

The polarization-based collimated beam combiner and the proposed NOVA fringe tracker (NFT) for the VLTI

Jeffrey A. Meisner^a, Walter J. Jaffe^a, Rudolf S. Le Poole^{ab}, Sylvania F. Pereira^c, Andreas Quirrenbach^d, David Raban^b, and Amir Vosteen^b

^aLeiden Observatory, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands

^bTNO Science and Industry, Stieltjesweg 1, 2628 CK Delft, The Netherlands

^cTechnische Universiteit Delft, Lorentzweg 1, 2628 CJ Delft, The Netherlands;

^dLandessternwarte Heidelberg, Königstuhl 12, D-69117 Heidelberg, Germany

ABSTRACT

The Polarization-Based Collimated Beam Combiner efficiently produces pairwise interference between beams from multiple telescopes. An important feature is achieving "Photometric Symmetry" whereby interference measurements have no first-order sensitivity to wavefront perturbations (or photometric variations following spatial filtering) which otherwise entail visibility measurements with increased error, bias, and nonlinearity in phase determination. Among other proposed applications, this topology has been chosen as the basis for the design of the *NOVA Fringe Tracker* (NFT), a proposed 4 or 6 telescope second-generation fringe tracker for the VLTI. The NFT takes advantage of the photometric symmetry thus achieved making it capable of tracking on stars resolved beyond the first visibility null, as well as interfering a telescope beam with one which is 20 times brighter, a design goal set by ESO. By not requiring OPD modulation for interferometric detection, the detector exposure time can be increased without performance reduction due to time skew nor is sensitivity reduced by removing optical power for photometric monitoring, and use of two-phase interferometric detection saves one half of the photons being diverted for detection of the other two (mainly) unused quadrature phases. The topology is also proposed for visibility measuring interferometers with configurations proposed for the achievement of balanced quadrature or 3-phase interferometric detection. A laboratory demonstration confirms $\gg 100:1$ rejection of photometric crosstalk in a fringe tracking configuration where atmospheric OPD fluctuations were simulated using a hair dryer. Tracking with a 30:1 intensity ratio between the incoming beams was performed while rejecting large introduced photometric fluctuations.

Keywords: Interferometry, Beam combiner, Fringe tracking, VLTI, Phase reference, Visibility estimation, Near infrared

1. INTRODUCTION

In this paper we present a topology for interferometric beam combination which has certain clear advantages for an astronomical interferometer particularly in addressing performance reduction stemming from what we will term *photometric crosstalk* (sect. 2).

1.1 Performance Criteria and Trade-offs

Optical interferometry for astronomy has a number of performance criteria to be simultaneously maximized. Of course not all interferometric instruments and scientific programs will require the same emphasis on each of the following criteria.

1. Sensitivity limitations and reducing the effect of detection noise.
2. Sufficient update rate relative to atmospheric (and other) OPD fluctuations.
3. Generation of data with higher spectral resolution.
4. Precision and reproducibility of measured visibility magnitudes and phases.
5. Ability to measure sources with a low visibility (low fringe contrast), not just a low level of correlated flux.

Send email correspondence to meisner@strw.leidenuniv.nl

6. Reducing the effect of wavefront fluctuations caused by the atmosphere.
7. Reducing the effect of photometric fluctuations following spatial filtering.

In almost all cases item (1), the sensitivity of an instrument, is of key importance since it determines the usefulness of the instrument, that is, how many different targets it can observe. The sensitivity of the optical detection process generally governs the resulting sensitivity of the instrument. In the visible it is possible to have photon counting detectors (whose noise is thus limited by the fundamental energy of a photon), and at longer infrared wavelengths background radiation in the field of view (or spatial mode) of the star presents the fundamental detection limit.

However at near-infrared wavelengths (arguably the most productive wavelength region for interferometry) detector noise generally sets the sensitivity limit for an instrument. In that case (1) competes with (2) as well as (3) since dividing the received photons into a greater number of detector channels or shorter detector exposures increases the effect of the detector noise (for the faintest sources). Thus it is advantageous to reduce the number of detector readings required to produce a visibility measurement (2), ideally to one.

However reducing the detector frame rate, for increased sensitivity, conflicts with the most common scheme for combating the effects of wavefront fluctuations (6) or, when using spatial filtering, the photometric fluctuations (7) produced by running such a wavefront through a single spatial mode filter, which is through fringe scanning or phase stepping. Interferometric detection schemes in which measurements are derived from successive readings of a detector in conjunction with OPD modulation have another basic drawback (sect. 2.4). The estimator used can only take into account the intentional OPD stepping in between the detector readouts. But in practice the entire system is not frozen during that period, for the atmospheric OPD itself is varying, as is the wavefront received by each telescope. These wavefront fluctuations cause fluctuations in the effective measured visibility or the photometric levels following a spatial filter. Increasing the precision of visibility measurements then requires OPD modulation (and thus detection) at higher rates, but all at the expense of sensitivity (1).

The configuration we present in this paper is particularly a solution to this dilemma, in which the precision of interferometric detection does not require a rapid detector readout, compromising the sensitivity of the instrument. Nor is an elementary measurement of visibility (magnitude and/or phase) based on more than a single detector exposure, so that the validity of the estimator does not rest on the (flawed) assumption of a “frozen” atmosphere during a period of time. This topology maximizes all of the above performance criteria except for (3), achieving higher spectral resolution, which is best accomplished through long coherent exposures in narrow spectral channels while fringe tracking on the broadband light. The immediate application for the concept is in fact such a fringe tracker for the VLTI.

1.2 Outline of This Paper

This paper will quantify the problem of *photometric asymmetry* and *photometric crosstalk* (sect. 2) pointing out the inadequacies of existing solutions. Then we introduce (sect. 3) the topology referred to as the “Polarization-Based Collimated Beam Combiner,” explain its operation and practical implementations.

That topology is at the heart of the *NOVA Fringe Tracker* (NFT) which has been proposed to ESO in a phase A study to recommend a second-generation fringe tracker for the VLTI. The design of the NFT is presented in some detail in sect. 4, as a concrete example in which the achievement of photometric symmetry solves performance requirements facing the implementation of a modern fringe tracking interferometer (with similar implications for a visibility measuring instrument).

Finally in sect. 5 we present laboratory results in which a realization of the concept using laser light subject to simulated “atmospheric turbulence,” and “photometric fluctuations,” has demonstrated the desired level of immunity to photometric crosstalk in a simple fringe tracking configuration.

