The GRAVITY fringe tracker

S. Lacour, R. Dembet, R. Abuter, P. Fédou, G. Perrin, E. Choquet, O. Pfuhl,
F. Eisenhauer, J. Woillez, F. Cassaing, et al.

The GRAVITY fringe tracker

S. Lacour[1,2,3], R. Dembet[1,4], R. Abuter[4], P. Fédou[1], G. Perrin[1], É. Choquet[5], O. Pfuhl[2], F. Eisenhauer[2], J. Woillez[4], F. Cassaing[6], E. Wieprecht[2], T. Ott[2], E. Wiezorrek[2], K. R. W. Tristram[4], B. Wolff[4], A. Ramírez[4], X. Haubois[4], K. Perraut[7], C. Straubmeier[8], W. Brandner[9], and A. Amorim[10]

[1] LESIA, Observatoire de Paris, Université PSL, CNRS, Sorbonne Université, Univ. Paris Diderot, Sorbonne Paris Cité, 5 place Jules Janssen, 92195 Meudon, France
    e-mail: sylvestre.lacour@obspm.fr
[2] Max Planck Institute for extraterrestrial Physics, Giessenbachstraße 1, 85748 Garching, Germany
[3] Department of Physics, University of Cambridge, CB3 0HE Cambridge, UK
[4] European Southern Observatory, Karl-Schwarzschild-Straße 2, 85748 Garching, Germany
[5] Aix Marseille Univ., CNRS, CNES, LAM, Marseille, France
[6] ONERA/DOTA, Université Paris Saclay, 92322 Châtillon, France
[7] Univ. Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France
[8] 1st Institute of Physics, University of Cologne, Zülpicher Straße 77, 50937 Cologne, Germany
[9] Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany
[10] CENTRA – Centro de Astrofísica e Gravitação, IST, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal

ABSTRACT

Context. The GRAVITY instrument was commissioned on the VLTI in 2016 and is now available to the astronomical community. It is the first optical interferometer capable of observing sources as faint as magnitude 19 in K band. This is possible through the fringe tracker, which compensates the differential piston based on measurements of a brighter off-axis astronomical reference source.
Aims. The goal of this paper is to describe the main developments made in the context of the GRAVITY fringe tracker. This could serve as basis for future fringe-tracking systems.
Methods. The paper therefore covers all aspects of the fringe tracker, from hardware to control software and on-sky observations. Special emphasis is placed on the interaction between the group-delay controller and the phase-delay controller. The group-delay control loop is a simple but robust integrator. The phase-delay controller is a state-space control loop based on an auto-regressive representation of the atmospheric and vibrational perturbations. A Kalman filter provides the best possible determination of the state of the system.
Results. The fringe tracker shows good tracking performance on sources with coherent K magnitudes of 11 on the Unit Telescopes (UTs) and 9.5 on the Auxiliary Telescopes (ATs). It can track fringes with a signal-to-noise ratio of 1.5 per detector integration time, limited by photon and background noises. During good seeing conditions, the optical path delay residuals on the ATs can be as low as 75 nm root mean square. The performance is limited to around 250 nm on the UTs because of structural vibrations.

Key words. instrumentation: interferometers – techniques: high angular resolution

1. Introduction

GRAVITY (Gravity Collaboration 2017) is an instrument used on the Very Large Telescope Interferometer (VLTI) situated at the Cerro Paranal Observatory. It can combine the light from four telescopes. These telescopes can either be the four Auxiliary Telescopes (ATs, with a primary mirror diameter of 1.8 m) or the four Unit Telescopes (UTs, with a diameter of 8 m). The specifications of the instrument were derived from the most demanding science case, which was to observe microarcsecond displacements of the light source causing the flares of the supermassive black hole Sgr A* (Genzel et al. 2010, and references herein). Such astrometric measurements are possible with 100 m baselines (Shao & Colavita 1992; Lacour et al. 2014) and were recently demonstrated on-sky by Gravity Collaboration (2018a,b,c).

In its quiescent state, Sgr A* can become fainter than $K = 18$ mag. Therefore, measuring its position reliably requires an integration time on the order of minutes. To enable such long integration times, it is important to correct the atmosphere effects in real time. The higher-order atmospheric wavefront distortions are compensated for by an adaptive optics (AO) system. However, the AO systems do not sense, and therefore cannot correct, the differential phase between the telescopes. This is the role of the fringe tracker: a phase-referencing target (IRS 16C in the case of Sgr A*) is used as a guide star. In real time, the optical path differences (OPD) between each pair of telescopes are computed, and are used to control the displacement of mirrors on piezoelectric systems. This is the counterpart of the AO system, but at interferometric scale.

To push the comparison a little farther: without fringe tracking, interferometry requires short integration times and deconvolution techniques. This was the time of speckle imaging (Labeyrie 1970; Weigelt 1977), when using bispectrum and closure phases was a good but not really sensitive technique. With fringe tracking, optical interferometry enters a new age: long detector integration times (DIT up to 300 s) give access to faint sources (Kmag of 19) and to the possibility of combining

spectral resolution (up to 4000) with milliarcsecond spatial resolution. This is the historical equivalent to the emergence of adaptive optics: it enables a new level of science.

Fringe tracking is not new, however. Small observatories demonstrated the concept, with PTI and CHARA (Berger & Monnier 2006). On the Keck Interferometer (Colavita et al. 2013), comparable astrophysical objectives (Woillez et al. 2012) pushed a similar development for phase referencing (Colavita et al. 2010). Previous projects also existed at the VLTI: at first, the FINITO fringe tracker (Le Bouquin et al. 2008) was used in combination with the AMBER instrument (Petrov et al. 2007). More recently, ESO developed the PRIMA fringe tracker (Delplancke et al. 2006).

However, unlike AO, fringe tracking was not yet mature, and many complex problems had to be investigated for GRAVITY. A first problem was how to deal with limited degrees of freedom (the piston actuators) while many more optical path differences are measured (Menu et al. 2012). A second problem, which does not exist in AO, is that the phase signals are known only modulo 2π. A third problem, partially addressed by the AO community (Petit et al. 2008; Poyneer & Véran 2010), is how to set a correct state space control system that optimally uses a Kalman filter to cancel the vibrations (Choquet et al. 2014). A fourth difficulty is that both group-delay (GD) and phase-delay (PD) tracking need to be used in a control system to keep the best of both. The last hurdle of the project consisted of dealing with multiple closing baselines, some of them resolved. The GRAVITY fringe tracker is now the best and most modern instrument in the field of fringe tracking for optical interferometry. Below we describe the algorithms and mechanisms.

This paper builds upon earlier works from Cassaing et al. (2008), Houairi et al. (2008), Lozi et al. (2011), Menu et al. (2012), and Choquet et al. (2014)[1]. The Menu et al. (2012) paper theoretically describes modal control of the phase delay. The Choquet et al. (2014) paper simulates the expected performance of the Kalman controller. The present paper wraps up the series by presenting the final implementation on the VLTI. Section 2 is an overview of the technical implementation of the fringe tracker and is followed by Sect. 3, where the basis of the fringe sensing is defined by the observables. The control algorithm is presented in three sections: Sect. 4 defines the operational modes, Sect. 5 presents the group-delay controller, and Sect. 6 presents the phase-delay controller. Section 7 gives examples and statistics of on-sky observations. Last, Sect. 8 concludes the paper by a discussion of possible improvements to increase the sensitivity and accuracy of the fringe tracker.

2. Overview of the fringe-tracking system

2.1. Hardware

GRAVITY is equipped with two beam combiners (Perraut et al. 2018) that perform fringe tracking and scientific observations in the K band. GRAVITY has two main operational modes. In the on-axis mode, the light of one star is split 50:50, that is, equal portions of the flux go to the fringe tracking and to the scientific channel. In off-axis mode, the field is split into two with the help of a roof mirror: one of the two objects serves as fringe-tracking reference, while the scientific channel carries out longer integrations on the typically fainter science target (Pfuhl et al. 2010, 2012, 2014).
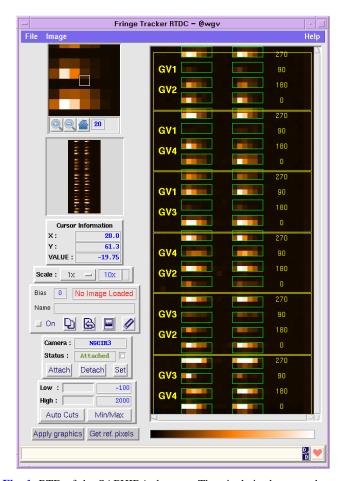
---

[1] Presented during the Final Design Review as document *VLT-TRE-GRA-15882-6701*.

Fig. 1. RTD of the SAPHIRA detector. The pixels in the green boxes are read by the fringe tracker and are used for tracking. They correspond to the six baselines, four ABCD outputs, two polarizations, and six wavelength channels. The names GV1 to GV4 correspond to the input beams. The values in yellow to the right correspond to the phase shift in degrees between the different ABCD outputs.

The two beam combiners are based on silica-on-silicium integrated optics (Malbet et al. 1999), optimized for $2\,\mu$m observations (Jocou et al. 2010). The beam combination scheme is pair-wise. Each telescope pair is combined using a static ABCD phase modulation. This means that for each of the six baselines, there are four outputs corresponding to a phase shift between the two beams of 0, $\pi/2$, π, and $3\pi/2$ radians. The total 24 outputs of the beam combiner can be seen on the real time display (RTD) of the instrument. The relevant pixels are delimited by the green rectangles in Fig. 1. Within each rectangle, the flux is dispersed in six spectral channels. The 24 lines of rectangles correspond to the 24 outputs, while the two columns of rectangles correspond to the two linear polarizations.

The detector is a HgCdTe avalanche photodiode array called SAPHIRA (Finger et al. 2016). It is running at 909 Hz, 303 Hz, or 97 Hz. It can also run at either low or high gain. The high gain corresponds to a gain of $\gamma = 7$ ADU per photodetection with a typical readout noise below $\sigma_{\rm RON} = 5$ ADU ($\approx 0.7\,$e$^-$) per pixel. The low gain does not amplify the photodetections ($\gamma = 0.5$ ADU/e$^-$) and is only used for very bright targets (K magnitudes below 5 on the UTs).

The flux is processed by a first local control unit (LCU) that yields values of the observables. The LCU is an Artesyn MVME6100 using an MPC7457 PowerPC®processor (Kiekebusch et al. 2010). The data are then transmitted to a

second LCU by means of a distributed memory system called reflective memory network (RMN). This second LCU processes the observables to control four tip-tilt piston mirrors on piezo-electric actuators from Physik Instrument. Each actuator has its own position sensor, driven in closed loop. The cutoff frequency of the piezoelectric delay lines is then higher than 300 Hz, with a maximum optical path delay of 60 μm (Pfuhl et al. 2014).

Real-time monitoring of the fringe tracking, including live display, is done on a separate Linux workstation connected to the LCUs through the RMN. This workstation processes the data, computes the best-fit control parameters (including the Kalman parameters), and updates the control parameters of the second LCU (Abuter et al. 2016).

2.2. Software

The two LCUs use the VxWorks operating systems. The computation is done using the so-called tools for advanced control (TAC) framework, which uses the standard C language environment. The TAC processing is triggered synchronously with the SAPHIRA, following the predefined frequency of the detector.

The first LCU computes the necessary estimators for the controller:

– Four flux values (F_i); see Sect. 3.2.
– Six phase delays ($\Phi_{i,j}$); see Sect. 3.3.
– Six phase-delay variance (Var($\Phi_{i,j}$)); see Sect. 3.4.
– Six group delays ($\Psi_{i,j}$); see Sect. 3.5.
– Four closure phases ($\Theta_{i,j,k}$); see Sect. 3.6.

Inside this LCU, only a few parameters can be changed: the number of DITs over which each of these quantities can be averaged. The default values are presented in Sect. 3.

The second LCU is in charge of controlling the piezo-mirrors for adequate fringe tracking. Figure 2 is a block diagram of the controller algorithm:

– The group-delay control loop, based on an integrator controller, with a direct command to the piezo-actuator (in blue).
– A feed-forward predictor, based on the action to the actuators, to increase the gain of the closed-loop system (in green).
– The phase-delay Kalman controller (in red).
– Two peripheral blocks for searching the fringes and adding a π modulation to the phase delay.

The dual architecture of the controllers is made to obtain both sensitivity and accuracy. In case of high signal-to-noise ratio (S/N), the Kalman filter can determine and predict the state of both atmospheric and vibrational perturbation for the best possible correction. In case of low S/N, the Kalman filter relies on its predictive model, which in the worst case, can be as simple as a constant value. In this case, the group-delay controller is still working efficiently and provides coherence instead of fringe tracking.

The third and last software element is on the Linux workstation. It is a python script that runs every 5 s on the last 40 s of data calculated on the first LCU. It computes the parameters that are then used by the second LCU, which does the real time control. It includes the parameters for the predictive control and the best gain for the Kalman filter.

