Delft University of Technology

Asgard/NOTT

L -band nulling interferometry at the VLTI: I. Simulating the expected high-contrast performance

Laugier, Romain ; Defrère, Denis; Woillez, Julien ; Courtney-Barrer, Benjamin; Dannert, Felix A. ; Matter, Alexis ; Dandumont, Colin ; Gross, Simon ; Loicq, J.J.D.; More Authors

DOI
10.1051/0004-6361/202244351

Publication date
2023

Document Version
Final published version

Published in
Astronomy & Astrophysics

Citation (APA)
Laugier, R., Defrère, D., Woillez, J., Courtney-Barrer, B., Dannert, F. A., Matter, A., Dandumont, C., Gross, S., Loicq, J. J. D., & More Authors (2023). Asgard/NOTT: L -band nulling interferometry at the VLTI: I. Simulating the expected high-contrast performance. *Astronomy & Astrophysics*, *671*, Article A110. https://doi.org/10.1051/0004-6361/202244351

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

A&A 671, A110 (2023)
https://doi.org/10.1051/0004-6361/202244351
© The Authors 2023

Astronomy & Astrophysics

Asgard/NOTT: *L*-band nulling interferometry at the VLTI

I. Simulating the expected high-contrast performance

Romain Laugier[1], Denis Defrère[1], Julien Woillez[2], Benjamin Courtney-Barrer[2,10], Felix A. Dannert[3,4], Alexis Matter[5], Colin Dandumont[6], Simon Gross[7], Olivier Absil[6], Azzurra Bigioli[1], Germain Garreau[1], Lucas Labadie[9], Jérôme Loicq[6,8], Marc-Antoine Martinod[1], Alexandra Mazzoli[6], Gert Raskin[1], and Ahmed Sanny[7,9]

[1] Institute of Astronomy, KU Leuven, Celestijnenlaan 200D, 3001 Leuven, Belgium
e-mail: `romain.laugier@kuleuven.be`
[2] European Southern Observatory Headquarters, Karl-Schwarzschild-Straße 2, 85748 Garching bei München, Germany
[3] ETH Zurich, Institute for Particle Physics & Astrophysics, Wolfgang-Pauli-Str. 27, 8093 Zurich, Switzerland
[4] National Center of Competence in Research PlanetS, 3012 Bern, Switzerland
[5] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, France
[6] STAR Institute, University of Liège, 19C allée du Six-Août, 4000 Liège, Belgium
[7] MQ Photonics Research Centre, School of Mathematical and Physical Sciences, Macquarie University, NSW 2109, Australia
[8] Faculty of aerospace Engineering, Delft University of Technology, 2629 HS Delft, The Netherlands
[9] I. Physikalisches Institut, Universität zu Köln, Zülpicher Str. 77, 50937 Köln, Germany
[10] Research School of Astronomy & Astrophysics, Australian National University, Canberra ACT 2611, Australia

Received 24 June 2022 / Accepted 18 November 2022

ABSTRACT

Context. NOTT (formerly Hi-5) is a new high-contrast L' band (3.5–4.0 μm) beam combiner for the VLTI designed with an ambitious aim to be sensitive to young giant exoplanets down to 5 mas separation around nearby stars. The performance of nulling interferometers in these wavelengths is affected both by fundamental noise from the background and contributions of instrumental noise. This motivates the development of end-to-end simulations to optimize these instruments.

Aims. The aim of this study is to enable a performance evaluation of NOTT and inform the design of such instruments with current and future infrastructures in mind, taking into account the different sources of noise and their correlation.

Methods. SCIFYsim is an end-to-end simulator for single-mode-filtered beam combiners, with an emphasis on nulling interferometers. We use it to compute a covariance matrix of the errors. We then use statistical detection tests based on likelihood ratios to compute compound detection limits for the instrument.

Results. With the current assumptions as to the performance of the wavefront correction systems, the errors are dominated by correlated instrumental errors down to stars of magnitude 6–7 in the *L* band, beyond which thermal background from the telescopes and relay system becomes dominant.

Conclusions. SCIFYsim is suited to anticipating some of the challenges of design, tuning, operation, and signal processing for integrated-optics beam combiners. The detection limits found for this early version of NOTT simulation with the unit telescopes are compatible with detections at contrasts up to 10^5 in the *L* band at separations of 5–80 mas around bright stars.

Key words. techniques: interferometric – techniques: high angular resolution – methods: data analysis – methods: statistical

1. Introduction

While thousands of exoplanets have been detected by indirect means, the direct detection of their emitted or reflected light is so difficult that only a relatively small number of giant, hot young planets on very wide orbits have been detected so far. This is in part due to the fact that the limited aperture of even optically perfect instruments produces an extended point spread function (PSF) as the image of point-like sources. As a result, the images of faint exoplanets are buried into the wings of the image of their host stars, and the associated photon noise dominates the signal-to-noise ratio (S/N). In the context of high-contrast detection, the resolving power can no longer be expressed with simply a minimum angle, but must be considered as a map of the attainable contrast as a function of the position of the planet relative to the star.

Coronagraphs have been developed and used to optically alter the response of instruments, effectively redirecting the on-axis light out of the science imaging channel. Such devices have reached a maturity in that design approaches and manufacturing tools are able to produce coronagraphs with inner working angles down to $0.5\lambda/D$ of the collecting aperture (Guyon et al. 2006).

Interferometry offers a way to increase the effective extent of collecting apertures, therefore increasing the maximum achievable angular resolution. However, the measurement of the interferometric fringes, much like the recording of images, is plagued by the same effect of photon noise carried by the (virtual) PSF of the considered aperture array. In the case of long baseline interferometry, this PSF evolves with the projection of the array on the plane orthogonal to the line of sight. Compared to the PSF produced by a monolithic telescope, the sparsity of this array

Open Access article, published by EDP Sciences, under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
This article is published in open access under the Subscribe to Open model. Subscribe to A&A to support open access publication.

results in a finer bright pattern with wings extending over the full effective field of view.

While pupil densification has been proposed to improve part of this slow falloff effect (Tcherniavski 2011; Guyon et al. 2005), recent works have shown (Lacour et al. 2019b; Nowak et al. 2020b,a; GRAVITY Collaboration 2021) that interferometry can be combined with efforts to optically discriminate the off-axis light from the starlight in the single-dish beams. In observations with the GRAVITY instrument (Lacour et al. 2019b), the spatial filtering provided by single-mode optical fibers in each of the beams is used to isolate the light of a known exoplanet from most of the light of its host before sending it to the beam combiner. On the detector, the planet light and the residual starlight form two distinct sets of fringes that can be effectively discriminated to reconstruct the spectrum of the planet.

Yet, to take advantage of the longer baselines as coronagraphs take advantage of larger diameter apertures, a different approach to beam combination must be taken that gives the on-axis light a special treatment. This is required in order to prevent the photon noise it carries from affecting the scientific measurement. This approach is called nulling interferometry (Bracewell 1978), and entails tuning the phases of the combined beams so as to send the dark fringe of the on-axis light to the main detector, while its bright fringe is measured separately. On the dark channel, called the interferometric null, light from an off-axis source may still get recorded, depending on its location in the field.

This type of beam combination is difficult to use, because the path length in the different beams must be matched to within a fraction of the wavelength. Now that the interferometric facilities are equipped with high performance fringe tracking systems, the use of a nulling beam combiner becomes appealing (Lacour et al. 2019a; Anugu et al. 2020; Pannetier et al. 2021).

The Nulling Observations of exoplaneTs and dusT (NOTT[1], formerly Hi-5) instrument is a proposed high-contrast visitor instrument for the VLTI (Defrère et al. 2018a,b, 2022) that will make use of nulling beam combination to allow the characterization of young giant planets in the L' band down to the snow line of their host system. It will leverage integrated photonic technology for beam combination.

Such photonic devices can implement optical waveguides integrated into monolithic glass substrates. Interference between the light from different apertures is obtained by coupling the wave guides, for example by bringing them close together so that their evanescent field can be coupled (Tepper et al. 2017), or by fusing them into sections that guide a larger number of modes (Soldano & Pennings 1995). These different approaches offer different characteristics, in particular with respect to their chromatic behavior. Such devices can be extremely compact, which offers the possibility to build complicated combiner architectures in small and stable packages.

NOTT is designed to operate as part of the Asgard suite (Martinod et al. 2022), a VLTI visitor instrument that also includes the HEIMDALLR high-precision fringe tracker (Ireland et al. 2018) and the Baldr wavefront sensor (based on Zernike wavefront sensor), which are currently under development.

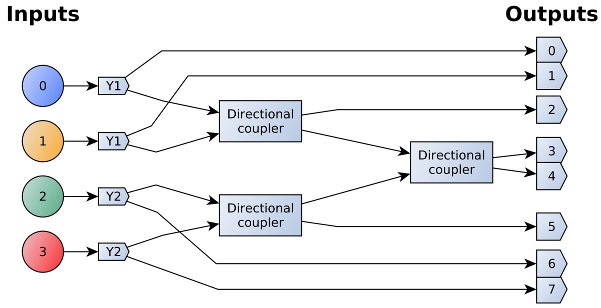


Fig. 1. Schematic of the photonic beam combiner including, from left to right, the four inputs of the chip, and photometric taps labeled Y1 and Y2, the first combination stage composed of two directional couplers, and the second stage consisting of a single directional coupler mixing the dark outputs of the first stage. The chip has eight outputs, including photometric outputs (0, 1, 6, and 7), bright outputs (2 and 5), and two complementary dark outputs (3 and 4).

With the capability to leverage baselines of more than 100 m, other limitations have been identified. Indeed, as the stellar disk loses its spatial coherence for the baseline (i.e. as the stellar disk becomes resolved), some of this light leaks to the dark outputs, leading to confusion with the planet light. This has been an important driver for the evolution of nulling combiners from the original idea of Bracewell (1978), implementing more than two apertures with the Angel & Woolf combiner (Angel & Woolf 1997; Velusamy et al. 2003), and then even more complicated designs (Guyon et al. 2013). Nevertheless, with the constraints of observing from the ground, the measurement errors due to varying optical path differences (OPD) are a dominating factor. More recent approaches have focused on this problem, producing closure phase from the leakage light of a nuller (Lacour et al. 2014), or designing specific combinations that cancel out the second derivative of the phase errors in kernel nulling (Martinache & Ireland 2018). In light of the analysis proposed by Laugier et al. (2020), one can argue that the desirable robustness properties of kernel nullers apply to the sin-chop configurations of the Angel & Woolf combiners. Indeed the two nulled outputs are constructed with contributions from all four inputs multiplied by vectors of phasors that are enantiomorph (i.e. that mirror one another), like the pairs of outputs of kernel nullers or the outputs of complementary phase-chopping states investigated by Lay (2004) and Defrère et al. (2010). This gives these two output intensities correlated responses to instrumental errors, and gives their difference a desirable robustness to instrumental errors.

