_3 \sin\phi \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \cos^2 2\theta_N + \sin^2 2\theta_N \cos\phi & \cos 2\theta_N \sin 2\theta_N(1-\cos\phi) & \sin 2\theta_N \sin\phi \end{pmatrix}\begin{pmatrix} I^i \\ Q^i \\ U^i \\ V^i \end{pmatrix}$$

The Stokes parameters are functions of the observed intensities $\mathbf{I}^o$ and the matrix elements of **M’**. In the classical approach to Stokes polarimetry sets $N = 4$, and the column vector $\mathbf{S}^i$ is found by inverting the square matrix **M’**. If the determinant of the measurement matrix is zero, no unique solution exists for the four Stokes parameters. As the magnitude of this determinant increases from zero, the Stokes parameters are less sensitive to measurement errors associated with $\mathbf{I}^o$, θ, and ϕ. Unless careful choice of the four angles θ_1, θ_2, θ_3, and θ_4 is made, the problem will be ill conditioned, and not all sequences of retarder angles produce valid results. These situations can produce large errors in the final results.

14.6. A Least Squares Approach

An alternate approach chooses $N > 4$, and casts the problem of retrieving the input Stokes vector $\mathbf{S}^i$ as a least squares problem. Restating the last equation as

$$\mathbf{I}^o = \mathbf{M}'\mathbf{S}^i,$$

the least squares estimate of $\mathbf{S}^i$ corresponds to minimizing the norm

$$\|\mathbf{e}\|_2^2 = \left\|\mathbf{I}^o - \mathbf{M}'\mathbf{S}^i_{est}\right\|,$$

which is found by using the pseudo-inverse of the measurement matrix,

$$\mathbf{S}^i_{est} = \left(\mathbf{M}'^T\,\mathbf{M}'\right)^{-1}\mathbf{M}'^T\,\mathbf{I}^o.$$

14.7. Error Analysis

If measurement errors in $\mathbf{I}^o$ dominate, then the covariance matrix of the input Stokes vector is given by

$$\mathrm{cov}\left(\mathbf{S}^i_{est}\right) = \left(\mathbf{M}'^T\,\mathbf{M}'\right)^{-1}\sigma^2_{\mathbf{I}^o}.$$

In the current circumstances we are also interested in the limitations due to errors in the measurement matrix. In which case a small error in $\mathbf{S}^i$, written as $\Delta\mathbf{S}^i$, can be expressed in terms of errors in the measurement matrix, $\Delta\mathbf{M}'$, and errors in the observed flux $\Delta\mathbf{I}^o$ using

$$\mathbf{I}^o + \Delta\mathbf{I}^o = \left(\mathbf{M}' + \Delta\mathbf{M}'\right)\left(\mathbf{S}^i + \Delta\mathbf{S}^i\right).$$

The conventional linear algebra method to analyze $\mathrm{cov}\left(\mathbf{S}^i_{est}\right)$ would be to form the matrix *p*-norms. Matrix norms are only defined for square matrices, and therefore this approach is only useful for $N = 4$. Inequalities relating norms can then be used to investigate the case where errors in the measurement matrix are not insignificant, e.g.,

$$\frac{\left\|\Delta\mathbf{S}^i\right\|}{\left\|\mathbf{S}^i\right\|} \le \left(\frac{\kappa}{1-\kappa\left\|\Delta\mathbf{M}'\right\|/\mathbf{M}'}\right)\left(\frac{\left\|\Delta\mathbf{I}^o\right\|}{\left\|\mathbf{I}^o\right\|} + \frac{\left\|\Delta\mathbf{M}'\right\|}{\left\|\mathbf{M}'\right\|}\right),$$

where $\kappa = \left\|\mathbf{M}'\right\|\left\|\mathbf{M}'^{-1}\right\|$ is the condition number for the measurement matrix. This expression yields an upper limit on the error $\left\|\Delta\mathbf{S}^i\right\|$. The term in the second set of parenthesis shows that he error in the estimate of the incident Stokes vector is proportional to the error of the detected flux vector and the error in the measurement matrix. Both these terms need to be minimized to minimize the error in the retrieved Stokes parameters. In addition, the first factor in parenthesis must be minimized. When the condition number is large the sensitivity to these errors is large. Conversely, when the condition number is small, the system sensitivity to errors is reduced. The magnitude of the condition number of the measurement matrix depends on the angles chosen: careful choice of this set of angles can decrease the sensitivity to errors by factors of ~ 2-3.

14.8. Methodology & Example: Single-Channel Polarimeter

The use of norms is attractive, because this analytic error analysis method yields a single number to consider. However, interpretation of the error norm, $\left\|\Delta \mathbf{S}^i\right\|$, is problematic and it cannot be extended to the non-square matrices that occur for $N > 4$. A more physical approach is to consider a figure of merit such the error in the degree of linear polarization, p, and the position angle, ψ. In Monte Carlo realizations we compare estimates of $\mathbf{S}^{\mathrm{i}}$ based on inaccurate knowledge of the measurement matrix **M**' with the true value of $\mathbf{S}^{\mathrm{i}}$. In our idealized example the fidelity of the reconstruction of $\mathbf{S}^{\mathrm{i}}$ depends on the errors in the position angle of the wave plate, θ, and its differential phase retardance of ϕ. We first construct an input Stokes, $\mathbf{S}^{\mathrm{i}}$, vectors based on the signals typical of likely astronomical sources. This is converted into output intensities $\mathbf{I}^o$ using the actual value of the measurement matrix, where **M**' includes errors due to uncertainty in the orientation of the wave plate and uncertainties in the wave plate retardance. The output intensities $\mathbf{I}^o$ are then converted back into an estimate of $\mathbf{S}^{\mathrm{i}}$ using the pseudo-inverse of the ideal measurement matrix. The difference between the true and estimates values of $\mathbf{S}^{\mathrm{i}}$ then provides a measure of the impact of errors in **M**'. For each realization we generate a set of N angle errors $\delta\theta_1$, $\delta\theta_2$, $\delta\theta_3$, …, $\delta\theta_{\mathrm{N}}$, and one value for the error in the retardance, $\delta\phi$. The errors $\delta\theta_{\mathrm{j}}$ simulate a lack of precision in the rotation mechanism as it scans through the N positions, while $\delta\phi$ represents our uncertainty in characterizing the retardation properties of the wave plate. Some results are shown in Figure 14-3 and Figure 14-4.

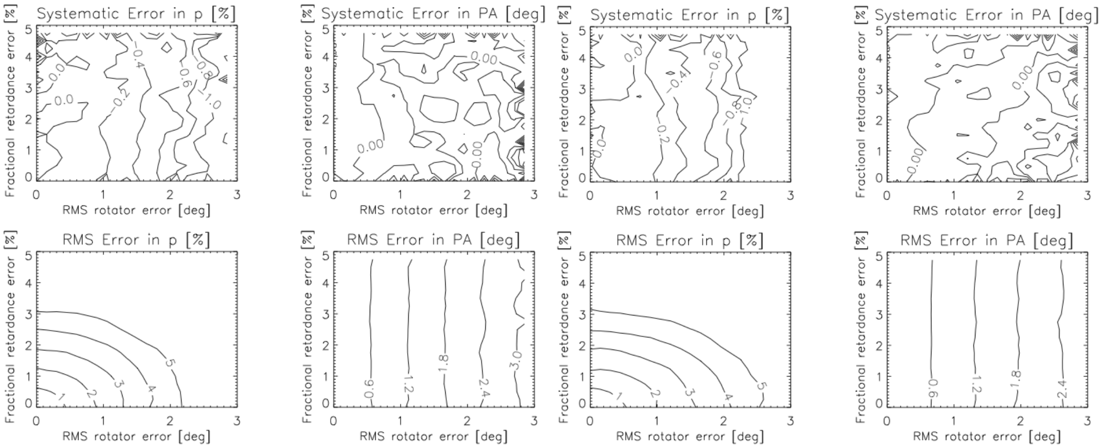

Figure 14-3: ***Left:*** Systematic and random errors in retrieved fractional linear polarization and PA for a quarter-wave plate for $N = 5$ with $\theta = 0, 36, 72, 108$, and 144°. The x-axis is the random RMS error in positioning the wave plate, and the y-axis is the RMS uncertainty in measured the retardance of the wave plate. The source population has fractional linear polarization between 0 and 0.5 distributed uniformly. The PA is distributed uniformly between –90° and 90°. ***Right:*** Same as the left panel, except that $N = 7$ with $\theta = 0, 25.71, 51.43, 77.14, 102.86, 128.57$, and 154.28°.

It is evident from these simulations that quarter wave and half wave retarders have distinctly different error properties if degree of linear polarization and PA constitute the performance metric. For a quarter-wave plate the degree of linear polarization is sensitive to the accuracy of the calibration of the wave plate phase retardance, whereas measurements of fractional linear polarization are robust against uncertainties in the properties of the half-wave plate. In both cases the results show only a $\sqrt{N}$ improvement with the number of wave plate angles, suggesting that the least squares approach is more robust against encountering ill-conditioned problems than the classical approach that calls for inversion of a square measurement matrix. Figure 14-3 also confirms one's intuition that the rms error in the PA of the wave plate rotator translates directly in an error in the inferred PA of polarization.

Comparison of Figure 14-3 and Figure 14-4 suggests that a half-wave plate is the best choice for measuring linear polarization in debris disks. To achieve 1% *accuracy* in p with a half-wave plate the wave plate retardance must be known to better than 5%. To achieve a *precision* of better than 1% in p the rotator mechanism must have an rms positional uncertainty of ≤ 0.5°. This analysis suggests that measurement of PA of polarization with good precision (< 0.5°) in requires that the wave plate be positioned with a precision of better than 0.5°.