3. Estimators of observables

3.1. Visibility extraction

The real-time processing load consists mostly of matrix multiplications. The pixel-to-visibility (P2VM) principle was first used for data reduction (Tatulli et al. 2007) on AMBER. AMBER

used spatial modulation for fringe coding, but the formalism was subsequently adapted to work with ABCD beam combiners (Lacour et al. 2008). Using a similar notation, the relation between the incoming electric field E_i and the outgoing electric field S_k can be expressed as

$$S_k = \sum_i T_{k,i} E_i, \tag{1}$$

where $T_{k,i}$ is the complex transfer function from the input i of the beam combiner to its output k. In the case of GRAVITY, $i_{\max} = 4$ and $k_{\max} = 24$. Averaged over the DIT, the flux is then equal to the average of its instantaneous intensity:

$$\langle |S_k|^2 \rangle_{\mathrm{DIT}} = \langle \left| \sum_n T_{k,i} E_i \right|^2 \rangle_{\mathrm{DIT}}, \tag{2}$$

or alternatively, after decomposition:

$$\langle |S_k|^2 \rangle = \Re \left[\sum_n |T_{k,i}|^2 \langle |E_i|^2 \rangle + 2 \sum_i \sum_{j>i} \langle T_{k,i} T_{k,j}^* E_i E_j^* \rangle \right]. \tag{3}$$

The temporal average of the electric field leads to two types of coherence losses. One depends on the optical path inside the beam combiner, the other on the spatial brightness distribution of the astrophysical object. The first, $C_{k,i,j}$, is intrinsic to the device and has to be calibrated. The second, $V_{i,j}$, is the reason why we built the interferometer. The relation with the mean electric field intensity is then approximated by the following equation:

$$\langle T_{k,i} T_{k,j}^* E_i E_j^* \rangle = |T_{k,i} T_{k,j}^*| \, C_{k,i,j} \, \sqrt{\langle |E_i|^2 \rangle \langle |E_j|^2 \rangle} \, V_{i,j}. \tag{4}$$

Hence, Eq. (3) can be written as a matrix product:

$$\begin{pmatrix} \langle |S_1|^2 \rangle \\ \vdots \\ \langle |S_{24}|^2 \rangle \end{pmatrix} = \Re \left[\mathsf{V2PM} \cdot \begin{pmatrix} \langle |E_1|^2 \rangle \\ \langle |E_2|^2 \rangle \\ \langle |E_3|^2 \rangle \\ \langle |E_4|^2 \rangle \\ \sqrt{\langle |E_1|^2 \rangle \langle |E_2|^2 \rangle} V_{1,2} \\ \sqrt{\langle |E_1|^2 \rangle \langle |E_3|^2 \rangle} V_{1,3} \\ \sqrt{\langle |E_1|^2 \rangle \langle |E_4|^2 \rangle} V_{1,4} \\ \sqrt{\langle |E_2|^2 \rangle \langle |E_3|^2 \rangle} V_{2,3} \\ \sqrt{\langle |E_2|^2 \rangle \langle |E_4|^2 \rangle} V_{2,4} \\ \sqrt{\langle |E_3|^2 \rangle \langle |E_4|^2 \rangle} V_{3,4} \end{pmatrix} \right], \tag{5}$$

where, using $^\top$ as the transpose operator:

$$\mathsf{V2PM}^\top = \begin{pmatrix} |T_{1,1}|^2 & \cdots & |T_{24,1}|^2 \\ \vdots & \ddots & \vdots \\ |T_{1,4}|^2 & \cdots & |T_{24,4}|^2 \\ |T_{1,1}T_{1,2}^*|C_{1,1,2} & \cdots & |T_{24,1}T_{24,2}^*|C_{24,1,2} \\ \vdots & \ddots & \vdots \\ |T_{1,3}T_{1,4}^*|C_{1,3,4} & \cdots & |T_{24,3}T_{24,4}^*|C_{24,3,4} \end{pmatrix}. \tag{6}$$

Everything related to the transfer function of the instrument is inside the 10 columns and 24 rows of the V2PM matrix. The V2PM is calibrated during daytime on the internal source of the instrument. It is regularly computed for verification, but has been proven to be very stable over several months. It is part of the calibration files that are needed by the first LCU (Sect. 2.2).

The P2VM is the pseudo-inverse matrix of V2PM. It can be obtained by splitting the V2PM into real and imaginary parts,
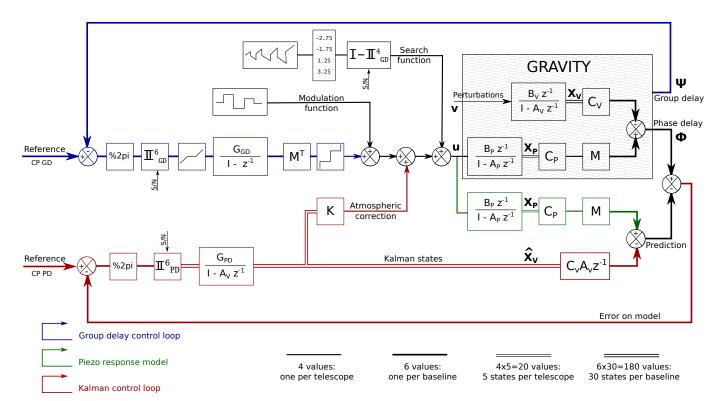
Fig. 2. Block diagram of the GRAVITY fringe-tracking controller. The gray area corresponds to the GRAVITY open-loop transfer function, and the rest corresponds to the two controllers. The group-delay integrator controller is shown in blue, the phase-delay state controller in red, and the actuators predictive model is plotted in green. The group-delay controller is the main controller: it continues to track the fringes even if the instantaneous S/N is too low for phase-delay tracking. The phase-delay state controller is a closed-loop system that determines the atmospheric perturbations $\hat{X}_V$. A proportional controller (K) corrects for the effect of the atmosphere. The last block of the group-delay controller, the quantization step, ensures that the group-delay control signal is always a multiple of 2π: the change in OPD caused by the group-delay controller is not seen by the phase-delay controller.

and taking the inverse using a singular-value decomposition (SVD). The P2VM is used to retrieve the astrophysical information from the flux observed on the pixels. This information consists of the flux F and the coherent flux Γ. Both are computed from the pixel flux $q_{k,\lambda}$ and P2VM according to Eq. (7):

$$
\begin{pmatrix}
F_{1,\lambda} \\
\vdots \\
F_{4,\lambda} \\
\Re[\Gamma_{1,2,\lambda}] \\
\vdots \\
\Re[\Gamma_{3,4,\lambda}] \\
\Im[\Gamma_{1,2,\lambda}] \\
\vdots \\
\Im[\Gamma_{3,4,\lambda}]
\end{pmatrix}
= \text{P2VM}_\lambda \cdot
\begin{pmatrix}
q_{1,\lambda} \\
\vdots \\
q_{24,\lambda}
\end{pmatrix}.
\tag{7}
$$

In the above equation, we used a slightly different nomenclature, which is used hereafter. $q_{k,\lambda} = < |S_{k,\lambda}|^2 >$ is the number of photons observed over one DIT of the fringe tracker on a given pixel. $F_{i,\lambda} = < |E_{i,\lambda}|^2 >$ is the energy per DIT of the incoming beam i at wavelength λ. Last, the complex coherent flux $\Gamma_{i,j,\lambda}$ corresponds to the visibility before normalization by the flux. It is obtained from the flux $F_{i,\lambda}$ and the visibilities $V_{i,j,\lambda}$:

$$
\Gamma_{i,j,\lambda} = \sqrt{F_{i,\lambda} F_{j,\lambda}}\, \Re[V_{i,j,\lambda}] + i\,\sqrt{F_{i,\lambda} F_{j,\lambda}}\, \Im[V_{i,j,\lambda}],
\tag{8}
$$

using $\sqrt{F_{i,\lambda} F_{j,\lambda}} = \sqrt{\langle |E_i|^2\rangle \langle |E_j|^2\rangle}$. All these values are computed in real time for each of the six wavelengths in the K band. In case of split polarization (when the Wollaston is inserted), the calculation is also done independently for both polarizations.

3.2. Flux estimator

The flux F_i is the value extracted after each DIT from Eq. (7), summed over the $N_\lambda = 6$ spectral channels:

$$
F_i = \sum_{\lambda=1}^{N_\lambda} F_{i,\lambda},
\tag{9}
$$

where i corresponds to the input beam number.

3.3. Phase delay estimator

The phase delay $\Phi_{i,j}$ is derived from the complex coherent flux, but after a first step to correct for the phase curvature caused by the dispersion:

$$
\Gamma'_{i,j,\lambda} = \Gamma_{i,j,\lambda} \exp\left(iD \left[1 - \frac{2.2\,\mu\text{m}}{\lambda} \right]^2 \right),
\tag{10}
$$

which is only a first-order approximation of the dispersion. It is caused both by the atmosphere and by the fibered differential delay lines (FDDL)[2]. Therefore, the corrective term D is a time-variable parameter that depends on the position of the star and the position of the FDDLs. The phase delay is then extracted by coherent addition of the six spectral channels:

$$
\Phi_{i,j} = \arg\left(\sum_{\lambda=1}^{N_\lambda} \Gamma'_{i,j,\lambda} \right).
\tag{11}
$$

---

[2] For more information, see the GRAVITY/ESO Final Design Review document VLT-TRE-GRA-15882-6401.

It is worth noting that the phase delay $\Phi_{i,j}$ is wrapped: it lies between $-\pi$ and π. No unwrapping effort is made at this stage.

3.4. Phase variance estimator

Computing the S/N of the fringes is essential for the fringe tracker. It ensures that the controller does not track on noise. It is also needed for the state machine to know if it has the fringes locked or if it must start looking for the fringes elsewhere. The S/N is calculated from the variance of the phase delay: $S/N_{i,j} = 1/\sqrt{\text{Var}(\Phi_{i,j})}$. For each DIT, from the photon and background noise, the variance on each pixel is estimated by the relation:

$$\text{Var}(q_{i,\lambda}) = \sigma_{\text{sky}}^2 + \gamma(q_{i,\lambda} - q_{i,\lambda,\text{sky}}), \qquad (12)$$

where σ_{sky} and $q_{i,\lambda,\text{sky}}$ are the noise and flux, respectively, observed during sky observations. The term γ is the detector gain in ADU per detected photon. The covariance matrix of the real and imaginary part of the Γ terms can be obtained from the P2VM:

$$\Sigma_\Gamma = \text{P2VM}_\lambda \cdot \begin{pmatrix} \text{Var}(q_{1,\lambda}) \\ \vdots \\ \text{Var}(q_{24,\lambda}) \end{pmatrix} \cdot \text{P2VM}_\lambda^\top, \qquad (13)$$

where $^\top$ is the transpose operator. To save processing time, only the diagonal values of the variance matrix are calculated. They correspond to the variance of the real and imaginary parts of $\Gamma_{i,j,\lambda}$. This assumes that the covariance between the real and imaginary parts is negligible (a good assumption for an ABCD with phase shift of $0, \pi/2, \pi$, and $3\pi/2$). In the end, for simplicity, we estimated the variance of the phase by the following equation:

$$\text{Var}(\Phi_{i,j}) = \frac{\sum_\lambda \langle \text{Var}(\Re\Gamma_{i,j,\lambda}) + \text{Var}(\Im\Gamma_{i,j,\lambda}) \rangle_{5\text{DIT}}}{2 \left| \sum_\lambda \langle \Gamma'_{i,j,\lambda} \rangle_{5\text{DIT}} \right|^2}, \qquad (14)$$

which is the variance of the amplitude of the coherent flux averaged over five DIT. The five-DIT average is a way to increase the precision of the calculation. However, this is done at the expense of accuracy: the coherent averaging of the complex coherent flux can add a negative bias to the phase variance estimator.

3.5. Group delay estimator

The group delay, $\Psi_{i,j}$, is also obtained from the complex coherent flux. Because it consists of a differential measure of the phase as a function of wavelength, this estimator is more noisy than the phase delay (Lawson et al. 2000). To increase its S/N, for each one of the $N_\lambda = 6$ spectral channels, the $\Gamma_{i,j,\lambda}$ is first corrected for dispersion, cophased, and averaged over 40 DITs. The result is then used to derive the group delay by multiplying the phasor of consecutive spectral channels:

$$\Gamma''_{i,j,\lambda} = \langle \Gamma'_{i,j,\lambda} \exp(-i\Phi_{i,j}) \rangle_{40\text{DIT}} \qquad (15)$$

$$\Psi_{i,j} = \arg\left(\sum_{\lambda=1}^{N_\lambda - 1} \Gamma''_{i,j,\lambda+1} \Gamma''^*_{i,j,\lambda} \right). \qquad (16)$$

As for the phase delay, the group delay is estimated modulo 2π. In terms of optical path, however, $\Psi_{i,j}$ corresponds to a value R

times smaller than $\Phi_{i,j}$, with $R = 23$, the spectral resolution of the GRAVITY fringe tracker. This is explicit in open-loop operation where the phase delay and group delay are compared for a response to top-hat piezo commands. Because both estimators wrap at 2π, this means that the estimator is valid over a long range equal to 23 times the wavelength. This is the main advantage of the group-delay estimator: to be able to find fringes far away from the central white-light fringe (the fringe of highest contrast). However, because it uses individual spectral channels, the group delay calculated on a single DIT would be extremely noisy. Thus the 40 DIT summation is a way to increase the S/N of the group delay, at the cost of losing response time. In the end, the group-delay estimator is a reliable, but slow, estimator of the optical path difference.