Other advantages can be argued in favor of the simpler Angel & Woolf design, like a higher raw throughput, and the possibility of external tuning, making it a technological stepping-stone towards the more complicated kernel-nullers that offer three times the number of observable quantities. The combination scheme anticipated for the photonic combiner of NOTT is illustrated in Fig. 1 and includes photometric taps on each of the inputs for monitoring the injection rate.

Our goal is to estimate the contrast performance of such a nulling combiner instrument and its sensitivity to the presence of a planet around nearby stars. In the literature, this has been done in different ways. Some studies targeting space-based missions (e.g., Mugnier et al. 2006; Dannert et al. 2022) focus exclusively on the fundamental noise sources, such as photon noise, background noise, and static leakage sources, and limit their conclusions to regimes that are dominated by these contributions.

Important efforts were led by Lay (2004) in the context of TPF-I to account for instrumental errors. These authors made use of second-order expansion of the effect of the perturbations on the interferometric outputs and remain a reference for this type of study. However, this is limited to the regime of small aberrations where first- and second-order terms dominate, and to the case of ideal beam combination with a continuously rotating array. In the case of a ground-based instrument working at smaller wavelengths, the atmospheric contributions are expected to play a dominant role in overall errors as was highlighted in the cases of GENIE (Absil et al. 2006) and ALADDIN (Absil et al. 2007). Furthermore, the instrumental noise is expected to be highly correlated between different wavelength bins, an effect that has in large part been neglected.

---

[1] Nott is the personification of night and darkness in Norse mythology.

Works by Ireland (2013), Ceau et al. (2019), and Kammerer et al. (2020) highlighted the performance benefits of accounting for correlations in interferometric data. In Thiébaut & Mugnier (2006), Mugnier et al. (2006) and Dannert et al. (2022), only the fundamental noise sources (such as photon noise) are considered. These are independent in classical spectrographs, as the spectral channels are recorded in different pixels. Lay (2004) mentions the temporal correlations between spectral channels but does not take them into account in the determination of the overall performance. The straightforward performance analysis offered by Martinache & Ireland (2018) considers broadband observations (a single spectral channel), and does not account for correlations that would arise between the three kernel-null outputs. To our knowledge, the correlations in the nulling interferometry data have only been considered by Lay (2006) in an approach that relies on this correlation to propose an ad hoc approach of polynomial fitting and subtraction. The author then concludes that the success of this approach depends on the geometry of the array and the resulting temporal behavior of the signal of interest.

To enable the exploration of the opportunities offered by advanced combiners, we developed a new simulation tool specifically tailored to complicated photonic beam combiners, and compatible with the demands of nulling. Here, we describe SCIFYsim[2], a new end-to-end simulator package written in Python and including Fourier optics tools for its bulk optics components and complex amplitude matrix manipulation for its single-mode components. We then show some of the early results that it provides in the effort to evaluate the performance of NOTT.

Nulling self-calibration (Hanot et al. 2011; Defrère et al. 2016; Mennesson et al. 2016; Norris et al. 2020; Martinod et al. 2021) is a different approach that makes use of the empirical distribution of the interferometric outputs to produce a statistical inference. To our knowledge, it has never been deployed for nullers more complex than the Bracewell nullers. Considering its implications in the context of NOTT is outside the scope of this work.

In Sect. 2 we describe the computation produced by SCIFYsim before listing in Sect. 3 the aberration affecting the input beams. We then examine the distribution of the differential nulled output in Sect. 4 and describe statistical tests for their interpretation in Sect. 5. In Sect. 6, we use these tests to evaluate the performance of the NOTT instrument at VLTI for different stars and different configurations, before proposing a discussion and conclusion in Sects. 7 and 8.

2. The SCIFYsim end-to-end simulator

2.1. The architecture of the simulator

Here, we give a short overview of the architecture of the simulator. More details on some of its components are given in the following sections. The simulator articulates around a number of modules and classes that are combined and used by a `director` class in order to compute and hold the parameters of a model observatory, instrument, and observing session. Here is a partial list:

- `observatory` holds the geometric configuration of the array, and relies on the astroplan and astropy packages to compute the location of the target in the sky at a given time, the resuling alt-azimuthal pointing, and the projection of the array along – and orthogonal to – the line of sight (see Sect. 2.5).
- `sources` is used to compute spectra of absorption and emission of the targets, and the parasitic contribution of the earth atmosphere and optical train (see Sect. 2.4).
- `correctors` contains tools for delay lines and longitudinal dispersion compensators (see Sects. 3.2 and 3.3).
- `combiners` is used to build and hold models of the beam combiner integrated optics (see Sect. 2.2).
- `spectrograph` contains classes to reproduce the behavior of the spectrograph and detector pixels (with the `integrator` class), with associated quantum efficiency, dark current, and readout noise (see Sect. 2.3).
- `analysis` contains tools to handle the data reduction, and compute the performance of the instrument (see Sects. 4 and 6.2).

Each astrophysical source starts as a vector of a real valued spectrum of emission collected by each telescope. In accordance with its position in the field of view, it is associated with a complex phasor encoding, for each telescope,

- the phase effect due to its geometric position away from the optical axis of the array;
- the phase and amplitude errors due to fluctuating optical aberrations (Sects. 3.2 and 3.3);
- the amplitude effect due to the near-Gaussian transmission map of the spatial filtering by the single mode waveguide, limiting the field of view;
- and the amplitude effect due to optical transmission of the sky and optics.

The complex amplitude of each source is multiplied by this phasor to obtain the complex amplitude of the light entering the integrated beam combiner. This is then multiplied by the matrix of the beam combiner (see Sect. 2.2) to obtain the complex amplitude at the outputs. This is computed for each wavelength bin, and for subexposures of a few milliseconds, which are summed over the detector integration time. Values can be retrieved here under different forms; for example as a total sum including dark currents and photon and readout noise as in a realistic observation, or as exact values for each contributing source in order to compare useful signals from error terms and evaluate performance (see Sects. 4 and 5). Due to their diffuse and incoherent nature, the parasitic thermal background sources are not computed at each time step, but are instead updated only when necessary in order to save computing time.

2.2. Chromatic combiner models

The flexibility of SCIFYsim comes in part from the use of a complex-amplitude transfer matrix to model the beam combiner. Such matrices have been used to model the most complex beam-combiner architectures, especially in the context of nulling (Guyon et al. 2013; Martinache & Ireland 2018; Laugier et al. 2020), and are well suited for modeling the interference of single-mode beams; they take as input a vector containing the complex amplitude of the incoming beams, and produce the vector of complex amplitudes at the outputs.

---

[2] https://github.com/rlaugier/SCIFYsim

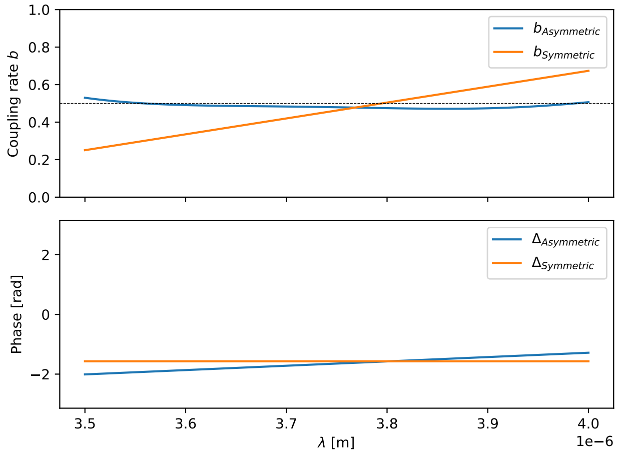


Fig. 2. Behavior of the asymmetric combiner used as a building block for the instrument in comparison to the symmetric alternative. Asymmetry is used in the evanescent wave coupler to produce the flatter coupling rate at the cost of a chromatic phase offset.

In SCIFYsim, they can be built in different ways, either based on `sympy` symbolic expressions such as those used in Laugier et al. (2020) and the corresponding assortment of combinatorial, algebraic, and analytical tools[3], or assembled algebraically based on building-block couplers of Fig. 1, each represented by a matrix of the form:

$$\mathbf{M}_{\text{coupler}} = \begin{bmatrix} \sqrt{b} & -\sqrt{1-b}e^{-j\Delta} \\ \sqrt{1-b}e^{j\Delta} & \sqrt{b} \end{bmatrix}, \tag{1}$$

where b is the coupling rate, and Δ is the incurring phase offset, both of which are wavelength dependent.

This informs the model on its internal architecture, but also, by replacing the matrix of the building-block coupler by a chromatic model (i.e., a parametric matrix as a function of the wavelength λ) informed by the characterization of one or other combiner technology (Tepper et al. 2017; Sharma et al. 2020), one obtains the chromatic model of the combiner as a whole. The model used here was prepared with beam-propagation simulations and experiments made at Macquarie University for the relevant technology, and its behavior is illustrated in Fig. 2. The matrix is then evaluated for the relevant wavelength channel and stored in a 3D array of complex float for use in the numerical computations.

In the case of fringe-measuring instruments like PIONIER (Berger et al. 2010) and GRAVITY (GRAVITY Collaboration 2017), with their outputs dispersed by a spectrograph, the visibility can be computed independently in each wavelength so that intensity or phase imbalance arising from chromatic aberrations have no first-order impact on performance. In the case of nulling interferometry, the rejection of the photon noise happens from a precise control of the phase of the interference, which must be set throughout the bandwidth of the instrument. In order to represent and examine the resulting complex null configuration, we extend the representation used in Laugier et al. (2020) to make use of color maps to represent all the wavelength channels of the instrument simultaneously, as can be seen in Fig. 3. Here, arrows are replaced by colored spots, with shading indicating the different wavelengths. On these same plots, we represent the sum of the contributions as an additional spot in shades of gray for each wavelength. Gray spots more tightly packed around the origin represent deeper and more broadband nulls.

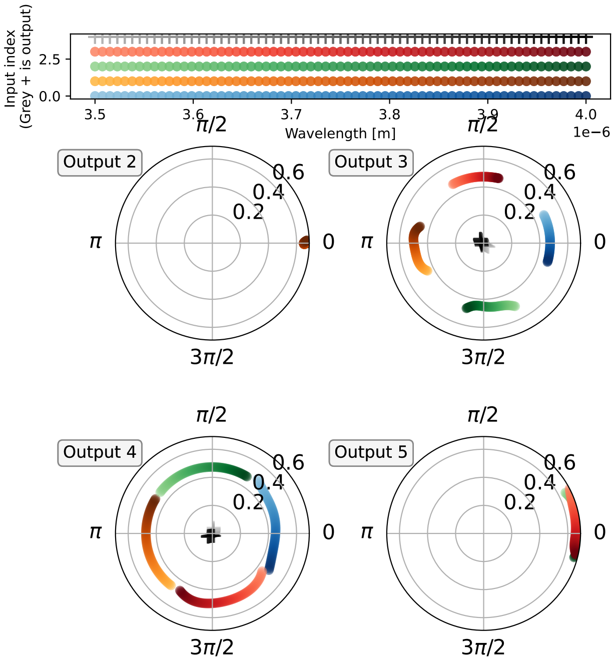


Fig. 3. Complex matrix plot of the matrix after correction. Each plot represents the complex plane for one output. Each input is represented by a hue and the density of the color is in proportion to the wavelength. A gray plus symbol marks the complex sum of the contribution, giving the nominal complex amplitude of the outputs. For the bright outputs, the colored spots land on top of one another, and the output plus symbol falls outside of the range.

