

Final report on LBTI/HOSTS sensitivity study

S. Ertel^{1,2}, D. Defrère³, B. Mennesson⁴, P. Hinz⁵, P. Willems⁴, S. Perera⁶, R. Zellem⁴, the HOSTS science team & the LBTI instrument team

Submitted: March 2nd, 2020

¹ Large Binocular Telescope Observatory, University of Arizona

² Steward Observatory, University of Arizona

³ STAR institute, University of Liège

⁴ Jet Propulsion Laboratory, California Institute of Technology

⁵ University of California, Santa Cruz

⁶ Max Planck Institute for Astronomy, Heidelberg

1 Executive Summary

We report here the results of our study of the as-built sensitivity of the LBTI instrument when measuring the levels of exozodiacal dust around nearby stars using nulling interferometry in the N band. Central to this study was a detailed error budget we built to account for all significant known sources of null depth error in a systematic and quantifiable way, using independently measured instrument parameters wherever possible. This error budget agrees well with the actual null depth uncertainties for most of the stars observed during the HOSTS survey; of the stars for which the error budget significantly differs from the measured results most were either measured early in the survey with less mature observing technique or under understood, suboptimal conditions. In addition to this theoretical error budget, we empirically analyzed the data from the HOSTS survey to search for correlations between various instrument and environmental conditions and supplementary engineering data to disentangle effects that could not be measured separately in the science data. Both the theoretical error budget and the empirical approach yield consistent predictions.

We find that most of the null depth uncertainty came from two sources. Differential optical pathlength variations due to low frequency vibrations in the telescope dominate the null uncertainty for brighter stars in the survey, and low frequency spatiotemporal background variations due to excess low frequency noise (ELFN) of the mid-infrared detector dominate the uncertainty for fainter stars. We estimate that plausible reductions in these two noise sources would improve the null uncertainty by a factor of two to three over the levels seen in the HOSTS survey. The low frequency vibrations could be reduced by damping the telescope secondary mirror support and by improving the feedforward of existing telescope accelerometer signals into the LBTI pathlength servo. The background variations could be reduced by replacing LBTI's existing Aquarius detector with a Teledyne H1RG detector; we note that experimental verification of the H1RG's background variations in the operating mode required for HOSTS was not possible at this time, and its lower background noise is still a conjecture.

We preliminarily estimate that a renewed exozodi survey with the improved detector and mitigated vibrations could yield a median exozodi estimate three times better than that provided by HOSTS.

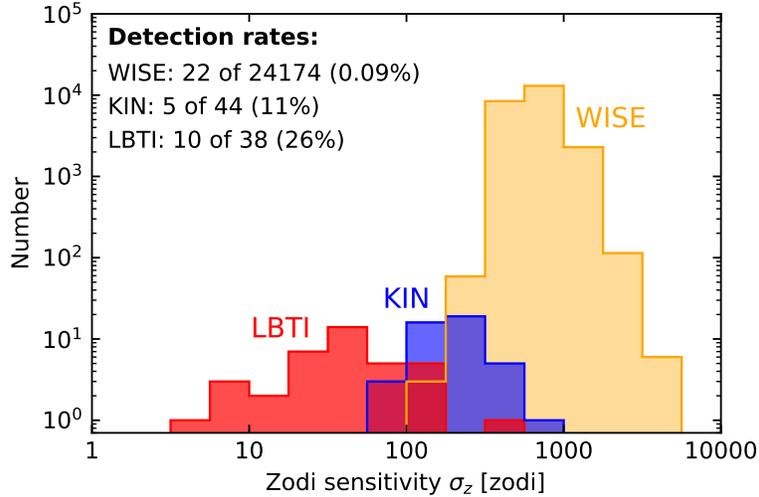


Fig. 1: Distribution of sensitivity to habitable zone dust and sample sizes of LBTI (Ertel et al., in press), KIN (Mennesson et al., 2014), and the WISE sample from Kennedy & Wyatt (2013). The unit of 1 zodi corresponds to the habitable zone dust surface density in the Solar system.

2 Introduction

The Large Binocular Telescope Interferometer (LBTI) was funded by NASA to search for exozodiacal dust (warm dust in the habitable zones of stars other than the Sun). Characterization of the typical amount of such dust around nearby stars is critical for the planning and success of future exo-Earth imaging missions, as it adds noise and confusion to such observations and thus limits their sensitivity. Nulling interferometry in the N band was used to suppress the bright star light and detect the faint circumstellar emission from the dust.

The survey was completed in spring 2018 after observing 38 stars. Our final survey paper has just been accepted for publication (Ertel et al., in press). The sensitivity reached was almost an order of magnitude higher than the LBTI's precursor, the Keck Interferometer Nuller (KIN), and about two orders of magnitude higher than space based spectro-photometry (Fig. 1). Critical input was provided to the HabEx, LUVOIR, and Starshade Rendezvous study teams, as well as to the European LIFE (Large Interferometer for Exoplanets, PI: S. Quanz) study team. The LBTI is a highly complex instrument employing a range of innovative techniques that had not previously been demonstrated. Under these circumstances, it is remarkable that the final performance of the instrument is within a factor of two of the initial predictions. Nonetheless, and an extended survey with better sensitivity would still be highly beneficial to NASA's exo-Earth imaging goals. Further improvement of the data quality prior to the HOSTS survey was not possible due to (a) the lack of a large data set for a statistical analysis and (b) the need to provide timely results for NASA's 2020 decadal survey. The LBTI team was thus instructed to carry out the survey with the current sensitivity rather than to focus on further improvements. After the completion of HOSTS, the LBTI team was tasked with a sensitivity study to understand the limitations to the survey sensitivity, and to identify potential improvements to the instrument, telescope, and observing strategy that would result in significantly better data quality and thus warrant an extended, more sensitive survey.

The sensitivity study has four main components: (1) Potential correlations of the data quality with atmospheric conditions are to be investigated. (2) Potential correlation of the data quality with telescope vibrations are to be investigated. (3) The SOUL (Single conjugated adaptive Optics Upgrade for the LBT) adaptive optics (AO) upgrade on the LBTI is to be completed and the suitability for nulling interferometry to be evaluated. (4) A new science detector available to the LBTI team is to be tested and its potential for improving the sensitivity of nulling observations to be investigated. In addition, we have explored the possibilities to use a larger dither pattern of the optical path delay setpoint to better break degeneracies in our data and using a cold stop at a different location in the instrument to reduce the thermal background. We report here our findings. An interim report was provided in June 2019. We include in condensed form the content of this interim report that has not been superseded by this final report.

We first briefly summarize the survey observations and data reduction in Sect. 3, including a brief description of the LBTI. In Sect. 4 we discuss the current sensitivity of the HOSTS survey data. We demonstrate that the uncertainties are well estimated by our data reduction pipeline (Sect. 4.1) and present and validate our theoretical error budget (Sect. 4.2) by demonstrating that it reproduces the available data well. We also determine in these sections the general nature of our main sources of uncertainties from both an empirical (Sect. 4.1) and a technical (Sect. 4.2) standpoint. In Sect. 5 we analyze our survey data and supplementary engineering data to measure directly the impact of specific error sources and their mitigation onto our data quality to complement and further validate our detailed theoretical error budget and to study realistic improvements to the system and observing strategy. We study the impact of weather conditions (Sect. 5.1), telescope and instrument vibrations (Sect. 5.2), and of detector instabilities (Sect. 5.3). In Sect. 6 we summarize additional studies carried out. These are the evaluation of the performance of the upgraded adaptive optics system (SOUL) for nulling interferometry (Sect. 6.1), using a larger setpoint dither during the observations (Sect. 6.2), and a possible way to reduce the telescope background emission in our data (Sect. 6.3). We then proceed in Sect. 7 to updating our theoretical error budget with the improvements studied and validated throughout this report and to making predictions on the expected errors for a range of scenarios. In Sect. 8 we briefly discuss the implications of our results for a renewed HOSTS survey and related exozodi observations with the LBTI.

3 Observations & data reduction

HOSTS observations have been carried out with the Large Binocular Telescope Interferometer (LBTI, [Hinz et al. 2016](#)) on the Large Binocular Telescope (LBT) on Mt. Graham, Arizona. The LBTI is located at a bent Gregorian focus between the two 8.4m primary mirrors of the LBT. Atmospheric wavefront distortions are measured with the LBTI's pyramid wavefront sensors in the visible (one for each telescope aperture) using the target star as a natural reference star. Adaptive optics (AO) correction is done with the LBT's adaptive secondary mirrors. The infrared light from the two telescope apertures enters the LBTI's cryogenically cooled universal beam combiner (UBC) from both sides. Interferometric optical pathlength (optical path delay, OPD) correction and tip-tilt stabilization is done using pathlength correctors in the UBC. Fast, low-amplitude pathlength correction (atmospheric piston and vibrations) can be performed on either side using a pupil mirror on a set of three stacks of piezoelectric devices for piston and tip-tilt correction. On the right side this assembly is in addition mounted on a translation stage for slow, large-amplitude pathlength correction (e.g., to find the coherence envelope between the two beams). The light from both sides is then sent to the Nulling and Imaging Camera (NIC) where the near-infrared light is used by our fringe tracker PhaseCam to measure the OPD (using phase delay

Table 1: List of nights during which nulling data considered in this study were obtained.

Observing semester	2016B	2017A	2017B	2018A
Dates [yyyy-mm-dd]	2016-10-16	2017-02-09	2017-11-10	2018-03-28
	2016-11-14	2017-02-10	2017-12-20	2018-03-29
	2016-11-15	2017-02-11	2017-12-23	2018-03-30
	2016-11-16	2017-04-03	2018-01-05	2018-05-23
	2017-01-29	2017-04-06		2018-05-25
		2017-04-08		
		2017-04-11		
		2017-05-01		
		2017-05-12		
		2017-05-21		

Note: LBT A semesters: February to July, B semesters: September to January.

tracking) and differential tip-tilt between the two beams. This information is then used to control the pathlength correctors in closed loop. The mid-infrared light is sent to the Nulling Optimized Mid-Infrared Camera (NOMIC) where the science data are recorded. In addition, NIC hosts LMIRCam, a camera capable of observing in the J to M band.

Nulling interferometry in N band was used to suppress the light from the central star, and to reveal faint, circumstellar emission. The OPD between the two apertures was stabilized using PhaseCam in the K band and optimized by minimizing the total flux transmitted onto NOMIC, where the null depth (the flux ratio between destructive and constructive interference) was then measured. Nodding (offsetting the telescope pointing by a small angle of 2.3 arcsec to place the source in a different position of the detector) was used to subtract the variable telescope and sky background. Nulling observations were taken in both nod positions. A sequence of 2,000 null frames was taken (integration time 45 ms per frame) per nod position and three pairs of nods (six sequences of 2,000 frames total) were obtained per observation of a star. In addition, a photometric observation and corresponding background exposures were taken to flux calibrate the null depth. A sequence of dark frames were also obtained each time a new star was targeted. Each observation of a science target (SCI) was paired with an identical observation of a reference star (CAL) to determine the instrumental null depth and to calibrate the science observations. CAL and SCI observations were typically concatenated to a CAL-SCI-SCI-CAL sequence and a nominal HOSTS observation of a science target consisted of two such sequences. Various reference stars were used to minimize the effects of imperfect knowledge about the reference stars (e.g., uncertain diameter, potential circumstellar emission, companions) and no individual reference star was paired twice with the same science target.