3.6. Closure-phase estimator

The closure phases, $\Theta_{i,j,k}$, are calculated on all four triangles from the coherent flux. Before taking the argument, the bispectra are averaged over 300 DIT (corresponding to 330 ms at the fastest 909 Hz sampling rate). Closure phases are estimated from the phase delay:

$$\Theta^{\text{PD}}_{i,j,k} = \arg\left(\langle \sum_{\lambda=1}^{N_\lambda} \Gamma'_{i,j,\lambda} \sum_{\lambda=1}^{N_\lambda} \Gamma'_{j,k,\lambda} \sum_{\lambda=1}^{N_\lambda} \Gamma'^*_{i,k,\lambda} \rangle_{300\text{DIT}} \right), \qquad (17)$$

but also from the group delay:

$$\Theta^{\text{GD}}_{i,j,k} = \arg\left(\langle \sum_{\lambda=1}^{N_\lambda} \Gamma''_{i,j,\lambda} \sum_{\lambda=1}^{N_\lambda} \Gamma''_{j,k,\lambda} \sum_{\lambda=1}^{N_\lambda} \Gamma''^*_{i,k,\lambda} \rangle_{300\text{DIT}} \right). \qquad (18)$$

They are observables modulo 2π, no unwrapping is intended.

4. State spaces, projections, and state machine

4.1. OPD-state space

A main difficulty for the fringe tracker consists of dealing with the different dimensions of the vectors involved. The number of phase observables is six. The number of delay lines is four. Last, the number of degrees of freedom is three. In Menu et al. (2012), we proposed an $\mathbb{R}^3$ modal-state space orthogonal to the piston space. However, as also mentioned in Menu et al. (2012), this modal control has an important drawback: it cannot work in a degraded mode where one or more telescopes are missing. Instead, the GRAVITY controller uses the OPD-state space. In some sense, the implemented fringe tracker is a downgraded version of the state controller proposed in Menu et al. (2012). However, it facilitates managing flux drop-out as well as working with a reduced number of baselines.

4.2. Reference vectors

For the system to work properly, the OPD-state space must be colinear to the piston space. However, because the astronomical object is not necessarily a point source, the closure phases are not necessarily zero. Therefore, the OPD component orthogonal to the piston space must be removed from the measurement. This is done by subtracting a reference position, or set point, which is computed from the closure-phase estimators $\Theta^{\text{PD}}_{i,j,k}$ and $\Theta^{\text{GD}}_{i,j,k}$. Then, the error terms, that is, the differences between the measured OPD and the set points, are colinear to the piston space.

However, the devil is in the details. There are four closure phases, and only three can be used. The noisiest closure phase is

therefore discarded. The three other closure phases are applied on the three edges of the triangle forming the noisiest closure phase. The reference vector is therefore defined as:

$$\mathbf{Ref_\Phi} = \begin{pmatrix} \Theta^{PD}_{1,2,4} \\ \Theta^{PD}_{1,3,4} \\ 0 \\ \Theta^{PD}_{2,3,4} \\ 0 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} \Theta^{PD}_{1,2,3} \\ 0 \\ -\Theta^{PD}_{1,3,4} \\ 0 \\ -\Theta^{PD}_{2,3,4} \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 0 \\ -\Theta^{PD}_{1,2,3} \\ -\Theta^{PD}_{1,2,4} \\ 0 \\ 0 \\ \Theta^{PD}_{2,3,4} \end{pmatrix} \text{ or } \begin{pmatrix} 0 \\ 0 \\ 0 \\ \Theta^{PD}_{1,2,3} \\ \Theta^{PD}_{1,2,4} \\ \Theta^{PD}_{1,3,4} \end{pmatrix},$$

$$(19)$$

depending on which triangle has the lowest S/N, from left to right, the 123, 124, 134, or 234 triangle. The reference values for the group delay are calculated similarly:

$$\mathbf{Ref_\Psi} = \begin{pmatrix} \Theta^{GD}_{1,2,4} \\ \Theta^{GD}_{1,3,4} \\ 0 \\ \Theta^{GD}_{2,3,4} \\ 0 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} \Theta^{GD}_{1,2,3} \\ 0 \\ -\Theta^{GD}_{1,3,4} \\ 0 \\ -\Theta^{GD}_{2,3,4} \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 0 \\ -\Theta^{GD}_{1,2,3} \\ -\Theta^{GD}_{1,2,4} \\ 0 \\ 0 \\ \Theta^{GD}_{2,3,4} \end{pmatrix} \text{ or } \begin{pmatrix} 0 \\ 0 \\ 0 \\ \Theta^{GD}_{1,2,3} \\ \Theta^{GD}_{1,2,4} \\ \Theta^{GD}_{1,3,4} \end{pmatrix}.$$

$$(20)$$

The closure-phase changes as a function of time, causing the reference position to adapt to any change in the phase closures. The closure phase is smoothed over a long enough time (300 DIT) to avoid adding additional noise. However, the choice of which triangle is the noisiest is made only once between each scientific frame to avoid sharp jumps in the reference vector over the integration time of the science detector.

This reference scheme works most of the time. However, problems arise in two specific instances. First, when the object is so highly resolved that the used closure phases contain a baseline with zero visibility. If that happens, an undefined reference value is applied to a perfectly sane baseline and the system can diverge. Second, if the fringe tracker is tracking on two unconnected baselines (e.g., between telescopes 1–2 and 3–4), the closure phases are undefined, and using their values would mean losing one of the two locked baselines. To resolve these two problems, the closure phases are modified as follows: if one of the baselines of any of ij, jk, or ik have an S/N below the value $S/N^{GD}_{threshold}$, then $\Theta^{PD}_{i,j,k}$ used in Eq. (19) is a fixed value, and $\Theta^{GD}_{i,j,k} = 0$ is used in Eq. (20). The difference in treatment between the group and phase delay arises because the default group-delay tracking shall be zero, while the default phase-delay tracking can be any constant value.

4.3. Transfer matrices

After the reference values are subtracted from the OPD, we can freely project the data in piston-state space as well as back to the OPD-state space. Hereafter, we use the same nomenclature as in Menu et al. (2012). $\mathbf{P}$ corresponds to the four-dimension piston-state vector, while $\mathbf{OPD}$ corresponds to the six-dimension OPD-state vectors. The matrix M corresponds to the conversion between piston and optical path difference:

$$\mathbf{OPD} = \mathsf{M}\,\mathbf{P}, \tag{21}$$

where

$$\mathsf{M} = \begin{pmatrix} -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}^\top. \tag{22}$$

The conversion $\mathbf{OPD} \to \mathbf{P}$ is ill-constrained, however: the rank of matrix M is 3, not 4. This is because the global piston cannot be obtained from the differences in the optical path. Nevertheless, we can define a pseudo-inverse matrix:

$$\mathsf{M}^\dagger = \frac{1}{4} \begin{pmatrix} -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}. \tag{23}$$

where $\dagger$ denotes the pseudo-inverse operator.

4.4. Thresholds and S/N management

The system uses two distinct thresholds. A first threshold, $S/N^{GD}_{threshold}$, disables the baselines where the S/N is too low to make the baseline useful. This is the case when the fringes are not yet found, when the fringes are suddenly lost, or when the astronomical target is so highly resolved that the spatial coherence is close to zero. To detect these events, the $S/N^{GD}_{threshold}$ is compared to a moving average of the phase-delay variance estimator: $\langle \text{Var}(\Phi_{i,j}) \rangle_{40\,DIT}$. This group-delay threshold must be adapted to the target: it mustbe high enough to ensure that the fringe tracker does not track on side-lobes, but low enough to detect the fringes.

A second threshold, $S/N^{PD}_{threshold}$, is used solely for the phase-tracking controller. It disables the tracking on a given telescope in case of rapid S/N drop-off. The main target of this threshold is to be able to catch a drop in flux injection (caused by external tip-tilt), with the expectation that the S/N will increase soon hereafter. Typically, $S/N^{PD}_{threshold} = 1.5$ and $S/N^{PD}_{threshold} < S/N^{GD}_{threshold}$.

Hence, we defined two matrices $\mathbb{I}$ that convert OPD space error into a space of the same dimension colinear to the OPD space. They read

$$\mathbb{I}^6_{GD} = \mathsf{M}(\mathsf{M}^\top \mathsf{W}\, \mathsf{M})^{\dagger_{GD}} \mathsf{M}^\top \mathsf{W} \tag{24}$$

$$\mathbb{I}^6_{PD} = \mathsf{M}(\mathsf{M}^\top \mathsf{W}\, \mathsf{M})^{\dagger_{PD}} \mathsf{M}^\top \mathsf{W}, \tag{25}$$

where the only difference is on the pseudo-inverse operator $\dagger$. In both equations, the 6×6 weighting matrix W distributes the weights among the different OPDs:

$$\mathsf{W} = \text{diag}\left(w_{1,2} \quad w_{1,3} \quad w_{1,4} \quad w_{2,3} \quad w_{2,4} \quad w_{3,4}\right), \tag{26}$$

where

$$w_{i,j} = \begin{cases} 0 & \text{if } 1/\langle \text{Var}(\Phi_{i,j}) \rangle_{40DIT} < |S/N^{GD}_{threshold}|^2 \\ 1/\text{Var}(\Phi_{i,j}) & \text{otherwise} \end{cases}. \tag{27}$$

This step is important to remove the risk of tracking on noise: if the variance of the phase reaches this threshold, the pseudo-inverse discards that baseline in its calculation. The pseudo-inversion is done using a SVD:

$$\mathsf{M}^\top \mathsf{W}\, \mathsf{M} = \mathsf{U}\mathsf{S}\mathsf{V}. \tag{28}$$

The idea behind this decomposition is that U and V are two invertible orthonormal matrices (the left-singular and right-singular eigenvectors, respectively): $\mathsf{I} = \mathsf{U}\mathsf{U}^\top$ and $\mathsf{I} = \mathsf{V}\mathsf{V}^\top$, where I is the identity matrix. S is a square diagonal matrix where the values on the diagonal correspond to the square root of the eigenvalues: $\mathsf{S} = \text{diag}(s_1, s_2, s_3, 0)$. For a four-telescope operation, three eigenvalues are non-zero. The number of non-zero eigenvalues decreases when the system cannot track all telescopes.
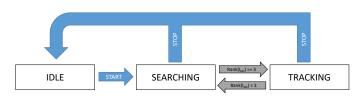
Fig. 3. State machine of the fringe tracker. There are only three states: IDLE, SEARCHING, and TRACKING. The blue transitions are commands from the operator. The gray transitions are automatic decisions. Switching from searching to tracking only depends on the rank of the $\mathbb{I}_{GD}^4$ matrix. If GRAVITY is running in a degraded mode, the TRACKING transition can happen for a rank of 2 or even 1.

The pseudo-inverse of matrix S is calculated differently for the group-delay and phase-delay control loop. For the group-delay control loop, we have $\mathsf{S}^{\dagger GD} = \mathrm{diag}(s_1^{\dagger GD}, s_2^{\dagger GD}, s_3^{\dagger GD}, 0)$, where

$$s_i^{\dagger GD} = \begin{cases} 1/s_i & \text{if } s_i > 0 \\ 0 & \text{otherwise} \end{cases} . \tag{29}$$

For the phase-delay control loop, we have $\mathsf{S}^{\dagger PD} = \mathrm{diag}(s_1^{\dagger PD}, s_2^{\dagger PD}, s_3^{\dagger PD}, 0)$, where

$$s_i^{\dagger PD} = \begin{cases} 1/s_i & \text{if } s_i > |S/N_{\text{threshold}}^{\text{PD}}|^2 \\ s_i/|S/N_{\text{threshold}}^{\text{PD}}|^4 & \text{otherwise} \end{cases} . \tag{30}$$

As a result, the two matrices $\mathbb{I}$ are calculated for each DIT from a new SVD and the equations

$$\mathbb{I}_{GD}^6 = \mathsf{M}\mathsf{V}^\top \mathsf{S}^{\dagger GD}\mathsf{U}^\top \mathsf{M}^\top \mathsf{W} \tag{31}$$

$$\mathbb{I}_{PD}^6 = \mathsf{M}\mathsf{V}^\top \mathsf{S}^{\dagger PD}\mathsf{U}^\top \mathsf{M}^\top \mathsf{W}, \tag{32}$$

where the difference between the two is that the eigenvalues in $\mathbb{I}_{PD}^6$ are weighted down when they are below a value equal to $|S/N_{\text{threshold}}^{\text{PD}}|^2$.