2.3. Spectrograph and detector

The spectrograph module takes the intensity computed at the outputs of the combiner chip and convolves it with a spectroscopic PSF that spreads the light depending on its wavelength on the 2D array of pixels of the detector. To accommodate the possibility of an undersampled spectroscopic focal plane, this is first done with a denser array and then binned to the final detector resolution.

While helpful for the validation of the spectrograph design and the validation of signal-extraction methods (Dandumont et al. 2022b), this is demanding in terms of computation resources. This step is bypassed when using large Monte-Carlo samples, relying instead on the assumption that each spectral channel is spread uniformly over a fixed number of pixels.

The integrator module simulates the behavior of the detector pixels; it adds the photon noise, readout noise, and dark current, and can also implement saturation effects or amplification and the associated excess noise factor such as will occur in the case of EMCCDs or avalanche photodiode detectors (Lanthermann et al. 2019). For NOTT, integrator module implements the properties of the Teledyne HAWAII2RG detector with 15 e^- of readout noise and 0.02 $e^- s^{-1} \mathrm{pix}^{-1}$ of dark current. Recordings of the OPD or other aberrations can also be saved to this object as metrology outputs.

2.4. Throughput and background signal

For flexibility, the throughput and thermal background signal are both evaluated based on a series of objects called

---

[3] `https://github.com/rlaugier/kernuller`

Table 1. Components of the radiometric chain given for the medium spectral resolution $R = 400$.

Source	Temperature [K]	Mean transmission	Contribution$^{(a)}$ $[e^{-} s^{-1}]$
Sky	285.0	0.92	1.20e+05
UT optics	290.0	0.31	2.91e+06
Warm optics	285.0	0.60	2.20e+06
Combiner	100.0	0.50	1.69e-04
Cold optics	100.0	0.56	2.65e-04
Enclosure	100.0	N.A.	4.23e+05
Dark current	60.0	N.A.	1.07e+01

Notes. $^{(a)}$On one dark output, cumulated over the whole wavelength band.

transmission–emission objects. Each is attributed a temperature and a transmission function (either flat or tabulated), in a way similar to what was done in GENIEsim (Absil et al. 2006) for a two-telescope Bracewell architecture. These include the sky, the unit telescopes (UT) or auxiliary telescopes (AT) optics (including the delay line train), the warm optics of the instrument, the combiner chip, and the cold optics (including the spectrograph). Here, these objects are linked into a chain by pointers so that they each provide methods to obtain, recursively, values such as the total throughput of the chain downstream from them, with commands such as `item.get_downstream_transmission()`.

As in GENIEsim, the thermal background is computed using the Planck law and an emissivity value taken as $1 - t$ where t is the functional transmission (reflectivity of mirrors and transmission of lenses). This implementation makes the approximation that the single mode filtering is done right at the end of the chain, so that the étendue obtained in each spectral channel corresponds to the étendue injected in the waveguides. This is an approximation that relies on the way the pupil of the spectrograph samples the Gaussian beam that propagates from the outputs of the combiner.

The exception to this is the enclosure term corresponding to the cryostat, which is purely radiometric and computes a flux of photons per detector surface, and therefore per pixel. The list of contributors is given in Table 1, with the total resulting current given for the sum of all spectral channels for $R = 400$, including the dark current. The total optical throughput is about 17% from the input of the Asgard table, or 5.2% from the VLTI telescopes, leading to an overall efficiency of about 3.2%, including the quantum efficiency of the detector. The total contribution of shot noise increases with the spectral resolution as the number of spectral bins increases.

2.5. Array geometry and source

The horizontal and projected geometry of the array is handled by a separate module called `observatory`. This is powered by `astropy` and `astroplan` to handle the coordinates of targets, pointings, and time. It also provides convenience functions with which to locate targets by name, or identify the best night of the year to make the observations; that is, where the transit of the meridian happens in the middle of the night. The pointing of the array also informs an airmass parameter inside the sky object.

This module computes the projection of the array onto the plane orthogonal to the line of sight, which is then used to compute the geometric OPD affecting the beams of each light source depending on its position (α,β) in the field of view:

$$\tilde{a}_l = a_l e^{\frac{2j\pi(X_l\alpha+Y_l\beta)}{\lambda}}, \tag{2}$$

where for each aperture l, a_l is the amplitude collected and X_l and Y_l are the coordinates of the apertures projected in a plane orthogonal to the line of sight and aligned to right ascension and declination. In the example presented here, the star is modeled as a dense collection of points on a disk with a black body spectrum. The module also computes the projection of the array onto the line of sight, therefore providing the relative path length to be corrected by the delay lines.

3. Description of the main sources of errors

3.1. Injection into single mode fibers

The injection into single-mode waveguides poses a challenge for any single-mode instrument. Indeed, aberrations in the beam represented by small amplitude or phase errors across the pupil produce small variations in the focal plane complex amplitude, which will result in decreased coupling into the mode field of the waveguide. These errors are the result of pointing errors and residuals from the correction of the atmospheric turbulence by the adaptive optics.

As illustrated in Fig. 1, the amount of light injected into the combiner is monitored for each input with an evanescent coupler sending 10% of the guided light to the photometric outputs. These enable the tuning of the injection and its monitoring in real time. While they will be invaluable in the field for flux calibration, their output is not used in the signal processing described here.

3.2. Residual OPD errors from fringe tracking

The optical path errors are introduced as pistons to each of the input beams. Random time series of these pistons are computed based on the power spectrum of measurements obtained with the fringe tracker of GRAVITY on a bright star. For this, we take the square root of this power spectrum and associate it with random phase following a uniform distribution between $-\pi$ and π, and then put it through an inverse Fourier Transform to obtain a random time series of the same power spectrum.

The amplitude of this residual on the unit telescopes (UTs) were found to exceed the expected values produced by the filtered atmosphere (Lacour et al. 2019a). This was attributed to vibrations in the telescope structure. An ongoing effort through SCIFY as part of the GRAVITY+ upgrade is expected to bring this residual down to 100 nm RMS of optical path difference, which corresponds to the level already reached on the ATs. For this reason, we compute this temporal power spectrum of the optical path from observations made with the ATs, and use them in simulations using the UTs.

3.3. Atmospheric dispersion and chromatic OPD

In addition to the “dry” (i.e., purely geometric) piston described in Sect. 3.2, the phase of the inputs is also affected by chromatic or “wet” contributions. This is important as the delay lines compensate – in ambient air – for a path difference that occurs in the vacuum of space, which results in longitudinal atmospheric dispersion. This chromatic effect is modeled in the same way as GENIEsim based on the tables by Mathar (2004). This static effect was not included in the early results presented here. NOTT will include ZnSe variable thickness corrector plates (Koresko et al. 2006) to compensate for this effect to the first order.

While the contribution of turbulent air is often only associated to a simple geometric piston as described in Sect. 3.2, the demands of a broadband nulling require the correction for a more thorough model, as in that described by Colavita et al. (2004) and Absil et al. (2006). However, relevant simulations of these effects will depend on the technical solutions retained for HEIMDALLR and Baldr modules of Asgard. Colavita et al. (2004) predicted the effect of water vapor seeing for Mauna Kea to be one-twentieth of the dry effect in the near-infrared and one-seventh at 10 μm. In this work, dynamic chromatic perturbations are neglected but they will be included in later work, based on expertise gathered from GRAV4MAT at VLTI and METIS for the ELT.

3.4. Tuning internal combiner chromatic errors

Chromatic biases in the combiner are modeled with polynomial fits of order six made from beam-propagation simulations and experimental results making available symbolic expressions of the beam-combiner matrix, and enabling flexibility in the choice of spectral channels. Options include some baseline implementations of symmetric and asymmetric evanescent wave couplers and multimode interference couplers. The result is an imperfect nulling combination due to chromatic phase and amplitude variations. As shown in Fig. 3, a common phase term affects the whole of the dark outputs, but is irrelevant to our measurement of intensity. For each wavelength, the orthogonal four-way pattern is respected, and close to symmetrical.

The variable thickness plates mentioned in Sect. 3.3 are used to compensate for the chromatic errors in the combiner. The model used is the same as that used in GENIEsim (Absil et al. 2006) and reproduces a vector $\mathbf{c}$ of complex phasors:

$$\tilde{c}_l(\lambda) = c_l e^{\frac{2j\pi}{\lambda}\left(n_{\mathrm{air}}(\lambda)p_l + n_{\mathrm{glass}}(\lambda)q_l\right)}, \tag{3}$$

where n_{air} and n_{glass} are the refractive index of air and glass, and p_l and q_l are the additional optical path in air and glass, respectively. The term c_l can be used to represent filters or adjustable stops, but is set to a value of 1 in this study. This leads to an effective combiner matrix that can be expressed as

$$\mathbf{M} = \mathbf{M_0}.\mathrm{diag}(\mathbf{c}), \tag{4}$$

which includes the effect of the phasor $\mathbf{c}$ on each of the inputs of the combiner, separately from $\mathbf{M_0}$, which is the matrix representing the static part of the beam combiner.

The air and glass path lengths are both adjusted by gradient descent in two independent steps involving different subsets of beams and different cost functions. The first can be interpreted as maximizing the null depth obtained by the first nulling stage by the relative phase between 0 and 1 and between 2 and 3. Using a gradient descent method, the ZnSe thickness and geometric path length are adjusted in order to minimize the cost vector $\mathbf{g}_{\mathrm{null}}$ based on the values of I_k, the intensity at output k:

$$\mathbf{g}_{\mathrm{null}} = \frac{\sum_{k\in\mathrm{dark}} I_k^-(\lambda)}{\sum_{k\in\mathrm{bright}} I_k^+(\lambda)}, \tag{5}$$

with one dimension per spectral channel. This is done using a Levenberg-Marquardt algorithm.

The second step can be described as tuning the second (mixing) stage and can be done from upstream by acting differentially on the input pairs (0, 1) – (2, 3). To do so, we need a cost function that reflects the desired symmetry quality between the complementary combinations that would guarantee the robustness quality of the observable, as outlined in Laugier et al. (2020). For this, we compute a shape parameter:

$$\Lambda = \frac{\min\limits_{\theta\in[-\pi,\pi]}\left(\|e^{j\theta}\cdot\mathbf{m}_h - \mathbf{m}_k^*\|^2\right)}{\left(\sum_{i=0}^{N}|m_{k,i}|\right)^2}, \tag{6}$$

with one dimension per spectral channel per differential output, where $\mathbf{m}_h$ and $\mathbf{m}_k$ are a pair of rows of the corrected combination matrix $\mathbf{M}$ corresponding to a pair of complementary dark outputs. θ is a dummy variable representing the fact that only the output intensity is recorded and the consideration must therefore be blind to an arbitrary phase offset between the two rows. A normalization is made by the sum of the coefficient in one of the rows, which represents the peak amplitude of the response map. Again, this is minimized numerically with a Levenberg-Marquardt algorithm by adjusting lengths in air and glass.