Due to this observing strategy, our data are composed of units of various sizes:

- One exposure takes 45 ms.
- One nod is a consecutive sequence of typically 2000 exposures in one nod position on the detector.
- One pointing is a sequence of typically six nods (three pairs) and a photometric observation obtained on a target star (SCI or CAL).
- One calibrated science observation is a sequence of at least one CAL and one SCI pointing.

Observations were carried out mostly in queue mode together with a variety of other observing

programs using the LBTI, including high-contrast direct imaging and integral field spectroscopic observations. This increased the pool of nights to choose from for the nulling observations. The nights during which nulling data were obtained are listed in Table 1.

Data reduction followed the strategy outlined by Defrère et al. (2016) with minor updates as described by Ertel et al. (2018). Various photometric aperture sizes were used to measure the null depth (see Ertel et al. 2018 for the relevance of each aperture), of which we select the one with a 13 pix (233 mas) radius for the analyses in this report as it optimizes the photon and read noise limited signal-to-noise for extended emission analogous to the Solar system zodiacal dust and is considered the standard aperture for the HOSTS survey. The raw null depth and its uncertainty were determined using the null self calibration method (NSC, Mennesson et al. 2011; Hanot et al. 2011; Defrère et al. 2016; Mennesson et al. 2016), combining all frames recorded within a nod for a statistical analysis. These measurements within an observing sequence of a science target were then combined and the corresponding calibrator observations were used to calibrate the null measurements.

4 Current survey sensitivity

A first step towards better understanding the achieved HOSTS sensitivity and its limitations is to analyze the available survey data and to compare them with predictions from a theoretical error budget. This is done in the following subsections.

4.1 Error estimation from the data and achieved sensitivity

We have estimated the measurement uncertainties as part of the survey data reduction and analysis. From the data reduction perspective, there are two dominant types of uncertainties:

- Statistical NSC uncertainties: The uncertainties in determining the uncalibrated null depth of a target within a nod are determined statistically by the NSC. This category contains short term effects such as phase jitter (piston) from telescope and instrument vibrations and the atmosphere, tip-tilt errors from the adaptive optics system (vibrations, seeing), uncertainties from determining the error made while minimizing the OPD, and photon and read noise from the flux measurements on the individual, nulled frames.
- Systematic NSC uncertainties: These uncertainties limit the repeatability of the null measurements between nods. This category contains long term effects such as background variations, low frequency detector noise, and OPD optimization errors. They are invisible to the NSC and constitute systematic uncertainties of the NSC results (although they are randomized between nods by our observing strategy), so that they are not included in the error bars estimated by the NSC.

Our data reduction uses both the statistical NSC uncertainties and the scatter of the measurements obtained by the NSC for individual nods (which contains both contributions) to statistically estimate the final uncertainty of the null measurements of a full calibration sequence.

Fig. 2 shows the distribution of final, calibrated null significance measurements (null measurement divided by its uncertainty) and the distribution of the uncertainties themselves for the photometric apertures used for HOSTS. The null significance of our measurements follows a Gaussian distribution with a standard deviation close to one (considering the small number statistics) and a tail of detections at significance >3 . This is to be expected if our uncertainties are neither over- nor under-estimated for

most stars. The distribution of the uncertainties is also reasonable with a sharp peak at low uncertainties and a tail of higher uncertainties for stars with a smaller amount of data or sub optimal data quality. We conclude that the uncertainties in our survey data are generally well estimated by our data reduction routines. Our achieved median null accuracy for the 13 pix aperture is 0.09% (900 ppm).

We can estimate the relative contribution of the statistical and systematic NSC uncertainties to our final uncertainties. An ideal, calibrated observation of a science target is composed of $N_{\text{nods}}=24$ nods on that science target and $N_{\text{nods}}=24$ nods on four different reference stars (six nods each). The statistical NSC uncertainties σ_{NSC} on a target can be combined to a single, more accurate one in the absence of systematic NSC uncertainties. The systematics free uncertainty of a complete science target observation can thus be estimated as:

$$\sigma_{\text{stat}} = \sqrt{2} \times \frac{\sigma_{\text{NSC}}}{\sqrt{N_{\text{nods}}}} \quad (1)$$

The $\sqrt{2}$ comes from the fact that the uncertainties from the 24 calibrator nods have to be added in quadrature to those of the 24 science target nods. Fig. 3 compares the distribution of the NSC uncertainties for each nod position with the final uncertainties per science target. For a median NSC uncertainty of 0.12% (1,200 ppm), this results in a systematics free accuracy of a full observation of 350 ppm. This compares to an actual median uncertainty of a full observation of 900 ppm, which must be dominated by systematics of the order of 830 ppm ($350^2 + 830^2 = 900^2$). This additional scatter is well visible in the example night shown in Fig. 4. We will determine in Sects. 4.2 and 5.2 that telescope vibrations are the dominant source of this uncertainty and discuss realistic ways to mitigate their impact on our data quality.

Fig. 4 also shows another trend that is visible in many data: The scatter and error bars of the calibration stars (and thus the measurement uncertainties) are smaller than those of the science targets. A statistical analysis of all observations shows that the statistical NSC uncertainties are on average by a factor 0.7 lower for calibrators than for science targets and the scatter of the measurements is different by the same factor. There are several differences between calibrators and science targets. Calibrators are typically chosen to be slightly brighter than science targets so our accuracy is not limited by the brightness of the calibrators. Calibrators are redder across the visible and near-infrared (giant stars) than science targets (main sequence stars), but this is not the case in the N band (Rayleigh-Jeans regime of all stars' emission). At the same N band brightness this would result in the calibrators being *fainter* in the visible and K band resulting – if anything – in worse AO and PhaseCam performance. Even our brightest stars are not saturating the AO or PhaseCam, so that saturation or non-linearity cannot be an issue for the majority of our stars. Calibrators have a larger diameter (typically one to a few milli-arcseconds) than science targets (well below 1 mas). Their diameters are less than an order of magnitude smaller than the typical differential tip-tilt jitter from the AO (see next section), which may to some extent mitigate this effect. None of the above effects can fully explain the fact that calibration observations appear more accurate than science observations. A combination of these effects is likely at play but a clear explanation is not available.

We conclude that our uncertainties are dominated by systematic NSC uncertainties that limit the repeatability of null measurements from individual nods. As discussed above, these systematics mostly originate from long term effects such as background subtraction and low frequency detector noise. Reducing these effects thus carries the strongest potential to increase our overall sensitivity. Even without any other improvements, reducing this contribution by 50% can reduce our overall uncertainties by 40% (a factor 1.7 improvement in sensitivity). This would put LBTI's accuracy to 550 ppm, close to

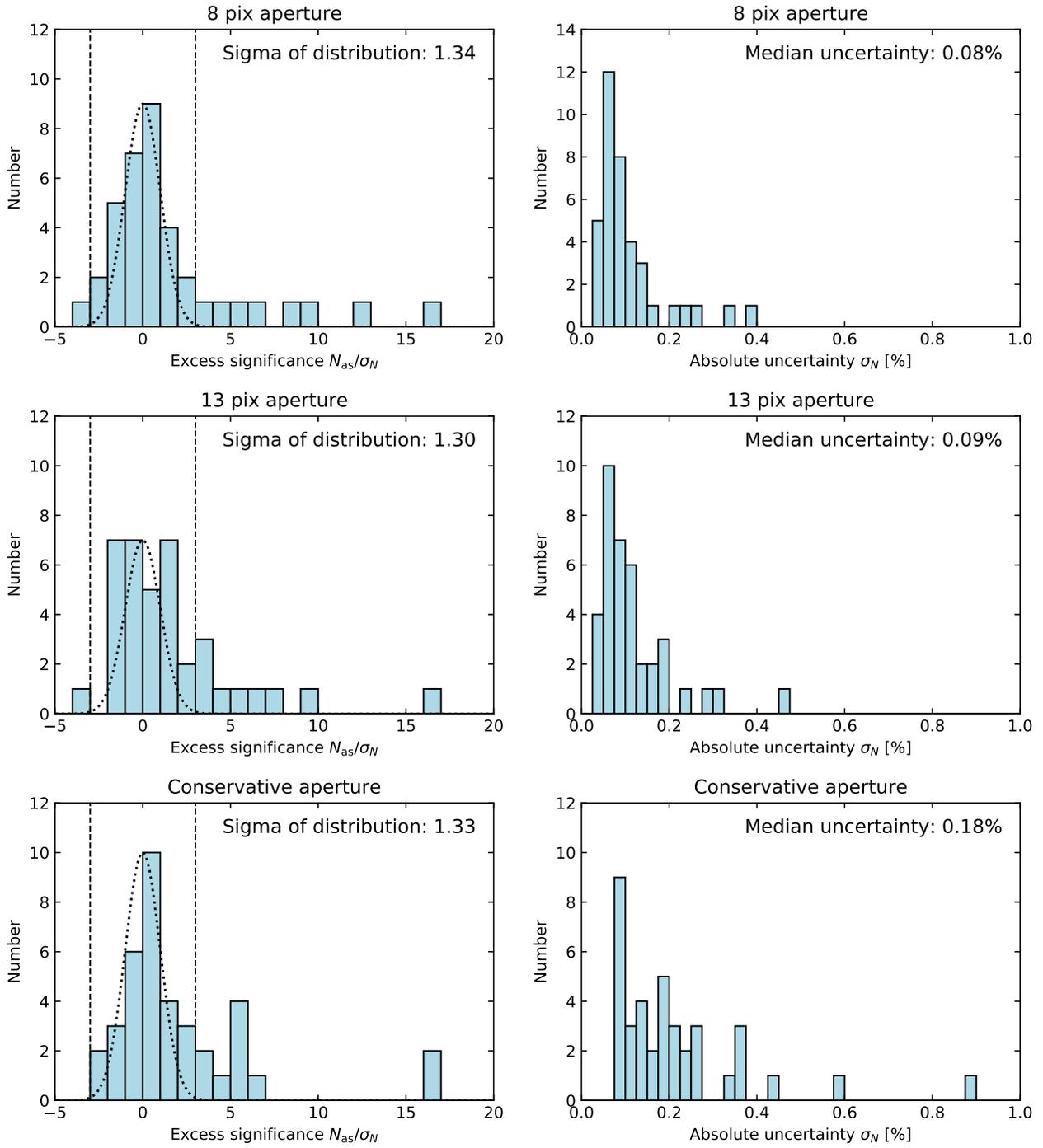


Fig. 2: Null significance (*left*) and uncertainty (*right*) of the HOSTS survey observations. The dotted line represents a Gaussian distribution with a standard deviation of 1 and a peak scaled to the peak of the distribution of measurements. It is for guiding the eye only. The vertical, dashed lines mark the -3σ to $+3\sigma$ range, within which deviations from zero are considered insignificant.