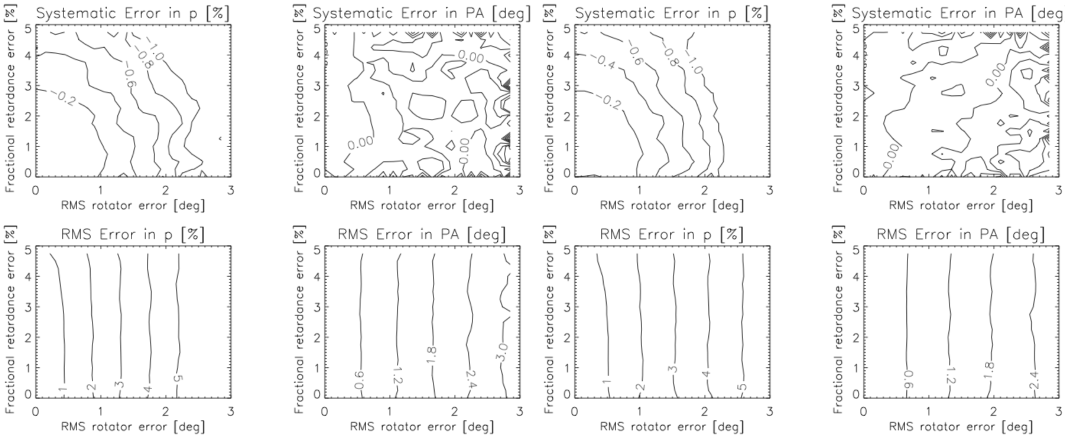

Figure 14-4: ***Left:*** Systematic and random errors in retrieved fractional linear polarization and PA for a half-wave plate for $N = 5$ with $\theta = 0, 36, 72, 108$, and 144°. ***Right:*** Same as the left hand panel, except that $N = 7$ with $\theta = 0, 25.71, 51.43, 77.14, 102.86, 128.57, 154.28$°. A half-wave plate is superior to a quarter-wave plate for achieving high precision measurement of the degree of linear polarization.

Inspection of the expression for a single-channel polarimeter explains why certain choices of ϕ lead to favorable error propagation. For example, for a wave plate that yields approximately a half-wave of retardance, $\phi = \pi + \delta\phi$, the Taylor series for the measurement expression is,

$$\begin{aligned} 2I^O(\theta_j,\phi) = {} & I^i + \\ & \left[\cos^2 2\theta_j - \left(1-\delta\phi^2/2\right)\sin^2 2\theta_j\right]Q^i + \\ & 2\left(1-\delta\phi^2/4\right)\cos 2\theta_j \sin 2\theta U^i + \\ & -\delta\phi \sin 2\theta_j V^i + O\left(\delta\phi^4\right) \end{aligned} .$$

If the circular polarization of the source is small, and $V^i \approx 0$, then errors in the phase retardance of the wave plate only enter to second order.

14.9. A Dual Channel Polarimeter

Suppose that the analyzer is a perfect polarizing beam splitter, and that both the e- and o-rays can be detected simultaneously. The Mueller matrix for the o-channel analyzer is

$$\mathbf{M}_{LP^o} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

and the Mueller matrix for the e-channel is

$$\mathbf{M}_{LP^e} = \frac{1}{2}\begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The corresponding o- and e-intensities are

$$\begin{aligned} 2I^{Oo}\left(\theta_j,\phi\right) = I^i + & \\ & \left(\cos^2 2\theta_j + \sin^2 2\theta_j \cos\phi\right)Q^i + \\ & \left(\cos 2\theta_j \sin 2\theta\left(1-\cos\phi\right)\right)U^i + \\ & \sin 2\theta_j \sin\phi V^i \end{aligned}$$

and

$$\begin{aligned} 2I^{Oe}\left(\theta_j,\phi\right) = I^i - & \\ & \left(\cos^2 2\theta_j + \sin^2 2\theta_j \cos\phi\right)Q^i - \\ & \cos 2\theta_j \sin 2\theta\left(1-\cos\phi\right)U^i - \\ & \sin 2\theta_j \sin\phi V^i \end{aligned}.$$

The primary purpose of the dual channel polarimeter is to make the system robust against uncorrected wavefront errors, which will show up as speckles in both channels. If we form the observables $I^+ = I^{Oo} + I^{Oe}$, and $I^+ = I^{Oo} - I^{Oe}$ unpolarized speckles are removed from the estimate of the polarized flux:

$$I^+\left(\theta_j,\phi\right) = I^{Oo}\left(\theta_j,\phi\right) + I^{Oe}\left(\theta_j,\phi\right) = I^i$$

and

$$
\begin{aligned}
I^{-}(\theta_j,\phi) = I^{Oo}(\theta_j,\phi) - I^{Oe}(\theta_j,\phi) = \; & 0I^i + \\
& \left(\cos^2 2\theta_j + \sin^2 2\theta_j \cos\phi\right)Q^i + \\
& \cos 2\theta_j \sin 2\theta_j (1-\cos\phi)U^i + \\
& \sin 2\theta_j \sin\phi V^i
\end{aligned}
$$

Evidently, $\boldsymbol{I}^{+}$ is instantaneous measurement of the Stokes I^i arising from the uncorrected starlight and any contribution from a circumstellar disk, although of course this estimate is subject to the instantaneous PSF. In contrast, in the ideal case the term I^- includes no contribution from I^i and hence only polarized, scattered disk light is present in this image. The I^- image will be contaminated by non-common path errors between the o- and e-channels of the polarimeter. The degree to which I^i cancels to zero in I^- depends on how accurately $p^o{}_x = 1$ and $p^o{}_y = 0$ in the o-channel and $p^e{}_x = 0$ and $p^e{}_y = 1$ in the polarizing beam splitter, and the flat field accuracy. As extinction ratios, $p_y{}^2/p_x{}^2$, in polarizing beam splitters are typically $< 10^{-4}$ it is likely that flat fielding errors will be the limiting factor.

A dual channel system does not eliminate systematic errors in the Stokes parameters due uncertainty in the wave plate position angle or the retardance of the wave plate. Is also does not eliminate sensitivity to transparency variations. Inspection of the expressions for I^o for a single-channel polarimeter and I^+ and I^- for a dual-channel polarimeter show the same dependence on the wave plate position angle and the wave plate retardance. Therefore the dual channel design does not relax any of the specifications on these angles. If the wave plate rotation is accurate, but imprecise, observation of I^+ and I^- at a large number of position angles can be used to average out his noise source.

An instantaneous measurement of I^i is made at every wave plate position, but in general I^- is a linear combination of Q^i and U^i. However, if transparency or Strehl variations are a concern then Q^i/I^i or U^i/I^i can be measured instantaneously by selecting favorable wave plate orientations and then combined to measure the degree of linear polarization,

$$p = \sqrt{(Q/I)^2 + (U/I)^2} \, .$$

However, because the position angle, $\psi = \frac{1}{2}\tan^{-1} U/Q$, depends on U/Q, this angle remains uncertain. For example, if a half-wave plate is employed then at $\sin 2\theta = 0$ or $\cos 2\theta = 0$ I^- measures only Q^i and $\sin^2 2\theta = 1/2$ measures only U^i. For this to be an effective strategy the wave plate rotator and retardance must be both precise and accurate.

14.9.1. Dual–channel with double differencing

It is convenient to operate a dual-channel polarimeter in a mode that collects data in pairs in which the half-wave plate is rotated by π/4 (45°) between exposures. This is useful because rotating a half-wave plate by this amount appears to rotate the plane of linear polarization by π/2 (90°) with respect to the analyzer. In the example shown in Figure 14-5 rotation of the wave plate exchanges +*Q* and –*Q* between the o- and e-channels.

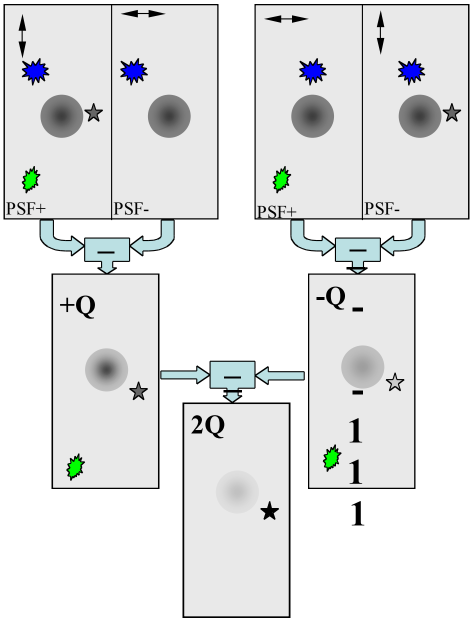

Figure 14-5: Illustration of a double differencing with an dual-channel polarimeter with perfect π retardance. In this example, the modulator is a half-wave plate that is rotated exactly by π/4, which has the effect of swapping +*Q* and –*Q* (indicated by the horizontal and vertical double headed arrows) between the o- and e-channels of the analyzer. The technique suppresses speckles due to common path wavefront errors (blue speckle) and non-common path errors (green speckle). In this example the uncorrected seeing halo is unpolarized, while the astrophysical signal (grey star) is polarized). Errors in wave plate retardance enter as second order, but finite instrumental polarization enters to first order, allowing I^i to leak into measurements of +*Q*,and –*Q*, i.e., the cancellation of the unpolarized halo in the second row is not perfect. If the instrumental polarization occurs upstream of the wave plate then this term varies with wave plate angle, leading to a residual in the double difference.

The double difference is defined as

$$I^{dd-}(\theta_j,\Delta\theta,\phi)=\left[I^{Oo}(\theta_j,\phi)-I^{Oe}(\theta_j,\phi)\right]-\left[I^{Oo}(\theta_j+\Delta\theta,\phi)-I^{Oe}(\theta_j+\Delta\theta,\phi)\right].$$

With reference to Figure 14-5, note that any double difference scheme cancels unpolarized speckles arising from both common and non-common path errors. The specific example shown in Figure 14-5, which uses a half-wave plate leads to an estimate of *Q*. Different choices of ϕ and $\Delta\theta$ will not change this, but in general the result is come linear combination of Stokes parameters that can be extracted using the least squares approach.

To investigate the sensitivity of a dual channel polarimeter to retardance errors in the wave plate, suppose that we construct the double difference with $\phi = \pi + \delta\phi$ and $\Delta\theta = \pi/4$. Thus, the double difference for a half-wave wave plate is

$$\begin{aligned} I^{dd-}\left(\theta_j,\pi/4,\phi\right) = {} & 0I^i + \\ & 2\cos 4\theta_j\left(1-\delta\phi^2/4\right)Q^i + \\ & 2\sin 4\theta_j\left(1-\delta\phi^2/4\right)U^i + \\ & \left(\cos 2\theta - \sin 2\theta\right)\delta\phi V^i \end{aligned} .$$

The sum of all four channels gives $I^{dd+} = 2I^i$. As asserted above a dual channel system used in double differencing mode does not eliminate systematic errors in the Stokes parameters due uncertainty in the wave plate position angle or the retardance of the wave plate. However, the errors are second order in the error in the retardance when a half-wave plate is used.