In order to use this parameter in practice, the matrix of the combiner would have to be measured using the calibration source. This can be obtained through a process comparable to the one used to measure V2PM matrices (Lacour et al. 2008; Cvetojevic et al. 2021) for ABCD combiners. Further work should make this process reliable and accurate.

The null depth and shape parameter for the initial and final state plotted in Fig. 4 show the wavelength range split into three different regions by two troughs of null depth and shape parameter. These deeper minima in this broadened favorable region have an impact on the instrumental noise described in Sect. 4.3.

4. Examination of the resulting errors

4.1. Time series of the raw combiner outputs

To illustrate the behavior of the simulated instrumental errors, we track the temporal variations of the contributions to input piston and amplitude. These are displayed in Fig. 5 for one of the four inputs. We note that, for this illustration, a single piston is considered rather than an optical path difference between the collectors of a baseline. The temporal spectrum for the simulated effects of injection contains little high-frequency content, as expected for averaging of large-diameter apertures, but the temporal spectrum of the fringe tracking error contains significant high-frequency content, perhaps inflated by measurement noise, because it was generated from real data.

The effects of these aberrations on the dark outputs are shown in Fig. 6 for two different spectral bins: 3.75 μm at the center of the band, and 3.56 μm down one of the troughs identified in Sect. 3.4. This representation highlights the practical effect of the self-calibration of instrumental errors in the differential null. Correlation of the error terms is higher at 3.56 μm as anticipated from the combination closer to its nominal scheme, leading to smaller errors in the differential output. We note that the values shown here are pure measurements of the intensity on the detector simulated at relatively short time steps of 5 ms. The effect of shot noise, dark current, read noise, and longer integration time are not shown here but are included in the rest of this study.

4.2. The distribution of errors

As noted by Martinache & Ireland (2018), a significant advantage of the differential nulls over the simpler Bracewell nulls is their

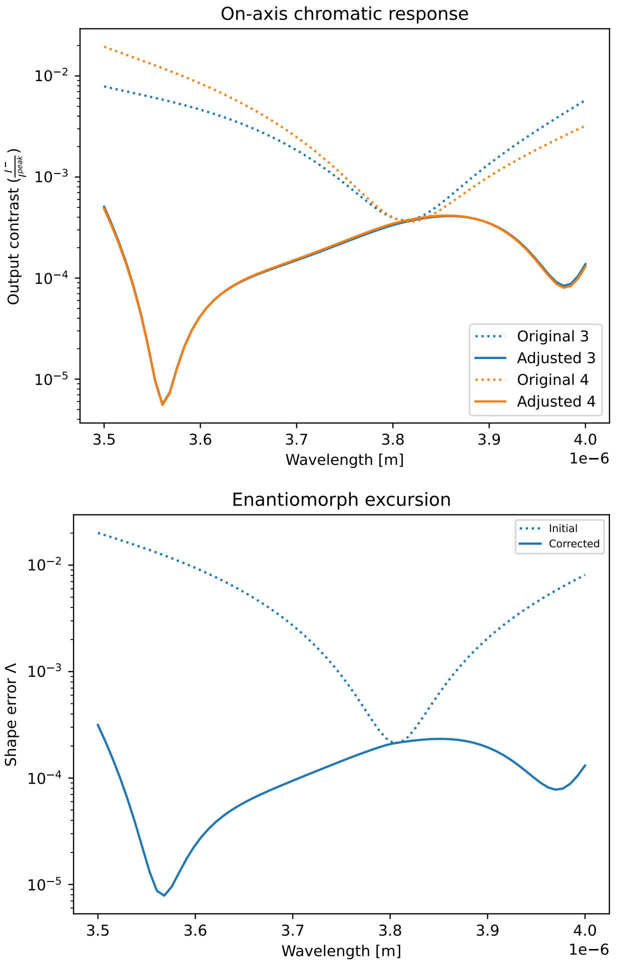


Fig. 4. Cost function for the adjustment of the delay lines and dispersion corrector. The null depths of the two dark outputs are labeled 3 and 4. These guide the optimization for the first stage of the combiner (top). The shape error guides the optimization for the second stage (bottom). Before and after adjustment are shown as the dotted line and plain line, respectively.

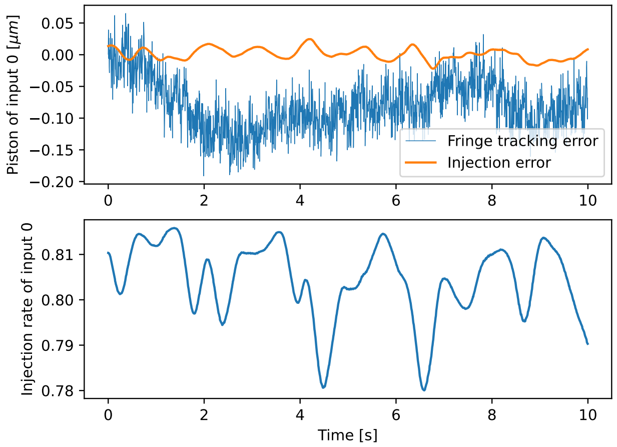


Fig. 5. High-temporal-resolution series of the instrumental noise contribution on one input. Outputs 3 and 4 are the two nulled outputs. This shows the time resolution of the simulation, which includes two different sources of phase errors: from fringe-tracking and from wavefront errors at the injection; as well as the injection rate error.

distribution. Indeed, the stochastic instrumental errors, which are dominated by the variation of the leakage light under varying optical path differences and amplitude fluctuations, are symmetrical. When the errors are small, the distribution approximates to a Gaussian. SCIFYsim can be used to evaluate the limits of this approximation.

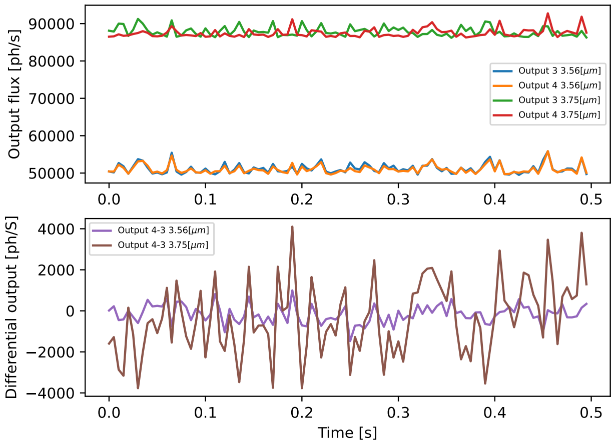


Fig. 6. High-resolution time series of the dark outputs (top) and their difference (bottom) for two different wavelength bins offering different qualities of combination, leading to different mean leakage and correlation levels. Correlation within each pair is the effect leading to self calibration in the differential output. Correlation between spectral channels is what motivates the data whitening approach. The time series shown here are only indicative, because in practice integrations are made over longer times and are contaminated by sensor noises before they can be numerically subtracted.

We examined the distribution of the differential null, as illustrated in Fig. 7 for two different spectral bins. The quality of the distribution is quantified with the p-value of a Shapiro-Wilk test, giving the probability of incorrectly rejecting the null hypothesis (Gaussian hypothesis). We consider that when this number drops below five percent, the Gaussian hypothesis should be rejected. For the parameters considered here, the PDF appears Gaussian in linear scale but the test starts to reject the Gaussian hypothesis with about 400 samples of the 3 s detector integration time (DIT) for most spectral channels. However, the Gaussian distribution remains a good approximation, especially for the mean of a large number n_{DIT} of samples as used here, and is used in the rest of this work, in particular for the hypothesis tests detailed in Sect. 5.

4.3. Parametric model of errors

SCIFYsim is primarily an end-to-end simulator designed to provide a realistic end product of the observation. However, producing Monte-Carlo simulations for each of the relevant pointing directions and star magnitudes would require prohibitive amounts of processing power, and therefore instead we evaluate the different contributions to errors from a single Monte-Carlo time series so that they can be scaled independently.

Let F_{source} be the total flux of a given source across the wavelength range of interest. We refer to the vector containing the spectrum of this source as collected on the spectral bins of the spectrograph as $\mathbf{S}_{\mathrm{source}} = [S_{\mathrm{source,bin}}]$. Leveraging linearity, to scale both signal and errors in this model with brightness of the source and integration time, we write

$$\mathbf{S}_{\mathrm{source}} = F_{\mathrm{source}} \int_{\mathrm{DIT}} \mathbf{r}(\mathbf{s}_{\mathrm{source}}) \mathrm{d}t, \tag{7}$$

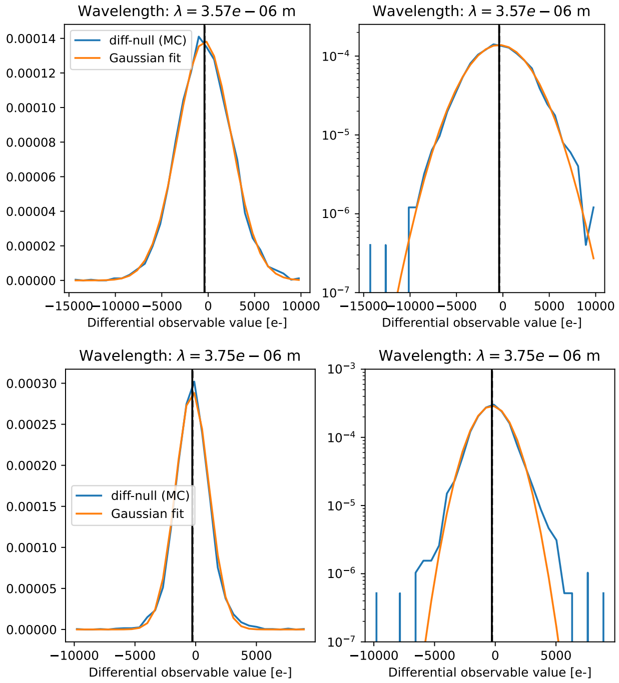


Fig. 7. Samples of the empirical probability density function of the total error for the 3.57 (top) and 3.75 (bottom) µm bins computed with 3000 samples of 3 s DITs with spectral resolution $R \approx 400$. The density function is displayed in linear scale (left) and in log scale (right). Some kurtosis is noticeable in log scale, causing the Shapiro-Wilk test to reject the Gaussian hypothesis in parts of the spectrum but the Gaussian distribution remains a relevant approximation.

where $\mathbf{s}_{\text{source}}$ is the spectral distribution of the source and $\mathbf{r}$ is the instantaneous instrumental response function which acts independently on all spectral bins. It depends, as described earlier, on the source location of the field of view, the projection of the collecting array, the phase and amplitude input errors and the combiner transfer matrix. For spatially incoherent sources, like the thermal background, it is constant.