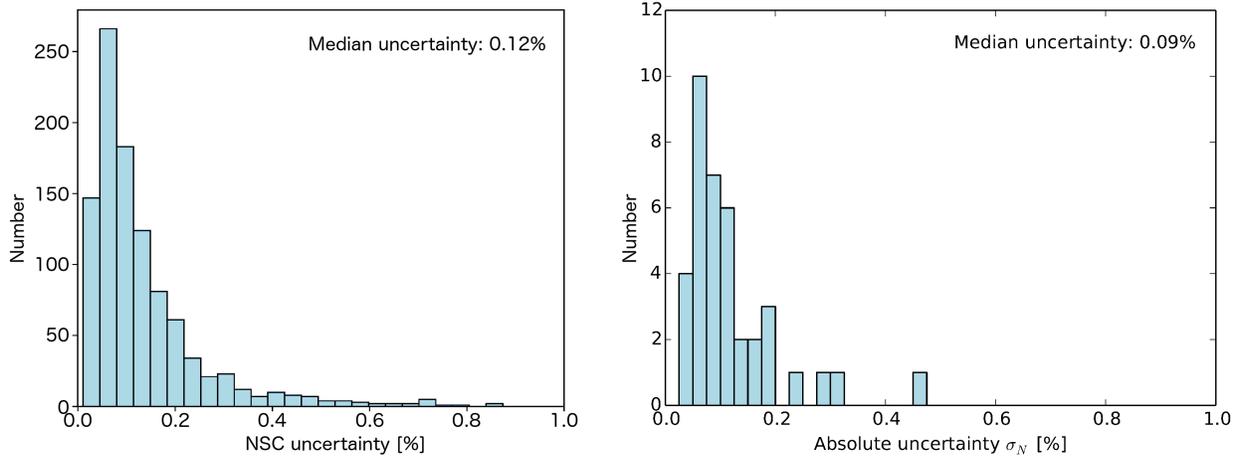


Fig. 3: Comparison of the largely statistical null uncertainties from the NSC per nod with the final uncertainties per star of calibrated science observations.

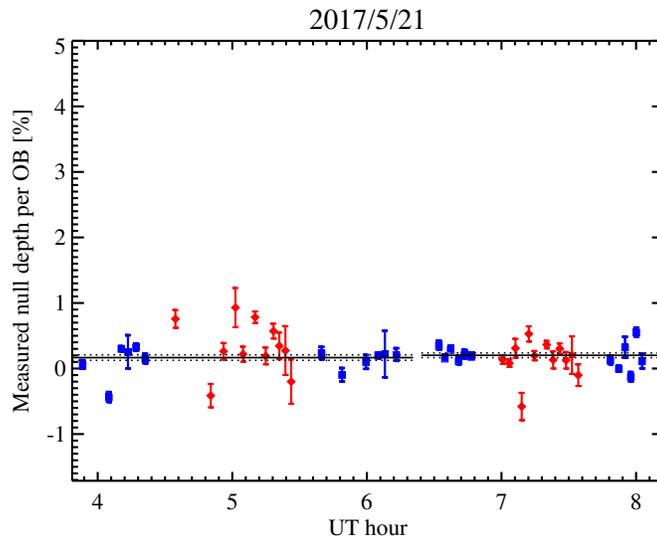


Fig. 4: Illustration of the systematic errors affecting the repeatability of the null measurements from nod to nod on a representative night. Blue points are nods on calibrators, red on science targets. The scatter of the data points from the individual nods is not well represented by their statistical error bars estimated by the NSC.

the original specifications of 400 ppm presented at the Operation Readiness Review (ORR) and in line with the accuracy achieved for very bright stars (e.g., β Leo commissioning data presented at the ORR), for which background subtraction errors have a smaller contribution to the total error budget.

4.2 Theoretical error budget

We now use a theoretical error budget, informed by measurements of the main technical properties of the LBTI, to refine our understanding of the limitations to our survey sensitivity. We validate this error budget by comparing the predicted uncertainties to the ones measured from our observations.

We separate the parameters of the budget into a range of categories:

- Fixed telescope and instrument parameters such as primary mirror diameter.
- Observing parameters such as observing wavelength.
- Target parameters such as brightness.
- Parameters related to noise sources.

The items in the first two categories are predetermined for a given observation. They are generally well known, have been measured to good accuracy during the design and commissioning of the LBTI, or have been decided during the observations. The parameters in the first category are primary mirror diameter, interferometric baseline, detector quantum efficiency, detector gain, detector pixel size, instrument and telescope optical throughput, instrumental null depth, and relative intensity mismatch between the two sides of the telescope and instrument. The second category includes observing wavelength and bandwidth, integration time per frame, number of frames per nod, number of nods per observation, number of SCI and CAL observations per science target, and size of the photometric aperture used for the data reduction (we focus here on our standard 13 pix aperture). Target parameters include the N band flux for the science target and calibrator stars (generally similar and assumed to be the same here for simplicity), the angular diameters of the science target and calibrators, and null level of a face-on, 1 zodi disk (for conversion of null measurements to zodi levels, see [Kennedy et al. 2015](#) for the detailed definition of one zodi).

Parameters related to noise sources deserve particular attention in the context of the error budget. Well known quantities are the background flux and detector read noise. The uncertainties on target and calibrator diameters are found to be negligible as neither are significantly resolved (nonetheless, their diameters are taken into account in the calibration). Standard noise sources such as background and target photon noise can either be measured directly or be determined from the above parameters. Uncertainties due to photometric calibration are conservatively estimated to be of the order of 10%, and are generally negligible for our nulling observations. The remaining noise sources are harder to predict. These are discussed individually in the following.

Relative tip-tilt variations between the two overlapped beams: This can be determined from the tip-tilt variations of the individual apertures, which were determined to be of the order of 10 mas for the LBT's AO systems. Higher tip-tilt variations are seen, e.g., due to wind shake, but experience has shown that in this case it is difficult to maintain the phase lock so that nulling observations are made difficult. The impact of relative tip-tilt variations is propagated through the error budget. It is currently found to be negligible compared to other uncertainties as it is very small compared to the N band PSF ($FWHM = 313$ mas). Piston (phase) variations introduced by the AO correction are tracked by PhaseCam together with atmospheric variations and telescope and instrument vibrations which are

discussed below.

Mean phase setpoint and the uncertainty in determining it with the NCS: The phase setpoint is minimized and tracked at the beginning of each sequence of integrations in a nod position. However, this process is not perfect and a residual phase offset is to be expected. This offset and the uncertainty of determining it are measured by the NSC for each nod and the average of each observing night is propagated through the error budget. The median offset throughout the HOSTS survey was $0.25 \mu\text{m}$ (0.14 rad at a wavelength of $11 \mu\text{m}$) with nightly median values ranging from 0.05 rad to 0.3 rad . The median uncertainty of determining this offset with the NSC was 0.02 rad with nightly medians from 0.007 rad to 0.09 rad . The effects are typically small but not entirely negligible for our brightest stars.

Low frequency phase jitter and the uncertainty in determining it with the NCS: The phase is stabilized with PhaseCam, but this process is not perfect. The jitter at frequencies lower than the frame rate of the science exposures is estimated by the NSC for each nod and the average of each observing night is propagated through the error budget. Sources of this jitter can be both low frequency telescope vibrations not corrected by the fringe tracker and atmospheric water vapor. We will show in Sects 5.1 and 5.2 that the former dominates the low frequency jitter while the latter is negligible. During the HOSTS survey, the median low frequency phase jitter RMS was $0.16 \mu\text{m}$ (0.09 rad at a wavelength of $11 \mu\text{m}$) with nightly median values ranging from 0.05 rad to 0.24 rad . The median uncertainty on determining this parameter with the NSC is 0.008 rad . This is a significant source of uncertainty for bright stars (N band flux $>$ a few Jy).

High frequency phase jitter and the uncertainty in determining it: Similar to the low frequency phase jitter, suppressing the high frequency phase jitter at frequencies higher than the frame rate of the science exposures with PhaseCam is not entirely possible. This jitter is, however, not visible to the NSC and needs to be determined from the PhaseCam telemetry directly and provided to the NSC analysis. The PhaseCam telemetry taken during the HOSTS survey shows that this value only changes by less than a factor of 2 and is of the order of 650 nm for the jitter. Since this is measured directly, its uncertainty is very small. The average over the whole HOSTS survey is used and propagated through the error budget. This contribution is generally negligible.

Systematic background estimation error (background bias): This noise source is related to the fact that after suppressing the bright star light we are left with the challenge of measuring photometry of very faint sources compared to both the background and instabilities of infrared detectors. The flux from the (nulled) source on NOMIC is measured using aperture photometry on each individual exposure. Temporal variations of the background are removed using a background annulus. Spatial variations are removed by estimating the background in the star's position on the detector during exposures in which the star has been moved away (e.g., opposite nod position). This method, however, is ineffective in removing the temporal variations *in the star's position at the time the star is observed in this position* (spatio-temporal background variations). The latter variations may be caused by telescope and sky background variations (e.g., clouds, moving optics, and temperature changes of the system) or spatio-temporal instabilities of the detector (excess low frequency noise, ELFN, Hoffmann et al., 2014). We will show in Sect. 5.3 that detector instabilities are the dominating factor because of our fully cryogenically cooled instrument and because HOSTS data have been obtained in clear conditions (with very few exceptions). The magnitude of the background bias has been determined by a long sky exposure to be of the order of 2 mJy per nod for a typical HOSTS observation. This bias is the dominant source of systematic NSC uncertainties, while the previous contributions are estimated by the NSC and constitute

statistical NSC uncertainties.

We validate the error budget by comparing the predicted uncertainties with the measured ones for each star. The result is shown in Fig. 5. We can see that the predicted and measured errors mostly agree within 50%, which is a reasonable target given the complexity of the error budget. There are a few exceptions, in particular at the beginning of the survey (note that the stars are listed in the order they have been observed), which have been eliminated by improving the observing strategy as we gained experience (e.g., implementation of setpoint dithering, avoid even thin clouds). In the remaining cases, the errors are neither under- nor over-predicted in a systematic way throughout the survey and predictions are met well across the full range of target fluxes. The latter is particularly important as it shows that the relative contributions of the individual components of the error budget across the full range of fluxes (vibrations for bright stars vs. background bias for faint stars), are well understood. We conclude that the error budget successfully predicts the absolute uncertainties for LBTI nulling observations as well as the relative contributions of individual error sources and can be used to predict the possible improvements that can be achieved from individual changes to the instrument, telescope, and observing strategy. A spread sheet containing the error budget used for this report (both for the current HOSTS data and for an upgraded LBTI) is provided with this report.

The theoretical error budget predicts that for faint targets the main source of uncertainty is the background bias, while for the brightest stars the dominating contributions come from telescope vibrations. The two contributions are equal at an N band flux of approximately 2.8 Jy. This is consistent with the previous analysis suggesting that systematic NSC uncertainties dominate, as 60% of our targets are fainter than that. However, it also shows that reducing vibrations in addition to the background bias is a critical aspect of improving the nulling sensitivity of the LBTI as reducing the background bias will quickly move the majority of suitable HOSTS stars to the vibration limited regime.

5 Data based analysis of noise sources

We have determined that both vibrations and background bias are important noise sources for our nulling observations. Our error budget gives us a handle on how improvements of individual noise contributions will affect the final uncertainties of our nulling observations. In addition to the uncertainties considered before, the impact of atmospheric conditions always needs to be considered for ground based observations. In this section, we use an alternative, data based approach to determine more directly the impact of these noise sources on our data. We first confirm that changes of the atmospheric conditions within the range that was deemed suitable for HOSTS observations do not significantly affect our data quality. We then make use of the fact that the vibrational environment of the telescope and instrument is variable to correlate our data quality with the strength of vibrations and to determine what level of improvement is realistically possible. Finally, we use new engineering data to determine the contribution of ELFN to the background bias.