14.10. A dual-channel polarimeter with instrumental polarization

If the polarimeter introduces linear polarization, i.e., $p_{Ix}^2 \neq p_{Iy}^2$, then a spurious signal proportional to I^i leaks into I^-, which otherwise is sensitive to only to polarized light. Instrumental polarization therefore compromises the ability to cancel the unpolarized halo of uncorrected light from the star (see Figure 14-5).

In this prescription, the orientation of the instrumental linear polarization is fixed in either the +*Q* or –*Q* directions. More generally, instrumental polarization could be oriented at an arbitrary position angle. However, the results of this section apply independent of the orientation of the instrumental polarization. Later we will present the results of a simulation of the effects of arbitrary instrumental polarization, including arbitrary phase retardance. For now ignore instrumental circular polarization. Suppose that the instrumental polarization occurs prior of the wave plate, the system Mueller matrix is $\mathbf{M}_{SYS}(\theta,\delta) = \mathbf{M}_{LP}\mathbf{M}_{ROT}(-\theta)\mathbf{M}_{WP}(\phi)\mathbf{M}_{ROT}(\theta)\mathbf{M}_I$. This circumstance would occur if the wave plate were located just ahead of the polarizing beam splitter. If the instrumental polarization is in the form of linear polarization, then

$$\begin{aligned} I^+\left(\theta_j,\phi\right) = I^{Oo}\left(\theta_j,\phi\right) + I^{Oe}\left(\theta_j,\phi\right) = {} & \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)I^i + \\ & \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)Q^i + \\ & 0U^i + \\ & 0V^i \end{aligned} ,$$

and

$$\begin{aligned} I^-\left(\theta_j,\phi\right) = I^{Oo}\left(\theta_j,\phi\right) - I^{Oe}\left(\theta_j,\phi\right) = {} & \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)\left(\cos^2 2\theta_j + \sin^2 2\theta_j \cos\phi\right)I^i + \\ & \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)\left(\cos^2 2\theta_j + \sin^2 2\theta_j \cos\phi\right)Q^i + \\ & p_{Ix}p_{Iy}\cos 2\theta_j \sin 2\theta\left(1-\cos\phi\right)U^i + \\ & p_{Ix}p_{Iy}\sin 2\theta_j \sin\phi V^i \end{aligned} .$$

In an ideal instrument $p_{Ix} = p_{Iy} = 1$. To the extent that instrumental polarization is zero I^- contains no contribution from the unpolarized, uncorrected halo. However, the I^- signal is modulated when there is instrumental polarization. Moving the wave plate further upstream in the optical train, so that it is placed before the bulk of the instrumental polarization, will reduce this modulation.

To see how moving the wave plate reduces the modulation of I^- suppose that the wave plate is located upstream of the instrumental linear polarization and the system Mueller matrix is $\mathbf{M}_{SYS}(\theta,\delta) = \mathbf{M}_{LP}\mathbf{M}_{I}\mathbf{M}_{ROT}(-\theta)\mathbf{M}_{WP}(\phi)\mathbf{M}_{ROT}(\theta)$. For example, if the wave plate were mounted at the entrance aperture of the instrument. Hence, for a dual channel polarimeter, with the wave plate mounted ahead of any polarizing optics, e.g., oblique reflections, we have

$$\begin{aligned} I^+(\theta_j,\phi) = I^{Oo}(\theta_j,\phi) + I^{Oe}(\theta_j,\phi) = {} & \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)I^i + \\ & \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)\left(\cos^2 2\theta_j + \sin^2 2\theta_j \cos\phi\right)Q^i + \\ & \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)\cos 2\theta_j \sin 2\theta(1-\cos\phi)U^i + \\ & \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)\sin 2\theta_j \sin\phi V^i \end{aligned}$$

and

$$\begin{aligned} I^-(\theta_j,\phi) = I^{Oo}(\theta_j,\phi) - I^{Oe}(\theta_j,\phi) = {} & \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)I^i + \\ & \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)\left(\cos^2 2\theta_j + \sin^2 2\theta_j \cos\phi\right)Q^i + \\ & \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)\cos 2\theta_j \sin 2\theta(1-\cos\phi)U^i + \\ & \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)\sin 2\theta_j \sin\phi V^i \end{aligned}$$

Comparison of these two cases shows that the residual signal in I^- due to leakage of I^i is the same order of magnitude in both cases. If we assume that in integrated light the fractional linear polarization is << 1 then $I^- \approx \frac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)I^i$. However, when the rotating wave plate is located at the entrance aperture to the instrument the factor $\frac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)I^i$ contributes equally to the measured signal at all wave plate positions.

In the case where the rotating wave plate is located immediately prior to the polarizing beam splitter, and after the instrumental polarization has been accumulated, the signal is modulated according to

$$I^- \approx \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)\left(\cos^2 2\theta_j + \sin^2 2\theta_j \cos\phi\right)I^i.$$

Thus in either case the instrumental linear polarization term $\left(p_{Ix}^2 - p_{Iy}^2\right)/2$ enters like a differential throughput or flat-fielding factor, but the location of the wave plate determines whether or not this term is modulated. Ideally, we require $p_{Ix} = p_{Iy} = 1$. However, it is likely that the instrumental polarization will be greater than a few percent, unless extraordinary steps are taken to minimize it. Moreover, and more fatally, I^i and Q^i are modulated by the same term:

$$\cos^2 2\theta_j + \sin^2 2\theta_j \cos\phi.$$

Therefore, I^{i} and Q^{i} are completely covariant and cannot be disentangled by any combination of retardance or modulator position angle. As a consequence, the instrumental polarization must be measured separately, and the covariance between the instrumental polarization and the astrophysical polarization must be considered when evaluating the systematic errors that occur when there is significant instrumental polarization downstream of the modulator.

Not surprisingly, linear polarization that arises in the source or induced along the line of sight ahead of the modulator, i.e., interstellar or instrumental polarization, cannot be distinguished. Sidereal rotation changes the projection of the Stokes vectors in the source relative to the instrument, however probably not on times scales fast enough to be of practical use to suppress the uncorrected seeing halo.

As a consequence it is important to understand how to measure p_{Ix} and p_{Iy} (and the relative throughput of the o- and e-channels) with sufficient accuracy to characterize the leakage term. If significant instrumental polarization is present before the wave plate, and the intensity I^{i} contribution to Γ is modulated, then it is important to know the wave plate retardance and its angle with good accuracy and precision.

These results have important consequences for understanding the operation in double differencing mode: if all the instrumental polarization is accumulated after the wave plate, then the contribution of I^{i} to Γ is independent of the wave plate angle, and therefore cancels in a double difference. Assuming that the AO correction is constant, then the unpolarized halo can be subtracted to degree possible by Poisson statistics and detector noise. If uncorrected tilt-tilt errors are present then flat fielding and interpolation errors may also be important.

Although instrumental polarization most likely also includes retardance, we have neglected this term in this analysis. Unknown instrumental retardance introduces errors in measuring Q, U and V, but does not modify the (1, 1) element of the system Mueller matrix. Therefore, instrumental retardance does not influence how effectively the seeing halo is suppressed.

14.10.1. Measuring the degree of polarization

To what extent are systematic errors due to the instrumental polarization canceled or suppressed in a dual channel system? Suppose that the dual channel is configured with a perfect half-wave plate, and we can ignore errors in the wave plate rotator mechanism. Consider the most favorable situation where the modulator is located upstream of the instrumental polarization. In this case

$$\begin{aligned} I^{+}\left(\theta_j, \pi\right) = {} & \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right) I^i + \\ & \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)\cos 4\theta_j Q^i + \\ & \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)\sin 4\theta_j U^i + \\ & 0V^i \end{aligned},$$

and

$$\begin{aligned} I^-(\theta_j,\pi) = {} & \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)I^i + \\ & \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)\cos 4\theta_j Q^i + \\ & \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)\sin 4\theta_j U^i + \\ & 0V^i \end{aligned}$$

By choosing $\theta = 0, \pi/4, \ldots$ we can measure Q^i, or by choosing $\theta = \pi/8, 3\pi/8, \ldots$ we can measure U^i, i.e.,

$$\begin{aligned} I^+(0,\pi) &= \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)I^i + \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)Q^i \\ I^+(\pi/8,\pi) &= \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)I^i + \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)U^i \\ I^-(0,\pi) &= \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)I^i + \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)Q^i \\ I^-(\pi/8,\pi) &= \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)I^i + \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)U^i \end{aligned}.$$

Evidently, this set of four equations has five unknowns, and the instrumental polarization cannot be eliminated. The introduction of additional wave plate angles makes is possible to solve for simultaneously solve for the I^i, Q^i and U^i, and the instrumental polarization. However, the problem is a non-linear one.

Even though it is not possible to make an estimate of the polarized flux, i.e., $P = \sqrt{Q^2 + U^2}$, which would be valuable to separate the unpolarized residual seeing halo from light scattered by circumstellar dust. An alternate approach is to try and extract the degree of linear polarization $p^2 = \left(Q^2 + U^2\right)/I^2$. In this case the values of the instrumental polarization are irrelevant, rather what we care about is the extinction ratio, $k = p_{Iy}^2 / p_{Ix}^2$. If we express I^+ and I^- in arbitrary flux units our expressions become

$$\begin{aligned} I_0^+ &= (1+k)I^i + (1-k)Q^i \\ I_{\pi/8}^+ &= (1+k)I^i + (1-k)U^i \\ I_0^- &= (1-k)I^i + (1+k)Q^i \\ I_{\pi/8}^- &= (1-k)I^i + (1+k)U^i \end{aligned}$$

then

$$p^2 = \frac{\left[I_0^+\left(I_{\pi/8}^+ - I_0^+\right) + I_0^-\left(I_0^- - I_{\pi/8}^-\right)\right]^2 + \left[I_{\pi/8}^+\left(I_{\pi/8}^+ - I_0^+\right) + I_{\pi/8}^-\left(I_0^- - I_{\pi/8}^-\right)\right]^2}{\left(I_0^- I_{\pi/8}^+ - I_{\pi/8}^- I_0^+\right)^2},$$

and the extinction ratio is

$$k = \frac{I_0^- - I_0^+ + I_{\pi/8}^+ - I_{\pi/8}^-}{I_0^+ + I_0^- - I_{\pi/8}^+ - I_{\pi/8}^-}.$$

Thus, for an ideal half-wave plate system, which is insensitive, to V^i and with zero instrumental retardance the fractional polarization can be extracted directly. The degree of linear polarization contains an unknown fraction of uncorrected light from the star.