Here we intend to account for the covariance in the errors. The covariance matrix of the error in a single DIT contains different contributions:

$$\mathbf{\Sigma}_{\text{DIT}} = \mathbf{\Sigma}_{\text{RON}} + \mathbf{\Sigma}_{\text{photon}} + \mathbf{\Sigma}_{\text{instrumental}}, \tag{8}$$

where $\mathbf{\Sigma}_{\text{RON}}$, the covariance of the readout noise, and $\mathbf{\Sigma}_{\text{photon}}$, the covariance of the photon noise, are diagonal matrices carrying the variance of the respective contributions, but the covariance of the instrumental noise $\mathbf{\Sigma}_{\text{instrumental}}$ is not diagonal. While this latter normally contains contributions of all the sources in the field of view, it can be simplified, as the incoherent background sources have only a static contribution (the variability of the background light is currently neglected by SCIFYsim for simplicity) and the planet luminosity is orders of magnitudes less luminous than the star, leaving only the fluctuations of the light from the central star. (Instrumental noise on an off-axis source would be difficult to account for, as it is correlated, non-Gaussian, and variable depending on the position in the map.) The effect of both injection errors and OPD errors produces variations of the starlight on the interferometric output across the different wavelength channels in a deterministic manner. The amplitude of this effect is directly proportional to the luminosity of the host star:

$$\mathbf{\Sigma}_{\text{instrumental}} \approx F_{\text{star}} \mathbf{\Theta}_{\varphi}. \tag{9}$$

Here, we evaluate this normalized covariance matrix $\mathbf{\Theta}_{\varphi}$ empirically through MC simulations on a reference star of flux F_{ref} giving the recorded signal $\mathbf{S}_{\text{ref}}$:

$$\mathbf{\Theta}_{\varphi} = \frac{1}{F_{\text{ref}}} \text{Cov}(\mathbf{S}_{\text{ref,MC}}). \tag{10}$$

The readout noise term can be computed as the sum of the variances of the n_{pix} pixels contributing to the differential signal. This total number includes the number used along the spectral dispersion n_{bin}, the number of outputs contributing n_{out} (typically 2), and the number of contributing polarization channels n_{pol} :

$$n_{\text{pix}} = n_{\text{bin}} \times n_{\text{out}} \times n_{\text{pol}}. \tag{11}$$

The photon noise term $\mathbf{\Sigma}_{\text{photon}} = \text{diag}(\sigma^2_{\text{photon}})$ follows a Poisson distribution and can be approximated to a normal distribution in the case of a large number of photoelectrons. For each bin, the variance writes:

$$\sigma^2_{\text{photon,bin}} \approx n_{\text{out}} t_{\text{DIT}} (S_{\text{thermal,bin}} + \langle S_{\text{leakage,bin}} \rangle), \tag{12}$$

and is a function of the thermal background $\mathbf{S}_{\text{thermal}}$ and the stellar leakage $\mathbf{S}_{\text{leakage}}$.

In this first approach, we assume that measurements made over each DIT are statistically independent. For convenience in the processing we decompose the observing sequence into n_{chunk} chunks (observing blocks) for which the signal from the n_{DIT} measurements is averaged, resulting in a covariance of:

$$\mathbf{\Sigma}_{\text{chunk}} = \frac{1}{\sqrt{n_{\text{DIT}}}} \mathbf{\Sigma}_{\text{DIT}}. \tag{13}$$

We combine the result over the full observing sequence into a single vector $\mathbf{z}$ by concatenation of the observable vectors of each of the observing blocks. Assuming the errors between the blocks are uncorrelated, the covariance matrix of $\mathbf{\Sigma}$ of $\mathbf{z}$ is a block-diagonal matrix containing the $\mathbf{\Sigma}_{\text{chunk}}$ on the diagonal (see illustration in Fig. 8).

Readout noise is added to each integration. To maximize the use of the dynamic range and minimize the added readout noise, t_{DIT} is chosen so that the flux on the dark outputs – here dominated by the static thermal flux – comes close to saturation:

$$\text{Max}_{\text{channels}}(t_{\text{DIT}} S_{\text{thermal}}) \approx S_{\text{full well}}. \tag{14}$$

The values used are detailed in Table 2. An example of this covariance for the case of a star at magnitude $L' = 4$ is displayed in Fig. 8, and shows how the independent diagonal terms are swamped by the correlated instrumental noise. The relative scale of the different contributors to the errors is represented in Fig. 9.

Closer examination of the covariance reveals a property of the three spectral regions mentioned in Sect. 3.4. The errors in the inner part of the spectrum ($3.57 < \lambda < 3.98$ µm are correlated together, as indicated with the red parts of the covariance matrix, but are anti-correlated with the outer regions ($\lambda <$ 3.57 µm and $\lambda >$ 3.98 µm), as indicated by the blue parts of the matrix.

5. Statistical detection tests

In order to evaluate the performance of the instrument for the detection of sources, we compute the detection limit of statistical tests. We use the tests described by Ceau et al. (2019) for the detection in kernel phases, but work here in a similar fashion.

Table 2. Observation parameters for the example sensitivity map.

Astrophysical parameters	
T_{star}	5263 K
R_{star}	0.77 $R_{\odot}$
d	10.8 pc
Magnitude L'	4.1
Star spectrum model	Black body
Planet spectrum	Flat
Instrumental parameters	
Collectors	UTs
Telescope input order	0, 2, 1, 3
Spectral resolution R	≈400
DIT	3 s
Combiner and spectro. temperature	100 K
Detector temperature	40 K
Readout noise	15 e^-
Dark current	0.02 $e^- s^{-1}\text{pix}^{-1}$

This is applied by considering the vector of observables that concatenates the different spectral channels of the instrument, and the different observation chunks in the session. The resulting vector has a length of $n_{\text{chunk}} \times n_{\text{channels}} \times n_{\text{outputs}}$. We neglect the temporal correlations between observation chunks. As a consequence, the resulting covariance matrix is block-diagonal, with square blocks of dimension n_{channels}.

Following the same principles, a comparison between model and observation writes, in the subspace of the differential null,

$$\mathbf{z} = \mathbf{z}^0 + \boldsymbol{\varepsilon}', \tag{15}$$

where $\mathbf{z}^o$ is the theoretical signal produced by a model object. When the model matches reality, $\boldsymbol{\varepsilon}' \sim \mathcal{N}(0, \boldsymbol{\Sigma})$. In order to correctly express the likelihood of this with correlated errors, we use a change of variable through the same whitening transform used in Ceau et al. (2019).

We produce a whitening matrix,

$$\mathbf{W} = \boldsymbol{\Sigma}^{\frac{-1}{2}}, \tag{16}$$

to produce – from the original vector of differential null $\mathbf{z}$ – a new output vector $\mathbf{y} = \mathbf{W}.\mathbf{z}$ which is to be compared with a model signal $\mathbf{x} = \mathbf{W}.\mathbf{z}^o$. The new model-fitting equation writes

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\varepsilon}, \tag{17}$$

this time with $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \mathbf{I})$.

The conclusions from Ceau et al. (2019) suggest that a more practical T_B (generalized likelihood ratio) test would have an expected performance lying in between those bounds. However, this case is not investigated here, as its sensitivity cannot be determined analytically. Instead, we examine the case of T_E, the energy detector, and T_{NP} the Neyman-Pearson test. While T_E offers a conservative estimate of the instrument performance, T_{NP} provides a theoretical upper bound to detection performance.

The amplitude of the signal of interest is computed so that it corresponds to the threshold ξ of the test that satisfies both targeted P_{FA}, the false-alarm probability, limiting the number of false detections, and targeted P_{Det}, the detection probability, ensuring the sensitivity of the test. This is solved in Sects. 5.1 and 5.2.

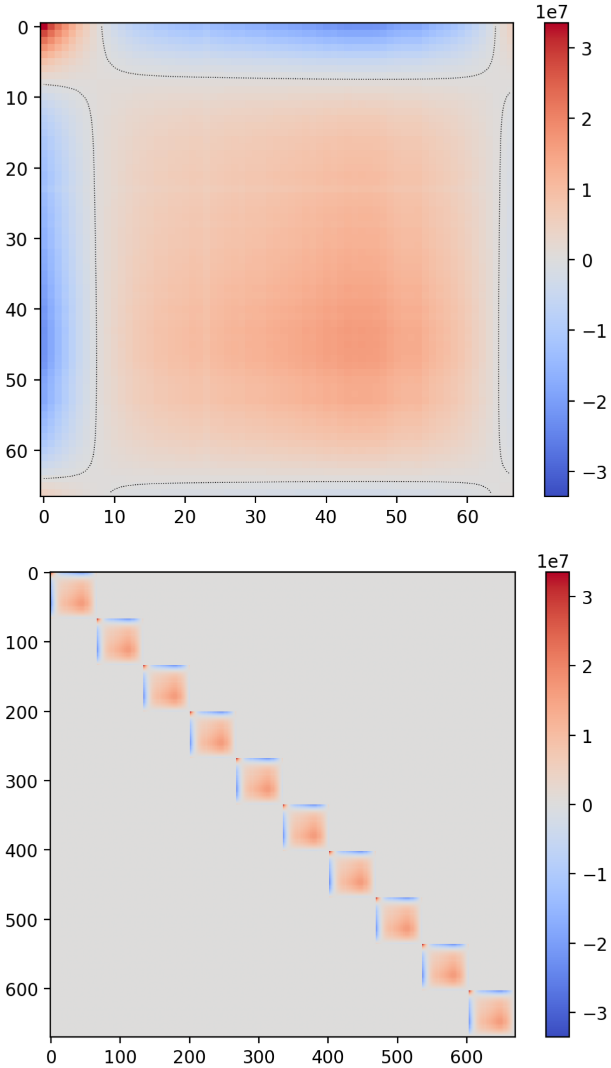


Fig. 8. Color-mapped representation of the covariance matrix $\boldsymbol{\Sigma}_{\text{chunk}}$ of the errors for the case of a star of magnitude 4. The values of the diagonal, resulting from independent, pixel-level noise (photon noise and read-noise) are dwarfed by the correlated instrumental noise. The dotted line separates the positive and negative covariance. $\boldsymbol{\Sigma}$ is the block-diagonal matrix populated by the positive covariance.

5.1. The energy detector test

For the test T_E, we use a numerical method to solve the following system for $d = \mathbf{x}^T\mathbf{x}$:

$$\begin{cases} P_{FA}^{T_E} = 1 - \mathcal{F}_{\chi^2_p(0)}(\xi) \\ P_{\text{Det}}^{T_E} = 1 - \mathcal{F}_{\chi^2_p(b)}(\xi) \end{cases}, \tag{18}$$

where $\mathcal{F}_{\chi^2_p(x)}$ is the cumulative distribution function (CDF) of the χ_2 distribution with p degrees of freedom in x.

With the linearity of the system, for a given off-axis relative position, observing sources of flux F and F_0 gives observable vectors of flux $\mathbf{x}$ and $\mathbf{x}_0$ respectively, so that

$$\mathbf{x} = \frac{F}{F_0}\mathbf{x}_0. \tag{19}$$

As a consequence, the flux of a star to solve this system writes

$$F = F_0 \frac{d}{\mathbf{x}_0^T \mathbf{x}_0}, \tag{20}$$

with $\mathbf{x}_0$ being the value of the differential map computed for a flux F_0.

