

Coronagraphy with HST/NICMOS: detectability is a sensitive issue

Glenn Schneider^{*}, Murray Silverstone
Steward Observatory, University of Arizona

ABSTRACT

HST provides an unparalleled venue for high contrast imaging which has enabled new observational domains in exoplanet and debris disk imaging. Unburdened by atmospheric "seeing", NICMOS and STIS achieve very low levels of background contamination from the wings of stellar point spread functions (PSF). Coronagraphy provides additional contrast gains approaching an order of magnitude at small angular distances from occulted stars. The stability of the platform allows scattered and diffracted light to be further reduced by two additional orders of magnitude through PSF-subtraction. The non-destructive read-out modes of the NICMOS detectors permit sampling the PSF, with its strong radial brightness gradient, over a dynamic range exceeding 5×10^7 in a single spacecraft orbit. In H-band, sub-stellar companions of $\Delta H \approx 8 + 2 \times$ (angular separation in arcseconds) are unambiguously detected in twenty minutes of integration (e.g., a 10 Myr "hot Jupiter" at 2.5 arcseconds from an occulted star). Raw *sensitivity* metrics, such as presumptively static Strehl ratios, are often invoked in comparing the performance of different instrumental systems but belie the true *detectability* levels which are dominated by systemic non-repeatable PSF variations (not photon statistics). Such variations can give rise to false detections of companions (and circumstellar disks) and introduce very significant photometric errors. The ability to rotate the HST field with high precision about the target axis and acquire temporally stable reference PSFs readily permits the identification and rejection of rotationally-invariant optical artifacts. We discuss the repeatable, quantifiable performance limits routinely reached by HST (currently unachievable on ground-based systems), for which PSF stability is critical.

Keywords: Coronagraphy, NICMOS, Hubble Space Telescope, sensitivity, detectability

1. INTRODUCTION

In less than a decade the question of the existence of exo-planetary systems has been put to rest as we have moved from the realm of speculation to census-taking. Exo-planet detection, by inference through radial velocity variations of their host stars¹, has matured to the level where we expect the announcement of newly identified systems on a regular basis. However, the direct imaging of planetary companions has remained an elusive, though by no means insurmountable, challenge. Recent years have seen great advances in high contrast imaging capabilities with new technologies, instrumental systems, and observing platforms on both space- and ground-based telescopes which will undoubtedly answer this challenge in the very near future². Such capabilities have been effectively demonstrated with the Hubble Space Telescope's (HST) Near Infrared Camera and Multi-Object Spectrometer (NICMOS), which has now been returned to service after a more-than three year hiatus³.

HST provides a unique venue for high contrast imaging through the combination of diffraction limited imaging at optical and near-infrared wavelengths (with Strehl ratios panchromatically exceeding 98%) and a highly stable opto/mechanical platform, resulting in a very high degree of temporal invariability in the structure of the PSF sampled by the instrument detectors. These characteristics have been exploited advantageously by HST's second generation instruments which incorporate coronagraphic optics. While neither the NICMOS or STIS (Space Telescope Imaging Spectrograph) coronagraphs are fully optimized, both provide significant gains in the suppression of instrumentally scattered and diffracted starlight in the wings of an occulted target's PSF. The temporal repeatability of the coronagraphic PSF routinely permits self-subtractions of the underlying stellar signature (importantly the spatial frequency components comparable to an unocculted PSF core) in high-contrast fields with an efficacy exceeding the capabilities of current ground-based adaptive optics (AO) augmented systems. This capability is enabled, in part, by HST's pointing control system which allows for highly precise intra-orbit field rotations about coronagraphically occulted targets (typically with target pointings repeatable to better than 10 milliarcseconds), permitting ready

^{*} gschneider@as.arizona.edu; phone 1 520 621-5865; fax 1 520 621-1891; <http://nicmosis.as.arizona.edu:8000/>; Steward Observatory, 933 N. Cherry Ave., University of Arizona, Tucson, AZ, USA 85721

discrimination between optical artifacts and features of celestial origin. Such discrimination is critically for exo-planet detection.

2. SENSITIVITY vs. DETECTABILITY

In assessing the effectiveness of an instrumental system for companion detection, the traditional sensitivity measures often adopted significantly underestimate the limiting effects of non-random systematics. In doing so, the assumption is implicitly made that the "noise" (e.g., pixel-to-pixel variance or dispersion) of the uncompensated background is both random and spatially uncorrelated. With such an assumption, one expects an improvement in sensitivity (i.e., signal-plus-noise to noise ratio) with the square root of the integration time under a fixed photometric extraction aperture. This is not the case, however, when instrumental artifacts are induced over multiple pixel scales, and is conflated when such artifacts have characteristic spatial frequencies on similar scales as the "signal" which is being sought. None-the-less, such simple metrics are frequently invoked in order to compare different instrumental systems, and inevitably fall short of delivering a true representation of what one might expect in practice by overestimating the detection capabilities of a system under evaluation. Moreover, many ground-based AO-augmented systems suffer from significant non-repeatabilities, driven by uncompensated time-variable atmospheric turbidity, which significantly limits performance. Hence, even an intrinsically well defined and *prima facie* unambiguous metric such as an optical system's "typical" Strehl ratio may not be a good predictor of performance when the Strehl ratio is temporally unstable.