14.10.2. Double-difference with instrumental linear polarization

To complete this section consider in full the double differencing scheme explicitly so that we can establish limits on instrumental polarization upstream of the modulator. We consider only linear polarization upstream of the modular because this term causes the intensity of the unpolarized, uncorrected seeing halo of the star to be modulated in brightness, and causes single and hence double differences not to cancel to zero. Linear polarization in the instrument between the modulator and the analyzer leads to errors in the Stokes parameters, but we ignore this calibration error in considering the detection problem.

The double sum signal is

$$\begin{aligned} I^{dd+}\left(\theta_j,\Delta\theta,\phi\right) = {} & \left(p_{UIx}^2 + p_{UIy}^2\right)I^i + \\ & \left(p_{UIx}^2 - p_{UIy}^2\right)Q^i + \\ & 0U^i + \\ & 0V^i \end{aligned},$$

and the double difference is

$$\begin{aligned} I^{dd-}\left(\theta_j,\Delta\theta,\phi\right) = {} & \tfrac{1}{2}\left(p_{UIx}^2 - p_{UIy}^2\right)\left[\cos^2\left(2\theta_j + 2\Delta\theta\right) - \cos^2\left(2\theta_j\right)\right]\left(1-\cos\phi\right)I^i \\ & \tfrac{1}{2}\left(p_{UIx}^2 + p_{UIy}^2\right)\left[\cos^2\left(2\theta_j + 2\Delta\theta\right) - \cos^2\left(2\theta_j\right)\right]\left(1-\cos\phi\right)Q^i + \\ & p_{UIx}p_{UIy}\left[\cos\left(2\theta_j + 2\Delta\theta\right)\sin\left(2\theta_j + 2\Delta\theta\right) - \cos 2\theta_j \sin 2\theta_j\right]\left(1-\cos\phi\right)U^i + \\ & p_{UIx}p_{UIy}\left[\sin\left(2\theta_j + 2\Delta\theta\right) - \sin 2\theta_j\right]\sin\phi V^i \end{aligned}.$$

The degree of instrumental polarization, $1 - p_{UIy}^2 / p_{UIx}^2$, the error in the measured value of the instrumental polarization, and errors in positioning the wave plate and knowledge of its retardance all influence the degree to which the unpolarized halo can be suppressed. The matrix element $M_{1,1}$ corresponding to the modulation factor in the first line of this expression for I^{dd-} is shown in for a half-wave plate and a angle increment of 45° between the difference sets.

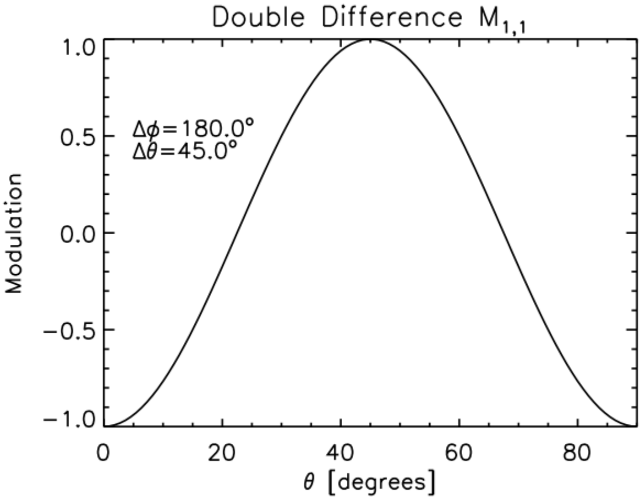

Figure 14-6: The modulation of system Mueller matrix $M_{1,1}$ for a double difference system versus wave plate angle for a half wave retardance. The angle increments is 45° between pairs. The angular modulation of $M_{1,1}$ and $M_{1,2}$ is identical.

Suppose that the instrumental polarization is modeled more generally as a reflection from a single metal surface with reflection coefficients, ρ_x^2 and ρ_y^2 and phase retardance of Δ. The retardance $\Delta=\pi$ radians at normal incidence on metal surface. Unlike the phase change which occurs for a dielectric, Δ falls smoothly through $\pi/2$ at the principal angle of incidence. The corresponding Mueller matrix describing instrumental polarization would be written as

$$\mathbf{M}_{IP} = \mathbf{M}_{ROT}(\theta_I)\frac{1}{2}\begin{pmatrix} \rho_x^2+\rho_y^2 & \rho_x^2-\rho_y^2 & 0 & 0 \\ \rho_x^2-\rho_y^2 & \rho_x^2+\rho_y^2 & 0 & 0 \\ 0 & 0 & 2\rho_x\rho_y\cos\Delta & -2\rho_x\rho_y\sin\Delta \\ 0 & 0 & 2\rho_x\rho_y\sin\Delta & 2\rho_x\rho_y\cos\Delta \end{pmatrix}\mathbf{M}_{ROT}(\theta_I),$$

where the instrumental polarization rotation angle θ_I allows for the possibility that the optics downstream of the instrumental polarization are rotated with respect to the coordinate system of the reflection. When using this formalism, note that the Mueller matrix for reflection implies a change in coordinate system that requires a change in sign of the angle for the rotation matrix.

The results of a full simulation, including measurement uncertainties in instrumental polarization are shown in Figure 14-7. In this example, a half-wave plate is assumed. The position angle and retardance of the wave plate are subject to nominal errors (0.5° is PA and 0.5° in retardance.) The instrumental polarization is allowed to vary, as is the precision with which it is known. The instrumental phase retardance is assumed to be exactly π. The source in this example has 1% linear polarization. Statistical fluctuation of 1% in I^- and I^+ are assumed (this could represent photon noise, flat fielding errors and any variation due Strehl fluctuations).

Systematic errors arise in the estimate of p because of the covariance between I^i and Q^i. These are shown on the left hand panel of Figure 14-7. Statistical errors for this example are shown on the right.

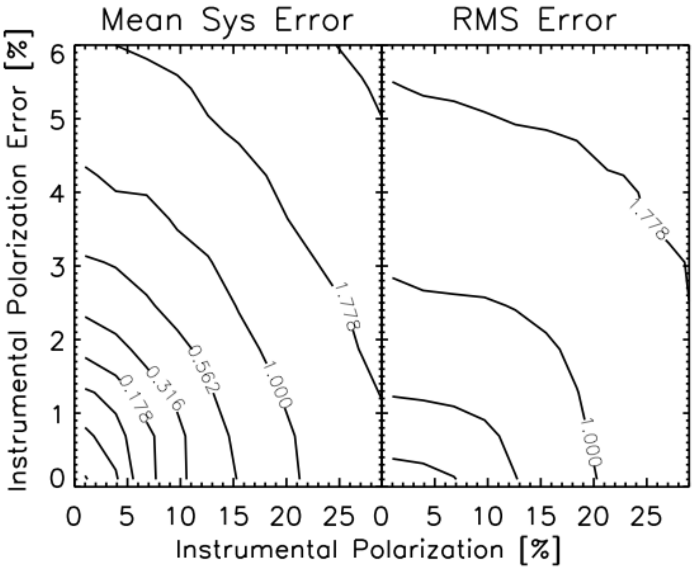

Figure 14-7: *Left*: Contours of percentage mean systematic error in polarized fraction in a dual channel polarimeter in the double difference as a function of instrumental polarization and the uncertainty in the instrumental polarization (both are in units of percentage linear polarization. ***Right:*** Contours of percentage rms error in polarized fraction. This example is for a half-wave plate. The polarimeter had nominal errors, i.e., 0.5° phase error in wave plate retardance and the rms error in the rotation of the wave plate is 0.5°. The wave plate is rotated to nine equally spaced position angle between 0 and π/2 radians and the increment for double differencing is π/4 radians. The instrumental polarization is oriented at a PA of 15° and in addition to the uncertainties plotted here the PA of the instrumental polarization is uncertain by 5°. The intensity is measured with a fractional error of 1% in each double difference and each cycle is repeated once. The input Stokes vector is **I** = [1, 0.00707, 0.00707, 0] or 1% linear at a PA of 22.5°.

Inspection of Figure 14-7 shows that systematic errors are comparable to statistical errors when the instrumental linear polarization is large, or when the measurement of the instrumental polarization has a large uncertainty. To reach the domain where unpolarized light is suppressed by a factor of 100 or more the systematic errors should be kept to a fraction of 1%. This requires that the instrumental linear polarization upstream of the wave plate should be less than 10%. At this level, the instrumental polarization need to me measured to a precision of ± 0.8%. If the instrumental polarization is 4%, it only needs to be measured ± 2%. In both these examples, it is assumed that the PA of the instrumental polarization is known to ± 5°.

Figure 14-15 shows the instrumental polarization that would accrue from five in-plane reflections from Al mirrors with Al_2O_3 oxide coatings for an angle of incidence of 45°. Such a configuration may be representative of the AO relay (although the angle of incidence may be exaggerated). The resultant linear polarization, about 4% at 1.5 µm would, in the light of these calculations be an acceptable degree of instrumental polarization if it can be measured to better than ± 2%

14.11. Measuring instrumental polarization

Consider first the downstream wave plate and a scheme for measuring p_{Ix} and p_{Iy} that involves observing a source of unpolarized source of light $\mathbf{I} = (I_c, 0, 0, 0)^T$. Because instrumental polarization occurs before the wave plate, rotating the wave plate modulates the instrumental signature. The calibration signal is

$$I^{+}(\theta_j,\phi) = \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right) I_c^i$$

and

$$I^{-}(\theta_j,\phi) = \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)\left(\cos^2 2\theta_j + \sin^2 2\theta_j \cos\phi\right) I_c^i .$$

Thus the calibration procedure to find p_{Ix} and p_{Iy} can be cast as:

$$\begin{pmatrix} I_1^+ \\ I_1^- \\ I_2^+ \\ I_2^- \\ \vdots \\ I_N^+ \\ I_N^- \end{pmatrix} = \tfrac{1}{2} I_c^i \begin{pmatrix} 1 & 0 \\ 0 & \cos^2 2\theta_1 + \sin^2 2\theta_1 \cos\phi \\ 1 & 0 \\ 0 & \cos^2 2\theta_2 + \sin^2 2\theta_2 \cos\phi \\ \vdots & \vdots \\ 1 & 0 \\ 0 & \cos^2 2\theta_N + \sin^2 2\theta_N \cos\phi \end{pmatrix} \begin{pmatrix} p_{Ix}^2 + p_{Iy}^2 \\ p_{Ix}^2 - p_{Iy}^2 \end{pmatrix} .$$

The results of an experiment using half or quarter-wave plates to measure instrumental polarization by observing an unpolarized source and rotating the wave plate are show in Figure 14-8. In each case it is assumed that the wave plate rotator is set at $N = 5$ position angles per measurement, and the rms error in setting the wave plate is given as the *x*-axis of the plots. It is also assumed that the wave plate retardance only known to some precision, which is shown on the *y*-axis.