5.2. The Neyman-Pearson test

For the Neyman-Pearson test, the law can be inverted analytically to obtain the sensitivity for a given false-alarm rate. This amounts to solving Eq. (17) in Ceau et al. (2019) for $\mathbf{x}^T\mathbf{x}$. Starting from their Eq. (16), we write

$$\begin{cases} \frac{\xi-\mathbf{x}^T\mathbf{x}}{\sqrt{\mathbf{x}^T\mathbf{x}}} = \mathcal{F}^{-1}_{\mathcal{N}(\mu=\mathbf{x}^T\mathbf{x},\,\sigma^2=\mathbf{x}^T\mathbf{x})}\left(1-P_{\mathrm{Det}}\right) \\ \frac{\xi}{\sqrt{\mathbf{x}^T\mathbf{x}}} = \mathcal{F}^{-1}_{\mathcal{N}(\mu=0,\,\sigma^2=\mathbf{x}^T\mathbf{x})}\left(1-P_{FA}\right) \end{cases}, \tag{21}$$

where $\mathcal{F}^{-1}_{\mathcal{N}(\mu=0,\,\sigma^2=\mathbf{x}^T\mathbf{x})}$ is the inverse of the cumulative distribution function of the normal distribution centered on zero of variance $\mathbf{x}^T\mathbf{x}$. From substitution, one can isolate the term representing the amplitude of the signal of interest:

$$\sqrt{\mathbf{x}^T\mathbf{x}} = \mathcal{F}^{-1}_{\mathcal{N}(\mu=0,\,\sigma^2=\mathbf{x}^T\mathbf{x})}\left(1-P_{FA}\right) - \mathcal{F}^{-1}_{\mathcal{N}(\mu=\mathbf{x}^T\mathbf{x},\,\sigma^2=\mathbf{x}^T\mathbf{x})}\left(1-P_{\mathrm{Det}}\right). \tag{22}$$

This signal amplitude is proportional to the luminosity of the off-axis source to be detected and depends on its position in the field of view and its spectrum. After computing the signal $\mathbf{x}_0$ obtained from a reference source of luminosity F_0 in each point of a map, one can rely on the linearity

$$\frac{\sqrt{\mathbf{x}^T\mathbf{x}}}{F} = \frac{\sqrt{\mathbf{x}_0^T\mathbf{x}_0}}{F_0} \tag{23}$$

to compute the limit luminosity F satisfying the criteria for the test:

$$F = \frac{F_0}{\sqrt{\mathbf{x}_0^T\mathbf{x}_0}}\left(\mathcal{F}^{-1}_{\mathcal{N}(\mu=0,\,\sigma^2=\mathbf{x}^T\mathbf{x})}\left(1-P_{FA}\right) - \mathcal{F}^{-1}_{\mathcal{N}(\mu=\mathbf{x}^T\mathbf{x},\,\sigma^2=\mathbf{x}^T\mathbf{x})}\left(1-P_{\mathrm{Det}}\right)\right). \tag{24}$$

This sensitivity value will be offset by the spectral energy distribution of both the planet or the star and will be different from the model. For simplicity, we use a flat spectrum here for the planet, which is a good approximation in the L' band for young giant planets at temperatures around 1000 K. We also note that deviations from the Gaussian hypothesis mentioned in Sect. 4 will lead to biases in the actual false-alarm rate. A large kurtosis in particular will lead to underestimation of P_{FA}. More thorough Monte-Carlo simulations may later be used to refine these thresholds by measuring the distribution of the final test statistic empirically.

6. Application and results

6.1. Example application case

NOTT is designed to operate as part of the Asgard suite with HEIMDALLR as high precision fringe tracker and the Baldr as wavefront sensor (based on Zernike wavefront sensor), which are currently under development. Here, we adjust the performance of wavefront control to match existing systems available at the Paranal observatory. We use the measured performance of the GRAVITY fringe tracker, and adjust the performance of adaptive optics to match the tip-tilt performance of NAOMI for the ATs with 25 mas RMS (Woillez et al. 2019) and SPHERE for the future GRAVITY-plus adaptive optics (GPAO) system with 3 mas RMS (Le Bouquin, priv. comm.). The spatial sampling of the pupil is also matched with 4 and 40 control elements across for NAOMI and GPAO respectively.

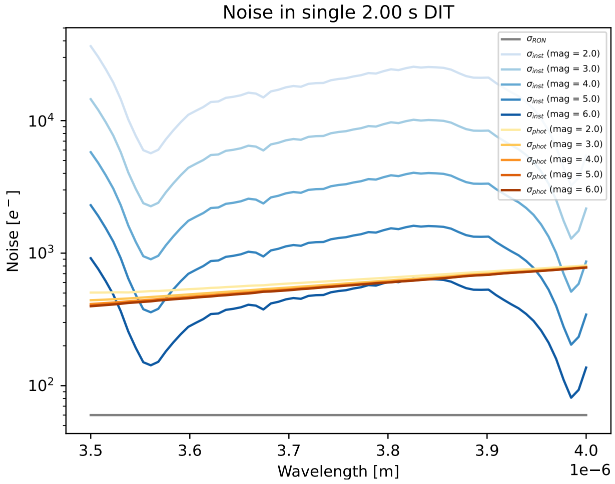


Fig. 9. Standard deviation of the different error contributions for different star luminosity as a function of wavelength. The read noise term is assumed constant. The photon noise term is dominated by the background at longer wavelengths, but the effect of leakage light can be seen at the shorter wavelengths. The instrumental noise dominates for the brighter stars and follows the overall trend of Fig. 4, with small ripples caused by atmospheric absorption lines.

The observed star is modeled after Gl 86 A, a nearby K dwarf of $L' \approx 4.1$ located at a declination of around −50 deg. Across the different curves in Figs. 10 and 11, the angular diameter and temperature of the star are kept constant, and only magnitude is adjusted. We assume a sequence of 20 observation blocks equally distributed over 6h around its passage of the meridian. Total integration time on target is 3h (overheads are 50%), yielding 180 DITs per block. Sensitivity is given for the detection of a planet with a flat spectrum for $P_{FA} = 0.01$ and $P_{\mathrm{Det}} = 0.9$. The relevant parameters for Figs. 12 and 13 are given in Table 2.

6.2. Results

The simulation of several hundred frames provides the covariance of the observables as a function of star magnitude. Using the transmission map of the instrument and the equations of Sect. 4.3, this covariance is used to produce the sensitivity map for each of the detection tests for any star magnitude. Both look alike with a different scale, and the map for T_E is shown in Fig. 12. The way this map is affected by the position on sky is discussed in Appendix A.

At the smallest separations, the lobes of the transmission map at different wavelengths overlap, creating a first broadband lobe. This lobe offers the most transmission and therefore gives the most sensitivity and performance; namely up to ≈ 1 magnitude above the field. However, on the brighter stars, when the instrumental noise dominates, performance decreases in the first lobes compared to the outer regions. In the first peak, a flat-spectrum planet would produce a broad spectrum signal in the data, which is consistent with highly correlated instrumental noise. However, at larger separations, the signal of a

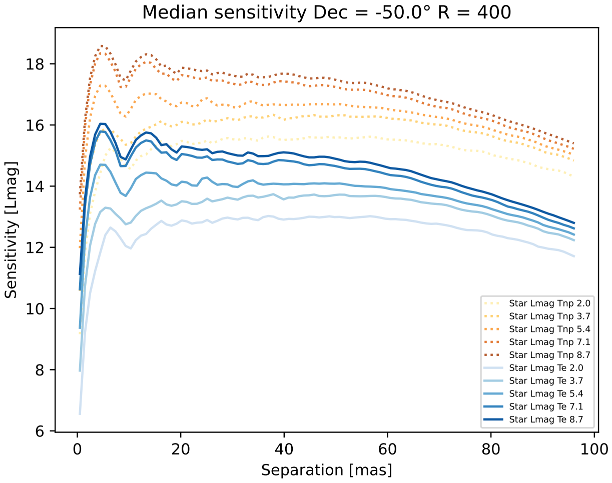


Fig. 10. Aggregate of the median plots from Fig. 13 for various stellar magnitudes and for the two tests. The dashed vertical line represents a separation of $2\lambda/D$ for the ELT.

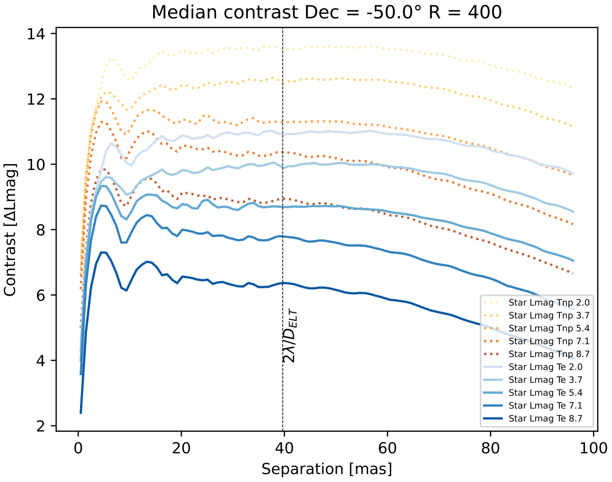


Fig. 11. Aggregate of the median plots from Fig. 13 shown in contrast for the two tests.

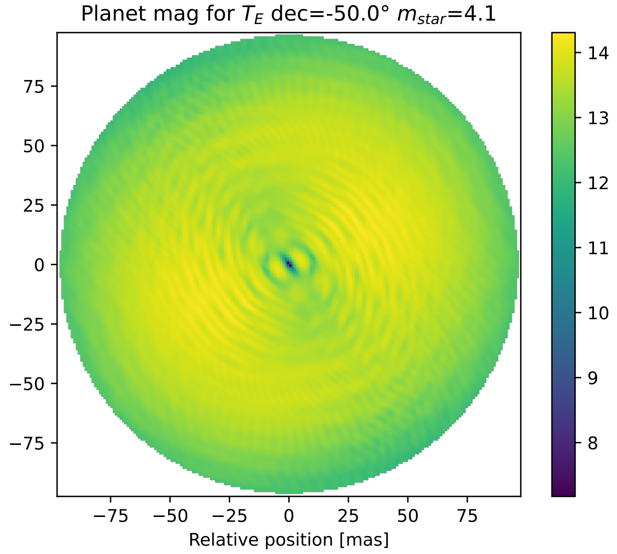


Fig. 12. Sensitivity map showing the maximum magnitude of a detectable companion with the energy detector test with $P_{FA} = 0.01$ and $P_{\mathrm{Det}} = 0.9$, at a declination of −50.0 degrees and around a star of $L = 4.1$. The map evolves in shape for different declinations, but not significantly in overall sensitivity.