5.1 Correlation with atmospheric conditions

A systematic analysis of the correlation of the data quality with the atmospheric observing conditions was carried out using separately the uncertainties of the calibrated null measurements on the science targets recorded for each nod position (not combined to one final measurement to have a finer time sampling and better statistics) and those of the uncalibrated individual null measurements per nod position as provided by the NSC for all stars observed. The former allows us to search for low frequency and systematic effects,

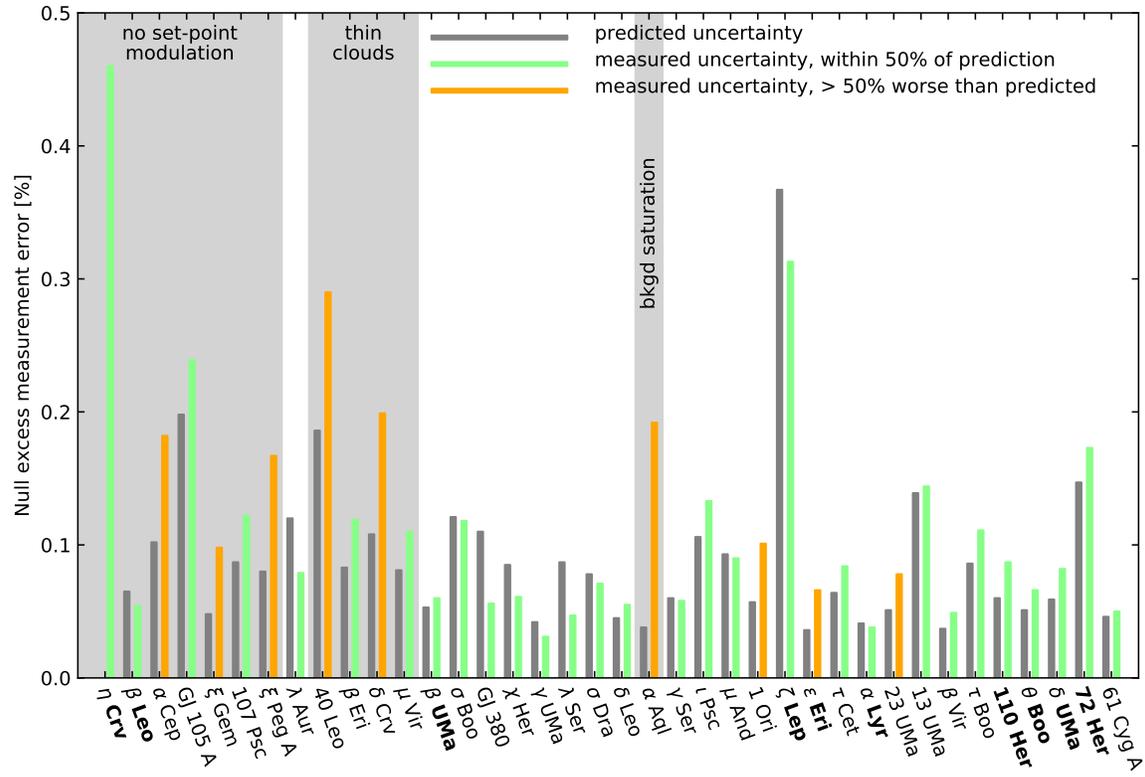


Fig. 5: Comparison between predicted and measured uncertainties for the HOSTS observations. Stars are arranged by the date they have last been observed, increasing left to right. Cases where the main cause for the deviations from the predicted uncertainties is known are highlighted. Stars with detected excess are highlighted in bold face.

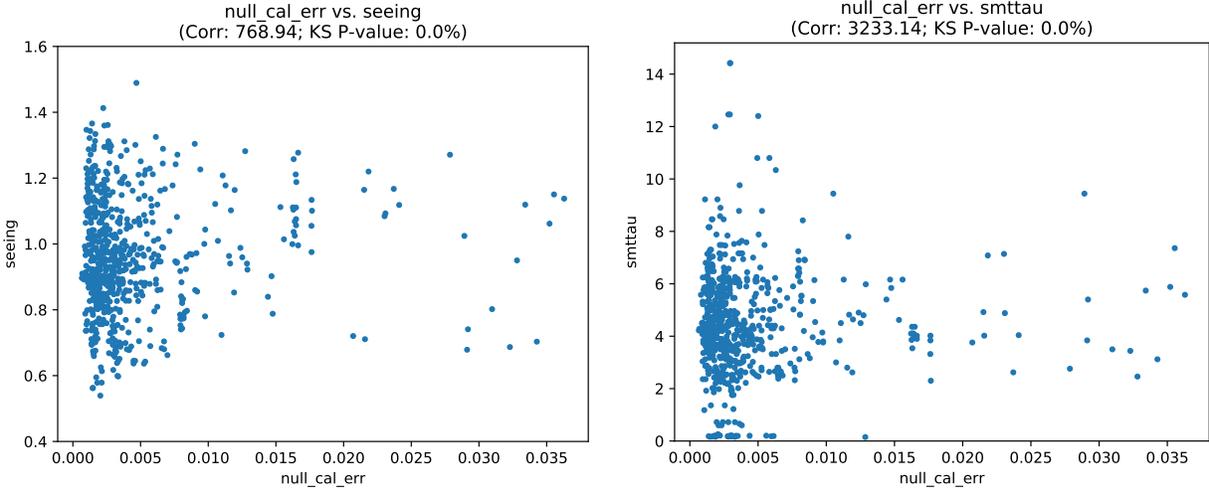


Fig. 6: Comparison of the calibrated nulling errors with seeing (*left*) and PWV (*right*). No correlation is visible.

while the latter can reveal high frequency effects that are visible within individual nods. Correlations with the seeing, precipitable water vapor (PWV), and wind speed and direction (relative to telescope pointing direction) were investigated. Fig. 6 shows the comparison between the uncertainties of the calibrated data with seeing and the precipitable water vapor (PWV), respectively. No correlation was found with any of the weather telemetry values. This confirms that weather conditions do not significantly affect our data quality as long as they are within the range considered suitable.

Various breaking conditions were triggered almost always before the data quality was affected: Nulling data can only be obtained if the seeing is low enough to close the phase loop (around 1.3 arcsec or better). In this case the AO correction on our bright target stars and at the N band is excellent and no limitation to our data quality. Nulling data are only obtained in clear nights, during which the PWV is almost always below 6 mm at Mt. Graham. Below this level, no significant impact from the absolute PWV level was found in the real time quality control and later reduction of the data. High winds cause vibrations of the adaptive secondary mirrors which make it impossible to close the phase loop. A correlation of data quality with vibrations that are attributed to more moderate wind speeds is discussed in Sect. 5.2.

In addition to the above parameters, there are other atmospheric parameters that may well be relevant for the nulling data quality but are not routinely tracked at Mt. Graham and thus not readily available in the FITS headers. Coherence time is a measure of the speed of the atmospheric turbulence. It is difficult to measure, thus no data are available for Mt. Graham. Similar to high seeing, short coherence time is expected to break the PhaseCam loop. The LBTI team is working on a method to estimate the atmospheric coherence time from the adaptive optics telemetry. Thus, in future observations the coherence time will be measured and compared to the nulling data quality. If a correlation is found, real time estimates of the coherence time will allow us to determine whether it is suitable for nulling observations or not. The variability of the PWV column density toward the target appears to be more related to atmospheric turbulence than the absolute amount of PWV and can strongly impact the data quality. If small, this effect is calibrated well by the NSC and the OPD variation can in fact help determining and correcting errors in minimizing the OPD during the statistical analysis of the NSC.

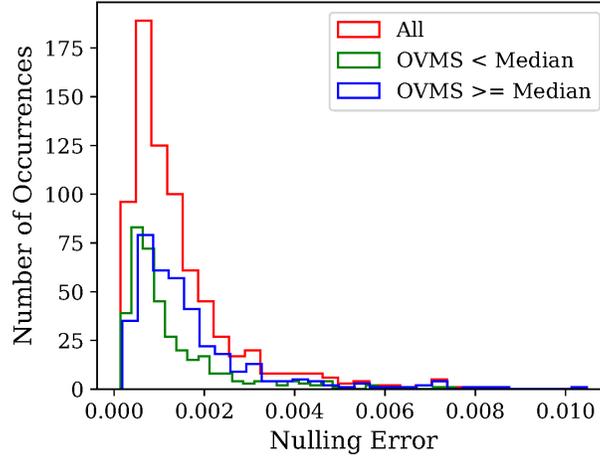


Fig. 7: Histograms of the nulling errors for the whole data set and for data points separated into those with above-median and below median vibrations as measured by the RMS of the OVMS predicted piston. The points with higher vibrations show a wider distribution toward larger errors. The median accuracy of the low vibration sample is better than for the whole sample by a factor of approx. 30%.

Larger variations are well visible during our real time quality control and can thus be avoided. Even thin cirrus result in strong temporal and spatial variations of the sky background, but are well visible and can be avoided.

5.2 Telescope and instrument vibrations

Telescope vibrations are expected to largely impact the quality of the individual data points as they have a continuous effect on the instrument (in contrast, e.g., to changing weather conditions on longer time scales). Strong vibrations simply break the phase loop and thus prevent data acquisition all together. More moderate vibrations result in a variable OPD on short time scales (within one integration or from one integration to another depending on frequency) that PhaseCam may only be able to partly stabilize. This may then effectively result in an unstable or less deep instrumental null which would limit our data accuracy. This effect is best visible in the uncertainties of the NSC results for each individual nod. We thus analyze these data and correlate them with PhaseCam telemetry obtained and stored in parallel with the nulling data.

Among the telemetry data stored, the most relevant set is the one containing the OPD variations predicted by the OPD and Vibration Monitoring System (OVMS) from the vibrations of the primary, secondary, and tertiary mirrors. Other interesting values are the phase measured from the PhaseCam fringes and the commanded piston of the pathlength corrector based on a combination of these data and an additional vibration filter. The latter two sets contain information on how well the individual contributions are suppressed by the phase loop. To estimate the magnitude of the variation in each parameter, we compute its root mean square (RMS).