2. PHOTOMETRIC CROSSTALK: DESCRIPTION OF THE PROBLEM AND CLASSICAL SOLUTIONS

Interferometric beam combination schemes as employed by astronomical interferometers fall into one of two basic types. In order to place the light from two (or more) telescopes into a single spatial mode, so that they can fully interfere, a beamsplitter (partially transmitting mirror) or comparable guided wave device is employed as is illustrated in fig. 1. The second type of beam combination does not attempt to place the light into a single spatial mode but allows two (or more)

beams in distinct spatial modes to intersect on a surface where a classical fringe pattern (as obtained with Young's double slit experiment) is formed and detected over a number of pixels*. The latter type of beam combination requires more detector pixels which makes it inefficient using near infrared detectors in which we are trying to meet the requirements listed in sect. 1.1. Thus we will only consider beam combiners of the first type.

2.1 Single Mode Beam Combiners and Measurement of Interference

Let us quickly clarify the scope of the interferometric hardware being discussed, of which the configuration of fig. 1 is but one example. The starlight from two (or more) sources entering a so-called *on-axis* beam combiner interfere because they are in the same spatial mode. This might be insured by having spatial filters in each incoming beam so that all light is in one spatial mode, or could simply be accomplished by pointing the telescopes so that the light from one star finds itself in the same spatial mode when it enters the beam combiner. In the latter case there may be a spatial filter following the beam combining surface, but this is not essential. It's also possible for the beam splitting surfaces to combine a field with several stars whose spatial modes are matched to those of the other input beam. In the case of a fiber coupler or other guided wave device, then there is only one spatial mode propagating to consider. What all of these configurations have in common is that the interference from one star can be efficiently detected using but a single detector element in each output path, ignoring any spatial information. The correlated power detected is dependent on the amount of light from each arm in one spatial mode regardless of what light, if any, is present and/or detected in other modes.

We will refer to the amount of light in the incoming beams which eventually gets detected in one or the other output of the beam combiner as *photometric*. This has become jargon in this field in referencing the total power of the beam whether or not it is correlated and can be easily and unambiguously quantified by examining cases in which there is no correlation between the incoming beams (zero visibility) or cases in which one beam is blocked by a shutter, for instance. A primary reason that spatial filters are of interest is that they provide a simple relationship between the "photometry" and the correlated power, namely that the correlated power[†] $I_{corr} = 2\sqrt{I_A I_B} V$ which defines the effective visibility V . I_A and I_B refer to the "photometric" powers of the interfering beams. That means that if I_A and I_B are known and I_{corr} is measured, the visibility V can be directly calculated.

2.1.1 "Photometric crosstalk"

However our discussion will proceed without any restrictions regarding the use of spatial filtering or monitoring of the photometric powers. Rather we are concerned with the presence of crosstalk from the photometry (in the above sense) into the measurement of interference. Consider again the configuration of fig. 1 which is sufficiently general to apply to all the hardware variants implementing single mode interference mentioned above when interfering only two telescopes (or performing pair-wise interference). If the beam splitting surface reflects exactly half of the light from either incoming beam and transmits the other half, then incoherent light from either beam will illuminate both detectors equally. When the detectors' readings are subtracted, there will be no net response and we describe this (desirable) beam combiner as having "photometric symmetry." The only signal generated by the detectors after subtraction is then due to the correlated power, that is, the interferometry we are trying to measure.

Conversely, if the reflection and transmission coefficients of the beamsplitter are *unequal* then there generally *will* be a non-zero result even in the absence of any correlated component. This may thus masquerade as an interferometric signal so we refer to this spurious component of the measurement as "photometric crosstalk." Obviously the presence of photometric crosstalk in a beam combiner having "photometric asymmetry" can only degrade the measurement of visibility. Herein lies the problem at hand.

2.2 Quantification of Photometric Asymmetry

There are two approaches we can use to quantify photometric asymmetry, which will have identical implications. We can invoke a model of an ideal lossless beamsplitter and analyze it, or we can take an existing beamsplitter using real photodetectors and make measurements. In practice we will need to follow the latter phenomenological approach, however

*Note that especially in older literature the first sort of combiner was frequently referred to as *pupil-plane* and the second as *image-plane* interference. However those terms do not accurately describe the distinction and, thankfully, are coming into disuse.

[†]Rather than this abstract formulation, one could say more specifically that the difference between the two detected intensities $\Delta I = 2\sqrt{I_A I_B} |V| \cos(\phi)$ where ϕ is a phase angle dependent on various path lengths etc. At some particular phase, namely $\phi = 0$, ΔI and $|I_{corr}|$ are equal.

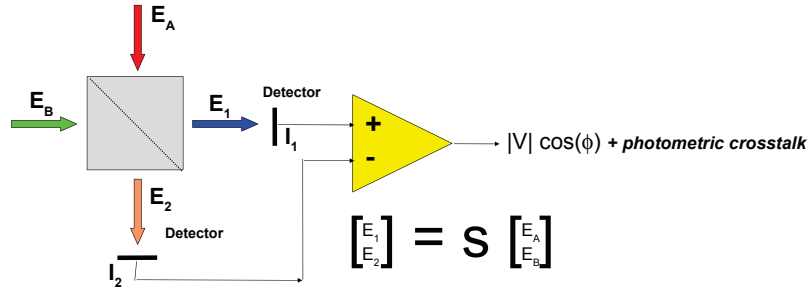


Figure 1. Single spatial mode beam combiner using a partially transmissive mirror. The output electric fields E_1 and E_2 are a linear combination of the input electric fields E_A and E_B according to the scattering matrix s . Subtracting the detected intensities $I_1 = |E_1|^2$ and $I_2 = |E_2|^2$ yields an estimate of the visibility. When the beam combiner lacks photometric symmetry, the estimate will include “photometric crosstalk” due to any imbalance between the incident intensities $I_A = |E_A|^2$ and $I_B = |E_B|^2$.

it will be reassuring to know that there is an idealized model which it could be related to, despite possible departures from that ideal behavior. We will restrict the analysis to considering a beam combiner with two inputs and two outputs. The following discussion will strictly apply to a single-mode optical system, but a typical optical system such as a partially silvered mirror which transmits an image will generally perform similarly for every spatial mode in one polarization (usually with a different response applying to the spatial modes of the other polarization). The following treatment will be taken to be limited not only to a single spatial mode in a single polarization, but also to a single wavelength. The parameters describing the system will thus generally be dependent on the optical wavelength and which polarization[‡] is considered.