Our experience with HST differential coronagraphic imaging data, which are intrinsically more stable than ground-based AO systems can currently produce, indicate that it is critically important to divorce raw *sensitivity* measures, in the conventional sense, from a more germane *detectability* metric. Reliance on noise statistics without considering the characteristic spatial frequencies of the background variations, the sources to detect, and how both are sampled and measured will yield unreliable (and often overly optimistic) results. In the presence of both temporal instabilities (non-repeatabilities) and non-random noise sources we suggest it is necessary to implant and recover high-fidelity models of "artificial stars" into observed data to truly understand the detection limits of any system, and the temporal instabilities of such limits themselves.

3. NICMOS CORONAGRAPHY

The ability to detect faint point sources near bright objects is enhanced by instrumental reduction of the brightness of the central star through the use of a coronagraph. The NICMOS coronagraph reduces the instrumentally scattered and diffracted background by factors of ~ 10 at edge of the coronagraphic hole to ~ 4 at a radius of $\sim 1''$ where the per-pixel intensity in H-band is $\sim 10^{-4}$ of the peak intensity without the coronagraph⁴. Differential (i.e., PSF-subtracted) coronagraphic imaging with HST/NICMOS consistently results in the diminution of background diffracted and instrumentally scattered circumstellar light on the order of 10^6 per pixel at an angular distance of $1''$ from an occulted target in H-band, and 10^5 area-integrated over a $2''$ – $3''$ radius annulus at $1.1\mu\text{m}$ (e.g., important for imaging circumstellar disks)⁵. The unocculted wings of the target PSF remain fixed on the detector when the spacecraft/telescope is rotated about the target in the coronagraph. The sub-orbital stability of the HST+NICMOS optical system is such that additional background reduction factor in the unocculted wings of the target PSF of ~ 50 is realized by self-subtraction of differentially rotated coronagraphic images.

Such performance levels preclude the direct imaging of planets in systems as old as (and similar to) our own, where even planets of Jovian mass would be fainter than their central stars by $\sim 10^{-9}$ and detected primarily by reflected starlight. However, young Jovian mass companions to "nearby" (i.e. within 100 pc) stars, at distances where they might form or migrate to through planet-planet interactions in a (possibly dissipated) circumstellar disk can be imaged. Young planets, still thermally emissive from the residual heat of their formation processes, are much more readily imaged in the near-infrared than at optical wavelengths. After formation, a planet's bolometric luminosity (L) will decline with time (t) as it cools⁶, roughly as $L \sim t^{-1.2}$. For a 10 million year old planet (generally accepted as sufficiently old for gas giants to form⁷) these detection limits correspond to a 1 Jupiter-mass object separated from its host star by $2''$. Planetary companions an order of magnitude younger (if they should exist) would then be detectable in the sub-Jovian mass range. Such candidates might be found, for example, in the newly-discovered young stellar associations in the southern hemisphere.

4. ASSESSING NICMOS CORONAGRAPHIC DETECTABILITY PERFORMANCE

During HST Cycle 7/7N (March-December 1998) the NICMOS instrument definition team (IDT) conducted a survey of young and nearby stars to search for close sub-stellar (i.e., brown dwarf and giant planet) companions. We have evaluated the effectiveness of the HST/NICMOS system for sub-stellar companion detection using a suite of fifty differential coronagraphic survey image pairs for target stars in the brightness range $4 < H < 11$. Each coronagraphically occulted target was observed at two field orientations, differing by 29.9° by rolling the spacecraft about the target axis[†] within a single visibility period, allowing ~ 24 minutes of total integration time for each target. NICMOS detections of close companions, however, are limited not by photon statistics but rather by the systematics due to small-scale structural variations in the HST+NICMOS PSF. Thus, while the integration time (and sampling strategy employed) sets the sensitivity floor (to about $H = 23$) in the absence of background light, this limit is not directly scalable to assess the detectability of a close companion.

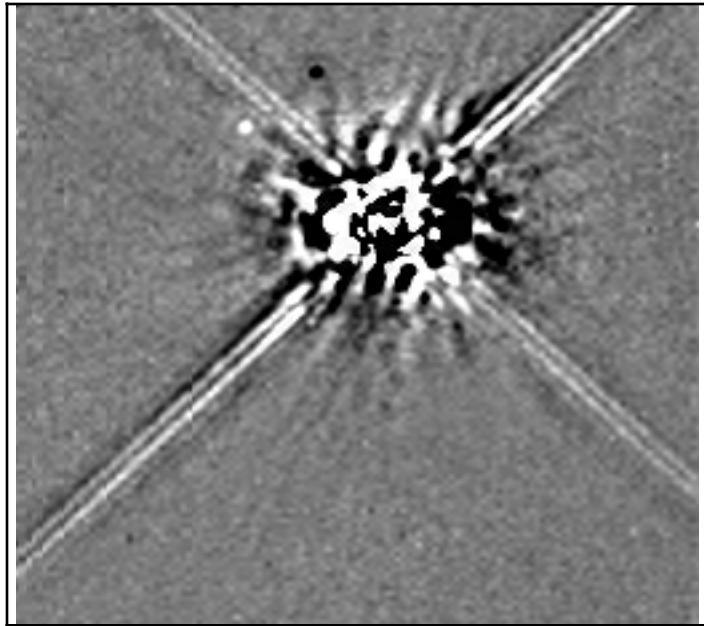


Figure 1. Coronagraphic difference image for the coronagraphic quadrant of the TWA6 field (stretch: $\pm 0.5\mu\text{mJy/pixel}$).