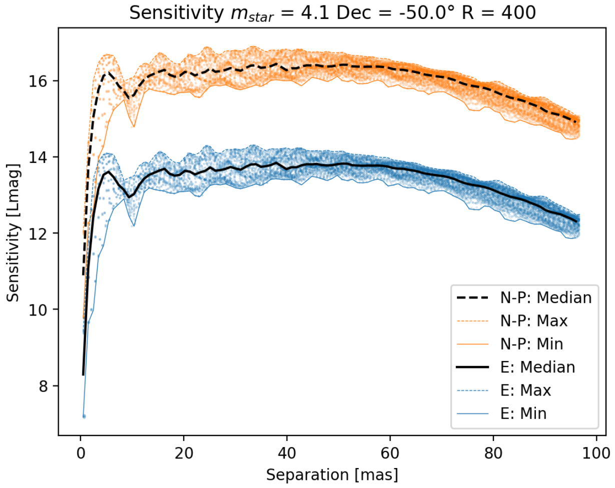


Fig. 13. Radial scatter plot representation of each pixel of the sensitivity map. The map shows the test T_E in blue and the Neyman-Pearson test T_{N-P} in orange. Bounds of minimum, maximum, and median are also plotted. The two tests show very similar behavior except for their different sensitivity.

flat-spectrum planet would appear as multiple positive and negative bumps in the differential null, which is distinct from the correlated instrumental noise. This shows the importance of accounting for the correlated instrumental noise, which allows us to still salvage the signal that does not correlate in the same way as the data.

7. Discussion

The dependency we highlight between the high contrast sensitivity and the stellar luminosity was examined assuming that the performance of wavefront control remains constant. In practice, this latter will also vary with the luminosity of the star, its spectral type, the atmospheric conditions, and the airmass, but a finer study of this performance is beyond the scope of the present work, and will concern the wavefront control provided by the HEIMDALLR (Ireland et al. 2018) for fringe tracking and Baldr for refined adaptive optics, which should yield improved performance.

The atmospheric dispersion is not taken into account in this first round of simulations. The static part of its effect will be compensated for by using a combination of air and ZnSe additional optical path via forward modeling. The dynamical part of this dispersive effect, often referred to as water-vapor seeing, is not taken into account here, and its impact on the performance will be the topic of future work.

The effect of transverse atmospheric dispersion is not expected to be a large contributor of errors, with an effect of the order of 3 mas in pointing. Nevertheless, this will be implemented in the simulator in the near future so as to correctly model its effect on the correlations of the nulled outputs.

The effect of polarization is not taken into account in this work. Characterization of the polarization in VLTI is currently incomplete (Le Bouquin et al. 2008; Lazareff et al. 2014), but most of the problems will be handled in the instrument with a

Wollaston prism in the spectrograph and birefringence compensation plates in each beam (Lazareff et al. 2012; Le Bouquin et al. 2011).

The confidence estimates offered by the detection test suggested in Sect. 5 are contingent upon the Gaussian distribution of the differential nulled output. In practice, with the level of wavefront correction used as a baseline here, a deviation from this hypothesis is measurable with a Shapiro-Wilk test on a few hundred Monte-Carlo samples. The temporal correlations in the measurement affect both our estimation of the amplitude of errors on the mean and its distribution. This does not invalidate the usefulness of the likelihood ratio framework as a whole. Accounting for these effects is beyond the scope of this work. Future efforts will include frame selection, relying on metrology and the measurements of the photometric outputs to better constrain the distribution by removing some outliers. Adapting the tests to use a distribution that better models the actual distribution of the differential output is also possible.

Here we use a fixed stellar spectrum and a fixed flat planet spectrum in order to produce results as a function of simple luminosity. Deviations of the planet spectrum impact the signal level $\sqrt{\mathbf{x}^T\mathbf{x}}$. Deviation on the stellar spectrum impacts the covariance matrix of errors, which therefore impacts the sensitivity result through the whitening matrix.

8. Conclusion

SCIFYsim offers a new framework to model integrated optics a beam combiner for interferometry, both in aperture-masking, and long-baseline configurations. It models the complex effects of fringe-tracking residual injection into the waveguides, as well as the chromatic amplitude and phase aberrations intrinsic to the integrated optics component. This emphasis on instrumental errors is suitable for the demanding case of nulling interferometry, and allows evaluation of the covariance of the errors on the outputs. Its flexibility makes it ready to work with new sophisticated combiner architectures such as kernel-nulling.

We show how the knowledge of the covariance of errors can be used to determine the performance of the instrument in a realistic scenario. The approach takes into account the rotation of the Earth, the combination of all spectral channels, and the covariance of the instrumental errors. To our knowledge, this is a first for nulling interferometry observations. The approach provides upper and lower bounds for the sensitivity of matched subspace detectors to single companions around the central object for optimal detection tests depending on the initial hypotheses on spectrum and position.

Initial results show performance in contrast detection comparable to that of ground-based coronography and to that of GRAVITY in off-axis mode down to a few λ/D of the single-dish telescopes. Our simulations show that nulling interferometry can attain this performance down to an angular separation of λ/B for long-baseline interferometers, which is beyond the reach of extremely large telescopes. Future work will incorporate atmospheric dispersion and its compensation into the simulator, and elaborate a target selection strategy to maximize the scientific yield. The tools we present here are designed to be flexible and could be used to investigate the on-sky performance of other single-mode beam combiners, such as GRAVITY, GLINT, or future missions such as LIFE.

Acknowledgements. SCIFY has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement CoG - 866070). This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004719. The authors wish to thank Stephane Lagarde, Fatmé Allouche, and Philippe Berio for their helpful suggestions regarding the design of MATISSE.

References

Absil, O., Hartog, R. D., Gondoin, P., et al. 2006, A&A, 448, 787
Absil, O., Coudé du Foresto, V., Barillot, M., & Swain, M. R. 2007, A&A, 475, 1185
Angel, J. R. P., & Woolf, N. J. 1997, ApJ, 475, 373
Anugu, N., Le Bouquin, J.-B., Monnier, J. D., et al. 2020, AJ, 160, 158
Berger, J.-P., Zins, G., Lazareff, B., et al. 2010, Proc. SPIE, 7734, 773435
Bracewell, R. N. 1978, Nature, 274, 780
Ceau, A., Mary, D., Greenbaum, A., et al. 2019, A&A, 630, A120
Colavita, M. M., Swain, M. R., Akeson, R. L., Koresko, C. D., & Hill, R. J. 2004, PASP, 116, 876
Cvetojevic, N., Norris, B. R. M., Gross, S., et al. 2021, Applied Optics, 60, D33
Dandumont, C., Laugier, R., Emsenhuber, A., et al. 2022a, SPIE, 12183, 779
Dandumont, C., Mazzoli, A., Laborde, V., et al. 2022b, SPIE, 12183 700
Dannert, F., Ottiger, M., Quanz, S. P., et al. 2022, A&A 664, A22
Defrère, D., Absil, O., Den Hartog, R., Hanot, C., & Stark, C. 2010, A&A, 509, A9
Defrère, D., Hinz, P. M., Mennesson, B., et al. 2016, ApJ, 824, 66
Defrère, D., Absil, O., Berger, J. P., et al. 2018a, Exp. Astron., 46, 475
Defrère, D., Absil, O., Berger, J.-P., et al. 2018b, Proc. SPIE, 10701, 10701U
Defrère, D., Bigioli, A., Dandumont, C., et al. 2022, SPIE, 12183, 184
GRAVITY Collaboration (Abuter, R., et al.) 2017, A&A, 602, A94
GRAVITY Collaboration (Abuter, R., et al.) 2021 , A&A, 647, A59
Guyon, O., Pluzhnik, E. A., Galicher, R., et al. 2005, ApJ, 622, 744
Guyon, O., Pluzhnik, E. A., Kuchner, M. J., Collins, B., & Ridgway, S. T. 2006, ApJS, 167, 81
Guyon, O., Mennesson, B., Serabyn, E., & Martin, S. 2013, PASP, 125, 951
Hanot, C., Mennesson, B., Martin, S., et al. 2011, ApJ, 729, 110
Ireland, M. J. 2013, MNRAS, 433, 1718
Ireland, M. J., Defrère, D., Martinache, F., et al. 2018, Proceedings of the SPIE, 10701, 10
Kammerer, J., Mérand, A., Ireland, M. J., & Lacour, S. 2020, A&A, 644, A110
Kervella, P., Arenou, F., Mignard, F., & Thévenin, F. 2019, A&A, 623, A72
Kervella, P., Arenou, F., & Thévenin, F. 2022, A&A, 657, A7
Koresko, C., Colavita, M. M., Serabyn, E., Booth, A., & Garcia, J. 2006, Adv. Stellar Interferometry, 6268, 626817
Lacour, S., Jocou, L., Moulin, T., et al. 2008, Opt. Infrared Interferometry, 7013, 701316
Lacour, S., Tuthill, P., Monnier, J. D., et al. 2014, MNRAS, 439, 4018
Lacour, S., Dembet, R., Abuter, R., et al. 2019a, A&A, 624, A99
Lacour, S., Nowak, M., Wang, J., et al. 2019b, A&A, 623, L11
Lanthermann, C., Anugu, N., Le Bouquin, J.-B., et al. 2019, A&A, 625, A38
Laugier, R., Cvetojevic, N., & Martinache, F. 2020, A&A, 642, A202
Lay, O. P. 2004, Appl. Opt., 43, 6100
Lay, O. P. 2006, Proc. SPIE, 6268, 62681A
Lazareff, B., Le Bouquin, J.-B., & Berger, J.-P. 2012, A&A, 543, A31
Lazareff, B., Blind, N., Jocou, L., et al. 2014, Proc. SPIE, 91460X, 15
Le Bouquin, J.-B., Rousselet-Perraut, K., Berger, J.-P., et al. 2008, SPIE, 70130 70130F
Le Bouquin, J. B., Berger, J. P., Lazareff, B., et al. 2011, A&A, 535, A67
Martinache, F., & Ireland, M. J. 2018, A&A, 619, A87
Martinod, M.-A., Norris, B., Tuthill, P., et al. 2021, Nat. Commun., 12, 2465
Martinod, M.-A., Defrère, D., Ireland, M. J., et al. 2022, SPIE, 12183, 1218310
Mathar, R. J. 2004, Appl. Opt., 43, 928
Mennesson, B., Defrère, D., Nowak, M., et al. 2016, SPIE, 9907, 99070X
Mugnier, L., Thiébaut, E., & Belu, A. 2006, EAS Pub. Ser., 22, 69
Norris, B. R. M., Cvetojevic, N., Lagadec, T., et al. 2020, MNRAS, 491, 4180
Nowak, M., Lacour, S., Lagrange, A.-M., et al. 2020a, A&A, 642, L2
Nowak, M., Lacour, S., Mollière, P., et al. 2020b, A&A, 633, A110
Pannetier, C., Mourard, D., Berio, P., et al. 2021, Proc. SPIE, 11446, 14460T
Sharma, T. K., Labadie, L., Strixner, D., et al. 2020, Proceedings of the SPIE, 11446, 1144618
Soldano, L. B., & Pennings, E. C. 1995, J.f Lightwave Technol., 13, 615
Tcherniavski, I. 2011, Opt. Eng., 50, 033206
Tepper, J., Labadie, L., Diener, R., et al. 2017, A&A, 602, A66
Thiébaut, E., & Mugnier, L. 2006, IAU Colloq., 200, 547
Velusamy, T., Angel, R. P., Eatchel, A., Tenerelli, D., & Woolf, N. J. 2003, ESA SP, 2003, 631
Woillez, J., Abad, J. A., Abuter, R., et al. 2019, A&A 629, A41