4.5. State machine

The rank of matrix $\mathbb{I}_{GD}^6$ (the number of non-zero eigenvalues) drives the decision-making of the state machine. For a four-telescope operation, a rank of 3 means that the position of the delay lines on all the telescopes is constrained. When only three telescopes are tracked, the rank is 2. The rank is 1 for two linked telescopes.

The state machine (Fig. 3) therefore has only three states: IDLE, SEARCHING, and TRACKING. When the operator starts the fringe tracker, it switches to the state SEARCHING. As soon as the rank of the $\mathbb{I}_{GD}^6$ matrix is 3, the fringe tracker transitions to state TRACKING. When the rank of the matrix decreases and remains low for a period of 1 s or more, then the system automatically switches back to SEARCHING mode. In both states, the group-delay and phase-delay controllers are running. This means that whether in SEARCHING or TRACKING state, the system still tracks the fringes on the baselines with sufficient S/N.

5. Group delay tracking

5.1. Group-delay block diagram

Figure 2 represents the block diagram of the full control architecture. Within this design, the prevalent control loop is the group-delay loop (in blue). It has to be very reliable: the signal on
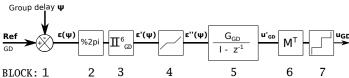

Fig. 4. Block diagram of the group-delay controller. The controller is centered on an integrator (block 5), but also includes several features to work in conjunction with the phase-delay controller.

the feedback signal Ψ is therefore enhanced by averaging over 40 DITs in Eq. (15). The control algorithm of the group delay is presented in Fig. 4. It generates a control signal $\boldsymbol{u}_{GD}$ whose unit is the radian of the phase delay. It is made of seven distinct blocks that are described below.

The measured group delay is a vector defined as

$$\boldsymbol{\Psi}_n = \begin{pmatrix} \Psi_{1,2} & \Psi_{1,3} & \Psi_{1,4} & \Psi_{2,3} & \Psi_{2,4} & \Psi_{3,4} \end{pmatrix}_n . \tag{33}$$

The first block is a comparator that extracts the error between a reference vector ($\mathbf{Ref}_\Psi$) and the measured group delay. The logic would be to have all six setpoints equal to zero to track on the white-light fringes. However, as explained in Sect. 4.2, this is not possible in the presence of non-zero group-delay closure phase Θ^{GD}. The use of a setpoint vector as defined by Eq. (20) therefore ensures that all baselines track the fringes with the highest contrast that do not contradict each other:

$$\boldsymbol{\varepsilon}_{\Psi,n} = \mathbf{Ref}_\Psi - \boldsymbol{\Psi}_n. \tag{34}$$

The second block is there because the phase measurement is only known modulo 2π. Because $\boldsymbol{\varepsilon}_{\Psi,n}$ can have any value, this block adds or subtracts an integer number of 2π to ensure that the error is between $-\pi$ and π.

The third block uses the $\mathbb{I}_{GD}^6$ matrix to weight the errors between the different baselines. Following that matrix, the OPD error vector is now strictly, in a mathematical sense, colinear to the piston space. Moreover, the error on any baseline with no fringes is either estimated from other baselines or set to zero. After the third block, the error group-delay vector is now

$$\boldsymbol{\varepsilon}'_{\Psi,n} = \mathbb{I}_{GD}^6([\mathbf{Ref}_\Psi - \boldsymbol{\Psi}_n]\%2\pi), \tag{35}$$

where the percent sign corresponds to the modulo function.

The fourth block is a threshold function that quenches the gain of the control loop if the absolute value of $\boldsymbol{\varepsilon}'_{\Psi,n}$ is lower than $2\pi/R$. This value corresponds to an OPD of one wavelength, meaning the group-delay controller cannot converge on an accuracy better than $2.2\,\mu$m. This is necessary to let the phase-delay controller track within a fringe:

$$\boldsymbol{\varepsilon}''_{\Psi,n} = \begin{cases} \boldsymbol{\varepsilon}'_{\Psi,n} - \pi/R & \text{if } \boldsymbol{\varepsilon}'_{\Psi,n} > \pi/R \\ \boldsymbol{\varepsilon}'_{\Psi,n} + \pi/R & \text{if } \boldsymbol{\varepsilon}'_{\Psi,n} < -\pi/R \\ 0 & \text{otherwise.} \end{cases} \tag{36}$$

The fifth block is the integrator. In the time domain, it writes

$$\boldsymbol{u}'_{GD,n} = \boldsymbol{u}'_{GD,n-1} + \mathsf{G}_{GD}\,\boldsymbol{\varepsilon}''_{\Psi,n}. \tag{37}$$

The same controller gain G_{GD} is applied to all baselines. It minimizes the closed-loop response time and maximizes robustness. The open-loop transfer function is the same for each baseline. It
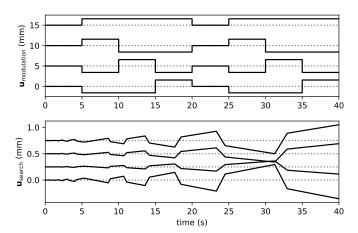
Fig. 5. *Upper panel*: signal from the modulation block, in radians (the Volts-to-radians gain is not accounted for here). At each science exposure (here 5 s), the $\boldsymbol{u}_{\mathrm{modulation}}$ values change to make π offsets between the baselines. The signal repeats every four science exposures so that each baseline is observed with as many $+\pi/2$ as $-\pi/2$ offsets. *Lower panel*: modified sawtooth signal to search for fringes when the fringe tracker is in SEARCHING state. Here the research is made on all telescopes, meaning that the rank of matrix $\mathbb{I}_{\mathrm{GD}}^{6}$ is 0.

is mostly a pure delay caused by the moving average of 40 DITs, as stated by Eq. (15) in Sect. 3.5. The -3 dB cutoff frequency therefore depends on the sampling rate of the fringe tracker. It is 13, 4.5, and 1.5 Hz for sampling rates of 909, 303, and 97 Hz, respectively.

The sixth block is the matrix $\mathsf{M}^{\dagger}$ to transpose the control signal from OPD space to piston space.

The last block is a quantization function. Practically, it means that the group-delay controller, when it detected a group-delay error larger than a fringe, only makes 2π phase-delay jumps until it comes to the reference fringe, without disturbing the long-term phase measurement. The command issued from the group-delay controller is therefore for each piezo-actuator the closest value that is a multiple of 2π. Hence

$$\boldsymbol{u}_{\mathrm{GD},n} = \mathrm{round}_{\{2\pi\}}\left(\mathsf{M}^{\dagger}\,\boldsymbol{u}'_{\mathrm{GD},n}\right). \tag{38}$$

When $\mathsf{M}^{\dagger}\,\boldsymbol{u}'_{\mathrm{GD},n}$ is between $-\pi$ and π, the control signal of the group-delay controller is constant, and the phase delay can work without interferences caused by the group-delay loop.

We note that the final control signal (as shown in Fig. 2) also includes the modulation function, the fringe-search function, and the phase-delay control signal:

$$\boldsymbol{u} = \boldsymbol{u}_{\mathrm{GD},n} + \boldsymbol{u}_{\mathrm{modulation}} + \boldsymbol{u}_{\mathrm{search},n} + \boldsymbol{u}_{\mathrm{PD},n}. \tag{39}$$

5.2. Modulation function and 2π phase jumps

The rounding of the group-delay command ensures that the same phase is always tracked even though jumps between fringes are required when the white-light fringe is searched for. However, this means that the science detector always records the fringes at the same phase delay, and when the sky is not well subtracted, for example, this can bias the visibility. This can be explained in the case of a perfect ABCD beam combiner. Assuming the recorded flux on each one of the four pixels is q_A, q_B, q_C, and q_D, the P2VM matrix calculates the raw complex visibility in this way:

$$V = q_A - q_C + i(q_B - q_D). \tag{40}$$

When the sky removal on q_A is not perfect, we have an additional flux that biases the measurement: $\hat{q}_A = q_A + \varepsilon_A$. To remove this error term, the solution is to record the fringes with π offsets:

$$\hat{V} = q_A + \varepsilon_A - q_C + i(q_B - q_D) \tag{41}$$

$$\hat{V}_\pi = q_C + \varepsilon_A - q_A + i(q_D - q_B), \tag{42}$$

giving

$$\frac{\hat{V} - \hat{V}_\pi}{2} = \frac{2q_A - 2q_C + i(2q_B - 2q_D)}{2} = V. \tag{43}$$

This temporal π modulation is added to the group-delay command (Fig. 5). It is not seen by the group-delay control loop because of the quenching block (block 4 in Fig. 4). This command is synchronized with the reset of the science detector, and it is sequentially:

$$\boldsymbol{u}_{\mathrm{modulation}} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \mathrm{or} \begin{pmatrix} \pi/2 \\ \pi/2 \\ -\pi/2 \\ -\pi/2 \end{pmatrix} \mathrm{or} \begin{pmatrix} \pi/2 \\ -\pi/2 \\ \pi/2 \\ -\pi/2 \end{pmatrix} \mathrm{or} \begin{pmatrix} \pi/2 \\ -\pi/2 \\ -\pi/2 \\ \pi/2 \end{pmatrix}. \tag{44}$$

Only with a minimum of 4 DIT with each of the $\boldsymbol{u}_{\mathrm{modulation}}$ above can we have, on each baseline, as many exposures with 0 and π offsets. Therefore the number of science DITs within a GRAVITY exposure is recommended to be a multiple of 4.

5.3. Fringe search

The search is made through a modified sawtooth function of increasing amplitude (u_{saw}). This function is started when the fringe tracker enters SEARCHING state. It is disabled when the system transition to TRACKING state. This is the only difference between the two modes. The same sawtooth function is generated for each piezo-actuator, but with different scaling factors: -2.75, -1.75, 1.25, and 3.25 for the first, second, third, and fourth beam, respectively. The four commands are then multiplied by the kernel of the $\mathbb{I}_{\mathrm{GD}}^{4}$ matrix. This matrix is computed as

$$\mathbb{I}_{\mathrm{GD}}^{4} = \mathsf{M}\,\mathbb{I}_{\mathrm{GD}}^{6}\,\mathsf{M}^{\dagger} \tag{45}$$

giving the following signal:

$$\boldsymbol{u}_{\mathrm{search},n} = (\mathsf{I}^4 - \mathbb{I}_{\mathrm{GD}}^{4})\begin{pmatrix} -2.75 & -1.75 & 1.25 & 3.25 \end{pmatrix}^{\top} u_{\mathrm{saw}}, \tag{46}$$

where I^4 is the 4×4 identity matrix. This ensures that the baselines with sufficient fringe signal are tracking and are not modulated by the sawtooth function. When no more fringes are found, the kernel is the identity, and each baseline is modulated. This gives the trajectory shown in Fig. 5.

An example of fringe research and acquisition is presented in Fig. 6. In this example, at $t = 0$, the fringes are found and tracked on two baselines: the baseline between telescopes 1 and 3, and the baseline between telescopes 2 and 4. The group delay for the two fringes is zero, and the rank of the $\mathbb{I}_{\mathrm{GD}}^{4}$ matrix is 2. The fringe tracker is therefore in SEARCHING state. At $t = 0.8$ s, the fringes are found on baseline 34. The rank of the $\mathbb{I}_{\mathrm{GD}}^{4}$ matrix increases, and the system switches to TRACKING state. The group delays of baselines 12, 14, 23, and 34 are brought to zero. Signals are later found on all six baselines and the system reaches a nominal tracking state at $t = 0$ s.
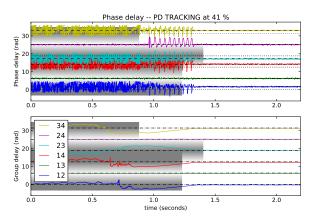
Fig. 6. Phase $\Phi_{i,j}$ (*upper panel*) and group $\Psi_{i,j}$ (*lower panel*) delays. All values are phases modulo 2π. The dashed lines correspond to the piezo command. From bottom to top, the phases correspond to baselines i, j equal to 12, 13, 14, 23, 24, and 34. At the beginning of this recording, the system only has fringes on two baselines (green and purple). At $t = 0.8$ s, the system found fringes on the yellow baseline, and later on all the other baselines. Gray areas correspond to the baselines whose S/N is below the value $S/N^{\mathrm{GD}}_{\mathrm{threshold}}$, meaning that no tracking is performed. After $t = 1.4$ s, the system nominally tracks on all baselines.

6. Phase-delay tracking

6.1. Principle

The GRAVITY phase control loop uses both piston-space and OPD-space state vectors. This is the easiest way to properly handle both piezo-actuators and the atmosphere dynamics (Correia et al. 2008). The vibrations and atmospheric perturbations are represented by six OPD-space state vectors labeled together $\boldsymbol{x}_V$. Each vector corresponds to a baseline. The piezo-actuators are characterized by four piston-space state vectors $\boldsymbol{x}_P$, one for each delay line. The phase delay, $\boldsymbol{\Phi}_n$, is a vector of measured phases:

$$\boldsymbol{\Phi}_n = \begin{pmatrix} \Phi_{1,2} & \Phi_{1,3} & \Phi_{1,4} & \Phi_{2,3} & \Phi_{2,4} & \Phi_{3,4} \end{pmatrix}_n, \quad (47)$$

which results from a linear combination of the state vectors $\boldsymbol{x}_{V,n}$ and $\boldsymbol{x}_{P,n}$ at a time n.