To search for statistical trends that are otherwise not visible due to other effects on the error such as target brightness, we separate the data into observations with the RMS of the OVMS above median

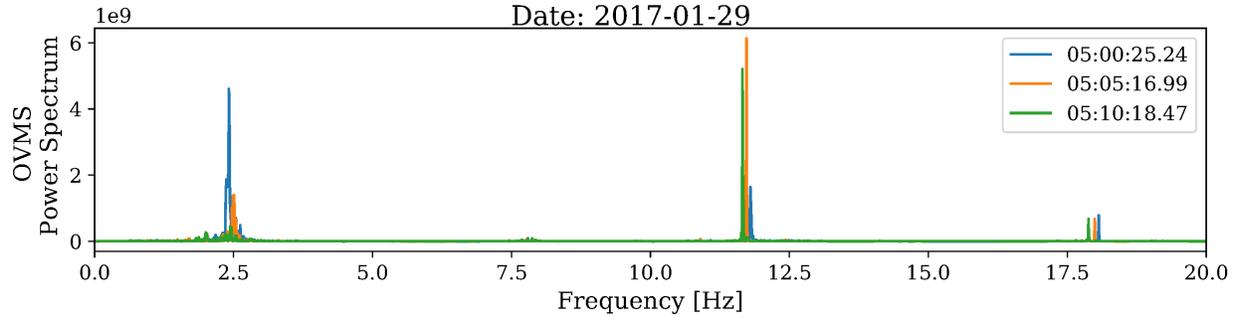


Fig. 8: Frequency analysis of the vibrations measured by OVMS for three times during one representative night.

and below median of $1.1 \mu\text{m}$. Fig. 7 shows histograms of the distributions of nulling errors for these subsamples. A clear correlation is visible where the nulling errors tend to reach higher values for higher OVMS RMS. The median of the error distribution for the strong vibrations sample is 1,500 ppm (0.15%), compared to 800 ppm (0.08%) for the weak vibrations sample. The median of the full sample is 1,200 ppm (0.12%). This indicates that obtaining data only during weak vibrations or preferably mitigating the vibrations can reduce the median error of the obtained data by approx. 30%. This only affects, however, the statistical NSC uncertainties described in Sect. 4.1. It would reduce those from 350 ppm to 230 ppm for a full observation of a science target.

It is interesting to note that achieving this improvement does not require a complete elimination of the vibrations but rather a reduction of the vibrations to be consistently at the level at which they already were during 50% of the HOSTS data acquisition. This means reducing the median OVMS RMS from $1.1 \mu\text{m}$ to $0.7 \mu\text{m}$. A similar correlation is not seen for the commanded piston or the phase, meaning that telescope vibrations must indeed be the dominant source of the increased error, not other effects that are not seen by PhaseCam such as PWV. We also see that the measured OVMS vibrations are of the same magnitude as the low frequency phase jitter discussed in Sect 4.2. This shows that telescope vibrations are the dominant source of this jitter, which in turn was determined by our theoretical error budget to be the main source of statistical NSC uncertainties. These results are all consistent with each other and show that the source of statistical NSC uncertainties is well understood and accurately represented in our error budget.

Fig. 8 shows a frequency analysis of the OVMS vibrations for a representative night. Most obvious are vibrations at a frequency of $\sim 2.5 \text{ Hz}$ and $\sim 12 \text{ Hz}$. These frequencies are common – despite varying strength – to all data sets. To further investigate if vibrations at any specific frequency are responsible for the degraded data quality, we repeat the above analysis after filtering out specific frequency ranges. We find that the correlation of the nulling error is specifically with the 12 Hz vibration. This vibration originates from the swingarms of the secondary mirrors of the telescope and is predominantly excited by wind. While tip-tilt variations and higher order aberrations at that frequency are well corrected by the AO system, the AO is blind to piston variations which are instead corrected by PhaseCam. Residuals to this correction are the main are what is being discussed here.

Comparing the 2.5 Hz vibration and the 12 Hz vibration, we find that the former is well suppressed by the OVMS feedforward, i.e., by compensating directly for the OPD changes predicted by the OVMS

system using the pathlength corrector. As a result, the 2.5 Hz vibration is significantly suppressed in the phase of the light measured by PhaseCam compared to the 12 Hz vibration. Nonetheless, the residual vibrations measured by PhaseCam show already a reduction by a factor of ~ 6 (from $\sim 1 \mu\text{m}$ to $\sim 0.16 \mu\text{m}$) compared to the vibrations measured by the OVMS system. This tells us that the 12 Hz vibration is suppressed by the OVMS feedforward, but improvements are possible to further increase its effectiveness. Discussion with J.-U. Pott, PI of the OVMS development, suggests that optimizing the latency correction of the OVMS feedforward via the readily implemented control parameter may allow us to better suppress vibrations at the critical 12 Hz frequency. Our analysis shows that suppression of this vibration by only 30% more may already have the desired effect, which appears entirely realistic. In addition, reducing this vibration through damping would be beneficial. LBTO is currently investing significantly in identifying and eliminating or damping sources of telescope vibrations. The fact that the 12 Hz vibration is critically affecting our nulling data accuracy has been communicated to LBTO. Another option to minimize this vibration is to observe only at very low wind and to point out of the wind. While as a general rule this would further limit the sky time that would be suitable for nulling observations it will enter into the real time target selection of future observations and may be a last resort if the vibrations cannot be suppressed or mitigated in any other way.

5.3 Detector instability

We have shown that the background bias is the most significant noise source in our data for faint stars. This bias in removing spatio-temporal background variations from the images for precise photometry of the faint, transmitted source flux in the null frames can have several sources:

- The sky background may be structured and vary over time.
- The background from the thermal emission of the telescope and instrument optics may vary spatially and over time.
- The ELFN of the detector may cause spatio-temporal variations of the measured counts without a physical variation of the detector.

The first two contributions are expected to be small due to the small field of view, the fact that HOSTS data were generally obtained under clear conditions, and the cryogenically cooled instrument. In contrast, ELFN is a known problem with the NOMIC detector ([Hoffmann et al., 2014](#)). It appears at frequencies below ~ 100 Hz in blocked-impurity-band (BIB) detectors such as our current Raytheon Si:As IBC Aquarius array as a randomly modulated photoresponse (for a detailed technical description of the phenomenon, see [Stapelbroek et al., 1984](#)). The effect is thus multiplicative with the flux on the detector. In this section we first describe an analysis of recent engineering data to understand the contribution of ELFN to the total background bias measured before as the combined effect of all contributions described above. This will allow us to evaluate the impact of mitigating the ELFN on our data quality. We then describe our evaluation of two alternative detectors that are candidates for replacing the current NOMIC detector for an improved background bias.

5.3.1 Evaluation of the contribution of ELFN to the background bias

We obtained a series of images with NOMIC in closed dome, most noteworthy:

- A ~ 2 h long series of darks with a detector setup resembling as closely as possible the HOSTS observations.

- A ~ 2 h long series of 'light' frames with the same setup as the darks but with a warm beam combiner and a closed gate valve as a stable screen. Filters were adjusted to reach about the same background as in the HOSTS science data.

These data were analyzed using an equivalent strategy to the HOSTS photometry. Since there was no source in any of the frames, the effects of background variation could be measured directly. Since no sky, telescope, or moving parts in the LBTI were observed, and the instrument temperature was stable when the engineering data were taken (no elevation changes, UBC at stable dome temperature), the remaining spatio-temporal flux variations can be directly attributed to detector effects that can then be characterized (in particular ELFN).

We estimate a lower limit of 1.8 mJy for the background bias caused by ELFN in these data compared to a total background bias of 2 mJy in our error budget (Sect. 4.2). Assuming that this contribution adds in quadrature to other effects (sky, telescope, and instrument background variations), this means that eliminating detector ELFN as a source of the background bias will reduce it to 0.9 mJy ($0.9^2 + 1.8^2 = 2.0^2$), i.e., by a factor of ~ 2 . We note that this is a difficult estimate to make as the strength of the dominant ELFN is the only quantity we can measure and only indirectly on engineering data. Because the individual effects add in quadrature, the contributions of less dominant effects that would remain after the ELFN is eliminated are quickly lost in the uncertainties of estimating the contribution of the ELFN. At the same time, these uncertainties are of little consequence for the final error budget. A $\pm 10\%$ error of the contribution of the ELFN would result in improvement factors of the background bias alone between 1.7 and 7! Because this is only one contribution to our null uncertainties, the ultimate difference between the lower boundary of 1.7 and our already conservative estimate of 2.0 is marginal. Improvements of the background bias by a factor 7 will not have a proportional effect on our sensitivity as background photon and detector read noise start to dominate before that. While a definite conclusion on the improvement of the background bias from eliminating the ELFN is thus not possible from our analysis, we note that the improvement is reasonably well constrained and almost certainly significant.

We also confirm that the effect is not visible in dark frames, which confirms the multiplicative nature of the noise with the number of photons on the detector, which is consistent with ELFN being the source of the noise.

5.4 Alternative detectors tested

The current NOMIC detector is a Raytheon 1024x1024 Si:As IBC Aquarius array which is known to exhibit ELFN (Hoffmann et al., 2014). There are two options being considered to replace the NOMIC detector with LW (3-13 μm sensitivity) HgCdTe (MCT) material:

- A Teledyne HAWAII-1RG (H1RG) device that was part of a test batch for developing LW detector material for NASA's NEOCam mission.
- A high-speed readout device dubbed GeoSnap, also from Teledyne.

The H1RG utilizes Teledyne's traditional low power, low noise source follower readout scheme and can be controlled using established readout electronics such as the SIDECAR ASIC and MACIE card. For high-background conditions such as ground-based thermal infrared observations, typical well depths of around 100,000 electrons or less tend to limit these devices applicability. A number of 13 μm H1RG detectors were manufactured as part of a test campaign to investigate the ability of MCT to perform

Table 2: Comparison of the current NOMIC detector and the two alternatives studied here.

Parameter	Aquarius SB383	GeoSnap 20561	H1RG 18508
Array size [pix]	1024×1024	1024×1024 ^a	1024×1024
Wavelength range [μm]	5–25	3–13	3–13
Quantum efficiency, QE [%]	40	65	65
Pixel size [μm]	30	18	18
Pixel size [mas]	17.8	10.7	10.7
Field of view [arcsec]	18	11	11
Read noise [electrons, e-]	400	170	100
Dark current [e-/sec]	100	300,000	1,000
Well depth [e-]	10 ⁶	1.2×10 ⁶	75,000
Gain [e-/ADU]	67	83	20
ADC bit depth	14	14	12
Min integration time [msec]	7.1	0.32	3.5
Operating temperature [K]	8	30–70	28–40

^a This particular GeoSnap device has a 2048×2048 pix format, but only a single quadrant is hybridized.

in the mid-IR with the intent of baselining these arrays for NASA's NEOCAM mission (PI: A. Mainzer). Based on the initial success of certain design elements, a series of 15 μm detectors with lower dark current have also been manufactured. We have acquired one of the original 13 μm (S/N 18508), which has been shown to be a good candidate for space applications running in the detector's Slow Mode operations (100 kHz pixel rate). However, given the low well depth, we need to evaluate the detector's Fast Mode operations at 10 MHz pixel rate, implying a frame time of ~ 7 msec (140 Hz frame rate).

GeoSnap is a new 2048×2048 imaging array readout recently developed by Teledyne for fast high-background infrared imaging. The large scale read-out integrated-circuit chip incorporates circuitry for timing and bias generation and 14 bit analog-to-digital signal conversion. Our unit provides two charge-well sizes, 10⁵ and 1.2×10⁶ electrons with read noise 30 and 170 electrons, respectively. The large-well conversion gain is ~ 83 electrons/digital-unit. The highest full-frame readout speed is 120 Hz, while typical operations are around 85 Hz. More recent ROIC versions have demonstrated well sizes up to 2.7×10⁶ electrons. Our unit has a 1024×1024 HgCdTe 13 μm wavelength cutoff detector bonded to one quadrant of the read-out. It is sensitive from 2 to 13 μm with quantum efficiency of 0.65 at 10 μm . The current device has not been CdZnTe thinned and is without antireflection coating, which should increase the QE to >85%. The array operates at temperatures between 30 and 70 K. The dark current is 3.3×10⁴ electrons/second at a temperature of 45 K with >98% pixels operative.