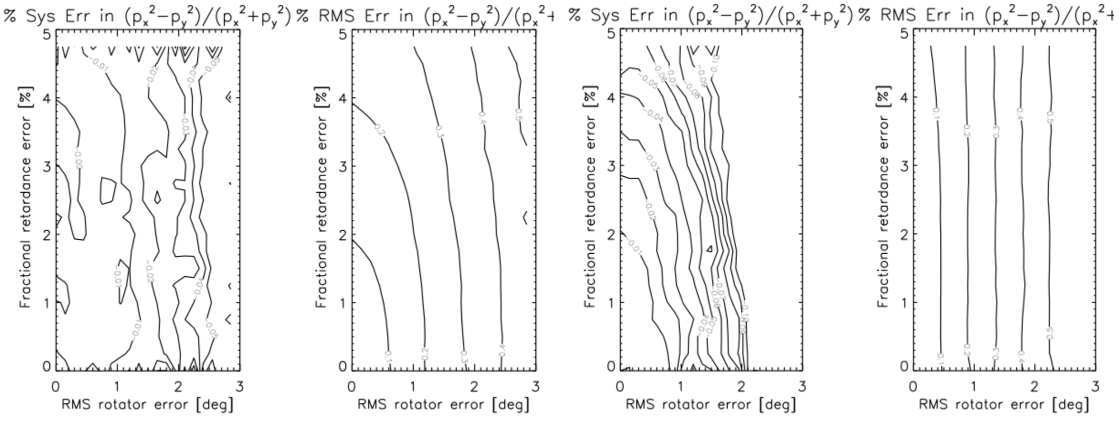

Figure 14-8: **Left:** Systematic and statistical errors in calibrating instrumental linear polarization with a down-stream quarter wave wave plate. This example is for $N = 5$ with $\theta = 0, 36, 72, 108$, and 144°. The instrumental polarization is 10%. The calibration source is assumed to be perfectly unpolarized and measurement errors in the flux are neglected. **Right:** Same for a down-stream half wave wave plate.

14.11.1.

Now consider an upstream wave plate and a scheme for measuring p_{Ix} and p_{Iy} that involves observing a source of unpolarized source of light $\mathbf{I_c} = (I_c, 0, 0, 0)^T$. The calibration signal is

$$I^+(\theta_j,\phi) = \tfrac{1}{2}\left(p_{Ix}^2 + p_{Iy}^2\right)I_c^i$$

and

$$I^-(\theta_j,\phi) = \tfrac{1}{2}\left(p_{Ix}^2 - p_{Iy}^2\right)I_c^i.$$

Thus the least-square formulation of the calibration procedure to find p_{Ix} and p_{Iy} can be written as as:

$$\begin{pmatrix} I_1^+ \\ I_1^- \\ I_2^+ \\ I_2^- \\ \vdots \\ I_N^+ \\ I_N^- \end{pmatrix} = \tfrac{1}{2}I_c^i \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p_{Ix}^2 + p_{Iy}^2 \\ p_{Ix}^2 - p_{Iy}^2 \end{pmatrix}.$$

Obviously, when the wave plate is positioned before the optical elements that introduce instrumental polarization, no modulation is introduced, and the signals are independent of the wave plate PA. Under these circumstances the errors in $\left(p_{Ix}^2 + p_{Iy}^2\right)$ and $\left(p_{Ix}^2 - p_{Iy}^2\right)$ are limited by the precision with which I_j^+ and I_j^- can be measured. As noted before, the polarization downstream of the wave plate is of less significance than the upstream polarization, as it is the instrumental linear polarization upstream of the wave plate that causes leakage of I^i into the double difference signal.

If the calibration source is completely unpolarized, $\mathbf{I_c} = (I_c, 0, 0, 0)^T$, then the measured signals are

$$I^+(\theta_j,\phi) = \tfrac{1}{2}\left(\rho_x^2 + \rho_y^2\right)I_c^i$$

and

$$\begin{aligned} I^-(\theta_j,\phi) = \tfrac{1}{2}\left(\rho_x^2 - \rho_y^2\right) & \\ & \left[\sin 2\theta_I(\cos\phi - 1)\sin 2\theta\cos 2\theta - \right. \\ & \left. \cos 2\theta_I(\cos\phi - 1)\cos^2 2\theta + \cos 2\theta_I\cos\phi\right]I_c^i \end{aligned} .$$

With a rotating wave plate we can measure ρ_x^2, ρ_y^2, and θ_I , but not Δ. To measure the retardance, we need a source of know polarization, e.g., a linear polarizer. For example, using a half-wave plate and a calibration source of known linear polarization $\mathbf{I_c} = (I_c, Q_c, U_c, 0)^T$ we have:

$$I^{+}\left(\theta_j,\phi\right)=\tfrac{1}{2}\left(\rho_x^2+\rho_y^2\right)I_c^i+\\ \tfrac{1}{2}\left(\rho_x^2-\rho_y^2\right)Q_c^i$$

and

$$\begin{aligned}I^{-}\left(\theta_j,\phi\right)=&\tfrac{1}{2}\left(\rho_x^2-\rho_y^2\right)\cos\left(2\theta_I+4\theta\right)I_c^i+\\ &\tfrac{1}{2}\left(\rho_x^2+\rho_y^2\right)\cos\left(2\theta_I+4\theta\right)Q_c^i+\\ &\tfrac{1}{2}\rho_x\rho_y\left[\sin\left(2\theta_I+4\theta+\Delta\right)+\sin\left(2\theta_I+4\theta-\Delta\right)\right]U_c^i\end{aligned}\quad .$$

14.12. Recommendations

- Wave plate stage
    - Precise and reliable angular motion with rms of ≤ 0.5°
    - 360° range in increments of ≤ 5°
    - Setting time ≤ detector readout time
- Wave plate location
    - The instrumental linear polarization upstream of the wave plate must be less than 10% and must stable and repeatable so that it can be measured to an rms error of 0.8% or better, i.e., the instrumental polarization is measured with a signal-to-noise ration of 12.5:1. If the instrumental polarization is lower then this requirement is relaxed according to Figure 14-7.
- Wave plates
    - Interchangeable half (π) and quarter wave (π/2) wave plates are needed for observing and calibration.
        - ♣ The primary function of the quarter wave plate is to quantify instrumental phase retardance. It might have relaxed wavefront error tolerances.
    - Wave plate retardance must be known to ≤ 0.5% in each operating band.
        - ♣ Absolute retardance of the half-wave plate must be to within ±10° of 180°so that errors in ϕ enter as second order: it is more important to know the retardance that it be exactly π.
- Calibration sources
    - Daytime source of unpolarized light (≤ 0.5% polarized) fills field of view
        - ♣ Must be photometrically stable on timescales required to accumulate a full polarization calibration set
    - Daytime source of polarized light of known Stokes vector, e.g., a high efficiency linear polarizer such as Polarcor™
    - Nighttime calibration can use bright stars
    - Ability to observe unocculted polarization standards
    - Twilight calibration of PA using Rayleigh scattered sky
- Instrumental phase retardance and instrumental circular polarization
    - Instrumental circular polarization should be < 10% to take advantage of the tolerance to retardance errors when using a half-wave plate

    - Polarization dependent phase errors due to reflections at dielectric coatings (AR coatings and dichroic beam splitters) must be quantified
        - Optics with proprietary dielectric coatings of unknown phase retardance must be in collimated space
    - Stress induced birefringence in coatings and transmissive components must be quantified
        - Induced and residual stresses in atmospheric dispersion compensation prisms and their coatings must be evaluated

14.13. Outstanding Issues

- Specify *total* instrumental linear polarization
    - The above figure only specifies the instrumental polarization *upstream* of the wave plate.
- Specify wavefront errors of polarizing components.

14.14. Additional Material on Calibration, Instrumental Polarization and Systematic Errors

14.14.1. Polarization Calibration using Scattered Sunlight

Several methods can be used to calibrate the polarization angle. A low-albedo asteroid at a phase angle that yields appreciable polarization is one option. For the degree of linear polarization to have reasonable values, the phase angle of the asteroid needs to be in the range 10 to 20 degrees. The electric vector is oriented parallel to the Sun-Asteroid-Earth plane and in the latter case it will be normal to this scattering plane.

Another obvious method is to observe the clear, twilight sky near the zenith. Sunlight is scattered predominantly in the earth’s atmosphere by Rayleigh scattering. This light is partially polarized and its polarization state of the radiation is straightforward to compute. The daytime sky, therefore constitutes a convenient polarization standard for calibrating an astronomical polarimeter. By symmetry any linear polarization from multiple scattering should on average out to zero, while the polarization angle of the singly scattered component is determined by the Sun's azimuth. Multiple scattered can arise from ocean or clouds and ocean waves or structure in cirrus can break this symmetry. Measurements of the plane of polarization of scattered sunlight at the AEOS telescope, Maui, HI are complicated by the site's location near a strongly reflecting ocean (Harrington, Kuhn, Whitman, 2006, PASP, 118, 845).

The electric field vectors of the polarized components are arranged in concentric circles about the original direction of propagation, as shown in Figure 14-9 (a). Figure 14-9 (b) represents a cross section of the spherical wave with the polarized electric field components tangent to the surface of the sphere.