Optical artifacts induced by HST+NICMOS optical/coronagraphic system (which might bear some resemblance to an unresolved point source) remain spatially fixed on the coronagraphic detector when the spacecraft is re-oriented. Such artifacts largely cancel when two such differentially-rolled images are registered at the occulted target point and subtracted. In addition to reducing the circumstellar glare from the occulted star, these difference images produce independent positive and negative image "conjugates" of real objects in the field. This adds a powerful constraint for artifact discrimination and object detection. To assert that a suspected object in a difference frame is likely a "real" unresolved source of celestial origin, we demand detection in both the positive and negative image conjugates of at least the same metrical level of statistical significance, require that they are both of equal photometric brightness, and that both have the same spatial brightness distribution as a point source. We further require that the image conjugates of a suspected object must lie along an arc centered on the occulted star whose angular length corresponds to the differential telescope rotation. Finally (though more subjectively), we require that the first Airy ring from a point source must appear in both the positive and negative conjugate images of a suspected companion. This requirement will be met when

[†]The coronagraphic channel in NICMOS is in camera 2 and provides an $f/45$ image at the 256×256 pixel detector. The field-of-view has a small ($\sim 0.9\%$) linear geometrical distortion resulting in $[X, Y]$ image scales (for the epoch of the IDT survey) of approximately $[76.2, 75.6]$ mas/pixel (i.e., a $19.49'' \times 19.33''$ FOV). The projection of the $0.3''$ radius coronagraphic hole was $[+73, -45]$ pixels, or $\sim [+5.56'', -3.40'']$, from the $[-X, +Y]$ corner of the FOV. Hence, the field coronagraphically surveyed was spatially asymmetric w.r.t. the occulted star. The re-orientation of the spacecraft around the target axis by 29.9° results in a 475 arcsec^2 surveyed area (as opposed to 377 arcsec^2), yielding an image overlap area of 280 arcsec^2 .

the statistical $(S+N)/N$ in the local region of the suspected companion, in an aperture enclosing the PSF core, reaches ~ 20 . Applying these imposed constraints greatly improves on both the rejection of artifacts and the detection of objects, beyond what would be implied strictly by optical/detector system sensitivities.

5. METRICAL ANALYSIS

We illustrate first by representative example, and then discuss our results for a large class of targets taken from the IDT survey. In Figure 1 we show a very typical NICMOS coronagraphic PSF-subtracted image, in this case for TWA6, which has an unresolved ($H=20.1$, $\Delta H=13.2$) object in the field at a distance of $2.5''$. The presence of such an object in close angular proximity to TWA6 serves well as a calibrator to cross-check our implantation and recovery of "artificial stars" into the observed field. In detail, we create a noiseless "template" PSF using TinyTim⁸, an on-orbit verified and calibrated NICMOS optical model. We then deploy rings of the template PSFs, flux scaled through the NICMOS on-orbit photometric calibration, around the location of the occulted and self-subtracted star varying both the flux scaling and radial distance (for example, see Figure 2).