2.2.1 Model

Consider a lossless beamsplitting surface as shown in fig. 1 with incident waves having (complex) amplitudes of E_A and E_B and output waves having amplitudes of E_1 and E_2 . The outputs are connected to the inputs through a 2×2 scattering matrix s which, describing a lossless system, necessarily is a *unitary matrix*[§] and can be decomposed as follows:

$$s = \begin{bmatrix} e^{j\phi_1} & 0 \\ 0 & e^{j\phi_2} \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1+\gamma}{2}} & \sqrt{\frac{1-\gamma}{2}} \\ -\sqrt{\frac{1-\gamma}{2}} & \sqrt{\frac{1+\gamma}{2}} \end{bmatrix} \begin{bmatrix} e^{j\phi_A} & 0 \\ 0 & e^{j\phi_B} \end{bmatrix}. \quad (1)$$

The external properties of the beamsplitter are contained in the middle matrix while the leading and trailing diagonal matrices contain phase factors whose actual values are dependent on the precise position of the reference planes at which the complex electric field would be measured (note that there is one free parameter among the 4 phase factors as presented in this decomposition).

Expressed in this manner, we term γ the *photometric asymmetry* coefficient where $-1 < \gamma < +1$. The special case of $\gamma = 0$ corresponds to a balanced beamsplitter in which an impinging wave in the A or B mode generates equal power in the output modes 1 and 2. In general a lossless beamsplitter will have its *power* (rather than amplitude) transmission into each output mode, assuming uncorrelated inputs, given by $\frac{1+\gamma}{2}$ and $\frac{1-\gamma}{2}$, as can be found by taking the squared magnitude of s in (1):

$$|s|^2 = \begin{bmatrix} \frac{1+\gamma}{2} & \frac{1-\gamma}{2} \\ \frac{1-\gamma}{2} & \frac{1+\gamma}{2} \end{bmatrix}. \quad (2)$$

These power coefficients are plotted with dashed lines in fig. 2A.

[‡]We will consider a principal state of polarization to be one such that excitation of E_A and E_B in the designated state of polarization will excite E_1 and E_2 in their designated state of polarization. The corresponding principal polarization state (and its orthogonal polarization state) need not be the same for each of the four beams considered. However in a practical beamsplitter with its beamsplitting surface vertically oriented, the principal polarization states will be the vertical or s polarization and the horizontal or p polarization, for all four spatial beams.

[§]A unitary matrix \mathbf{A} has the property that multiplying it by its Hermetian transpose obtains the identity matrix: $\mathbf{A} \times \mathbf{A}^\dagger = \mathbf{I}$.

2.2.2 Measurement of photometric asymmetry

Now let us suppose that we are using ordinary photodetectors and thus unable to directly measure the electric field. Even worse, let us assume that the photodetectors are uncalibrated and thus measure with arbitrary scale factors. We can still determine γ through four simple measurements. We attach a photodetector to each output path of the beamsplitter and measure the intensities (times unknown scale factors) $c_1 I_1$ and $c_2 I_2$ where c_1 and c_2 are unknown. Let us shine light (of a particular polarization, wavelength, and spatial mode) into input A of the beamsplitter (and no light into the B input) from which we measure (for instance) photocurrents i_{1A} and i_{2A} respectively. The input power leading to these two measurements was arbitrarily chosen, but we will only be interested in the ratio of the detector readings i_{1A}/i_{2A} . Then illuminating only input B , not necessarily with the same power, we measure (for instance) photocurrents i_{1B} and i_{2B} and their ratio. Then, without making any assumptions about the character of the beamsplitter, we can empirically determine the photometric asymmetry coefficient γ using:

$$\gamma = \frac{\sqrt{i_{1A}/i_{2A}} - \sqrt{i_{1B}/i_{2B}}}{\sqrt{i_{1A}/i_{2A}} + \sqrt{i_{1B}/i_{2B}}} , \quad \text{also} = \frac{\sqrt{i_{1A}/i_{1B}} - \sqrt{i_{2A}/i_{2B}}}{\sqrt{i_{1A}/i_{1B}} + \sqrt{i_{2A}/i_{2B}}} . \quad (3)$$

Having thus found γ , the photometric asymmetry, we could conjecture that the underlying operation of the beamsplitter (in the case that it is lossless) is as given in (1) given some choice of phase factors. However we will not need to consider that underlying identification and will mainly proceed on the basis of the intensity response observed from the two ratios i_{1A}/i_{2A} and i_{1B}/i_{2B} summarized in the parameter γ .

2.2.3 Visibility reduction

Although this is of rather minor importance, we will also mention that besides the issue of photometric crosstalk quantified in the next section, there is a modest decrease in the magnitude of the measured visibility itself when $\gamma \neq 0$. Assuming a beam combiner with a transfer function given by (1), then analyzing the optical power that would be detected in either output beam of fig. 1 we would find, for instance for I_1 :

$$I_1 = |E_1|^2 = \sqrt{1 - \gamma^2} \cdot \text{Re}\{E_A E_B^* e^{j\phi}\} + \text{photometric terms} \quad (4)$$

and likewise for I_2 . This small reduction of the detected interferometric power as a function of γ is shown in fig. 2A. So for instance with an asymmetric beam combiner whose $\gamma = \frac{1}{3}$, in which the splitting ratios are 1:2 and 2:1 respectively, the strength of the visibility signal is only reduced by 6%. This is of almost no concern in comparison with the photometric crosstalk caused by such an asymmetry. However the factor $\sqrt{1 - \gamma^2}$ will be propagated into various equations in which the signal level is compared with photometric crosstalk or noise, such as determinations of the visibility estimation error.