The specifications of the three detectors are listed in Table 2.

We have extensively tested the GeoSnap detector and have determined that our detector does not have the ELFN, but the read-out exhibits a 1/f noise which becomes significant below 1.0 Hz. We are working with the manufacturer to identify the cause and possible remediation of this noise.

We have also attempted to characterize the H1RG detector in the fast read-out mode. This work was delayed by the approval to carry out this testing as part of the sensitivity study which was received only in July 2019 and the long lead time for purchasing the necessary read-out electronics of six months.

The work to prepare the test cryostat has started on schedule in November 2019 and the detector was inserted in January. Despite the fact that this type of detector is designed to be able to operate in a fast read-out mode that would allow us to use it in the high background environment of ground-based infrared astronomy, we have not been able to operate the specific unit we have in this mode. The source of this problem was identified to be a single missing wire bond which according to Teledyne is likely an error in the bond program for specific H1RG test devices that no one has noticed until now. We are waiting for a cost estimate from Teledyne for adding this wire bond. However, Teledyne recently discovered that the latest revision of H1RG devices are now unable to correctly run Fast Mode at cryogenic temperatures, because the Fast Mode clocking and sampling develops a ‘race condition’ below a temperature of 100 K. We are discussing remedies for this with Teledyne. While this appears solvable, it may require manufacturing a new detector for our specific application if we were to move forward with a potential detector upgrade. Two unused HgCdTe wafers are available for free (W. Forrest, U. Rochester, personal communication), so that a suitable H1RG detector could be manufactured at only the cost of the read-out integrated circuit (ROIC) and the hybridization.

We are routinely operating with LMIRCam in the LBTI a comparable detector (H2RG, the ‘2’ standing for a size of 2048×2048 pixels, also the detector is not sensitive beyond $\sim 5 \mu\text{m}$) with the same readout electronics and in the Fast Mode we are hoping to operate the new NOMIC H1RG detector. We see no ELFN or $1/f$ noise on this detector. The tests of the GeoSnap detector with the same substrate as the H1RG have shown that the substrate does not produce the ELFN noise. We thus expect that a H1RG detector with the LW material will also not display any of those noise characteristics.

Another challenge is the low well depth of the H1RG detector (Table 2). We have directly scaled NOMIC counts measured from the HOSTS data to the H1RG detector parameters (gain, QE, pixel size, dark counts). We find that for the full frame readout of 7 ms, the background counts will be close to the detector’s saturation level. However, since the NOMIC field of view is smaller than the array size, we can enable vertical windowing to reduce the integration time by a factor of two, comfortably within the linear regime. We may be able to further tune the detector and to reduce the background (see Sect. 6.3), in combination with windowing, optimize the detector duty cycle. In any case, we confirm that we can operate the detector comfortably in the high background regime of NOMIC, once we are able to run it in Fast Mode as designed.

6 Other improvements studied as part of the sensitivity study

In addition to the analyses described in the previous section, we have studied and implemented other improvements to the instrument and observing strategy and the results from these investigations significantly inform our sensitivity study. This work is presented in the following.

6.1 SOUL adaptive optics upgrade

The LBTI’s adaptive optics system has been upgraded as part of the LBT wide SOUL (Single conjugated adaptive Optics Upgrade for the LBT) upgrade. It provides a higher number of correctable modes (in practice 500 modes vs. 300 modes for the old system), faster loop speed (up to 1.8 kHz compared to 1 kHz for the old system), and a 3 mag fainter limiting magnitude. LBTI’s left (SX) AO system has been upgraded in summer 2018 and has been successfully used on sky since semester 2018B. LBTI’s right (DX) AO system has been upgraded in February 2019 and has been successfully used on sky since semester 2019A. Both systems are operational for science and already provide better performance and

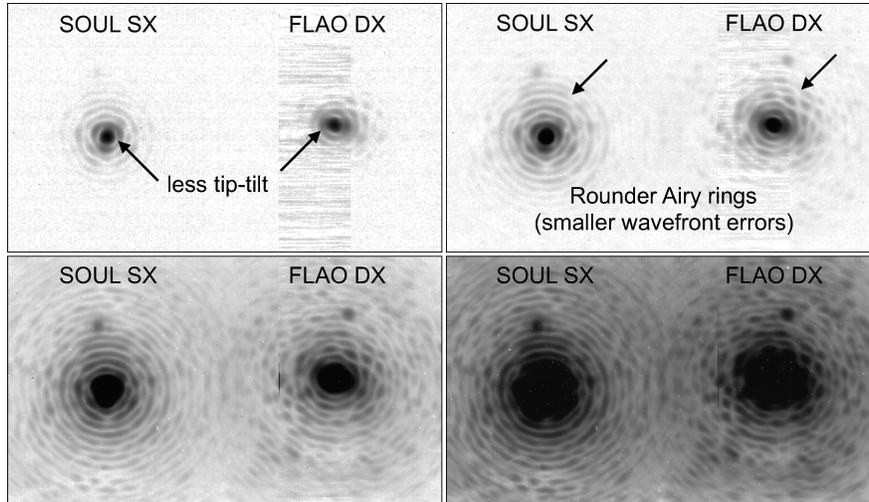


Fig. 9: Double sided images obtained with LBTI/LMIRCam in high contrast imaging mode of a bright star representative for the targets observed during HOSTS. The images of the star from the SX (left) aperture using the new SOUL system and DX (right) aperture using at that time still the old FLAO system were placed next to each other on the detector. Four different exposure times were used to show different scales of the point spread functions. The features illustrating the better performance of the SOUL system are highlighted in the images. The DX image was placed by accident on a noisy detector channel which is of no consequence for nulling as LMIRCam is not used there. Some non-common path aberrations are visible for SX (triangular shape of the first airy ring in the upper-left panel) which will be improved during further SOUL commissioning and is also of no significant consequence for nulling. The faint ghost above the star originates from the trichroic beam splitter that sends the L and M band light to LMIRCam and it's position between the two sides is different because the two beams hit this beam splitter at slightly different angles in this imaging mode.

stability (i.e., observing efficiency) than the old system. Final commissioning is ongoing.

Commissioning and scientific use of the new SOUL system was so far driven by the majority of the science projects in the queue, which were largely single or individual aperture AO imaging (no coherent combination of the light from the two apertures for interferometry). We have demonstrated the targeted improvements of the limiting magnitude and wavefront correction. Fig. 9 shows a comparison of the old system on DX and the new system on SX obtained on Dec. 24th 2018, where we pointed to the same star with both sides and placed the images next to each other on the detector. The L' filter (our most common imaging filter) was used. The magnitude of the target star is $R = 3.1$ and both AO systems were used at peak performance, both typical for HOSTS observations. Atmospheric conditions were also within specifications for nulling observations. We can see in these images the signs of improved tip-tilt correction and reduced wavefront errors. Both demonstrates that the new system provides at least the performance in two of the three main parameters relevant for nulling interferometry that the old system provided.

The characterization of the system for nulling interferometry has so far been limited by the small number of relevant observations in our observing queue, which were also assigned lower TAC priority

than other observations, and the limited availability of interferometric modes while various instrument upgrades were ongoing. A suitable nulling sequence on a calibration star was obtained on UT2019-04-20 with the goal to confirm performance of the SOUL system for nulling interferometry. Observing parameters such as loop speed and number of modes were chosen to be comparable to those used during the HOSTS survey, so that the performance could be compared. Despite relatively high wind speed, the OPD variations measured by PhaseCam (both at high and low frequency compared to the NOMIC frame rate) were typical for nulling observations carried out with the old AO system during the HOSTS survey. This confirms that the third parameter of the SOUL performance that is relevant for interferometric observations, the magnitude and frequency of piston variations due to vibrations, is also at least consistent with the old system and likely superior.

Since April 2019 we have made further progress with commissioning SOUL, including a vibration filter that is expected to further suppress a high frequency vibration that was detected when running the SOUL system at peak speed. Further characterization after the final commissioning of the SOUL system and based on a larger nulling data set is required for a final conclusion of the performance of the new system, but we can confirm that the new system is operational and suitable for nulling interferometry, at least delivering comparable performance to the old system.

6.2 Larger setpoint dither

The setpoint is the K band phase between the two beams on PhaseCam that is tracked during observations. The best setpoint is determined by minimizing the N band phase by minimizing the transmitted flux on NOMIC. This approach is prone to uncertainties and there is a degeneracy between a setpoint error and an astrophysical signal due to circumstellar dust. Setpoint dithering is used to add a periodic, well defined modulation to the setpoint. The transmitted flux in the 'off' position of the phase can then be used in the data reduction to break this degeneracy.

During the HOSTS survey, a phase offset of 0.2 rad in the N band was used. When the HOSTS survey was completed, we obtained an observing sequence with offsets of both 0.2 rad and 0.3 rad. The latter improved the accuracy of the determination of the setpoint error by a factor of two. Multiple on-sky nulling data sets are required to determine the ideal dither offset.

6.3 Improved cold stops for background reduction

The design of the LBT has a hole in the center of the primary mirror to host an instrument in the direct Gregorian focus. This hole has an emissivity of 1 compared to much lower emissivity (approx. 0.04 for the primary, secondary, and tertiary mirrors) of the telescope optics. This emission is currently not blocked by a cold stop. In addition, the current cold stop that blocks the light from outside the telescope pupil is located far downstream in the LBTI, close to the NOMIC detector, where only a blurry image of the pupil is formed. This results in a light leak from the outer edge of the telescope pupil which also increases the background.

From a geometrical model, we estimate the contribution from the primary hole to the total thermal background to be 18%. Imaging the pupil and measuring the contribution directly confirms this result Fig 10 (although the reliability of the latter is limited because of the way the pupil image is formed). The contribution from the light leak due to the sub-optimal location of the cold stop cannot be easily estimated.

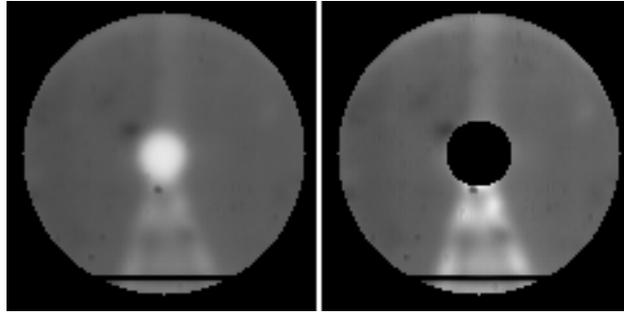


Fig. 10: Pupil image of NOMIC showing the bright central hole of the primary mirror in the left image (central, bright spot). The swingarm holding the secondary mirror is also visible. The horizontal line at the bottom is a detector channel edge (not a physically dark area). To the right image, a software mask was added with the size of the aperture used to measure the contribution of the central hole its contribution to the light on the pupil. An outer software mask was added to illustrate the area over which the total flux was computed.