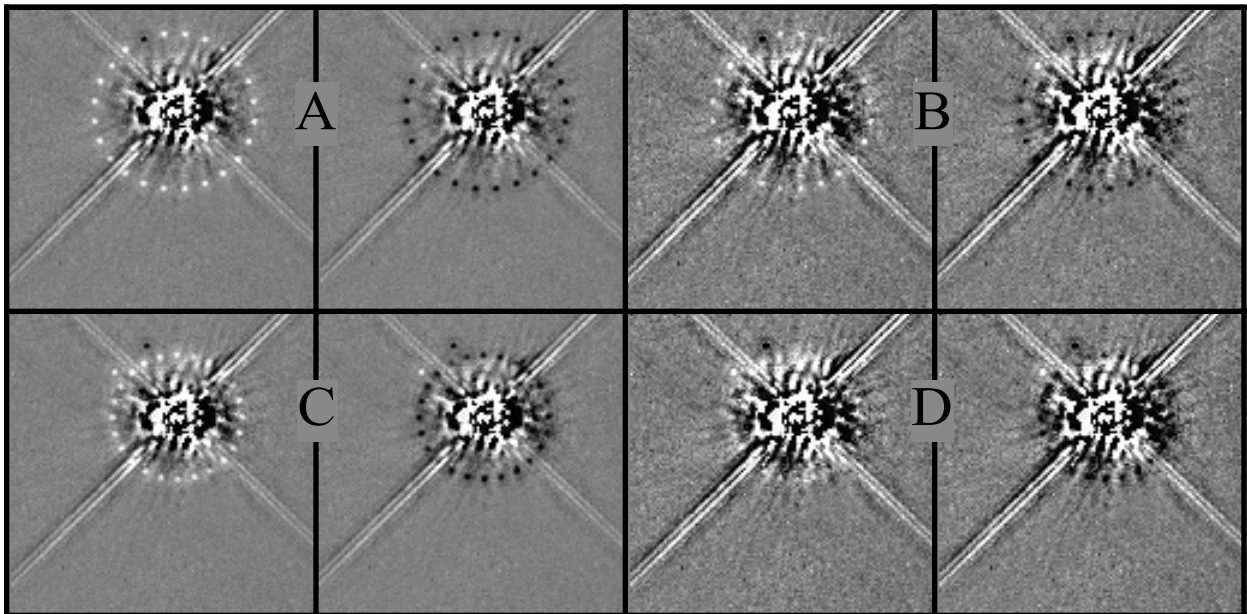


Figure 2. Positive and negative PSF implants in the TWA6 field. A & B: $r=2.5''$, C & D: $r=2''$, A&C: $\Delta H = 13.2$, B&D: $\Delta H = 14.2$.

We perform aperture photometry on each implant against the observed backgrounds, and of the pre-implanted background locations themselves, in a square aperture sized to fully enclose the PSF core (5×5 pixels for the F160W filter used in this survey). We measure both the statistical $(S+N)/N$ ratio (flux in implanted PSF core compared to the standard deviation of the background pixels) and how much flux was "recovered" from the PSF implants in the presence of the observed backgrounds (for example see Figure 3-left). At first glance there appears to be an inequity between the percentage of flux recovered and the statistical measure of the $(S+N)/N$ for corresponding samples. This actually bespeaks the fact that the presumption of random, uncorrelated noise, is invalid. The dispersion in the percentage of recovered flux does not come from measurement uncertainties due to small spatial scale (pixel-to-pixel, or PSF-scale) variations in the background, which affect the implant detectability. Rather, this dispersion arises from the uncertainty in establishing a true photometric background at each implant location due to much larger scale spatial residual gradients in the background after imperfect PSF subtraction. A star of a given flux density seen against a uniform, but non-zero, background would be as easily detectable if there was no background, though the amplitude of such a flux density offset would be imprecisely determined, particularly in the more realistic case of elevated (or depressed) backgrounds correlated over pixel scales comparable to a resolution element. This affects the uncertainty in the absolute photometry, but not an adopted metric for detectability. Typically, in the 5×5 pixel apertures used to assess the dispersion in the pixel-to-pixel variations on spatial scales of or smaller than the PSF, there are impressed much larger scale gradients which cause

"global" offsets from the nominally-nulled background levels. Azimuthally (at fixed radii), these offset levels vary such that using "neighboring regions" of comparable sizes poorly predicts the local backgrounds.

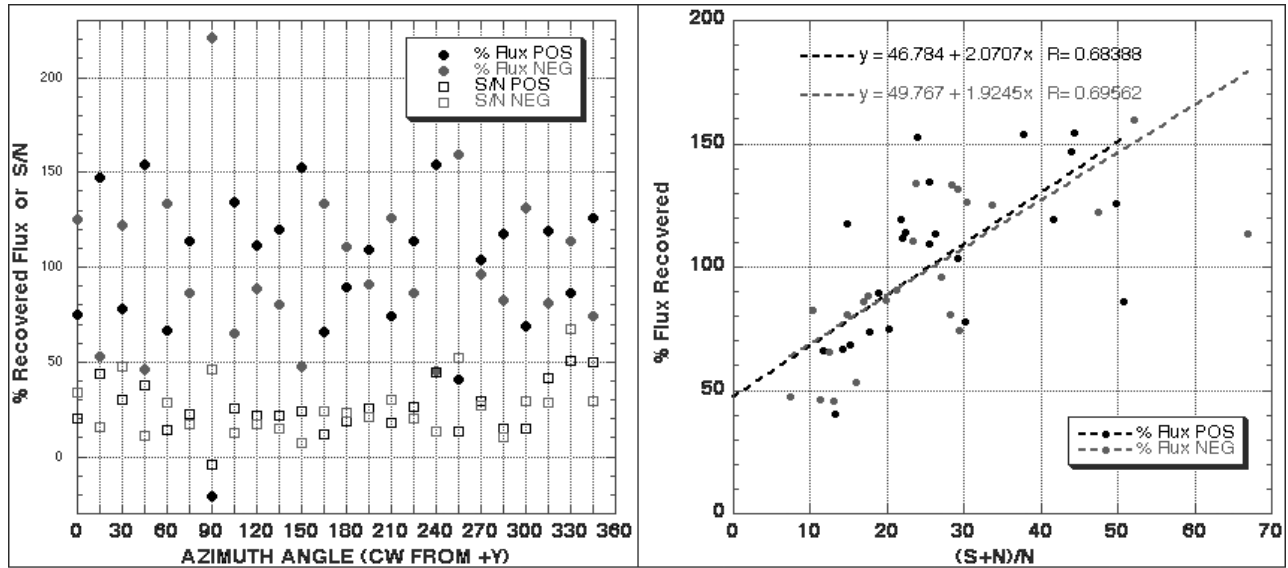


Figure 3. Azimuthal dependence (left) and (S+N)/N of recovered flux (right) for TWA6: $r=2.5''$, $\Delta H=13.2$ implants.