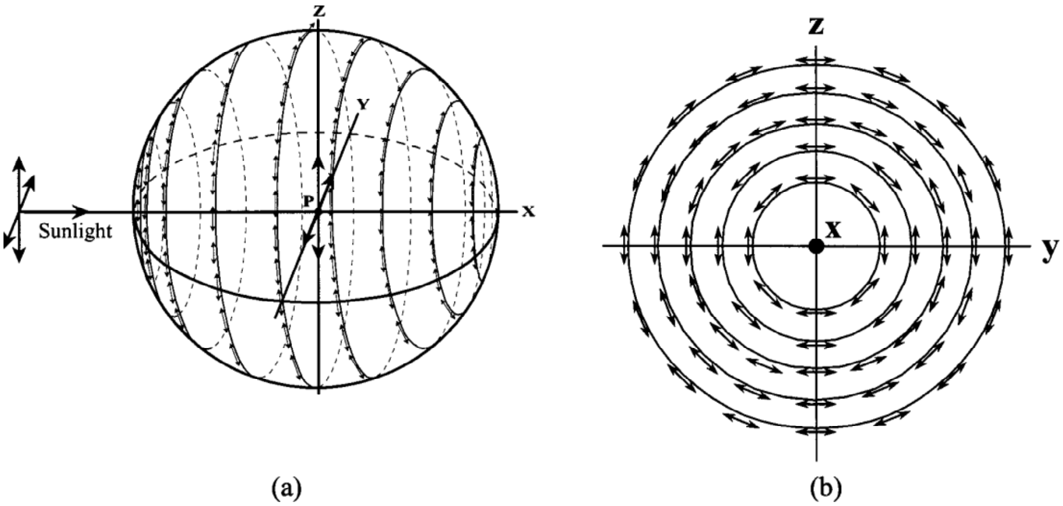

Figure 14-9: (a) Scattering produces linearly polarized skylight with polar axes arranged in concentric circles about the direction of propagation, and (b) a cross section of the spherical wave with the polarized electric field components tangent to the surface of the sphere (Matchko & Gerhart 2005).

14.14.2. Rayleigh scattering

The polarization of Rayleigh scattered sunlight is

$$p = \frac{\sin^2 \Omega}{1 + \cos^2 \Omega}$$

(Born & Wolf 1993). Figure 14-10 shows the scattering angle, Ω, in the observation plane containing the incident solar ray, scattering center, and observer.

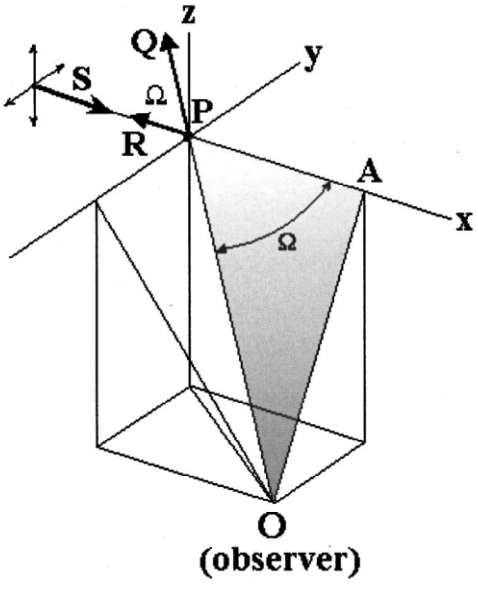

Figure 14-10: The observation plane *PAO* containing an incident solar ray *PA*, a scattering center *P*, the observer *O*, and the scattering angle Ω (Matchko & Gerhart 2005).

The scattering angle, Ω, can be found by computing the dot product RS•PQ expressed in the horizon coordinate system,

$$\cos\Omega = \cos\alpha\cos\theta\cos(A-\varphi) + \sin\alpha\sin\theta$$

where α is the altitude of the sun, A is the azimuth of the sun, θ is the altitude of the point of observation on the celestial sphere, and φ is the azimuth of the point of observation, where south corresponds to 0° and west to 90°.

14.14.3. Derivation of ψ

Figure 14-11 shows the geometry for deriving the polarization azimuth angle (Matchko & Gerhart 2005).

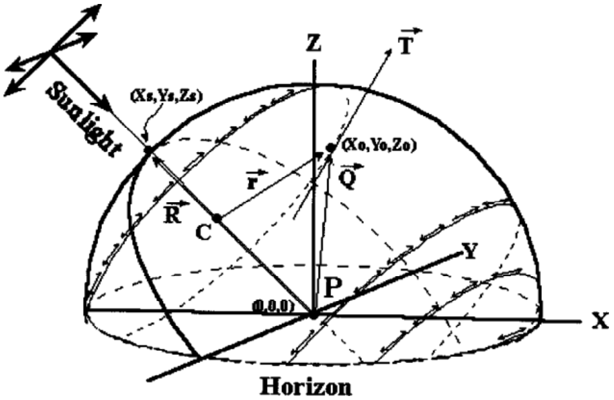

Figure 14-11: The geometry for finding the polarization azimuth angle for Rayleigh scattering (Matchko & Gerhart 2005).

The unit vector towards the sun is $\mathbf{R} = x_s\mathbf{i} + y_s\mathbf{j} + z_s\mathbf{k}$, where $x_s = \cos\alpha\ \sin A$, $y_s = \cos\alpha\ \cos A$, and $z_s = \sin\alpha$. A unit vector in the direction of observation is given by $\mathbf{Q} = x_0\mathbf{i} + y_0\mathbf{j} + z_0\mathbf{k}$, where $x_0 = \cos\theta\sin\varphi$, $y_0 = \cos\theta\cos\varphi$, and $z_0 = \sin\theta$.

Since a tangent, **T**, to the polarization circle passing through the observation point is perpendicular to **R** and **Q**, we have $\mathbf{T} = \mathbf{Q} \times \mathbf{R}$. A vector **N** perpendicular to the plane containing the z-axis and **Q** is given by $\mathbf{N} = \mathbf{Q} \times \mathbf{k}$. If **t** is a unit vector parallel to **T**, and **n** is a unit vector parallel to **N**, then $\mathbf{n} \cdot \mathbf{t} = \cos\psi$, where ψ is the angle between **n** and **t**. The angle ψ is the polarization azimuth angle observed at the observation point (x_0, y_0, z_0) along the direction of **Q**. The expression for ψ in the horizon coordinate system, is

$$\cos\psi = \frac{\sin\alpha\cos\theta - \sin\theta\cos\alpha\cos(\varphi - A)}{\pm\sqrt{1-[\sin\alpha\cos\theta + \sin\theta\cos\alpha\cos(\varphi - A)]^2}}\,.$$

14.14.4. Matrix Norms

Given a square complex or real matrix, $\mathbf{A}$, a matrix norm is a nonnegative number associated with having the properties: $\|\mathbf{A}\| > 0$ when $\mathbf{A} \neq 0$ and $\|\mathbf{A}\| = 0$ if and only if $\mathbf{A} = 0$; $\|k\mathbf{A}\| = |k|\|\mathbf{A}\|$ for any scalar k; $\|\mathbf{A}+\mathbf{B}\| \le \|\mathbf{A}\| + \|\mathbf{B}\|$; and $\|\mathbf{AB}\| \le \|\mathbf{A}\|\|\mathbf{B}\|$. The matrix p-norm is defined for $1 \le p \le \infty$. For example, the $p = 1$ norm

$$\|\mathbf{A}\|_1 = \max_j \sum_i |A_{ij}|,$$

which is the maximum absolute column sum norm, or the the $p = \infty$ norm

$$\|\mathbf{A}\|_\infty = \max_i \sum_j |A_{ij}|,$$

which is the maximum absolute row sum norm.

14.14.5. Instrumental Polarization

Suppose we model the instrumental polarization as reflection from a metal mirror with a dielectric coating, e.g., Al with a Al_2O_3 oxide coating. We can compute the reflectances R_s and R_p, and the phase retardance, Δ, from the standard relations for a thin dielectric film deposited on a metal surface (Born & Wolf 1980, § 13.4.2). The situation is shown in Figure 14-12. The dielectric layer has refractive index n_2 and thickness h. The metal has a refractive index n_3 with complex part k_3. In Figure 14-12 $\mathbf{E}_s$ (out of the page) and $\mathbf{E}_p$ indicate the TE and TM components of the incident electric field, respectively.

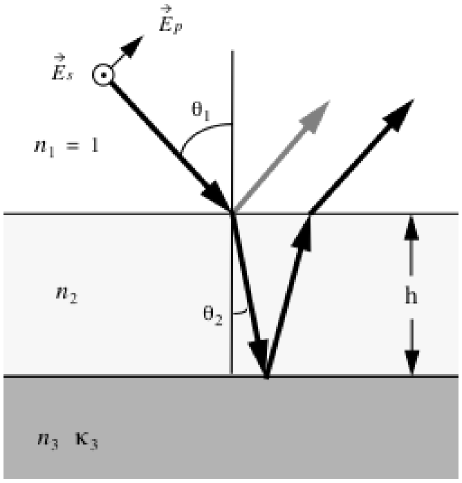

Figure 14-12: The electric field of the incident s-wave or TE component is denoted by $\mathbf{E}_s$ (coming out of the page), and $\mathbf{E}_p$ is the p-wave or TM component.

The ratio of the electric field amplitude for the reflected wave to that of the incident wave is

$$r = \frac{r_{12} + r_{23}\exp(i2\beta)}{1 + r_{12}r_{23}\exp(i2\beta)}.$$

Here r_{12} is the amplitude reflectance at the vacuum -dielectric interface, r_{23} is the amplitude reflectance at the dielectric-metal interface, and β is a phase factor related to the thickness of the dielectric layer. The intensity and amplitude reflection coefficients are related by $R = rr^*$, where the asterisk denotes complex conjugation.

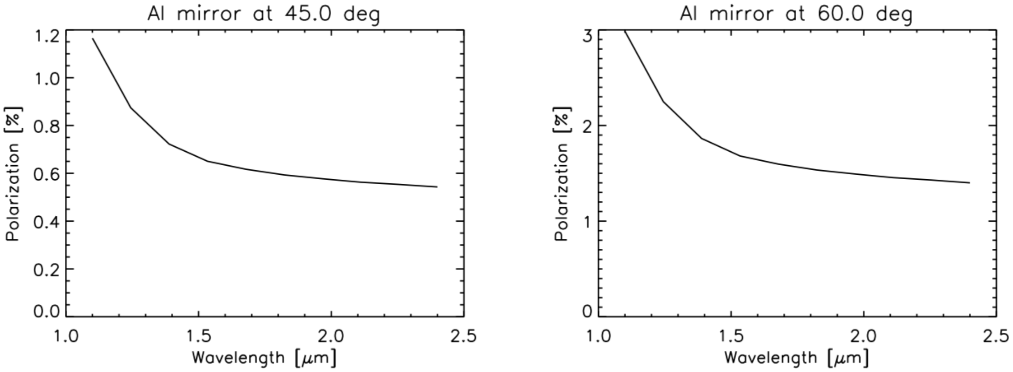

Figure 14-13: The linear polarization (M_{12}/M_{11}) versus wavelength due to reflection from a single Al surface with a 10 nm surface layer of Al_2O_3.