The real-time algorithm of the fringe tracker follows a sequence:

1. It predicts the future state of the system from previous state using the equation of state (Sect. 6.2),
2. It uses Kalman filtering to update the state vectors (Sect. 6.5), and
3. It uses the system state to command the piezo-actuators to correct for vibrations and atmospheric effects. (Sect. 6.6).

The phase-delay controller is summarized by the block diagram presented in Fig. 7.

6.2. Equations of state

The equations of state are

$$\boldsymbol{x}_{V,n+1} = \mathsf{A}_V \cdot \boldsymbol{x}_{V,n} + \mathsf{B}_V \cdot \boldsymbol{v}_n \quad (48)$$

$$\boldsymbol{x}_{P,n+1} = \mathsf{A}_P \cdot \boldsymbol{x}_{P,n} + \mathsf{B}_P \cdot \boldsymbol{u}_n, \quad (49)$$

where each matrix is three-dimensional. Because we assumed uncorrelated baselines, only the diagonals are populated:

$$\mathsf{A}_V = \mathrm{diag}\begin{pmatrix} \mathsf{A}_V^{1,2} & \mathsf{A}_V^{1,3} & \mathsf{A}_V^{1,4} & \mathsf{A}_V^{2,3} & \mathsf{A}_V^{2,4} & \mathsf{A}_V^{3,4} \end{pmatrix} \quad (50)$$

$$\mathsf{B}_V = \mathrm{diag}\begin{pmatrix} \mathsf{B}_V^{1,2} & \mathsf{B}_V^{1,3} & \mathsf{B}_V^{1,4} & \mathsf{B}_V^{2,3} & \mathsf{B}_V^{2,4} & \mathsf{B}_V^{3,4} \end{pmatrix} \quad (51)$$

Fig. 7. Block diagram of the phase-delay controller. The two control signals $\boldsymbol{u}_{\mathrm{GD}}$ and $\boldsymbol{u}_{\mathrm{PD}}$ are summed before they are applied to the actuators. In this control scheme, the state vectors $\hat{\boldsymbol{X}}_V$ are regulated. The control signal $\boldsymbol{u}_{\mathrm{PD}}$ is then issued from a matrix multiplication of the state vectors: $\boldsymbol{u}_{\mathrm{PD}} = \mathsf{K}\hat{\boldsymbol{X}}_V$.

$$\mathsf{A}_P = \mathrm{diag}\begin{pmatrix} \mathsf{A}_P^1 & \mathsf{A}_P^2 & \mathsf{A}_P^3 & \mathsf{A}_P^4 \end{pmatrix} \quad (52)$$

$$\mathsf{B}_P = \mathrm{diag}\begin{pmatrix} \mathsf{B}_P^1 & \mathsf{B}_P^2 & \mathsf{B}_P^3 & \mathsf{B}_P^4 \end{pmatrix}. \quad (53)$$

Here the upper index corresponds to the telescope or baseline numbers. In the block diagram of Fig. 2, the equations of state are written in the frequency domain, but the transfer function writes

$$\mathcal{Z}\{\boldsymbol{x}_{V,n}\} = \frac{\mathsf{B}_V\, z^{-1}}{1 - \mathsf{A}_V\, z^{-1}} \mathcal{Z}\{\boldsymbol{v}_n\} \quad (54)$$

$$\mathcal{Z}\{\boldsymbol{x}_{P,n}\} = \frac{\mathsf{B}_P\, z^{-1}}{1 - \mathsf{A}_P\, z^{-1}} \mathcal{Z}\{\boldsymbol{u}_n\}. \quad (55)$$

The evolution of the system is driven on one hand by white noise $\boldsymbol{v}_n$, and on the other hand by a user-controlled voltage applied to the piezo-actuators $\boldsymbol{v}_n$. When $\boldsymbol{v}_n$ is white noise, $\boldsymbol{x}_{V,n}$ has a colored noise, as highlighted by Eq. (54). The problem is similar for AO systems, where it has already been mentioned and corrected for (Poyneer & Véran 2010).

In Menu et al. (2012) and Choquet et al. (2014), we have shown that using several autoregressive (AR) models of order 2 in parallel was effective to correct for both vibration frequencies and atmospheric turbulence. However, practically, two issues made that implementation difficult: i) the determination of the vibration peaks for low S/N data, and ii) the need to change the space model when vibrations appear or disappear. Both problems can be technically resolved, but to ensure maximum robustness, we preferred a fixed well-defined space model. The idea is that the state x_V do change with time, but the space-state model does not. We therefore used an autoregressive model of order 30. This means that the state-space model corresponds to the 30 last values of the phase delay:

$$\mathsf{A}_V^{i,j} = \begin{pmatrix} v_1^{i,j} & v_2^{i,j} & v_3^{i,j} & \cdots & v_{29}^{i,j} & v_{30}^{i,j} \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \quad (56)$$

and

$$\mathsf{B}_V^{i,j} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}^\top. \quad (57)$$

The value 30 allows for complexity in the vibrational pattern while characterizing the perturbations with a sufficiently low degree of freedom. It was chosen in light of Kulcsár et al. (2012): they showed that it is a good compromise with respect to other state-space models to correct atmosphere and vibration for the tip-tilt of AO systems. The transfer matrix $A_V^{i,j}$ is determined

Table 1. Open-loop transfer function parameters of the fringe tracker running at 909 Hz.

(rad/Volts)	$i = 1$	$i = 2$	$i = 3$	$i = 4$
a_1^i	0.16	−0.03	−0.22	−0.1
a_2^i	−0.15	0.09	0.11	0.12
a_3^i	6.52	4.08	3.56	4.41
a_4^i	9.61	9.01	7.12	11.14
a_5^i	1.31	4.72	7.05	1.85
G_piezo^i	17.47	17.89	17.63	17.43

Fig. 8. Bode plot of the frequency response of the four pistons in open loop (including pure delay caused by integration and computation). The −90° cutoff frequencies are 61 Hz, 56 Hz, 54 Hz, and 59 Hz. The cutoff frequency is dominated by the pure delay caused by the control loop system.

every 5 s on a workstation that is connected on the reflective memory network.

The transmission matrix A_P^i is also an autoregressive model, but of order 5. It includes the response function of the full system (image integration time, processing time, inertia of the piezo mirror, etc.), from setting a control voltage $\boldsymbol{u}_n$ to measuring a phase delay $\boldsymbol{\Phi}_n$. It is measured monthly by injecting a top-hat signal into the system. The matrix for the AR5 model verifies

$$A_P^i = \begin{pmatrix} a_1^i & a_2^i & a_3^i & a_4^i & a_5^i \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{58}$$

and

$$B_P^i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix}^\top, \tag{59}$$

where the a_0^i, a_1^i, a_2^i, a_3^i, and a_4^i values, obtained experimentally on the fringe tracker are presented without normalization in Table 1. The same values can be represented in a Bode plot, see Fig. 8. The cutoff frequencies, calculated at a phase of −90°, are around 60 Hz. This is considerably below the bandwidth of the piezo-actuator measured by Pfuhl et al. (2010), who showed a cutoff frequency above 220 Hz. This is caused by the pure delays inside the system: detector integration time, processing time, and data transfer between the different units (LCUs). The static gain of the piezo $G_\text{piezo}^i = \sum_{k=1}^5 a_k^i$ is also an important property of the piezo-actuators because they are used both in the group- and phase-delay controllers. Practically, to remove the static gain of the control loop, and because $\boldsymbol{u}$ is in unit of OPD radians instead of Volts, the a_k^i values are normalized by the static gain of the G_piezo^i.

6.3. Observation equation

The observable, $\boldsymbol{\Phi}_n$, depends on the state vectors of the piezo-actuators $\boldsymbol{x}_\text{P}$ and on the state vectors of the atmosphere and vibrations $\boldsymbol{x}_V$. Both contribute to the phase delay. The equation for the measurement process of the phase delay is a linear combination of the two:

$$\boldsymbol{\Phi}_n = \mathsf{M} \cdot \mathsf{C}_P \boldsymbol{x}_{\text{P},n} - \mathsf{C}_V \boldsymbol{x}_{V,n} + \boldsymbol{w}_n + \mathbf{Ref}_{\boldsymbol{\Phi}}, \tag{60}$$

where $\boldsymbol{w}_n$ is a white measurement noise and C_V and C_P are three-dimensional measurement matrices with two-dimensional matrices on the diagonal for the vibrations and the piezo-actuator, respectively:

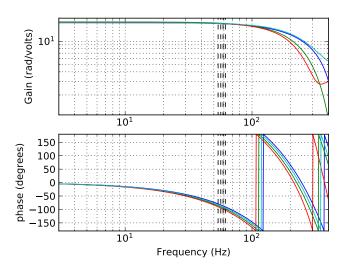$$\mathsf{C}_V^{i,j} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{61}$$

$$\mathsf{C}_P^i = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{62}$$

In the case of a resolved astronomical target, the phase-delay vector $\boldsymbol{\Phi}_n$ also includes a term that corresponds to the spatial signature of the target on the phase. It appears in the form of a non-zero closure phase ($\Theta^\text{PD} \neq 0$) and is included in the phase delay by the term $\mathbf{Ref}_{\boldsymbol{\Phi}}$, as defined in Eq. (19).

6.4. Parameter identification

The $a_{1 \le k \le 5}$ predictive terms for each of the piezo-actuators are determined offline during dedicated calibration laboratory measurements. This is not the case for the identification of the $v_{1 \le k \le 30}$ AR values. They are calculated on a distinct Linux workstation. The workstation senses in real time the phase delay that passes by the reflective network. Every 5 s, it collects the last 40 s of data (36 000 sample at 909 Hz) and computes the pseudo-open-loop phase delay $\boldsymbol{\Phi}_{\text{ATM},n}$ by removing the influence of the piezo command and closure phases. This phase delay corresponds to the phase delay produced by the atmosphere and the vibrations only:

$$\boldsymbol{\Phi}_{\text{ATM},n} = \boldsymbol{\Phi}_n - \mathsf{M} \cdot \mathsf{C}_P \boldsymbol{x}_{P,n} - \mathbf{Ref}_{\boldsymbol{\Phi}}, \tag{63}$$

from which it calculates the differences between consecutive values:

$$\Delta\boldsymbol{\Phi}_{\text{ATM},n} = \boldsymbol{\Phi}_{\text{ATM},n} - \boldsymbol{\Phi}_{\text{ATM},n-1}. \tag{64}$$

The $\Delta\boldsymbol{\Phi}_{\text{ATM},n}$ differential phase values are then processed baseline after baseline to derive 29 $p_{1 \le i \le 29}$ AR parameters that best represent the data. This uses the Python toolbox "Time Series Analysis". The generated parameters ensure stationarity (Jones 1980) and thus provide stability to the closed-loop algorithm. Last, the $v_{1 \le i \le 30}$ values are determined through

$$v_i = p_i - p_{i-1}, \tag{65}$$

after which the matrices A_V and A_P are sent by the RMN to the real-time LCU to adjust the parameters of the phase-delay controller.