In this particular case (discarding the single highly discrepant point at an azimuth angle of 90°) the average flux density measured over all positive and negative implants at $2.5''$ is 105.3% and $94.8\% \pm 32\%$ (1σ) of their true value, respectively. This ~ 0.3 stellar magnitude overall scatter is appropriate for expressing the photometric uncertainty for a single observation of the target, and the systematic nature of this background uncertainty does not improve as rapidly as the square root of the number of independent measures.

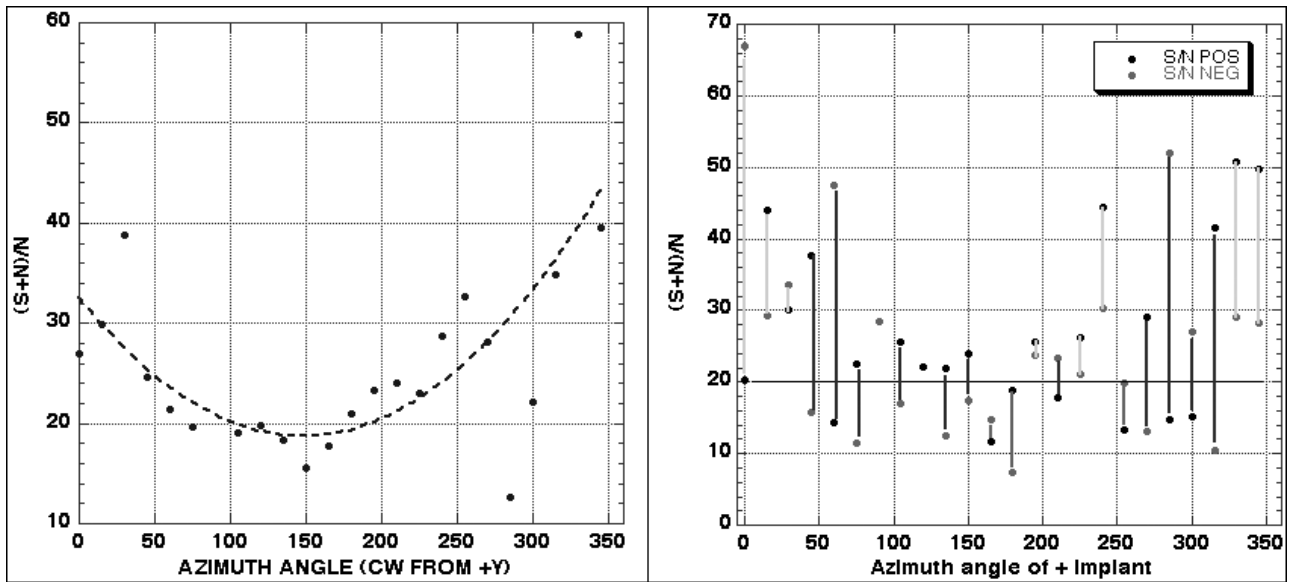


Figure 4. Left: (S+N)/N as a function of azimuth angle around the coronagraphic hole for the sample case considered (averaging positive and negative model PSF implants). Right: Similar plot for conjugate implant pairs at $r=1''$ and $\Delta H=11.5$.

As one might expect, the measured $(S+N)/N$ is correlated with the accuracy of the flux densities recovered from the implants. However, as shown in Figure 3-right, the flux densities are overestimated as the statistical $(S+N)/N$ increases. This is indicative of the fact that residual systematics dominate the measurement uncertainties (causing an overestimation in the total flux density) while the variance in the uncompensated zero-point offsets are not increasing in a Poissonian manner.

This behavior suggests that we might relate the $(S+N)/N$ in this non-random regime to a detectability metric. On examination we find that the $(S+N)/N$ for NICMOS differential coronagraphy in the non-photon-noise dominated regime is a strong function of azimuth angle around the coronagraphic hole, with the lowest $(S+N)/N$ in the 90° – 180° quadrant as shown in Figure 4-left. The observing strategy adopted in the IDT Cycle 7 survey obtained paired images differing by 29.9° of field rotation (the maximum routinely schedulable due to HST pointing and roll constraints). Hence, as the target is "clocked" in azimuth around the coronagraphic hole we evaluate the $(S+N)/N$ in each of the two independent target images in the difference frame. In the NICMOS/HST system this azimuthal dependence is highly repeatable, as it is caused by micro-scattering due to defects in the manufacture of the coronagraphic hole interacting with the HST PSF.