2.2.4 Photometric crosstalk due to non-zero γ

Rather the major effect of an unbalanced beam combiner is the unwanted introduction of *photometric crosstalk* into the interferometric signal. What would have been the balanced interferometric determination, $\Delta I = I_1 - I_2$, is subject to photometric crosstalk as follows:

$$\Delta I = I_1 - I_2 = 2\sqrt{1 - \gamma^2} \text{Re}\{E_A E_B^* e^{j\phi}\} + \gamma I_A - \gamma I_B . \quad (5)$$

Let us see what effect the crosstalk terms would have on an estimate of fringe visibility V . Using the complex visibility V we can rewrite the quantity $\text{Re}\{E_A E_B^* e^{j\phi}\}$ in (5) as $\text{Re}\{\sqrt{I_A I_B} V e^{j\phi}\}$. Without loss of generality let us assume $\phi = 0$ and call \widehat{V}_R an estimate of the real part of the visibility as follows:

$$\widehat{V}_R = \frac{i_1 - i_2}{2\sqrt{1 - \gamma^2} \sqrt{I_A I_B}} \quad (6)$$

where $i_1 = I_1 + n_1$ and $i_2 = I_2 + n_2$ are signals from the detection of I_1 and I_2 containing added (zero mean) detector noise terms of n_1 and n_2 respectively. Then we can evaluate that estimator using (5):

$$\widehat{V}_R = \text{Re}\{V\} + \frac{n_1 - n_2}{2\sqrt{1 - \gamma^2} \sqrt{I_A I_B}} + \frac{\gamma}{2\sqrt{1 - \gamma^2}} \frac{I_A - I_B}{\sqrt{I_A I_B}} . \quad (7)$$

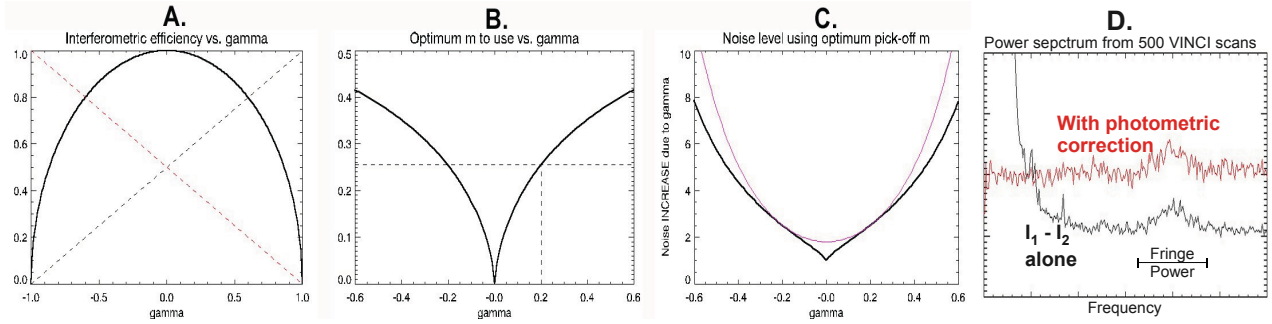


Figure 2. **A.** As a function of γ in the idealized model, the two power transmission coefficients (sect. 2.2.1) $|s_{11}|^2 = |s_{22}|^2 = \frac{1+\gamma}{2}$ and $|s_{12}|^2 = |s_{21}|^2 = \frac{1-\gamma}{2}$ are shown in dashed lines. The relative detected power due to the interferometric signal $|s_{22}s_{21}| = |s_{12}s_{11}|$ is plotted in bold (relative to its peak) as a function of γ . **B.** In order to correct for a photometrically asymmetric beam combiner, a portion m of each incoming wave is diverted toward its own detector instead of the beam combiner. The optimum m (assuming identical detector noise levels and the other assumptions stated in the text) is plotted as a function of γ . **C.** The noise power level σ_V^2 affecting a visibility estimate using (8) as a function of $\gamma = 0$ using the m determined in (B) is plotted, relative to that obtainable with a balanced beam combiner not requiring photometric monitoring ($\gamma=0, m=0$). Simply using $m = .255$ (optimum for $\gamma = .2$) yields the result plotted with a thin line. **D.** Incoherent power spectrum integrated from 500 scans of VINCI data on a highly resolved star (low visibility). Using photometric pick-off correction results in the upper plot. Crosstalk from the photometry (which dominates at low frequencies) has been eliminated. However at all higher frequencies (including the fringe frequency) the detector noise from the photometric channels introduced as a result of the photometric “cancellation” substantially increases the noise level, as is discussed in sect. 2.3.1.

Beyond the first term which is the actual visibility we are estimating, we have a second term due to the noise in each detector which is enhanced by $1/\sqrt{1-\gamma^2}$ due to the reduction of the signal strength that we noted in (4). The next fraction, however, contains the difference between the photometric levels of the input beams in relation to their geometric mean, with a leading coefficient increasing in magnitude with $|\gamma|$, the photometric asymmetry coefficient (as well as the less important $1/\sqrt{1-\gamma^2}$ factor). Since even in the best case I_A and I_B are independently fluctuating, there will be a photometric crosstalk term present unless $\gamma = 0$.

What’s more, in the case that the beams being interfered are photometrically imbalanced, $\langle I_A \rangle \neq \langle I_B \rangle$, (7) predicts a *bias* in the visibility determination unless $\gamma = 0$. In some cases this will be of major concern. And when there is a *large* discrepancy, say $I_A \gg I_B$, then in addition to that bias, even the random contribution due to photometric fluctuations in I_A is further enhanced due to $\langle I_A \rangle$ being much larger than its denominator $\sqrt{I_A I_B}$.

Also of special concern is the effect of the photometric crosstalk terms when measuring a very low underlying visibility $|V|$. Then the relative effect of both the fluctuating and bias terms is enhanced. Even though measures (such as fringe scanning) may be employed to counter the effect of such biases, the enhanced fluctuations will make measurement of the small underlying visibility more difficult, notably leading to a bias in the estimation of $|V|^2$ using incoherent estimators.

2.3 Achieving Photometric Symmetry Using Photometric Pick-off Detection

One way of defeating the effect of photometric crosstalk (causing either random or bias errors) in (5) is by gauging the intensities of the incoming beams, I_A and I_B (fig. 3A), and using that information to correct for the photometric crosstalk terms in (7) for instance.¹ Photometric pick-offs of this sort have become common among interferometers using spatial filtering,^{2,3} though this is also for the purpose of properly scaling a visibility estimator such as (6) in which $\sqrt{I_A I_B}$ appears in the denominator.