We are currently upgrading the pupil mirrors on the pathlength correctors in the LBTI (one for each telescope aperture). A much sharper pupil image is formed on these mirrors. The upgraded mirrors have masks for both the outer edge of the telescope pupil and the central hole. The right side mirror has been installed recently and a quick measurement of the background on sky has been obtained. This measurement suggests a reduction of the background of approx. 40% is achieved with the new mirror. This is a factor of two larger than our estimate from the central hole alone, suggesting that the rest of the emission comes from the light leak around the old cold stop. These measurements are encouraging but need to be confirmed during further commissioning of the new pupil mirrors.

This background improvement has several positive effects:

- The background photon noise is reduced.
- The background itself is reduced helping to operate the H1RG detector at longer integration times for better efficiency.
- Even if the ELFN noise on NOMIC is not reduced (no new detector can be installed), the reduction of the background will directly transform into a reduction of the ELFN noise (by the same factor as the background flux reduction) as the effect is multiplicative with the flux on the detector and the background is our main source of flux.

7 New error budget

Based on the potential improvements discussed in the previous sections, we now create an updated error budget for potential renewed HOSTS observations. We first summarize the modifications to the input parameters of the error budget:

- The low frequency phase jitter is reduced by 1/3 due to reduced telescope vibrations as discussed in Sect. 5.2. We adopt a value of 0.06 rad compared to a median of 0.09 rad during the original HOSTS survey. The uncertainty of determining this parameter is reduced by a factor of two, which has also been measured from the HOSTS data of the full sample and the low vibration sample described in Sect. 5.2.

- The background bias is reduced due to a new detector with no low frequency noise as discussed in Sect. 5.3. We adopt a conservative value of 0.9 mJy compared to 2 mJy for the original HOSTS survey. We also explore a slightly more aggressive improvement by a factor of 3 to 0.67 mJy.
- The uncertainty estimating the phase setpoint error is reduced by a factor of 2 due to larger setpoint dithering. The setpoint search algorithm is not expected to improve, so that the setpoint error itself is assumed to remain unchanged.
- The acquisition parameters are updated to a DIT of 7 ms (realistic DIT for the new detector given its low well depth but fast readout). Accordingly, we assume a total of 5000 frames per nod, a nodding frequency increased by a factor of 1.5, and 10 nods per pointing. This produces improved statistics due to the larger number of frames and nods without significantly increasing overheads for nodding and is thought to be close to the ideal way to operate with the new detector. The total execution time of a calibrated nulling observation does not change significantly using these parameters.
- The detector parameters (pixel size, quantum efficiency, read-out noise, gain) are updated following Sect 5.3.
- The background flux is reduced by a conservative 20% from the new pupil mirror masks (Sect. 6.3).
- We conservatively assume that the performance of the SOUL AO system is similar but not improved compared to the old system.

The two significant improvements in the above list are the improvement of the vibrations and the improvement of the background bias. We explore three cases. (1) The vibrations are improved, but the background bias cannot be improved. (2) The background bias is improved, but the vibrations are the same as during the HOSTS survey. (3) Both the background bias and the vibrations can be improved as expected. For the improvement of the background bias we show the two cases improvements by a factor of two and three. We illustrate the outcome of these scenarios on the example of repeating the same observations (same targets, same number of pointings per target) of the original HOSTS survey. We compare the *predicted* errors from our available data with those for future data. Fig. 11 shows the predicted improvements for each case. It is visible that the improvements of the vibrations affect mostly the sensitivity of stars brighter than 2-3 Jy. This is because the background bias dominates for the faintest stars. For stars brighter than 3 Jy, the improvement is typically a factor 1.5 and for the brightest stars it can reach a factor of a few. The scatter for brighter stars is due to the fact that they have been observed at a range of vibration environments, some already at low vibrations (smaller improvements) and some at high vibrations (larger improvements). This is visible when comparing current predicted errors and future ones because the current predicted errors use the actual vibration telemetry of the observing night. Improving the background bias has the strongest effect for faint stars and typically provides an improvement of a factor of two in sensitivity, slightly higher if the background bias can be reduced by a factor of three. If both can be improved, the resulting sensitivity is predicted to be about a factor of two or more better than the predictions for the available HOSTS data.

We conclude that the dominating limitation to our sensitivity is currently the background bias. If this bias can be reduced as expected, an improvement of a factor of two to three depending on the star is entirely realistic. Fig. 12 shows the comparison between our the prediction of our current uncertainties and the predicted uncertainties if both the vibrations and the background bias can be improved. It is important to note that some of our strongest improvements can be made for some of the most interesting stars such as Vega, ϵ Eri (both interesting detections to be characterized) and τ Ceti (an important nearby Sun-like star with known planets and a so far mediocre upper limit on habitable zone dust). These stars

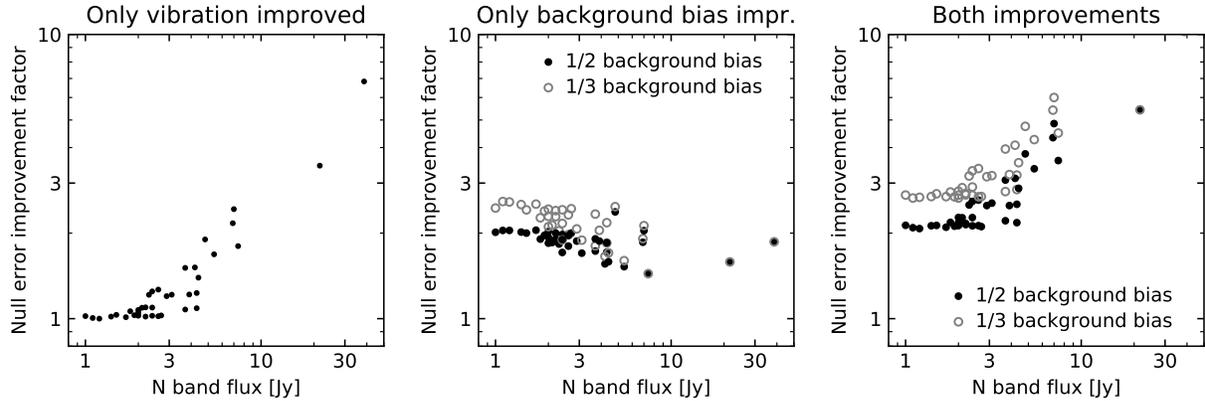


Fig. 11: Improvements to the predicted uncertainty of the null depth measurements (current prediction divided by prediction after improvement) for improved vibrations (*left*), improved background bias (*center*), and both (*right*). Predictions have been made for the stars observed by the original HOSTS survey assuming the same amount of data for each star has been obtained again. The cases of an improvement of the background error by both a factor of two and a factor of three are shown.

are all bright so that even an improvement of only the vibrations can already have an important impact.

8 Potential for a renewed HOSTS survey

In this Section we briefly discuss the implications of our predicted sensitivity improvements for a potential, renewed HOSTS survey. A detailed proposal for a renewed survey is however beyond the scope of this report.

The implications of the HOSTS results for future exo-Earth imaging missions have been discussed in detail by Ertel et al. (in press). We found a best-fit median habitable zone dust level of 3 zodis from our sample, where 1 zodi is the level in the Solar system. At this level, LUVOIR, HabEx, and the Starshade Rendezvous missions are expected to be capable of achieving the benchmark O_2 spectral detection at a wavelength of $0.76 \mu\text{m}$ with around their median sample target at a continuum signal-to-noise ratio of ten in reasonable integration times (less than 60 days). However, our measured median zodi level comes with a large uncertainty, so that the 1σ upper limit is 9 zodis and the 95% upper limit is 27 zodis. While at 9 zodis LUVOIR and HabEx can still achieve their O_2 measurements, WFIRST will have to relax its target signal-to-noise to ~ 8 . At 27 zodis HabEx would not reach its target signal-to-noise. A two to three times stronger median zodi constraint as achieved by the predicted sensitivity of improved LBTI observations would allow us to conclude with high confidence that exozodiacal dust is not a risk to the HabEx mission concept, and with good confidence that it has only a small impact for the Starshade Rendezvous concept.

It is important to note that new observations do not have to replace or supersede the current results, but instead can add to the constraints already achieved. Fig. 13 shows the full HOSTS sample from Weinberger et al. (2015). It is visible that in particular among the Sun-like sample there is a significant number of stars for which we are the most sensitive to habitable zone dust that have not yet been

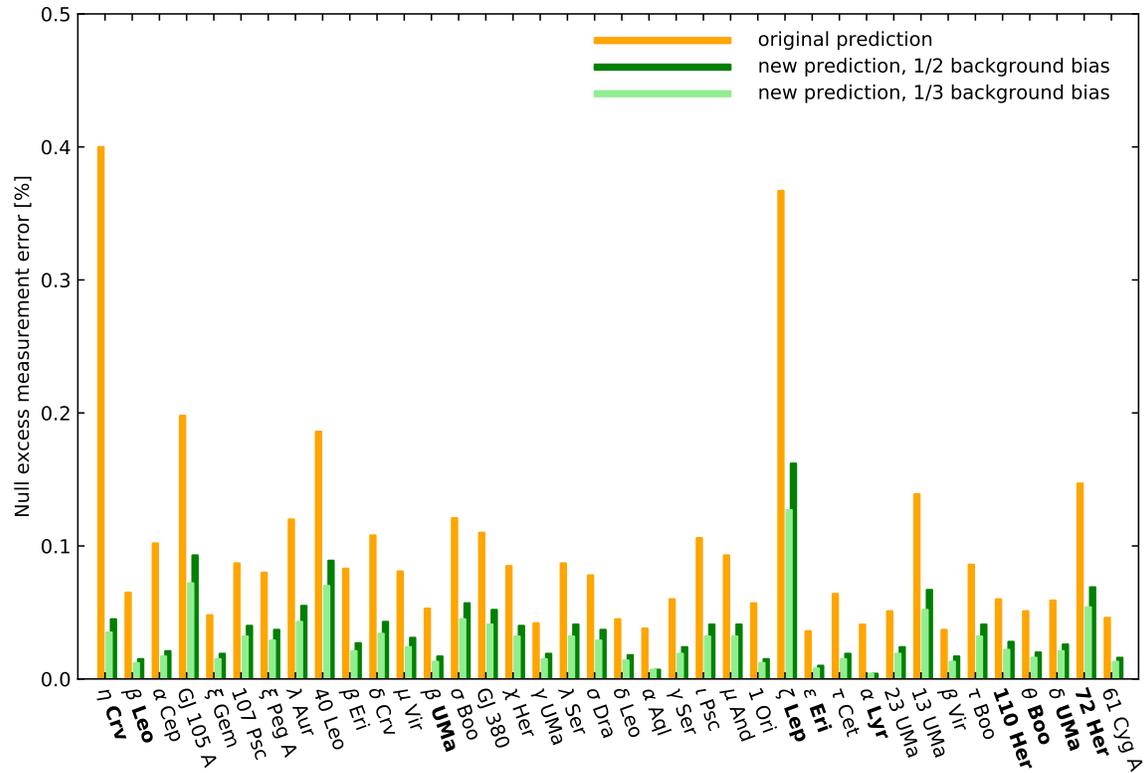


Fig. 12: Comparison of achieved null uncertainty from the HOSTS survey, expected uncertainty for the system during the survey, and predicted uncertainty after all improvements. Stars with detected excess are highlighted in bold face. The cases of an improvement of the background error by both a factor of two and a factor of three are shown. It is clear that the former has a critical impact, while the latter only results in a small additional improvement of the overall sensitivity.

observed. Even at the current sensitivity, observing those stars would add to the median zodi level and zodi luminosity function constraints. It would also allow us to vet a larger fraction of exo-earth imaging targets, removing the stars with the brightest exozodis as unsuitable. With an improved sensitivity, a combination of observing those stars and re-observing stars for which we are particularly sensitive and for which the strongest improvements over the available data can be made (e.g., the brighter ones for which the sensitivity improvement exceeds a factor of 3) can have the strongest impact and can improve the median zodi level and zodi luminosity function constraints beyond the naive improvements expected from the predicted higher sensitivity of the instrument. We have estimated the improvements based on the original HOSTS survey which include all the imperfections expected from real, ground-based observations. We conclude that a renewed survey that is optimized for complementing and improving over the existing data can likely provide a median exozodi constraint three times better than currently available. With the predicted sensitivity and using our best-fit luminosity function from the HOSTS survey, we expect to detect bright exozodis around 10% to 30% of these stars, identifying a significant fraction of unsuitable targets for future exo-Earth imaging missions.