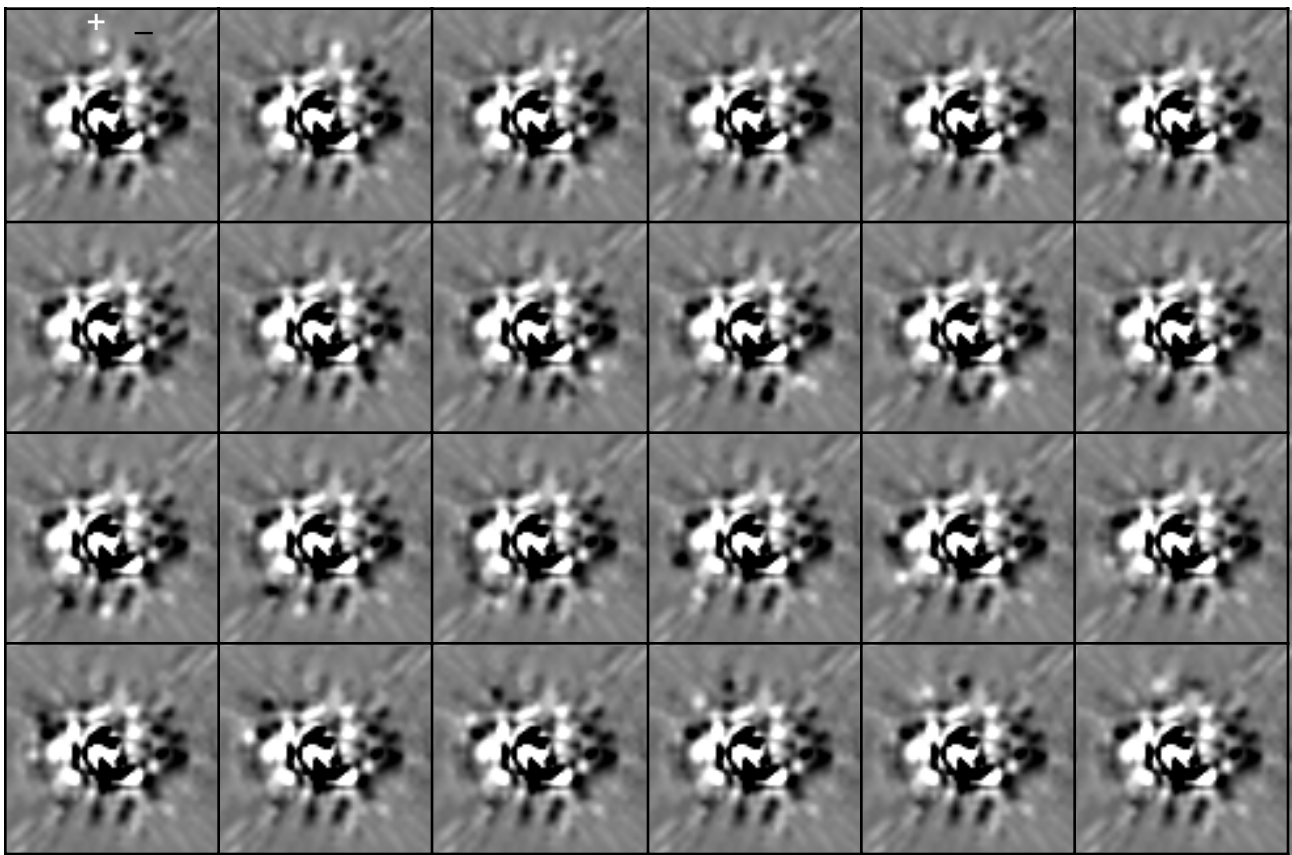


Figure 5. Pushing the limits: Image conjugate implants for $\Delta H=11.5$ at $r=1''$ used to derive $(S+N)/N$ measures in Figure 4-right. Implants are clocked 30° clockwise from the top-left to bottom-right. The 0° fiducial angle for these measures is along the $[-X,+Y]$ diagonal.

Establishing a quantitative threshold for "unambiguous" companion detection is somewhat subjective. We could arguably set the barrier quite high, for example corresponding to a statistical $(S+N)/N > 20$, measured as discussed above, though some may judge this criteria to be overly conservative (as shown by Figure 2A, from which the measures in Figure 4-left are derived). Given our requirement for statistically significant detections in both the (independent) positive and negative image conjugates, we consider this in a more challenging regime of detectability at an angular distance of $1''$ from the occulted target. For a $\Delta H=11.5$ at $r=1''$ that requirement is met by the implants for only $\sim 1/3$ of the azimuths around the coronagraphic hole (see Figure 4-right). The subjectivity of this threshold becomes apparent by examining the conjugate image pairs (Figure 5). If one image of the pair has a significantly higher $(S+N)/N$, whereas the

second is just marginally below the adopted threshold, a detection might still reliably be made. Overall, with this less stringent criterion, the implanted companion would have been found in $\sim 50\%$ of the cases, and hence may be more representative of the azimuthally averaged detection threshold.

6. DETECTABILITY LIMITS

We have performed similar implantation experiments for G–K stars from PSF-subtracted image-pairs (one hundred coronagraphic observations) acquired over the ten months of the IDT survey. These data were obtained over nearly the full range of allowable spacecraft Sun and Beta angles, and on varying timescales (sub- to multi-orbit), after small and large spacecraft slews (which effect the thermal stability of the HST optical telescope assembly (OTA)). From these observations we find an azimuthally averaged detection limit corresponding to a 50% probability of detection (which is in the noise regime strongly dominated by systematics from imperfect PSF subtractions) with a 1σ dispersion of ~ 0.3 magnitudes over all stars in the sample. Overall, we express this probabilistic detection threshold as $\Delta H(50\%) = 9.7 \pm 0.3 + 2.1 \times r$ ". We compare this in Figure 6 to a statistical $(S+N)/N = 25$, which is what was achieved in the illustrative case of TWA6 (though at $H=6.9$, becomes read-noise limited at $r > 3"$ for the integration time used).