6.5. Asymptotic Kalman filter

The vectors $\hat{\boldsymbol{x}}_{V,n|n-1}$ correspond to our best estimate of the state of the vibrations at a moment n from all the measurements available up to a time $n-1$. It can be estimated from $\hat{\boldsymbol{x}}_{V,n-1}$ according to the equation of state in Eq. (48):

$$\hat{\boldsymbol{x}}_{V,n|n-1} = \mathsf{A}_V \cdot \hat{\boldsymbol{x}}_{V,n-1}. \tag{66}$$

However, the goal of the Kalman filter is to update this estimate from the error between the new observable and this new estimate. This error can be derived from the measurement process in Eq. (60) and writes

$$\boldsymbol{\varepsilon}_{\boldsymbol{\Phi},n} = \mathbf{Ref}_{\boldsymbol{\Phi}} - \boldsymbol{\Phi}_n + \mathsf{M} \cdot \mathsf{C}_P \boldsymbol{x}_{\mathrm{P},n} - \mathsf{C}_V \cdot \hat{\boldsymbol{x}}_{V,n|n-1}, \tag{67}$$

which becomes, after wrapping around 0 (between $-\pi$ and π) and baseline weighting,

$$\boldsymbol{\varepsilon}'_{\boldsymbol{\Phi},n} = \mathbb{I}^6_{\mathrm{PD}}.([\boldsymbol{\varepsilon}_{\boldsymbol{\Phi},n}]\%2\pi), \tag{68}$$

where $\%2\pi$ corresponds to modulo 2π. The $\mathbb{I}^6_{\mathrm{PD}}$ matrix is here for two purposes. The first is to derive the errors on low S/N baselines through the baselines with higher S/N. The second purpose is to ensure that the low S/N data are either weighted down or discarded through the weighted pseudo-inversion in Eq. (32). The state estimator $\hat{\boldsymbol{x}}_{V,n}$ is finally updated through an integrator:

$$\hat{\boldsymbol{x}}_{V,n} = \mathsf{A}_V \hat{\boldsymbol{x}}_{V,n-1} + \mathsf{G}_{\mathrm{PD}} \cdot \boldsymbol{\varepsilon}'_{\boldsymbol{\Phi},n}. \tag{69}$$

The gain G_{PD} is not a scalar but a three-dimensional matrix. The correct estimate of G_{PD} is the basis of Kalman filtering. In GRAVITY, it is not identified at each DIT, but every 5 s on the sensing Kalman workstation. Therefore, it is obtained from the asymptotic Kalman equations. It is calculated from the two covariance matrices of the measurement noise Σ_w and the steady-state error Σ_∞ (Poyneer & Véran 2010; Menu et al. 2012):

$$\mathsf{G}_{\mathrm{PD}} = \Sigma_\infty \mathsf{C}_V^\dagger (\mathsf{C}_V \Sigma_\infty \mathsf{C}_V^\dagger + \Sigma_w)^{-1}. \tag{70}$$

The steady-state covariance matrix can be obtained from the algebraic Riccati equation, the vibrations input noise, and the vibration and atmospheric noise Σ_v:

$$\Sigma_\infty = \mathsf{A}_V \Sigma_\infty \mathsf{A}_V^\dagger - \mathsf{A}_V \Sigma_\infty \mathsf{C}_V^\dagger (\mathsf{C}_V \Sigma_\infty \mathsf{C}_V^\dagger + \Sigma_w)^{-1} \mathsf{C}_V \Sigma_\infty \mathsf{A}_V^\dagger + \Sigma_v. \tag{71}$$

The noise characteristic (Σ_v, Σ_w) makes the Kalman filter optimally adapted to the average noise on the system, but not to instantaneous noise variations. Adaptability is the role of the $\mathbb{I}^6_{\mathrm{PD}}$ matrix in Eq. (68).

6.6. Determination of the control signal

To use the Kalman filter as a controller, an optimal command is needed. The purpose of this section is to determine the control signal $\boldsymbol{u}_{\mathrm{PD},n}$ from the atmospheric state vectors: $\boldsymbol{u}_{\mathrm{PD},n} = \mathsf{K}\hat{\boldsymbol{x}}_V$. The task of this signal is twofold. First, it must cancel the atmospheric perturbations on the measured phase delay ($\boldsymbol{\Phi}_{\mathrm{ATM},n}$). Second, it must ensure that the phase delay converges to the group-delay control signal $\boldsymbol{u}_{\mathrm{GD},n} + \boldsymbol{u}_{\mathrm{modulation}} + \boldsymbol{u}_{\mathrm{search},n}$.

The first task uses the predictive power of the equation of state:

$$\boldsymbol{u}_{\mathrm{PD},n} = \mathsf{M}^\dagger \mathsf{C}_V \mathsf{A}_V^{n_{\mathrm{DIT}}} \hat{\boldsymbol{x}}_V, \tag{72}$$

where $\mathsf{A}_V^{n_{\mathrm{DIT}}}$ is the power of matrix A_V over n_{DIT} samples. The integer n_{DIT} is there to account for the delay between control

signal and its effect on the phase delay. Ideally, $n_{\mathrm{DIT}} = 1$, but because of the pure delay in the open-loop transfer function, it must be higher. This pure delay depends on the frequency of the fringe tracker. We currently use $n_{\mathrm{DIT}} = 3$ for the 909 Hz frequency and $n_{\mathrm{DIT}} = 2$ for the lowest frequencies.

The second task is achieved through the integrator in Eq. (69). This integrator updates the state vector of the atmospheric perturbations to minimize the error between estimation and observation. It causes the quantity $\boldsymbol{\varepsilon}_{\boldsymbol{\Phi},n}$ to converge to zero: $\boldsymbol{\varepsilon}_{\boldsymbol{\Phi},n\to\infty} = 0$. From Eq. (67), we can derive

$$\boldsymbol{\Phi}_{n\to\infty} = \mathbf{Ref}_{\boldsymbol{\Phi}} + \mathsf{M} \cdot \mathsf{C}_P \boldsymbol{x}_{P,n\to\infty} - \mathsf{C}_V \cdot \hat{\boldsymbol{x}}_{V,n\to\infty}. \tag{73}$$

Under a steady-state assumption, $\mathsf{C}_P \boldsymbol{x}_{P,n\to\infty} = \boldsymbol{u}_{n\to\infty}$ because the piezo-gains are normalized. From Eq. (72) we also have, under the same assumption, $\mathsf{C}_V \cdot \hat{\boldsymbol{x}}_{V,n\to\infty} = \mathsf{M} \cdot \boldsymbol{u}_{\mathrm{PD},n\to\infty}$ because the first row of A_V is normalized. Hence, we have what we desire, i.e.,

$$\boldsymbol{\Phi}_{n\to\infty} = \mathbf{Ref}_{\boldsymbol{\Phi}} + \mathsf{M} \cdot (\boldsymbol{u}_\infty - \boldsymbol{u}_{\mathrm{PD},n\to\infty}) \tag{74}$$

$$= \mathbf{Ref}_{\boldsymbol{\Phi}} + \mathsf{M} \cdot (\boldsymbol{u}_{\mathrm{GD},n\to\infty} + \boldsymbol{u}_{\mathrm{modulation}} + \boldsymbol{u}_{\mathrm{search},n\to\infty}). \tag{75}$$

6.7. Why not a simpler state controller?

The block diagram presented in Fig. 2 shows the complexity of the fringe tracker. A simpler version was used during the first commissioning in 2016. The phase and group-delay controllers were both different. First, the group-delay control signal was not used to directly command the piezo-actuators, but instead, it was used as a setpoint for the phase-delay controller. Second, the phase-delay controller was a proportional-integral controller. It was efficient and robust and a good commissioning tool. However, at low S/N, its performance was not adequate. The reasons were that

– the phase-delay estimator is noisy. The proportional controller (or worse, a derivative) directly injected the noise back into the control signal. This was especially problematic during flux dropouts when the S/N can be close to zero; and that

– the phase delay is modulo 2π. It needs to be unwrapped with respect to a prediction. Using the setpoint as the prediction resulted in many fringe jumps during poor atmospheric conditions.

We therefore decided to have the group delay directly command the piezo, skipping the phase-delay controller. In parallel to the group-delay controller, we included a Kalman filter on the phase-delay feedback and used the predictive model to unwrap the phase (the vector $\boldsymbol{\varepsilon}'_{\boldsymbol{\Phi},n}$). This gave the block diagram in Fig. 7.

7. On-sky observations

7.1. Operation

The simplicity of operating the fringe tracker lies in the simplicity of the state machine. With only three states, the operator interactions are limited to the transition between IDLE, SEARCH, and back to IDLE. When the fringes are detected, the system automatically switches to TRACKING and the instrument then starts recording data. However, there are two free parameters that could ask for the intervention of the operator: the $S/N_{\mathrm{threshold}}^{\mathrm{GD}}$ and $S/N_{\mathrm{threshold}}^{\mathrm{PD}}$ thresholds. An operational error could be, for example, to set a $S/N_{\mathrm{threshold}}^{\mathrm{GD}}$ too low and risk having the fringe-tracker tracking on the second lobe of a fringe packet.

When the thresholds are correctly set, the system is made to be fully autonomous, and able to deal with any glitch of the VLTI. For example, Fig. 9 shows the fringe tracker losing
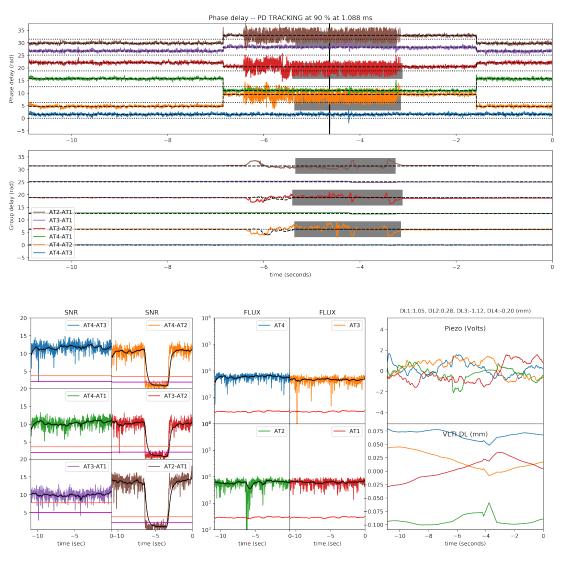
Fig. 9. Observation of star GJ 65 A on 21 November 2018. This example shows the case of a glitch on the AT2 adaptive optics system that resulted in the loss followed by the recovery of the fringe by the fringe tracker. *Upper two panels*: fringe tracker phase $\Phi_{i,j}$ and group delay $\Psi_{i,j}$ estimators. The color lines correspond to each of the six baselines. The $\pi/2$ phase jumps at -7 and -1.5 s are normal and correspond to the modulation synchronized with the 5 s DIT science camera (the high spectral resolution detector). The gray areas correspond to detection of low S/N fringes by the controller. *Lower left panels*: S/N for each of the baselines calculated as the inverse of the square root of the phase variance ($1/\sqrt{\mathrm{Var}(\Phi_{i,j})}$). The horizontal red lines correspond to the group-delay threshold ($S/N_{\mathrm{threshold}}^{\mathrm{GD}}$). If the black line dips below the red line, the FT stops tracking on that baseline. The horizontal purple lines correspond to the phase-delay threshold ($S/N_{\mathrm{threshold}}^{\mathrm{PD}}$). *Lower center panel*: flux F_i. Similarly, the red line is a threshold made by a moving average of all flux, which is used to detect the loss of a telescope. *Lower right panels*: commands to the piezo-actuators and VLTI delay line actuators. The piezo-actuators take care of the fast control signal, while the VLTI delay lines are used to offload and to search for fringes over large distances.

telescope AT2 and how it recovered. The figures are the same plots as those available to the support astronomers during real-time GRAVITY operation. The data were captured and plotted at $t = 0$ s. At $t = -6.39$ s, the AT2 looses its pointing target, and the injected flux in the fiber drops. Immediately, the S/N level drops below the purple line, and the FT discontinues tracking on all the three baselines that include AT2. At this point, the system is still in TRACKING state. At $t = -5.39$ s, after a time delay of 1 s, the system switches to SEARCHING state, and the VLTI delay lines start following the sawtooth function. At $t = -3.26$ s, the system recovers the fringes on the AT2-AT1 baseline and starts centering them. At this point, the rank of the $\mathbb{I}_{\mathrm{GD}}^4$ matrix is back to 3 and the fringe tracker switches back to TRACKING state. At $t = -3.12$ s, the fringes are detected on all six baselines, and the system again tracks nominally.

In summary, the complexity of dealing with multiple baselines is hidden behind the $\mathbb{I}_{\mathrm{GD}}^6$ and $\mathbb{I}_{\mathrm{PD}}^6$ matrices presented in Sect. 4.4. From the user's point of view, the fringe tracker transitions from SEARCHING to TRACKING state, but the engine behind the scene does not change the way it operates.

7.2. Sensitivity

The sensitivity is mostly a question of having enough photons on the detector to generate a feedback signal for the fringe tracker. In Fig. 10 we plot the flux versus magnitude of calibrators observed with GRAVITY during a period covering June 2017 to November 2018. The selection of the files with a tracking ratio above 80% leads to a total of 1117 exposures on 473 distinct calibrators. The orange dots correspond to the 814 AT
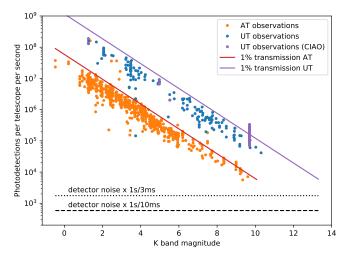
Fig. 10. Transmission plot, i.e., photons detected on the GRAVITY FT receiver, per telescope and per second, as a function of the K-band magnitude (tracking ratio >80%). The magnitudes of the targets are obtained from SIMBAD. Most of the observations are made on-axis, meaning that 50% of the flux is lost because of the beam splitter. However, the many $K = 9.7$ mag CIAO observations are taken off-axis, hence the higher flux. Targets observed with ATs can be as faint as 10 mag. Based on this plot, it is clear that because the flux observed with the UTs is more than 10 times higher, observations will be possible up to $K = 13$ mag.