Consider a beam combiner in which a fraction m of the intensity of each input beam has been diverted towards a photometric detector generating measurements $i_A = mI_A + n_A$ and $i_B = mI_B + n_B$ (where n_A and n_B are the detection noise terms), while $1 - m$ of each beam’s power proceeds to the interferometric beamsplitter whose outputs as before are I_1 and I_2 which are detected (with added detector noise). Now an estimator replacing (6) can be written which includes a portion of the detected i_A and i_B signals in order to cancel the photometric crosstalk contaminating $\Delta I = I_1 - I_2$:

$$\widehat{V}_R = \frac{1}{2\sqrt{1-\gamma^2}\sqrt{I_A I_B}} \left(\frac{i_1 - i_2}{(1-m)} + \frac{\gamma(i_B - i_A)}{m} \right). \quad (8)$$

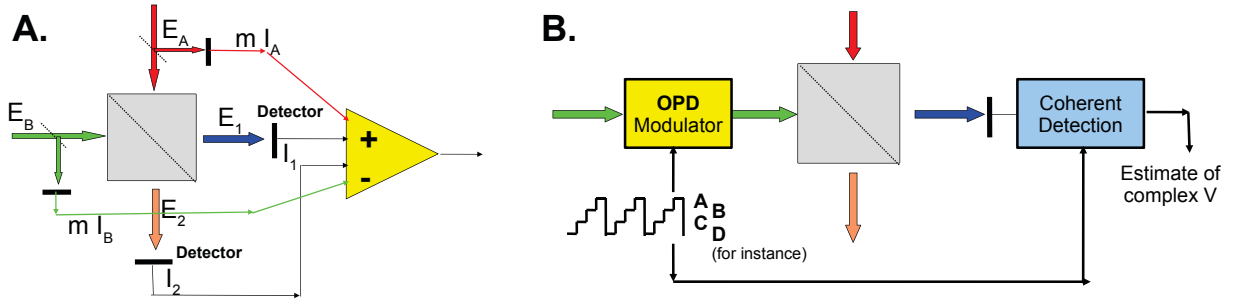


Figure 3. **A.** The beam combiner in fig. 1 with added *photometric monitoring* channels. Although the beam combiner itself lacks photometric symmetry, the addition of a specified portion of the input beams' power levels I_A and I_B can cancel the photometric crosstalk in the visibility estimator. **B.** Modulation applied to the OPD in one arm of an interferometer is used to transform the visibility into an AC signal. Switching between 4 quadrature phases can be employed as shown, in order to obtain 4 detector readings from which the complex visibility can be estimated. Another approach which is widely used involves scanning the OPD across the entire coherence envelope of the fringe packet, among other possibilities. In all cases the resulting detector readings are only interpreted *differentially* so that DC or low frequency components of the readings (including most of the energy from photometric crosstalk) do not affect the interferometric result.

2.3.1 Visibility estimation error using photometric pick-off detection

Let us call $\epsilon = \widehat{V}_R - \mathcal{R}e\{V\}$ the estimation error using (8). Because we have successfully subtracted the photometric crosstalk using (8), ϵ is zero mean and only reflects the effect of detector noise. The first term (inside the parenthesis) of (8) includes detection noise from the interferometric channels, whose effect is accentuated due to the portion of optical power, m , consumed by the photometric channels. The second term includes detection noise from the photometric channels, whose effect is mitigated by diverting more of the optical power toward the photometric channels (increasing m). What is the mean squared estimation error $\sigma_V^2 = \text{Var}\{\epsilon\}$? Assume that all 4 photodetectors have a mean squared noise level of N_0 . Then it can be seen that the resulting estimation error would be given by:

$$\sigma_V^2 = \frac{N_0}{2I_A I_B (1 - \gamma^2)} \left(\frac{1}{(1 - m)^2} + \frac{\gamma^2}{m^2} \right). \quad (9)$$

σ_V^2 in (9) can be minimized by adjusting m , the proportion of power split into the photometric monitoring beams. The optimum m is given by:

$$m_0 = \frac{|\gamma|^{\frac{2}{3}}}{1 + |\gamma|^{\frac{2}{3}}}. \quad (10)$$

This value is plotted as a function of γ in fig. 2B. For instance, with $\gamma = .2$ corresponding to a 60:40 beamsplitter, the ideal value of m (dashed line) is .255. About one quarter of the light is thus used for photometric correction, and three quarters is combined interferometrically.

Using the optimum pick-off factor according to (10), it can be shown that the mean squared visibility estimation error σ_V^2 in (9) becomes:

$$\sigma_V^2 = \frac{N_0}{2I_A I_B (1 - \gamma^2)} \left(1 + |\gamma|^{\frac{2}{3}} \right)^3. \quad (11)$$

The second factor, $(1 + |\gamma|^{\frac{2}{3}})^3$, represents the increase in noise power affecting the visibility estimator due to the use of photometric pick-off and correction, beyond the amount (in the first fraction) due to the noise contribution of the two interferometric detectors alone with no photometric pick-offs ($m = 0$). That noise increase factor, along with the noise increase due to the $\sqrt{1 - \gamma^2}$ visibility reduction (sect. 2.2.3) is plotted as a function of γ in fig. 2C.

Use of photometric monitoring with a pick-off ratio m different from the optimum according to (10) will lead to a noisier result when properly combined for rejection of photometric crosstalk. For instance, using $m = .255$, which we found to be optimum when $\gamma = .2$, yields a result with a noise level plotted with a thin line in fig. 2C. For $.1 < \gamma < .35$ the performance is hardly worse than if m had been chosen using (10).

Even given the optimum pick-off ratio m using (10), it can be seen that the magnitude of estimation error rises rapidly as γ departs from zero. Thus a beam combiner using a 60:40 beamsplitter ($\gamma = .2$) with photometric correction, in a detector noise limited instrument, loses half a stellar magnitude in sensitivity! The noise increase in processing VINCI scans by the addition of photometric monitoring is seen in fig. 2D, where it isn't really needed since VINCI already uses OPD modulation (sect. 2.4). It can still be seen that it does indeed cancel the photometric crosstalk which is dominant at lower frequencies, but again at the expense of an increased noise floor at all frequencies (and consequently an increased estimation error) as detailed above.

2.3.2 Inadequacy of the photometric monitoring approach

We have seen that even using the optimum power pick-off coefficient m_0 in (10), there will be a substantial noise penalty due to the photometric monitoring even to compensate a moderately asymmetric beam combiner. But beyond that, the success of any such scheme for removing the effect of photometric asymmetry in an interferometric beam combiner is predicated on γ having a known (perhaps after being measured) value. But in practice, γ will generally be a function of wavelength over the bandwidth of an instrument, as will the response of the photometric pick-off itself.² So any correction applied using the photometric monitoring cannot take these into account. The problem is further compounded when a system with two polarizations is considered, since the response of most beam combiners will be substantially different between the two polarizations.

This means that there will be residuals after the photometric correction unless it can be assured that the power in all wavelengths and both polarizations remains in a fixed proportion. That will generally not be the case. For instance, if the photometric fluctuations are a result of spatial filtering, then for each particular perturbed wavefront the net transmission at different wavelengths will be in a unique proportion. And if the spatial filter is a single mode optical fiber, then the entrance mode seen by the fiber will be different between polarizations.