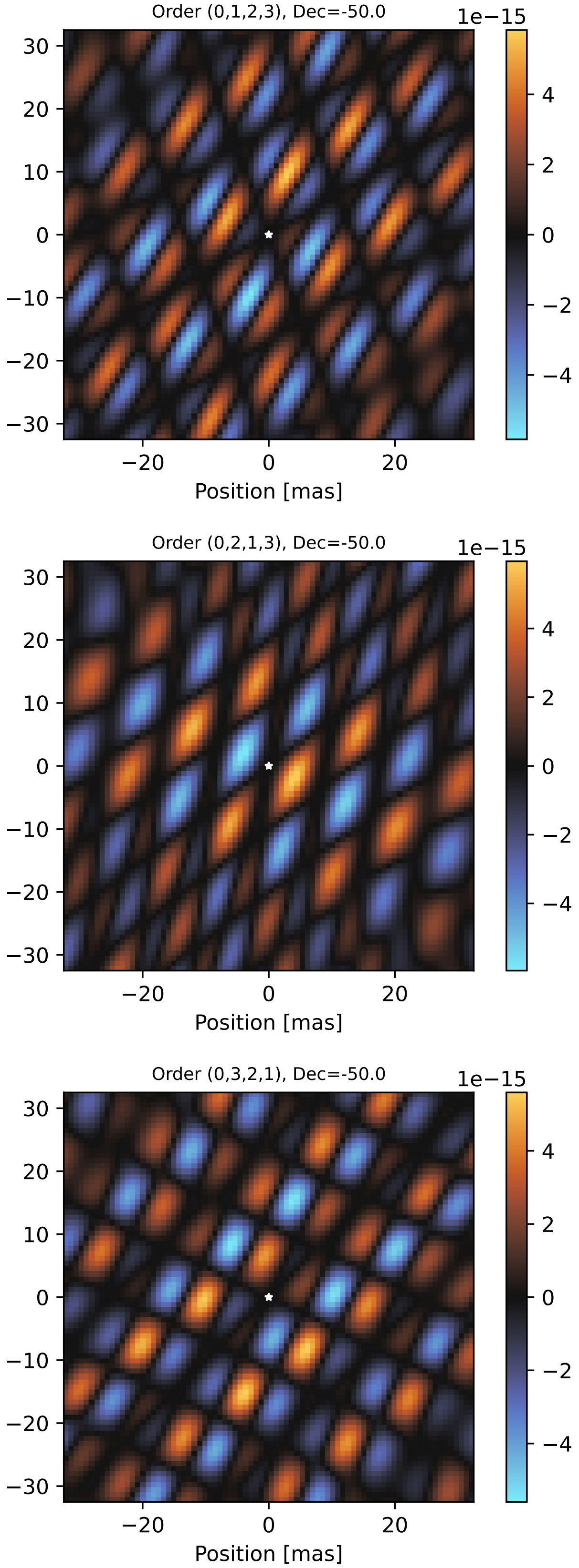
Appendix A: Variation of the sensitivity maps with configuration and sky position

The sensitivity map is affected by multiple parameters of the pupil. The first is the position of the collecting telescopes, through the length and orientation of the baselines it offers. Here we use the position of the four UTs as an example. The second is the diameter of individual apertures. Through the spatial filtering of the single-mode waveguide, they dictate the field of view and its progressive falloff at larger separations. The third is the permutation of the collecting telescopes with respect to the inputs of the instrument. The double Bracewell configuration of NOTT offers only one pair of dark outputs, which corresponds to one of the kernel pairs of the VIKiNG solution proposed by Martinache & Ireland (2018). However, one can produce the three different configurations of the VIKiNG nuller by observing with permutations of the input telescopes (Fig. A.1 for one given sky pointing).

The fourth is the pointing in the sky which changes the projected geometry of the array. Throughout an observing session, the rotation of the Earth between the different chunks of the sequence changes this geometry by a projection into the plane orthogonal to the line of sight, effectively stretching and rotating the maps of Fig. A.1. Within the parameters used here, the effect of airmass on throughput and added photon noise are negligible, as this case is not sensitivity limited. The effect of airmass on the performance of adaptive optics, fiber coupling, and fringe tracking is not taken into account here.

The overall sensitivity compounded throughout the session is affected. For comparison, we plot side-by-side the sensitivity map for targets of different declinations observed in the same six-hour series of 20 chunks spread throughout the passage of the meridian (see Fig. A.2). These maps are shown in Fig. A.3 and are all distinct in the location of higher and lower sensitivity. The various sensitivity lobes are spread out in a bow-like pattern, while the contribution of the different wavelengths contribute to a radial blurring at some of the larger separations.

For a blind search, one would argue that the overall sensitivity across the field of view appears equivalent, although some retain some better sensitivity around the first few mas around the host star, as has been shown in the first yield estimations by Dandumont et al. (2022a). However, for a targeted characterization observation with prior knowledge of the position of the planet, this would help to pick a configuration maximizing the S/N. Even astrometry information, such as that gathered by GAIA (Kervella et al. 2019, 2022), can offer hints on the position angle of a suspected planet by the acceleration excess measured, and could prompt the use of a particular beam permutation.


Order (0,1,2,3), Dec=-50.0

Order (0,2,1,3), Dec=-50.0

Order (0,3,2,1), Dec=-50.0

Fig. A.1. Raw transmission maps of the differential null for three different input orders and the same sky pointing, on the meridian for a target at Dec −50.0. The signal is summed over all spectral channels. Unlike the four-telescope kernel-nuller, NOTT can only obtain one at a time. The values take into account the transmission of the whole instrument, (including quantum efficiency etc.) and are given in units of étendue [sr.m^2], interpreted as the equivalent collecting capacity with regard to each elementary patch of sky (pixel).

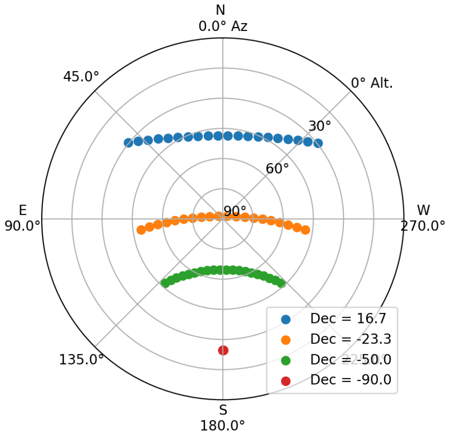


Fig. A.2. Map showing altitude and azimuth of the series of pointings combined to establish the maps in Fig. A.3.

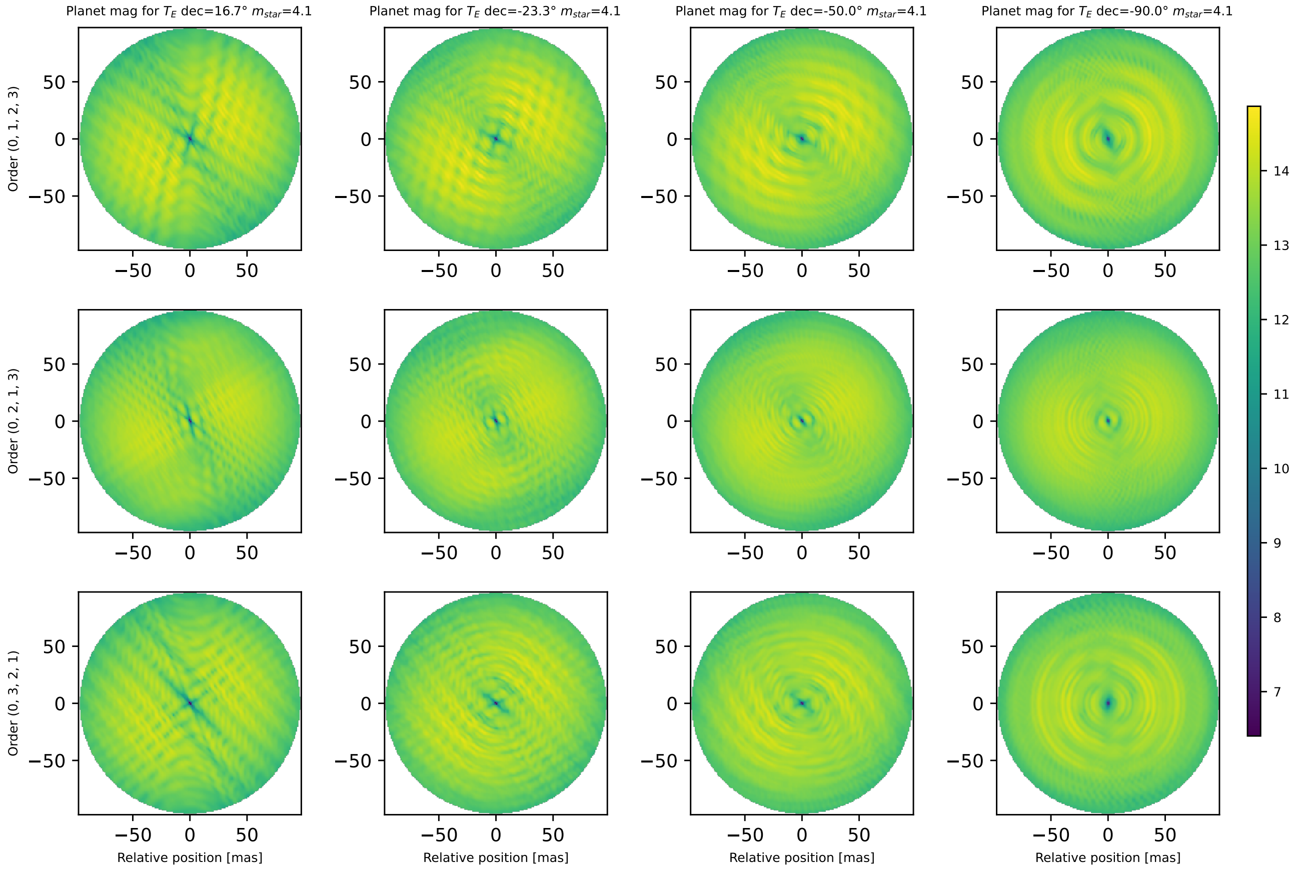


Fig. A.3. Sensitivity around a star of fixed magnitude,, for different declinations and input order. The distribution of the sensitivity changes significantly, but the overall performance is preserved. The order (0, 2, 1, 3) seems to offer better sensitivity at the smaller separations, owing to the slightly tighter position of the first inner lobes of the transmission map if Fig. A.2.