Fig. 11. Signal-to-noise ratio as a function of detected photon per telescope per DIT. The S/N is the computed by the real-time computer, defined by $1/\sqrt{\mathrm{Var}(\Phi_{i,j})}$, as stated in Eq. (14). The three colors correspond to the three different frame rates of the fringe tracker. The solid lines correspond to theoretical values assuming 100% fringe contrast. The vertical lines are theoretical flux assuming observations on UTs with 1% throughput. The horizontal line corresponds to an S/N of 1.5.

observations. The others correspond to the MACAO (visible AO) and CIAO (infrared AO) UTs observations. The solid lines correspond to a total transmission of 1% from telescope to the fringe-tracker detector. The dashed lines correspond to the theoretical detector noise, scaled by $\sqrt{N_p}$, with N_p the total number of pixels divided by the number of telescopes.

On the ATs, the faintest target observed so far was TYC 5058-927-1, a star with $K = 9.4$ mag. It was observed in the night of 5 April 2017, when the atmospheric conditions were excellent: seeing down to 0.4″ and coherence time up to 12 ms. The observations were made in single-field mode, where only half the flux is

sent to the fringe tracker. The fringe tracker efficiently tracked the fringes throughout the entire exposure time, at a frequency rate of 97 Hz, with OPD residuals between 250 and 350 nm.

Technically, if we were to extrapolate to the UTs the sensitivity observed on the ATs, we should be able to track stars of K magnitudes up to 12.5. However, the faintest star observed with the fringe tracker on the UTs so far is TYC 7504-160-1, a star with $K = 11.1$ mag. It was observed in the night of 2 July 2017 during good atmospheric condition: seeing between 0.4″ and 0.6″, coherence time between 4 and 7 ms. The frequency rate was 303 Hz, with OPD residuals between 350 and 400 nm. The relatively low UT sensitivity is still not understood. One explanation could be that the 97 Hz integration time cannot be used: the fringe contrast is attenuated by the vibrations of the UT structure. Another explanation could come from moderate AO performances. With a low Strehl ratio, difficulties in injecting the light in the fibers decrease the number of available photons, but also create flux drops that decrease the visibility and complicate the fringe tracking. One last possibility is a selection effect caused by the lower availability of the UTs.

7.3. Signal-to-noise ratio

The limit for fringe tracking could in theory be an S/N of 1 per coherence time. Below that number, the phase varies faster than our ability of measuring it: our knowledge of the phase decrease with time, hindering the convergence of the fringe tracker. However, practically, we observed that an S/N of 1.5 per DIT is required to keep the fringes in the coherence envelope. The signal is proportional to the number of coherent photons received. The noise is created by quantum noise on one hand and the background noise on the other hand. At 909, 303, and 97 Hz, we measured during sky observations a standard deviation of $\sigma_{sky} = 4$, 5, and 8 ADU per pixel, respectively. These values come from the quadratic sum of the noises caused by the scattered metrology light (7 ADU at 97 Hz), the sky and environmental background (6 ADU at 97 Hz), and the read-out noise (4 ADU).

The detector variance observed during sky observations is used to compute the real-time S/N shown in the lower left panel of Fig. 9. In Fig. 11, for a dataset covering the seven months between April and November, we have plotted this S/N as a function of the measured flux. The three solid lines correspond to the theoretical S/N calculated from σ_{sky} and photon noise. These values lie below these theoretical lines because of a loss of visibility contrast. This can be caused either by non-equilibrium of the flux between the different telescopes, or by OPD variations within a DIT of the fringe tracker detector.

The vertical lines correspond to the flux of a star of K magnitude of 8, 12, and 16 observed with the UTs under the assumption of 1% throughput. The drop at low flux of the theoretical S/N curves correspond to the effect of the observed sky noise. This noise is mostly detector noise at high frequency, and a combination of background and metrology noise at low frequency. The dotted lines are theoretical computations of the noise assuming only photon noise. Under this assumption, 1% throughput, and 100% visibilities, the UT sensitivity could technically reach magnitudes up to $K = 16$ mag.

7.4. OPD residuals and S/N

The fringe tracker performance degrades when the limiting sensitivity is approached. This is partly because the phase-delay control-loop gain decreases at low S/N because of the weighted inversion of matrix $\mathbf{S}^{\dagger PD}$ in Eq. (32). This is also partly caused by
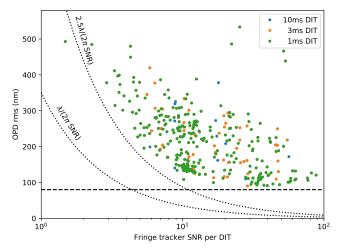
Fig. 12. OPD residuals as a function of the fringe tracker S/N. Above an S/N of 10, the best fringe-tracking residuals are limited to a constant value around 80 nm, caused by the intrinsic bandpass of the fringe tracker hardware (dashed line). At lower S/N, the fringe tracker performance decreases because it is difficult to predict and track the evolution of the perturbations. The phase error caused by a readout noise on the phase goes as $\lambda/(2\pi\,S/N)$ (lower dotted line). The actual performances of the fringe tracker shows that the degradation is more likely a factor 2 or 3 above that limit.
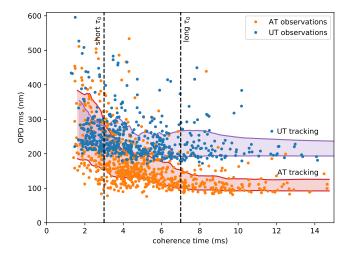

Fig. 13. OPD residuals as a function of coherence time (at 500 nm) for all calibrators and GC data taken between July 2017 and November 2018 in the 909 Hz mode. The dashed vertical lines correspond to the first quintile (3 ms) and last quintile (7 ms) of the coherence time distribution as observed by the ESO astronomical site monitor (ASM) at visible wavelength. The thick lines correspond to the first and last quintile of the OPD residual distribution (purple for UTs and red for ATs). The median fringe-tracking residuals are 150 nm on the ATs and 250 nm on the UTs.

the difficulty of estimating the correct fringe-tracking state from a noisy measurement. This decrease in performance is shown in Fig. 12 for the same dataset as presented in Fig. 11. Concretely, below a S/N of 4, the residuals are above 300 nm. These residual include the variance caused by the measurement noise, which is equal to $\lambda/(2\pi\,S/N)$: ≈ 90 nm for a S/N of 4. The fringe tracker residual does not correct the phase to that level, however: the residual remains a factor 2 or 3 above that value. At an even lower S/N (< 3), the Kalman filter has problems to identify the state of the atmosphere and sometimes cannot update them fast enough. Fringe tracking is then only possible when the atmosphere varies slowly, when the atmospheric conditions are best.

7.5. OPD residuals and τ_0

At high S/N (≤ 10), the accuracy of the fringe tracking depends on several environmental parameters: vibrations, wind speed, seeing conditions and coherence time (τ_0). Between wind seeing and coherence time, Lacour et al. (2018) showed that the strongest correlation is observed with τ_0. Using the same dataset as in Fig. 10 and using only the calibrators observed at 909 Hz, we plot in Fig. 13 the OPD residuals as a function of the coherence time as observed by the ESO site monitor at 500 nm. On average, the ATs perform better, with a median residual of 150 nm. Under the best conditions, the OPD residuals can be as low as 75 nm. The UTs residuals are higher, with a median value of 250 nm, and a minimum at 180 nm.

Depending on the seeing conditions, the fringe tracker shows different limitations. In the worst atmospheric conditions, which make up 20%, ($\tau_0 < 3$ ms), the UT and AT performances are limited by the coherence time of the atmosphere. Under these conditions, 20% of all observations have OPD residuals above 380 nm. The consequence is then that the jitter of the phase significantly affects the visibility in the science channel with long integration times. We can show that this contrast loss can be directly estimated from the variance of the phase according to the relation

$$V_{\text{residuals}} = \langle \exp\left(i\Phi\right)\rangle = \exp\left(-\sigma_\Phi^2/2\right). \tag{76}$$

With 380 nm rms jitters ($\sigma_\Phi \approx 1.1$ rad), the contrast of the fringes of the science detector will only be at $V_{\text{residuals}} \approx 55\%$ of its maximum.

During the best 20% of atmospheric conditions ($\tau_0 > 7$ ms), however, the fringe tracker can reach very low residuals. On the ATs, 80% of all observations are below 140 nm ($V_{\text{residuals}} > 93\%$). However, on the UTs, the environmental vibrations cause the residuals to remain between 220 and 320 nm. The explanation is the higher level of vibrations observed on the UTs.

7.6. Power spectral density

In Figs. 14 and 15 we show the power spectral density of the OPD for two astronomical objects: IRS 16C observed with UTs, and GJ 65 observed with ATs. The observation conditions are listed in Table 2. The IRS 16C galactic center target was observed with the CIAO off-axis AOs. The conditions were good to excellent, and the residual OPD was around 220 nm over the entire duration of the observations. The GJ 65 binary was also observed in good seeing conditions, without AO, and OPD residuals of 140 nm.

To compute the power spectral density, the phase delay Φ_n is unwrapped and the $\pi/4$ modulation function removed. The pseudo-open-loop phase delay $\Phi_{\text{ATM},n}$ is then computed from Eq. (63) using the piezo transfer function estimated during calibration. In Figs. 14 and 15, the six upper plots correspond to the square root of the power spectral density. The black curves correspond to the measured OPD estimated from the phase delay Φ, and the red curves show the power spectrum of the reconstructed atmosphere $\Phi_{\text{ATM},n}$. The atmospheric power spectrum (which also includes the vibrations) shows, as expected, the prevalence of the low frequencies in the atmospheric perturbations. Below 10 Hz, the power spectral density fits the relation

$$\text{PSD}_{\Phi_{\text{ATM},n}}(f < 10\,\text{Hz}) = 1f^{-2}\,\mu\text{m}^2\,\text{Hz}^{-1}, \tag{77}$$

very well, which is not far, but different, from the power of $-5/3$ of a Kolmogorov type atmosphere. After fringe-tracking
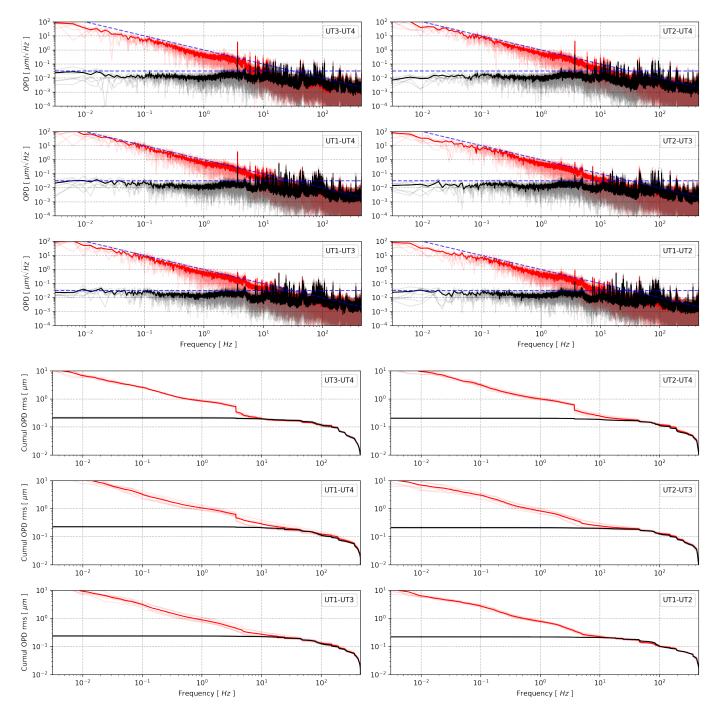
Fig. 14. Power spectrum density and cumulative sum of the power spectrum density of phase residuals ($2.2\,\mu\text{m}/2\pi \times \Phi_n$) toward target IRS 16C. The red curves are the pseudo-open-loop values, i.e., in the hypothesis of no fringe tracker ($2.2\,\mu\text{m}/2\pi \times \Phi_{\text{ATM},n}$). The blue curves correspond to power spectral densities of $1\,f^{-2}\,\mu\text{m}^2/\text{Hz}$ and $0.001\,\mu\text{m}^2/\text{Hz}$. The six lower plots are the reverse cumulative sum of the power spectrum. At zero frequency, the sum reaches 220 nm on average over the six baselines.

correction, the residual OPDs are attenuated below the spectral density:

$$\text{PSD}_{\Phi_n}(f < 10\,\text{Hz}) \le 0.001\,\mu\text{m}^2\,\text{Hz}^{-1}. \qquad (78)$$

The fringe tracker is therefore clearly a high-pass filter. The cutoff frequency of the fringe tracker correction is not well defined because it depends on the efficiency of the Kalman filter and the accuracy with which the equations of state reflect the system. On the UTs, the 3 dB cutoff frequency is on the order of 10 Hz. On the ATs, maybe because the vibrations are less frequent and the predictive model more accurate, the fringe tracker

performance has a higher cutoff frequency at 30 Hz. This 30 Hz shows that the system is optimized because it is close to the cutoff frequency of the open-loop system ($\approx$60 Hz in Fig. 8).