2.4 Defeating Photometric Crosstalk using OPD Scanning or Phase Switching

Most astronomical interferometers include an applied OPD modulation (fig. 3B) so that the detected visibility is rotated in phase (or that the point of measurement within the fringe packet is scanned). This overcomes *most* photometric crosstalk by shifting the interferometric signal to a frequency (typically hundreds of Hertz) somewhat beyond the time scale of atmospheric fluctuations, which can then be detected and interpreted synchronously. One can thus detect interference in quadrature phases by stepping in between four points⁴ separated by $\lambda/4$. Many interferometers scan across an entire *fringe packet*,^{2,3} thus detecting the fringe oscillation at one cycle per λ of applied OPD multiplied by the coherence envelope. And there are other possibilities.

One drawback of all such OPD modulating schemes is that the data obtained is somewhat delayed due to the fact that it is based on not just the last detector reading, but typically includes readings from several detector frames in the past. This is an issue in a fringe tracking application like the NFT (sect. 4) whose phase measurement is employed in realtime. Simply increasing the detector frame rate is not an acceptable solution, as discussed below.

The interpretation of data from an OPD modulated interferometer is always based on the idealized model in which the atmosphere is "frozen" during the duration of the detector exposures which contribute to the resulting measurement. Therefore the actual atmospheric OPD change during that finite time period will produce individual readings with fluctuations having nothing to do with detection noise. This effect can be reduced by increasing the detector frame rate to the point that the modulation cycle is *much* shorter than the atmospheric coherence time τ_0 .

Similarly, there will be a residual interjection of photometric noise due to variations in the incident telescopes' intensities I_A and I_B during the modulation cycle. In other words, with $\gamma \neq 0$ then frequency components of I_A and I_B which are near the fringe modulation frequency will be aliased and add noise to the estimators which cannot possibly take those fluctuations into account. Again this can be reduced by increasing the detector frame rate and OPD modulation frequency, although it is more difficult to quantify in this case since the time scale depends on the actual mechanism of photometric fluctuations. What's more, unlike the influence of atmospheric OPD discussed above, this effect is *additive* to the visibility estimator rather than multiplicative. Therefore it becomes especially severe in the case of trying to detect a low visibility source or in cases of a large discrepancy between the photometric levels of the two telescopes (both being performance goals facing the NFT).

The straight-forward solution in both cases, of just decreasing the detector frame time in order to allow for faster OPD modulation, is not an attractive option in a near infrared instrument. Since the limiting sensitivity of the instrument is

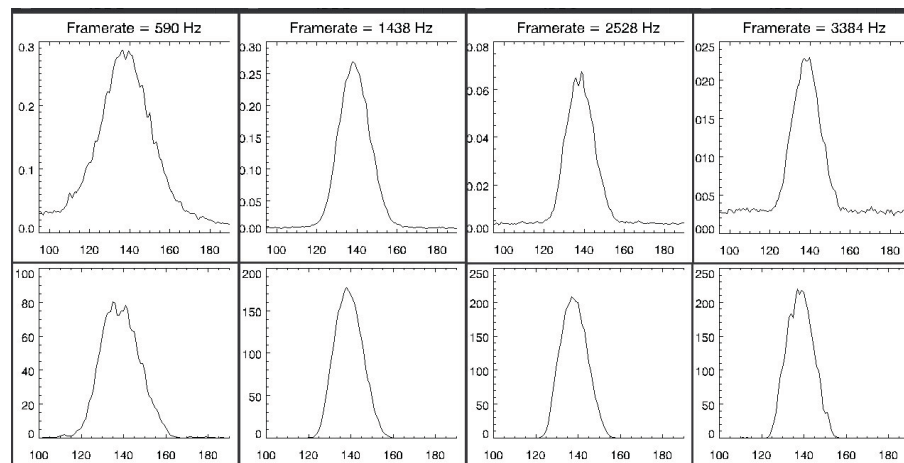


Figure 4. Power spectra from 200 VINCI scans of ϵ_{sco} integrated incoherently and averaged (top plots) and from the coherently integrated fringe (bottom) at four detector frame rates (and thus fringe detection frequencies) increasing from left to right. The x axis refers to optical frequency in THz.

determined by the detector readout noise, an increase in the detector frame rate thereby reduces the instruments's sensitivity. This could clearly be seen in the case of the VINCI instrument, in which there was simultaneous pressure to raise the frame rate in order to increase the coherence of the fringe scanning, and to reduce it in order to increase the SNR of each detector reading. An example of 4 technical observations on ϵ_{sco} in which the frame rate was varied between 590 and 3384 Hz (fringe frequency approximately 120 to 675 Hz) is shown in fig. 4 where the incoherent power spectra accumulated from 200 scans¹ are shown in the top plots, and the power spectra of the coherently integrated fringe⁵ shown in the bottom plots. The atmospheric OPD distortion of the fringe scanning at the low framerate of 590 Hz, on the leftmost plots, is seen to broaden the fringe power well beyond that attributable to actual K band radiation. The top plot at 590 Hz using incoherent integration of the power spectra also shows an increase in the noise floor on the low frequency side, and both plots show obvious irregularities due to the atmosphere when being sampled so slowly, leading to increased error (and often bias) in visibility determinations.

On the other hand, the smoother spectra obtained at somewhat higher frame rates deteriorate at the highest frame rate of 3384 Hz due to an decrease in the SNR per frame from detection noise. This is also clearly seen in the increased pedestal of the incoherently integrated power spectra proportional to the white noise from the detector operating with a shorter exposure. In the case of this very bright star (ϵ_{sco} , $m_K = -.55$) there is a sufficient margin in between the slow and fast frame rates in which accurate interferometry can be performed. With a second magnitude star (10 times dimmer), however, the reduced SNR seen at 3384 Hz would already occur at 340 Hz, well below the 590 Hz used in the leftmost plots, where the performance is already compromised. In conclusion, we have solved the original problem by creating new problems which reduce the data quality on all but the very brightest sources.