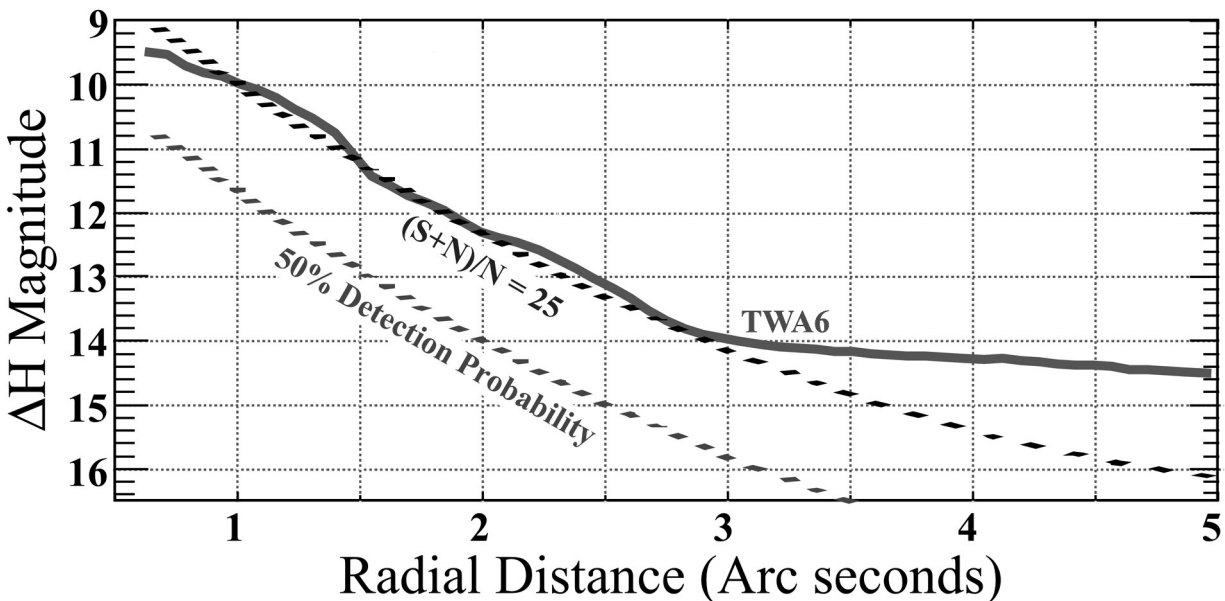


Figure 6. NICMOS H-band differential coronagraphic detection limits from an ensemble of 50 stars.

7. IMPROVEMENTS WITH POST PROCESSING

The noise statistics and detection limits we have discussed are derived from co-registered and subtracted 29.9° differentially oriented coronagraphic images. Additional detectability and contrast gains can be achieved by removing quantifiable image artifacts in post-processing. In particular, the diagonal "spikes" seen in NICMOS images are linear structures induced by diffraction from the OTA secondary mirror support structure and their obscurations in the NICMOS coronagraphic pupil mask, which can be modeled and largely removed. Once a composite image is made from the geometrically rectified difference image conjugates, residual brightness gradients and artifacts with spatial scales differing from that of the unocculted PSF core may be reduced by suitable high-pass filtering. We do this with a radially variable kernel characterized by the spatial frequencies of the coronagraphic PSF difference image itself. As a demonstration of the improvements such processing typically yield we show a post-processed image of the TWA6 field in Figure 7. The "final" statistical $(S+N)/N$ achieved for the $\Delta H=13.2$, $r=2.5"$ field star in the post-processed image was 35 (an improvement of $\sim 40\%$) in the presence of the remaining residuals, which are very typical for NICMOS differential coronagraphic images post-processed in this manner.

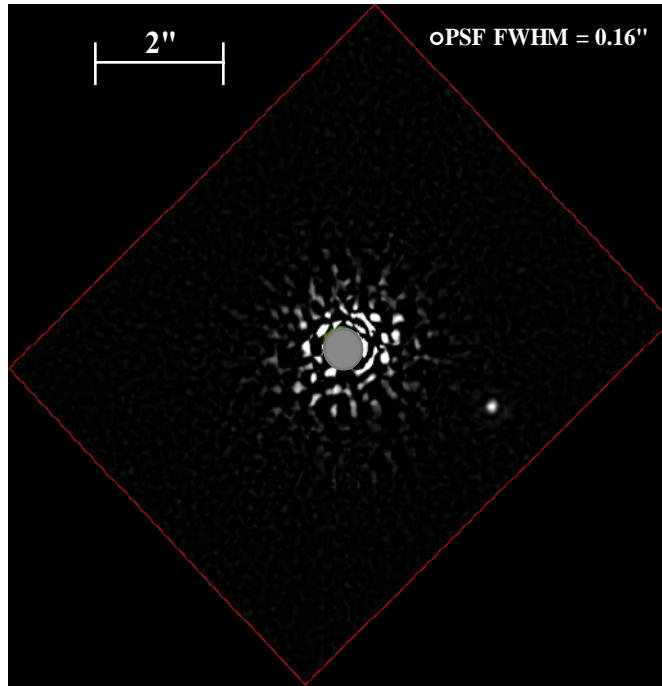


Figure 7. Composite image of TWA6 field (with north up) after removing residual image and background artifacts.