Another result from the HOSTS survey is that most of the detected, massive exozodiacal dust disks have been found around stars with massive cold disks detected by Spitzer and Herschel. This would suggest that the habitable zones of stars without known cold disks are less dusty than those in our whole sample. Better statistics and in particular a confirmation of this trend at lower HZ dust levels are needed to confirm this hypothesis. In addition, we can predict with high confidence that we will detect HZ dust around almost all stars with known cold dust, even with the current sensitivity of the LBTI. This opens up the possibility of a targeted survey of such disks with the goal to study them in detail. Such a study would produce a deep understanding of the origin of the habitable zone dust which will provide the means to predict the zodi levels around stars that could not be observed with the LBTI.

Finally, as discussed, e.g., by Ertel et al. (in press), the HOSTS results are not only enabling exo-Earth imaging missions. The HOSTS data have also the potential to critically complement the data from such missions by constraining the environment in which potentially habitable planets exist. The observations constrain the dynamics of the planetary and planetesimal systems observed, the presence and locations of asteroid belts, and the presence and influence of shepherding planets on the dynamics, and the strength of cometary activity. The latter will ultimately have a critical impact on whether a detected planet is habitable or not as cometary impacts may deliver water but also erode planetary atmospheres (Kral et al., 2018). The HOSTS data can thus critically contribute to the understanding of the detected planets. An increased sensitivity and observing more stars now will ensure that more systems with potential exo-Earth detections will be characterized at the time an exo-Earth imaging mission is launched.

9 Summary and conclusions

We have shown that our detailed error budget successfully predicts the observed null uncertainties measured during the HOSTS survey, especially in the latter two-thirds of the survey after the observing and data collection strategies had fully matured. Most of the input parameters in the error budget can be compared to independently measured quantities in the instrument and telescope, and are in agreement. This gives us good confidence that we can modify parameters in the error budget to accurately predict the performance of an upgraded instrument. In addition to validating the theoretical error budget by comparing it with our survey data, we have used the variable vibration environment of the telescope throughout the survey to empirically study the correlation between telescope vibrations and our null

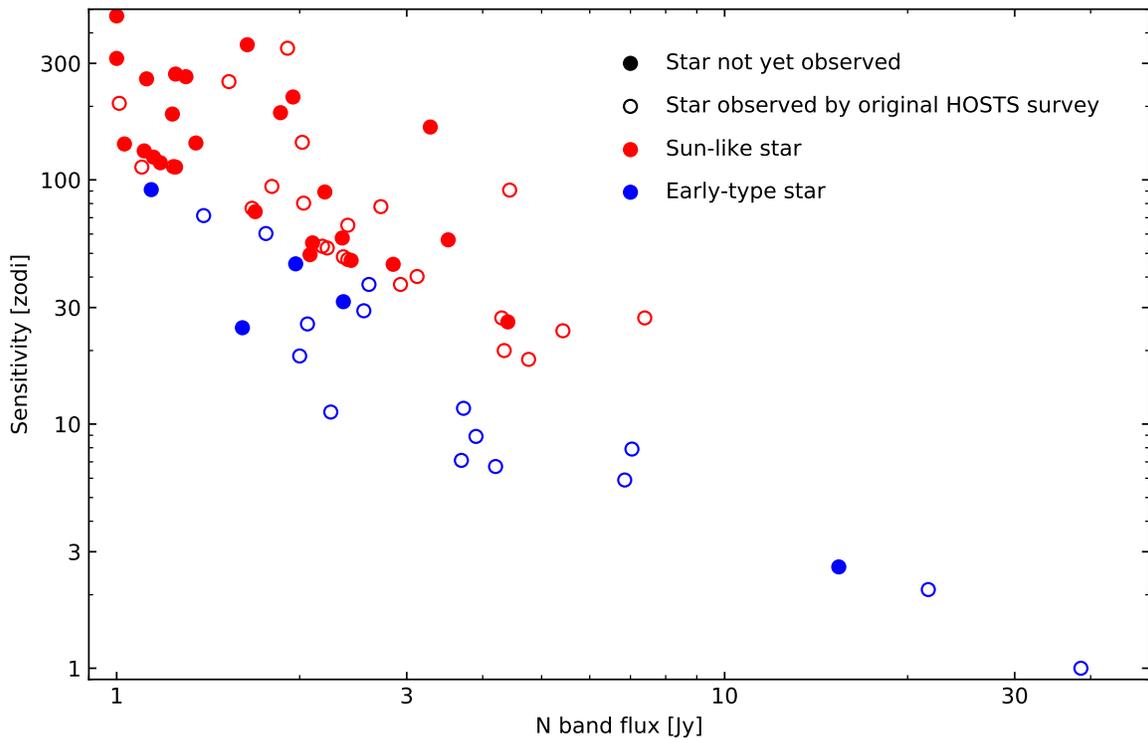


Fig. 13: Full HOSTS sample from [Weinberger et al. \(2015\)](#). The sensitivity predictions have been updated according following the results of our sensitivity study presented in this report.

accuracy and to investigate realistic mitigations of the vibrations and their impact on our data quality. The results from our theoretical error budget and the empirical analysis are in good agreement.

We also investigated whether the null uncertainty was correlated with observing parameters not easily characterized within the error budget, including seeing, wind speed, wind direction, and precipitable water vapor. None of these parameters showed any obvious impact on null uncertainty within the constraints set for suitable conditions during the survey. One less quantifiable parameter that did appear to impact null uncertainty is the presence of thin cirrus. Thin cirrus is readily apparent in the real time quality control of the data during acquisition as a strongly temporally varying sky background level, and is easily avoided during observations.

We found that the expected null uncertainty in exozodi measurements comes mostly from two sources. Differential optical pathlength variations due to low frequency vibrations in the telescope dominate the null uncertainty for brighter stars in the survey, and low frequency spatio-temporal background variations due to excess low frequency noise (ELFN) of the mid-infrared detector dominate the uncertainty for fainter stars. The LBTI pathlength control servo already includes feedback and accelerometer feedforward mechanisms to reduce the impact of the 12 Hz vibrations – the OVMS rms of $\sim 1 \mu\text{m}$ is reduced about six times by the servo to be $0.16 \mu\text{m}$ rms in the error budget – but these could be improved to further reduce the impact of vibrations. The vibrations themselves could also potentially be reduced at their source by adding damping to the secondary mirror swing arms. A further 30% reduction of the differential optical pathlength variations appears to be realistic.

The background variations due to ELFN in LBTI's Aquarius detector have been characterized independently from the HOSTS survey data, and the amount of variation measured is consistent with most of the null uncertainty seen for fainter stars. The only way to eliminate the ELFN is to use a different detector in the instrument. HgCdTe detectors provided by Teledyne are free of ELFN. The accompanying changes of the operating parameters (quantum efficiency, well depth, read noise, integration time, ...) are taken into account in our error budget. The H1RG detectors are also free of significant excess $1/f$ noise. Due to the limited time frame of the sensitivity study, we were not able to perform the characterization of the H1RG detector that would permit us to confirm all relevant detector characteristics. Nevertheless, a factor of two to three reduction in background bias appears realistic. We are continuing the characterization of the H1RG detector.

Modifying the error budget to account for these upgrades improves the expected null uncertainties by a factor of two to three, depending on stellar N band flux, and by more for the brightest stars.

Other instrument upgrades were considered within this study that had smaller impact on nulling uncertainty, but could improve performance still further or provide technical margin on the major improvements listed above. One upgrade, the SOUL adaptive optics upgrade, has already been installed and is working on the instrument. It has been demonstrated to provide similar performance for nulling interferometry compared to the old system used during the HOSTS survey, while possible improvements remain to be demonstrated. In particular, it could extend observations to conditions of worse seeing or to stars at lower altitudes. Larger set point dithers during observations can reduce the mean phase set point uncertainty by a factor of two, which could provide modest reduction in null uncertainty for brighter stars. Lastly, cold stops to block the thermal background from the primary mirrors' Gregorian ports and reduce leakage of thermal light around the edges of the pupil would reduce background levels potentially by 40%. Even if the NOMIC detector were not replaced, this lower background would reduce the effect of ELFN by a similar amount, yielding a modest performance improvement similar to (but

smaller than) that achieved by a detector upgrade.

A renewed HOSTS survey can likely provide a median exozodi constraint three times better than currently available. This would allow us to conclude with high confidence that exozodiacal dust is not a risk to the HabEx mission concept, and with good confidence that it has only a small impact for the Starshade Rendezvous concept.

References

- Defrère, D., Hinz, P. M., Mennesson, B., et al. 2016, ApJ, 824, 66
- Ertel, S., Defrère, D., Hinz, P., et al. 2018, AJ, 155, 194
- Hanot, C., Mennesson, B., Martin, S., et al. 2011, ApJ, 729, 110
- Hinz, P. M., Defrère, D., Skemer, A., et al. 2016, in Proc. SPIE, Vol. 9907, Optical and Infrared Interferometry and Imaging V, 990704
- Hoffmann, W. F., Hinz, P. M., Defrère, D., et al. 2014, in Proc. SPIE, Vol. 9147, Ground-based and Airborne Instrumentation for Astronomy V, 914710
- Kennedy, G. M., & Wyatt, M. C. 2013, MNRAS, 433, 2334
- Kennedy, G. M., Wyatt, M. C., Bailey, V., et al. 2015, ApJS, 216, 23
- Kral, Q., Wyatt, M. C., Triaud, A. H. M. J., et al. 2018, MNRAS, 479, 2649
- Mennesson, B., Serabyn, E., Hanot, C., et al. 2011, ApJ, 736, 14
- Mennesson, B., Millan-Gabet, R., Serabyn, E., et al. 2014, ApJ, 797, 119
- Mennesson, B., Defrère, D., Nowak, M., et al. 2016, in Proc. SPIE, Vol. 9907, Optical and Infrared Interferometry and Imaging V, 99070X
- Stapelbroek, M. G., Petroff, M. D., Speer, J. J., & Bharat, R. 1984, in Proc. IRIS Detector, Vol. 2, 63
- Weinberger, A. J., Bryden, G., Kennedy, G. M., et al. 2015, ApJS, 216, 24

Acknowledgements

The research was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004), and at the University of Arizona under subcontract #1561710 with the Jet Propulsion Laboratory.