3. THE POLARIZATION-BASED COLLIMATED BEAM COMBINER

Having now examined the problem of photometric crosstalk in a practical beam combiner, and some imperfect ways of compensating for that problem, we now wish to introduce the proposed beam combining topology which easily achieves perfectly balanced beam combination ($\gamma = 0$) while retaining most other advantages of beam combiners using co-collimated beams (rather than off-axis illumination resulting in a spatial fringe pattern). Rather than relying upon the precision of beamsplitting surfaces (such as the coating of the hypotenuse surface in fig. 1) the balanced response of the *polarization-based collimated beam combiner* derives from a simple geometric alignment. The balance thus achieved has no further dependence on wavelength or time dependence.

Although the topology looks very different, the effect of this configuration is to take two input beams as in fig. 1 and create two output beams whose electric field amplitudes are the sum and difference of the two incident electric field amplitudes (times $\sqrt{2}/2$ and any further factors due to optical attenuation), or more generally a scattering matrix of the form (1) with $\gamma = 0$. The concept is not limited to two-phase detection but can also be used to produce quadrature interference

outputs, or 3 phase outputs, as will be detailed. In all cases the distinguishing feature remains that for uncorrelated ($V=0$) input beams, each output beam includes a contribution from input beam 1 and input beam 2 in a specific ratio (usually 1:1 when the input beams have the same power) and that ratio is the same for each output beam. Note that this is the condition that would cause the numerator of (3) to vanish, just considering any two outputs at one time, again proving that photometric symmetry is achieved. The consequence of these ratios being identical between each output beam is that the detected powers of two (or more) output beams can be subtracted in a particular proportion so that the only signal remaining after subtraction is due to interferometric correlation; the effect of incoherent radiation in each input beam is cancelled.

This topology has been chosen for the basis of the design of the *NOVA Fringe Tracker* (NFT) as will be described in detail in sect. 4. In the following explanation of the topology we will make references to the NFT design as a particular application of the concept in which it addresses several of the design requirements for a 4 or 6 telescope fringe tracker. Among modern astronomical interferometer designs, the NFT is somewhat unusual in only employing two-phase interference (with no OPD modulation) and no spatial filtering. We emphasize that these are specific choices for the NFT, whereas use of the general topology in an interferometer designed for measuring visibilities almost certainly *would* employ spatial filtering and quadrature (or 3-phase) interferometric detection.

3.1 Single Polarization Limitation

Besides being only applicable to pairwise interference, the only further limitation of this topology is that it only combines one polarization of one beam with one polarization of the other beam. Therefore it is not able to duplicate the functionality of the two input two output beam combiner diagrammed in fig. 1 in which both input beams include both polarizations and both output polarizations in each output beam fall upon the same detector.

Practically speaking this is rarely a significant limitation. It only rules out detecting interference between two beams using two-phase detection using only two photodetectors. More commonly an interferometer will want to detect interference in quadrature phases, for instance, in which case two parallel systems can process the two polarizations but with a 90° relative interference phase, with the output beams landing on a total of 4 detectors. And many interferometers (such as the NOVA Fringe Tracker, sect. 4) are intended to use light from 3 or more telescopes at one time. In that case, pairwise interference requires that each input beam be split between two (or more) beam combiners, and that can be efficiently done by separating their polarization components. The practical design detailed in sect. 4 illustrates the efficient application of this principle.

3.2 Principle of Operation

In its essential form, the polarization-based collimated beam combiner consists of three optical functions, labelled 1, 2 and 3 in fig. 5.

1. **Polarization reversing**, where the horizontal and vertical polarizations are interchanged, but only for every second input beam.
2. The **polarization recombining stage** (PRS) where one polarization from one telescope is combined with one polarization from another telescope so that those two waves travel together in a single spatial mode but with opposite polarizations.
3. The **polarization analyzer** and detectors, where interferometry between the two waves is effected and complementary detector readings generated.

The essence of the operational principle is that a wave from one telescope in one polarization is superimposed spatially (2) onto a similar wave from another telescope but whose polarization has been rotated (1) by 90° . When this “polarization recombined beam,” following any number of polarization-insensitive optical elements, eventually reaches the polarization analyzer (3), it is divided into an “A” and “B” polarization component, where A and B refer to the electric field vector at 45° and -45° from the horizontal. This implements complementary and truly balanced interference between the original telescopes’ waves in one polarization, that is, with γ (as defined in section 2.2.1) truly approaching zero.

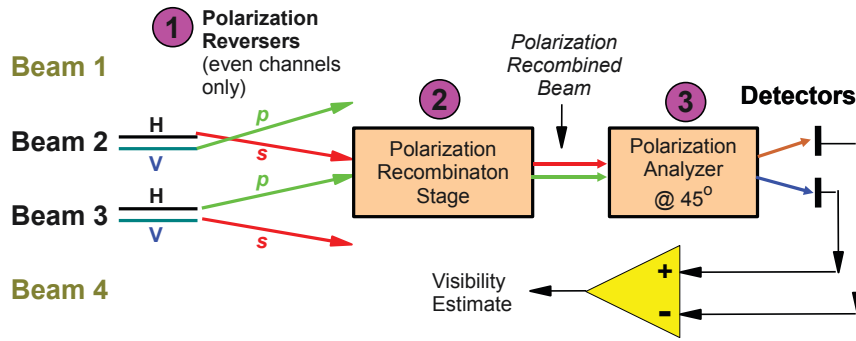


Figure 5. The essential optical components of the polarization-based collimated beam combiner. Input beams from a number of telescopes are combined pairwise, such as the horizontal polarizations of beams from telescopes 2 and 3. The beam from telescope 2 (and other even beams) has its polarizations reversed (1) so that its horizontal polarization is now vertically (*s*) polarized entering the polarization recombining stage (2). The resulting “polarization recombined beam” contains waves from telescopes 2 (*s* polarized) and telescopes 3 (*p* polarized) in the same spatial mode. The polarization analyzer (3) separates the polarizations as seen at $+45^\circ$ and -45° sending them to separate detectors (or pixels of a detector array). Subtracting these detector readings yields an estimate of visibility (at a particular phase) unaffected by photometric crosstalk.

3.3 Polarization Reversing

Since the polarization recombining stage implements an interferometric channel using the *s* polarization of beam M and the *p* polarization of beam $M + 1$, it is required that these two waves are traced back to the same polarization on the sky in order for those electric fields to be correlated. This is accomplished by reversing the *s* and *p* polarization of every second beam (whose M is even, in the diagrams shown). Thus in fig. 5, beam 2 has its horizontal polarization transformed to vertical (*s*) so that it can be paired with the horizontal polarization of beam 3 which stays horizontally (*p*) polarized. Meanwhile, the original vertical polarization of beam 2 becomes horizontally (*p*) polarized in order to be paired with the horizontal polarization of beam 1 (not shown).