Consider the TE or s-wave. Since there is no absorption in the dielectric, r_{12} is real and is given by the Fresnel equation,

$$r_{12}^s = \frac{n_1\cos\theta_1 - n_2\cos\theta_2}{n_1\cos\theta_1 + n_2\cos\theta_2},$$

where n_1=1 for vacuum and n_2 is the refractive index for the dielectric. The angle of incidence of the incoming wave is θ_1 and θ_2 is given by Snell's law,

$$n_1\sin\theta_1 = n_2\sin\theta_2.$$

The metal surface is characterized by complex quantity $n_3(1 + i\kappa_3)$, where κ_3 is known as the absorption index,. The "extinction coefficient" is sometimes used instead of the absorption index, and is defined as $k_3 = n_3\kappa_3$.The amplitude reflectance of the dielectric-metal interface for the s-wave is given by the complex quantity

$$r_{23}^s = \frac{n_2\cos\theta_2 - (u + iv)}{n_2\cos\theta_2 + (u + iv)}$$

where

$$2u^2 = Q + \sqrt{Q^2 + 4n_3^4 + \kappa_3^2}$$

$$2v^2 = -Q + \sqrt{Q^2 + 4n_3^4 + \kappa_3^2}$$

and

$$Q = n_3^2\left(1 - \kappa_3^2\right) - n_2^2 \sin^2\theta_2.$$

For the TM or p wave the Fresnel equation is

$$r_{12}^p = \frac{\cos\theta_1/n_1 - \cos\theta_2/n_2}{\cos\theta_1/n_1 + \cos\theta_2/n_2}$$

and the equation for the reflectance at the dielectric surface is

$$r_{23}^p = \frac{\left[n_3^2\left(1-\kappa_3^2\right) + 2in_3^2\kappa_3\right]n_2\cos\theta_2 - (u+iv)}{\left[n_3^2\left(1-\kappa_3^2\right) + 2in_3^2\kappa_3\right]n_2\cos\theta_2 + (u+iv)}.$$

The phase factor is related to the thickness of the dielectric layer by

$$\beta = 2\pi n_2\left(h/\lambda_0\right)\cos\theta_2$$

where λ_0 is the wavelength in vacuum. The relative phase retardance, $\Delta = \phi_s - \phi_p$ is derived from

$$\tan\phi = \frac{\mathrm{Im}(r)}{\mathrm{Re}(r)}.$$

Figure 14-13 shows the instrumental polarization (M_{12}/M_{11}) using the optical data from Table 14-1 and Table 14-2. This example assumes a single Al surface for angles of incidence of 45° and 60°. The corresponding Mueller matrix elements are plotted in Figure 14-14. The combined effect of five reflections (the number of elements in the AO relay) is shown in Figure 14-15. In this example, all the angles of incidence are 45° and the reflections are all in the same plane.

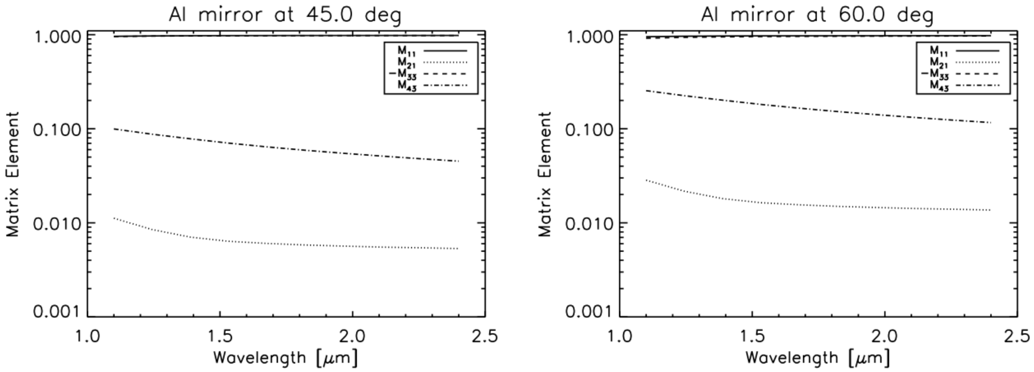

Figure 14-14: The Mueller matrix elements for reflection from a single Al surface with a 10 nm layer of Al_2O_3. A perfect mirror has M_{11} = 1, M_{21}=M_{12} = 0, M_{33} = -1 and M_{43} = -M_{34} = 0.

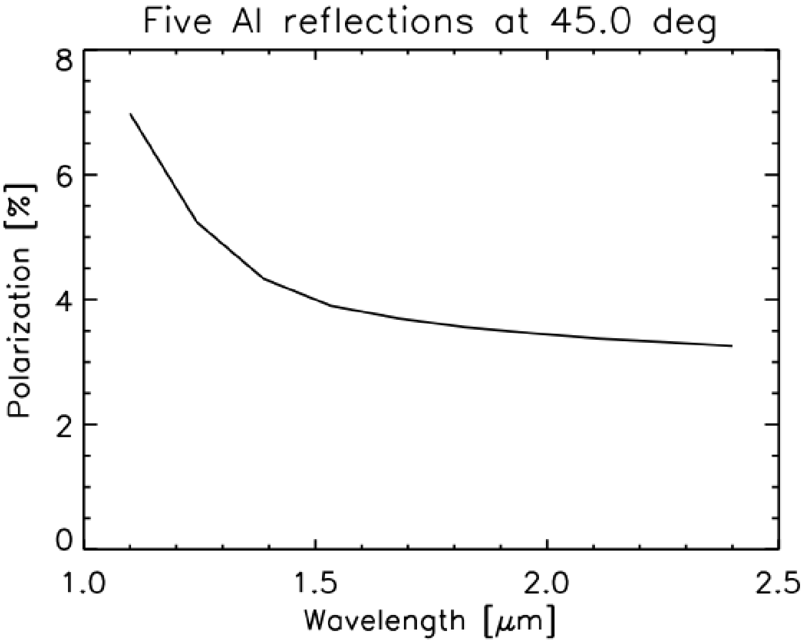

Figure 14-15: The instrumental polarization for a five in-plane Al reflections, each with an angle of incidence of 45°.

Table 14-1: Index of refraction & extinction coefficient for Al (Shiles et al. Phys. Rev. B 1980, 22, 1612).

Energy (eV)	Wavelength (µm)	*N*	*k*	**R**
0.5	2.48	3.072	25.581	0.9817
0.6	2.07	2.273	21.403	0.9806
0.7	1.77	1.770	18.328	0.9794
0.8	1.55	1.444	15.955	0.9778
0.9	1.38	1.264	14.021	0.9749
1.0	1.24	1.212	12.464	0.9697
1.1	1.13	1.201	11.181	0.9630
1.2	1.03	1.260	10.010	0.9521

Table 14-2: Refractive index of crystalline Al_2O_3

Wavelength (µm)	n_o	n_e
1.064	1.754	1.747
1.320	1.750	1.742
1.550	1.746	1.738
2.010	1.737	1.729
2.249	1.732	1.724
2.703	1.719	1.711

14.14.6. Polarization dependent aberrations

Polarization dependent aberrations are particularly pernicious. Conventional wavefront sensors and controllers (the deformable mirror) are not polarization sensitive, and a phase error that varies according to polarization state cannot be corrected. Two types of polarization dependent aberrations may be occur in GPI: 1) stress induced birefringence; 2) phase errors caused on reflection or transmission by dielectric beamsplitters.

The mechanical interface used to hold transmissive optical components, e.g., lenses and prisms, induces stress at the interface between the optical element and the mechanical assembly that can result in a change in the optical path. The magnitude of this effect depends on the stress-optical coefficients, or the "photoelastic constants". A change in the thermal environment of a transmissive optical component may alter the material properties and hence the mechanical stress.

Birefringence can be caused by inherent stress in lens material, which is a mechanical effect during the annealing process. Typical values of, *K*, the stress optical coefficient are $1\text{–}3 \times 10^{-6}$ mm^2 N^{-1} at 21°C, and at a wavelength of 589.3 nm (Sun & Edlou, 2006, Proc. SPIE, 6289, 14S).

When light is incident on an interface between media of different refractive index, part of the light is reflected and part of it is transmitted. If an optical system consists of several elements with high index of refraction or if the system has many interfaces, then the light that is reflected can be a significant problem both on the grounds that it is lost from the main beam, so that the image is reduced in intensity and also because some of the light that is reflected turns up at the final image plane in the wrong position. The reflection coefficient at normal incidence at a surface of refractive index n in air is given by $R = (n - 1)^2 / (n + 1)^2$. Typical optical refractive indices are between 1.45 and 4.0 and hence R = 3.4–36% for one surface. Applying an anti-reflection (AR) layer on a component, can achieve very low reflectivity down to $R < 0.05\%$. Dielectric coatings are likely to be employed in the atmospheric dispersion correcting prisms and in dichroic beamsplitters.

An important consequence of dielectric coating is the phase retardation between the s- and p-polarization states of both the reflected and transmitted waves. Although much effort is concentrated on optimizing transmission or reflection, depending on the nature of the component, often the phase retardation effects of a coating are neglected. Systems not in collimated space, especially those with high numerical apertures, have rays incident at large angles towards the edge of the clear aperture, and are more sensitive to coating phase retardation. An ideal dichroic beam splitter in a converging beam will therefore cause uncorrectable astigmatism. Inhomogeneities in the coating will cause higher order phase errors.

The system transmission is a product of transmission of each element including glass internal transmission and coating transmission. Transmissive phase retardation $\phi_s - \phi_p$, on the other hand, is a sum of individual phase retardation at a surface and at a specific pupil position, provided the lens elements are isotropic. Therefore, the phase retardations at different coating surfaces may have different signs, and multiple coatings can have a corrective controlling effect on phase retardation. This has important implications because it is possible to minimize the total transmissive phase retardation even if the retardation at each element is finite.