8. SUMMARY

We have assessed the characteristic limits of companion detectability using NICMOS H-band differential coronagraphy from a suite of one hundred coronagraphic observations obtained over a period of ten months and acquired with the HST OTA in many different thermal states. We find these performance limits would routinely permit the unambiguous detection and photometric characterization of Jovian mass planets on the order of 10 Myrs of age at distances as close as $\sim 1''$ from the host star. Further, we find the mid- and high frequency spatial structures of the coronagraphic PSF (which must be actively controlled on ground-based systems) highly stable with both time and spacecraft attitude/roll, resulting in highly repeatable and predictable companion detection limits.

The re-enabling of NICMOS coronagraphy, following HST servicing mission 3B, and the subsequent successful re-cooling of the instrument's detectors, is now underway. A detailed plan to re-verify coronagraphic performance is in place, but has not yet been executed. Preliminary evaluations from already completed target acquisition and coronagraphic focus testing indicate that the NICMOS coronagraphic system has remained essentially unchanged since it was last used in late 1998. The performance characteristics and detectability levels achieved in HST Cycle 7/7N are expected to be replicated once coronagraphic science is re-enabled in the very near future.

We have stressed that simple photometric noise statistics, which are germane to assess systemic sensitivities when measurement uncertainties are dominated by noise which is both random and spatially uncorrelated, are inappropriate in the presence residuals which arise from imperfect PSF subtractions. Here we have focused on metrics for companion detection, but the situation is similar for imaging circumstellar disks and has been considered elsewhere for both NICMOS⁵ and STIS⁹. We strongly urge the developers and users of other instrumental systems, particularly on ground-based AO-augmented telescopes, to adopt methods and metrics similar to those we have suggested here and which speak to the stability and repeatability of such metrics. Without the adoption of such a metrical lingua francae, the inter-comparison of space- and ground-based instrumental systems will be on very uncertain footing and will remain a sensitive issue.

ACKNOWLEDGEMENTS

The authors wish to thank the other members of the NICMOS instrument definition and guaranteed time observing (GTO) teams for their many contributions to this investigation, and in particular to Rodger Thompson for allocating the GTO orbits under which the scientific programs of the Environments of Nearby Stars group had been conducted. This work is based on observations with the NASA/ESA Hubble Space Telescope, obtained at the Space Telescope Science Institute, which is operated by AURA, Inc. under NASA contracts NAS2-655 and supported by NASA grants NAG5-3042 to the NICMOS IDT.

REFERENCES

1. Mayor, M. and Queloz, D., "A Jupiter-Mass Companion to a Solar-Type Star", *Nature*, **373**, 355, 1995.
2. Schneider, G., "High Contrast Imaging and the Disk/Planet Connection", *Hubble's Science Legacy: Future Optical-UV Astronomy From Space*, ed. C. Blades, ASP Conf Series, in press, 2002.
3. Schultz, A., et al. "Post-NCS Performance of the HST NICMOS", *IR Space Telescopes and Instruments*, Proc. SPIE vol. **4850**, in preparation, 2002.
4. Schneider, G., Thompson, R.I., Smith, B.A., Terrile, R.J., "Exploration of the environments of nearby stars with the NICMOS coronagraph: instrumental performance considerations", *Space Telescopes and Instrumentation V.*, ed. Pierre Y. Bely and James B. Breckinridge, Proc. SPIE vol. **3356**, 222, 1998.
5. Schneider, G., Becklin, E.E., Smith, B.A., Hines, D., Weinberger, A.J., and Silverstone, M., "NICMOS Coronagraphic Observations of 55 Cancri", *Astronomical Journal*, **121**, 525, 2001.
6. Burrows, A., et al., "A Nongray Theory of Extrasolar Giant Planets and Brown Dwarfs", *Astrophysical Journal*, **491**, 856, 1997.
7. Boss, A.P., "Modes of Gaseous Planet Formation", *Planetary Systems in the Universe*, eds. A. Penny, P. Artymowicz, A.-M. Lagrange & S. Russell, IAU Symposium **202**, Manchester, England, in preparation, 2002.
8. Krist, J., and Hooke, R.N., "NICMOS PSF variations and tiny tim simulations", *The 1997 Calibration Workshop with a New Generation of Instruments*, eds. S. Casertano, R. Jedrzejewski, and T. Keyes, 192, STScI, Baltimore, 1997.
9. Grady, C., et al, "The HST Legacy and Future for High Contrast Imaging of Disks and Outflows from Young Stars", *Hubble's Science Legacy: Future Optical-UV Astronomy From Space*, ed. C. Blades, ASP Conf Series, in press, 2002.