8. Discussion

8.1. Have we reached the ultimate sensitivity?

It is often assumed that the sensitivity of an optical interferometer must decrease as a function of N, the number of telescopes. This is not necessarily true. The limiting sensitivity is when each
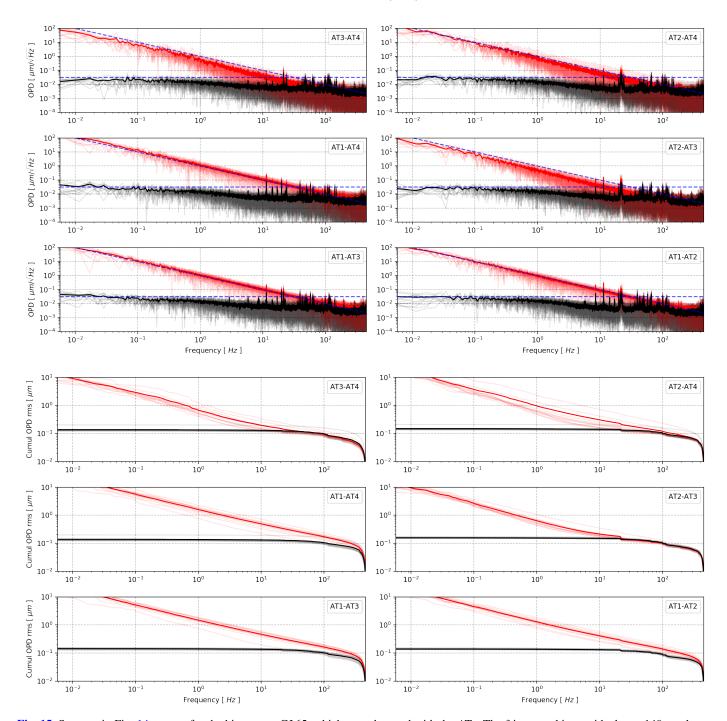
Fig. 15. Same as in Fig. 14, except for the binary star GJ 65, which was observed with the ATs. The fringe-tracking residuals are 140 nm, lower than what is observed with the UTs. The main difference is the lower level of vibrations in the 10−100 Hz range.

of the degrees of freedom of the fringe tracker reaches a variance of one during one coherence time of the atmosphere, and each of the degrees of freedom corresponds to a non-zero eigenvalue of matrix $\mathbb{I}$ (Sect. 4.4). Therefore the threshold on the phase-delay control loop ($S/N_{\text{threshold}}^{\text{PD}}$) is applied on the eigenvalues in Eq. (30). For $N \gg 1$, the signal grows like the flux ($\propto N$), the degrees of freedom like N, and the number of pixels like the number of baselines, N^2. For an increasing number of telescopes, the detector noise therefore becomes ever stronger. However, in the case of a system without background and detector noise, the S/N on the eigenvalues no longer depends on N (Woillez et al. 2017). The ultimate sensitivity of the fringe tracker can then be established at one photon per coherence time and per

telescope, regardless of the number of telescopes. For the VLTI, assuming 8 m telescopes, a coherence time of 10 ms, a through-put of 1%, and using the full K band, the ultimate sensitivity is Kmag = 17.5. Even assuming the need for an S/N of 1.5 per baseline and per coherence time, we should be able to reach a Kmag of 16 (Fig. 11).

What can be done? There are several paths forward to reach this magnitude. The first is to decrease the background and detector noise. The sky brightness in the K band at Paranal is about 12.8 mag/square arcsec[3]. Therefore, the fraction of sky

---

[3] Table 6 in https://www.eso.org/observing/etc/doc/skycalc/The_Cerro_Paranal_Advanced_Sky_Model.pdf

Table 2. GRAVITY dataset plotted in Figs. 14 and 15.

Object Configuration	MJD	FT freq. K mag	Seeing	τ_0 (ms)
IRS 16C UT1-UT2-UT3-UT4	58209	909 Hz 9.7	0.4″	0.007
GJ 65 A (&B) A0-G1-J2-K0	58367	909 Hz 5.7 (5.9)	0.6″	0.006

background light entering the 60 mas single-mode fiber is almost negligible ($K\mathrm{mag}_{\mathrm{sky}} \leq 19$). A more important light emitter is the VLTI thermal background because of the ≈75% absorption of the optical surfaces. All of these surfaces are typically about $T = 283$ K to 288 K and contribute to the majority of the background light. Regarding the detector noise, even if the SAPHIRA detector has a readout noise below $1\,e^-$, the number of pixels used by the fringe tracker per telescope is 36 (72 in split polarizations). For faint objects, this noise can therefore dominate. Solutions to use fewer pixels have been proposed (Petrov et al. 2014, 2016) and could lead to a better sensitivity. The development of infrared detectors could also be a promising path forward.

The second path is to increase the coherent throughput. On average, only 1% of the total flux reaches the fringe tracker detector. The 20 mirrors between M1 and the GRAVITY cryostat absorb up to 75% of the light. In addition, 50% of the light is scattered inside the IO beam combiner, 50% is lost by using the beamsplitter on the on-axis mode, and 50% because of the amplification of the detector (the so-called multiplicative noise). The last part (≈60% of the remaining flux) is lost when the light is injected into the single-mode fiber. Throughput improvements could therefore come from using fewer optical elements and possibly switching more of the mirrors involved from aluminium to gold coatings. Focus can also be placed on a more efficient light coupling into the single-mode fiber. This would mean a better AO system for a better Strehl ratio. The advantages of a good Strehl ratio are manifold. It increases the mean throughput. It also maximizes the fringe contrast by giving a better instantaneous flux equilibrium. Last, it avoids flux drop-out, which is detrimental for good fringe tracking and model prediction.

Third, but not least, special care will be taken in monitoring and removal of the vibrations. The vibrations cause two main problems. Because they are usually at high frequency, they are difficult to predict and affect the fringe-tracking performance (in contrast to the atmospheric perturbations, which are easier to correct because they are at a lower frequency). The main problem, however, is that they limit how slow the fringe tracker can run because the decrease in fringe contrast hurts more than the additional integration time.

8.2. Have we reached the ultimate accuracy?

A false assumption is that sensitivity can be gained by trading accuracy for sensitivity because coherencing (i.e., keeping the fringes within the coherent length) requires fewer photons per coherence time than fringe tracking. However, by simultaneously using the group delay and the phase delay, it is possible to have the best of both worlds: the group-delay controller does the coherencing, while the phase-delay controller works in parallel as far as the S/N permits (Fig. 12).

Despite the high sensitivity, we routinely track high S/N fringes within 100 nm residual rms with GRAVITY. However,

when the coherence time is short ($\tau_0 < 3$ ms at 500 nm), the fringe tracker performance degrades (Fig. 13). This is caused by the open-loop latency of the fringe tracker (of about 4 ms). Observing during these conditions could clearly benefit from a fringe tracker with a shorter response time. Could we still improve the control loop during good atmospheric conditions, however?

The answer is yes. It lies in the proper management of the S/N by the Kalman filter. For convenience, we used an asymptotic estimation of the Kalman gain from the Riccati equation. A better Kalman filter would propagate errors as well as the state by also applying the equation of state to the covariance matrix:

$$\hat{\Sigma}_{\boldsymbol{x},n|n-1} = \mathsf{A}_V \cdot \hat{\Sigma}_{\boldsymbol{x},n-1} \cdot \mathsf{A}_V^{\top}, \tag{79}$$

and derive the optimum gain each DIT. This could be achieved with additional computing power.

An additional amelioration could come from modal control. This was proposed in Menu et al. (2012) and could theoretically be implemented. However, we have currently not been able to find any practical implementation that would make it robust for a realistic environment.

References

Abuter, R., Dembet, R., Lacour, S., et al. 2016, in Optical and Infrared Interferometry and Imaging V, Proc. SPIE, 9907, 990721

Berger, D. H., & Monnier, J. D. 2006, in SPIE Conf. Ser., Proc. SPIE, 6268, 62683K

Cassaing, F., Duigou, J. M. L., Houairi, K., et al. 2008, in Optical and Infrared Interferometry, eds. W. C. Danchi, F. Delplancke, & M. Schöller, Proc. Soc. Photo-Opt. Instrum. Eng., 7013

Choquet, É., Menu, J., Perrin, G., et al. 2014, A&A, 569, A2

Colavita, M. M., Booth, A. J., Garcia-Gathright, J. I., et al. 2010, PASP, 122, 795

Colavita, M. M., Wizinowich, P. L., Akeson, R. L., et al. 2013, PASP, 125, 1226

Correia, C., Raynaud, H. F., Kulcsár, C., & Conan, J. M. 2008, in Adaptive Optics Systems, Proc. SPIE, 7015, 70151F

Delplancke, F., Derie, F., Lévêque, S., et al. 2006, in SPIE Conf. Ser., Proc. SPIE, 6268, 62680U

Finger, G., Baker, I., Alvarez, D., et al. 2016, in Adaptive Optics Systems V, Proc. SPIE, 9909, 990912

Genzel, R., Eisenhauer, F., & Gillessen, S. 2010, Rev. Mod. Phys., 82, 3121

Gravity Collaboration (Abuter, R., et al.) 2017, A&A, 602, A94

Gravity Collaboration (Abuter, R., et al.) 2018a, A&A, 615, L15

Gravity Collaboration (Abuter, R., et al.) 2018b, A&A , 618, L10

Gravity Collaboration (Sturm, E., et al.) 2018c, Nature, 563, 657

Houairi, K., Cassaing, F., Perrin, G., et al. 2008, in Optical and Infrared Interferometry, eds. W. C. Danchi, F. Delplancke, & M. Schöller, Proc. Soc. Photo-Opt. Instrum. Eng., 7013

Jocou, L., Perraut, K., Nolot, A., et al. 2010, in Optical and Infrared Interferometry II, Proc. SPIE, 7734, 773430

Jones, R. H. 1980, Technometrics, 22, 389

Kiekebusch, M. J., Chiozzi, G., Knudstrup, J., Popovic, D., & Zins, G. 2010, in Software and Cyberinfrastructure for Astronomy, Proc. SPIE, 7740, 77400T

Kulcsár, C., Massioni, P., Sivo, G., & Raynaud, H. F. G. 2012, in Adaptive Optics Systems III, Proc. SPIE, 8447, 84470Z

Labeyrie, A. 1970, A&A, 6, 85

Lacour, S., Jocou, L., Moulin, T., et al. 2008, in Optical and Infrared Interferometry, Proc. SPIE, 7013, 701316

Lacour, S., Eisenhauer, F., Gillessen, S., et al. 2014, A&A, 567, A75

Lacour, S., Dembet, R., Abuter, R., et al. 2018, in Optical and Infrared Interferometry and Imaging VI, SPIE Conf. Ser., 10701, 1070107

Lawson, P. R., Colavita, M. M., Dumont, P. J., & Lane, B. F. 2000, in Interferometry in Optical Astronomy, eds. P. Léna, & A. Quirrenbach, Proc. SPIE, 4006, 397

Le Bouquin, J.-B., Bauvir, B., Haguenauer, P., et al. 2008, A&A, 481, 553

Lozi, J., Cassaing, F., Le Duigou, J. M., et al. 2011, in Techniques and Instrumentation for Detection of Exoplanets V, ed. S. Shaklan, Proc. Soc. Photo-Opt. Instrum. Eng., 8151

Malbet, F., Kern, P., Schanen-Duport, I., et al. 1999, A&AS, 138, 135

Menu, J., Perrin, G., Choquet, E., & Lacour, S. 2012, A&A, 541, A81

Perraut, K., Jocou, L., Berger, J. P., et al. 2018, A&A, 614, A70

Petit, C., Conan, J.-M., Kulcsár, C., Raynaud, H.-F., & Fusco, T. 2008, Opt. Express, 16, 87

Petrov, R. G., Malbet, F., Weigelt, G., et al. 2007, A&A, 464, 1

Petrov, R. G., Elhalkouj, T., Boskri, A., et al. 2014, in Opt. and Infrared Interferometry IV, Proc. SPIE, 9146, 91462P

Petrov, R. G., Boskri, A., Bresson, Y., et al. 2016, in Opt. and Infrared Interferometry and Imaging V, Proc. SPIE, 9907, 99071F

Pfuhl, O., Eisenhauer, F., Haug, M., et al. 2010, in Opt. and Infrared Interferometry II, Proc. SPIE, 7734, 77342A

Pfuhl, O., Haug, M., Eisenhauer, F., et al. 2012, in Opt. and Infrared Interferometry III, Proc. SPIE, 8445, 84451U

Pfuhl, O., Haug, M., Eisenhauer, F., et al. 2014, in Opt. and Infrared Interferometry IV, Proc. SPIE, 9146, 914623

Poyneer, L. A., & Véran, J.-P. 2010, J. Opt. Soc. Am. A, 27, A223

Shao, M., & Colavita, M. M. 1992, A&A, 262, 353

Tatulli, E., Millour, F., Chelli, A., et al. 2007, A&A, 464, 29

Weigelt, G. 1977, Opt. Commun., 21, 55

Woillez, J., Akeson, R., Colavita, M., et al. 2012, PASP, 124, 51

Woillez, J., Lai, O., Perrin, G., et al. 2017, A&A, 602, A116