To explore this effect consider a multilayer periodic dielectric coating that might be employed as a dichroic beam splitter between light sent to the wavefront sensor and the science light. Suppose the refectivitiy is optimized for 600 nm and consists of two $\lambda/4$ layers of $n_2 = 2.3$, $n_3 = 1.25$ repeated four times between vacuum, $n_1 = 1$, and a substrate n_l=1.52. The reflection and transmission of such a coating at normal incidence is listed in Table 14-3 using the formalism of Born & Wolf §1.6. No effort has been made to optimize the transmission of the near science infrared light.

Table 14-3: Transmission and refection properties of a simple multilayer dielectric beam splitter consisting of four quarter wave (150 nm) high ($n_2 = 2.3$) and low ($n_3 = 1.25$) index layers.

Wavelength (µm)	0.60	1.25	1.65	2.2
Reflectivity (%)	96	4	15	4
Transmission (%)	4	96	85	96

The corresponding phase shifts for reflected light at 600 nm and transmitted light at 1.25, 1.65 and 2.20 µm are shown in the upper panel of Figure 14-16. The phase shift is converted into a physical wavefront error and plotted in nm. In reflection that phase shift is close to value of π, i.e., 300 nm. In this example the phase changes only very slowly with angle of incidence. The lower part of panel of Figure 14-16 shows the phase retardation for the transmitted wave. If there is an angular spread of a few degrees in the incident beam a differential phase corresponding to a few nm will accumulate between the s- and p-components.

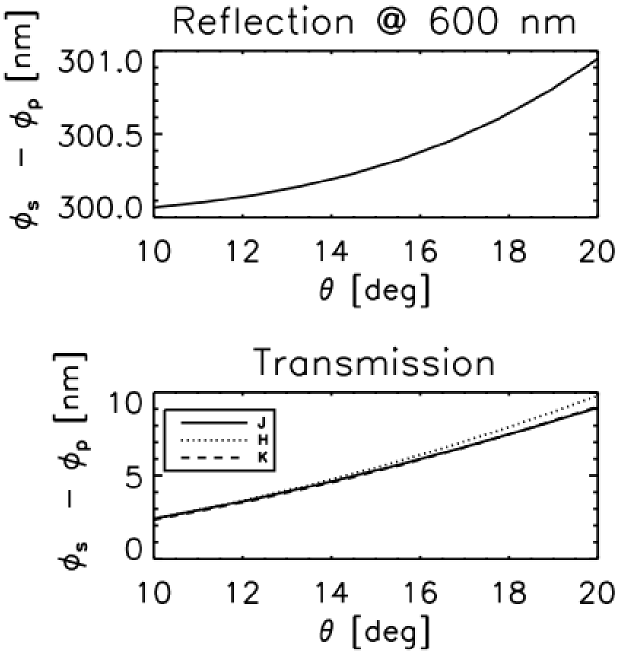

Figure 14-16: Plots of phase retardance between the s- and p-components of the electric field on reflection and transmission by the multilayer dielectric coating described in Table 14-3 as function of angle of incidence. In a collimated beam the phase retardance is constant. However, if an optic with a dielectric coating is located in a diverging or converging beam a polarization dependent aberration will occur that cannot be sensed and corrected by a conventional AO system.

Thin film coatings deposited at elevated temperatures may exhibit substantial residual stresses induced by growth strains and thermoelastic strains that develop during the cool-down period. A comprehensive description of these stresses includes not only the normal stresses in the film layers and the substrate but also the interfacial shearing stresses, which sometimes may cause delamination to occur.

15Appendix-System Mueller Matrix

James R. Graham (UCB), Jenny Atwood (DAO/HIA), & Kent Wallace (JPL)

v 1.3 April 26, 2007

15.1. Introduction

The system Mueller matrix is required understand the instrumental polarization and to explore approaches to calibrate the instrument in polarization mode. The system Mueller matrix must be known to establish that the design meets the requirements on instrumental linear polarization (conversion of I to Q and U) and phase retardance (conversion of Q and U to V).

Ultimately, the Mueller matrix must be know for the full range of operating wavelengths 0.96–2.4µm and for a representative number of field points. We present here a preliminary evaluation at the primary science wavelength (1.6 µm) and for the chief ray. This discussion will be refined prior to CDR. The range of field angles in GPI is small therefore we do not anticipate that strong field dependence of the polarization properties. However, it is inevitable that instrumental polarization and phase retardance will increase to shorter wavelengths, and poor performance is likely at wavelengths close to the crossover wavelength of the dichroic beam splitter. However, raising the level of detail in this calculation prior to a detailed design of the dichroic would be premature.

15.2. Evaluation of Mueller Matrix

The system Mueller matrix was computed using a combination of Zemax and custom MATLAB code. Zemax allows the user to propagate arbitrarily polarized light through any optical design. Visualizing the 4 × 4 Mueller matrix **M** as

$$\mathbf{M} = \begin{pmatrix} I \to I & Q \to I & U \to I & V \to I \\ I \to Q & Q \to Q & U \to Q & V \to Q \\ I \to U & Q \to U & U \to U & V \to U \\ I \to V & Q \to V & U \to V & V \to V \end{pmatrix},$$

it is evident that **M** can be found by illuminating the system with different types of polarized light. For example, if the input Stokes vector used in Zemax is $[1,0,0,0]^T$ (unpolarized light), then the output vector is just the first column of the Mueller matrix. Similarly if one computes the output for $[1,1,0,0]^T$ (100% horizontal linear polarization) and $[1,-1,0,0]^T$ (100% vertical linear polarization) then the difference is just twice the second column of **M**. To get column three compute the output for $[1,0,1,0]^T$ and $[1,0,-1,0]^T$ and take the difference and divide by two, and so forth.

15.3. Translation to Zemax

Zemax does not use Stokes vectors directly, but rather the magnitude and phase components of the Jones vector. Zemax's matrix conventions are as follows:[E_x, E_y, ϕ_x, ϕ_y]. Thus, [1,0,0,0] = linear horizontal polarization, [-1,0,0,0] = linear vertical polarization, [1,1,0,0] = linear polarization at +45°, [-1,1,0,0] =

linear polarization at –45°, [1,1,90,0] = right hand circular polarization (ϕ_x - ϕ_y = 90°), and [1,1,-90,0] = left hand circular polarization. Conversion to the Stokes vector is accomplished using the standard definition

$$\begin{aligned} I &= E_x^2 + E_y^2 \\ Q &= E_x^2 - E_y^2 \\ U &= 2E_x E_y \cos(\phi_x - \phi_y) \\ V &= 2E_x E_y \sin(\phi_x - \phi_y) \end{aligned}.$$

15.4. Implementation & Results

The Fresnel *s*- and *p*- reflection coefficients and the phase retardance of each surface was computed in Zemax assuming gold coatings on all mirrors, single layer MgF_2 on all transmissive optics, except for the dichroic beamsplitter which was implemented as a simple “cold mirror” consisting of four quarter wave (150 nm) high (n_2 = 2.3) and low (n_3 = 1.25) index layers. The reflection and transmission of such a coating at normal incidence is listed in Table 14-3 using the formalism of Born & Wolf §1.6. No effort has been made to optimize the transmission of the near science infrared light.

Table 15-1: Transmission and refection properties of a simple multilayer dielectric beam splitter consisting of four quarter wave (150 nm) high (n_2 = 2.3) and low (n_3 = 1.25) index layers.

Wavelength (µm)	0.60	1.25	1.65	2.2
Reflectivity (%)	96	4	15	4
Transmission (%)	4	96	85	96

The Zemax recipe uses the latest models for the AO relay (Rev. # 0.5), coronagraph (Rev. # 0.2), and back end (post occulter) (Rev. # 0.4). Two versions of the calculation are of interest: 1) one that includes the telescope mirrors, M1, M2 and M3, representing GPI located on a side looking port of the instrumental support module; 2) one that includes only telescope mirrors M1 and M2, representing GPI located on the up-looking port of the instrumental support module.

For (1) we find

$$\mathbf{M} = \begin{pmatrix} 0.5263 & 0.0078 & 0.0006 & 0.0000 \\ 0.0078 & 0.5263 & -0.0001 & 0.0063 \\ 0.0006 & 0.0012 & 0.5182 & -0.0920 \\ 0.0000 & -0.0062 & 0.0920 & 0.5181 \end{pmatrix},$$

or in normalized form

$$\mathbf{M}^{(N)} = \begin{pmatrix} 1 & 0.0148 & 0.0011 & 0.0000 \\ 0.0148 & 1 & -0.0019 & 0.0120 \\ 0.0011 & 0.0023 & 0.9846 & -0.1748 \\ 0.0000 & -0.0118 & 0.1748 & 0.9844 \end{pmatrix}$$

The corresponding instrumental linear polarization, $p = \sqrt{Q^2 + U^2} \big/ I$ = 1.5%. We note that the value of matrix m_{43} is significantly non-zero, and the eccentricity angle $\chi = \sin^{-1}(V/I)/2 \approx 0.088$ rad. (5°) for vertically polarized light and the corresponding eccentricity $e = \sqrt{1 - \tan^2 \chi}$ = 0.996 (see Figure 15-1).

For case (2) see Graham (2008)

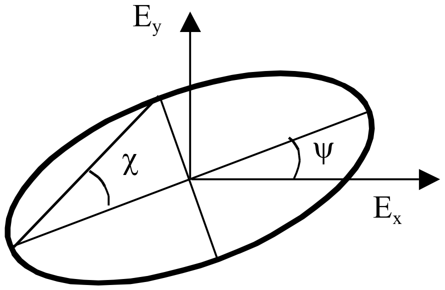

Figure 15-1: The polarization ellipse showing the eccentricity angle χ and position angle ψ. If *a* and *b* are the semi major and semi minor axis of the ellipse, then tanχ = *b*/*a*. For linearly polarized light χ = 0 and for circularly polarized light $\chi = \pm\pi/4$.

15.4.1. References

Born, M., & Wolf, E. 1998, Princples of Optics, Sixth Edition, Cambridge
Graham, J. R., 2008, PASP, *submitted*
Harrington, D. M., Kuhn, J. R., and Whitman, K. 2006, PASP, 118, 845
Matchko, R. M. & Gerhart, G. R. 2005, Opt. Eng, 44, 028001