One implementation of the polarization reversing function involves in the even beams a system of 3 mirrors each bending the beam by 90° in a principal plane, whereas the odd beams are similarly deflected but using a system with 2 mirrors. This results in a 90° relative *rotation* between the even and odd beams, and thus between their polarizations. A second implementation achieves the polarization reversal through the insertion of a $\lambda/2$ retarder oriented at 45° in the even beams. This is the solution used in the NFT, employing an almost achromatic half wave plate in the even beams and an identical plate in the odd beams but at 0° in order to compensate for its material dispersion. One advantage of this scheme is that the plates can all be rotated by 45° in order to switch the polarization reversing function to the odd beams. This could be used, for instance, in investigating the possibility of systematic errors (although considered unlikely) arising in the case of a polarized stellar source.

3.4 The Polarization Recombining Stage (PRS)

The polarization recombining stage (PRS) performs the key function of creating the *polarization recombined beams* from the incoming telescope beams. In this stage, the horizontal (*p*) and vertical (*s*) polarization components from each telescope beam are separated and reconfigured as diagrammed in fig. 6A. The *s* polarization of telescope beam M is reflected twice, placing it in the same spatial mode as the *p* polarization of telescope beam $M + 1$. The resulting polarization recombined beam has various desirable properties that we will take advantage of in the subsequent optical train. It is not truly an “interfered” beam since the incoming waves are still in separate electromagnetic modes.

But the incidence of the two polarizations on the final polarizing beamsplitter surface closely resembles the operation of the basic two-phase on-axis beam combiner of fig. 1. In particular, this is the point at which the crucial alignment of the incoming beams is required, with an offset in the directions of the incoming beams (or a lateral offset in the beams’ pupils) leading to reduced fringe contrast. Therefore (again referring to fig. 6A) it is required that telescope beam 3 be directed so as to match the spatial mode of the *s* polarization of telescope beam 2 following two reflections. Given that alignment for beam 3, then beam 4 must likewise be aligned to match its spatial mode with the reflected *s* polarization of telescope beam

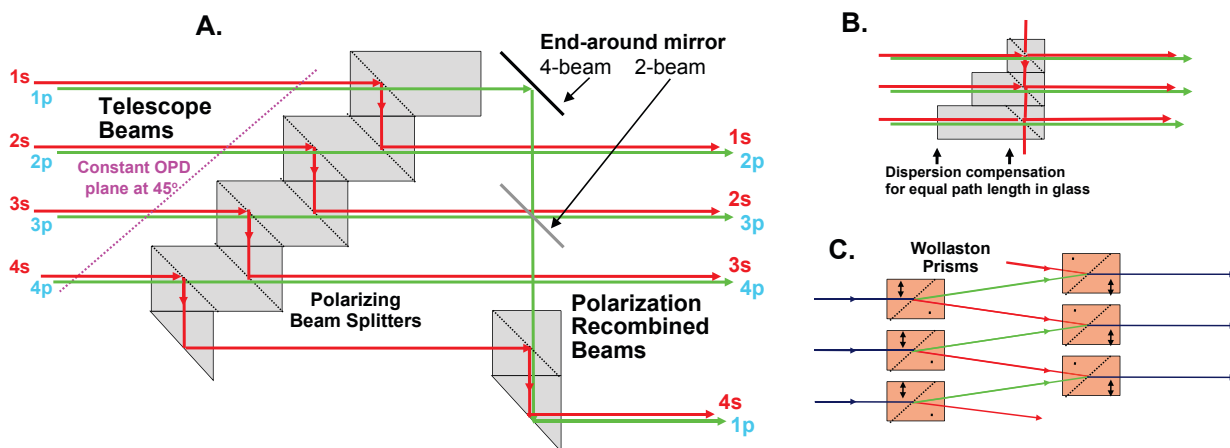


Figure 6. **A.** The *polarization recombining stage* (PRS) implemented using polarizing beam splitting surfaces in a monolithic prism assembly. The *s* (vertical) polarization of telescope beam M is shifted down one level to be placed in the same spatial mode as the *p* polarization of telescope beam $M + 1$. **B.** An implementation using only one polarizing beamsplitter per telescope beam, both for separating and recombining the *s* waves. **C.** Using Wollaston prisms, the two polarizations are directed at an upwards angle and at a downwards angle until the beams from adjacent telescopes overlap. The second set of Wollaston prisms, operating in reverse, restores the original directions of the two waves, now superimposed, from the two telescopes.

3, and so on. In addition to the spatial alignment, this is also the point at which the relative OPD (phase shift) between the two beams is established.

However unlike the simple on-axis beam combiner of fig. 1, there are not two distinct exiting beams, but rather the radiation all leaves in one direction. And moreover the light from the two telescopes has not exactly undergone interference, for that will not truly occur until the recombined beam enters the polarization analyzer, the next essential stage of the interferometric system.

In between, we have a *polarization recombined beam* having unique properties. Since the two incoming beams are in separate (orthogonal) polarization modes, their radiation is not actually interfered and could still be separated from each other. However as a single “beam,” these two components that soon *will* be interfered can proceed along the same optical path, through the same optical components, and remain, in a sense, “bound” to each other both spatially and in their relative phase. That property is only violated by birefringent or other polarization-sensitive optical devices, which are generally prohibited from the subsequent optical system (except where specifically intended to act on the two waves differently, such as in the eventual polarization analyzer). *This unique character of the polarization recombined beams enables a variety of optical functions to be applied identically to the two component waves*, as will be illustrated in some of the following sections and in the design of the NFT.

3.4.1 PRS implementations

Figure 6A shows a practical implementation of the PRS function using cube polarizing beamsplitters in a compact arrangement, that is, with these elements implemented in a monolithic prism block. This avoids surface reflection losses at the interfaces between discrete elements, and also has the advantage of relative mechanical stability both with respect to beam alignment and also OPD referencing. This configuration has been selected for the NFT design, presented in sect. 4.

It should be mentioned that modest departures of the polarizing beamsplitters from the ideal (i.e. some reflection of the *p* polarization or transmission of the *s* wave) do not generate significant crosstalk or spurious response in the eventual interferometric detection beyond the simple loss of throughput entailed.

A similar configuration is shown in fig. 6B, however this one requires extra dispersion compensating glass sections as shown. The implementation in fig. 6C uses Wollaston prisms instead of polarizing beamsplitters in a rather different geometry.

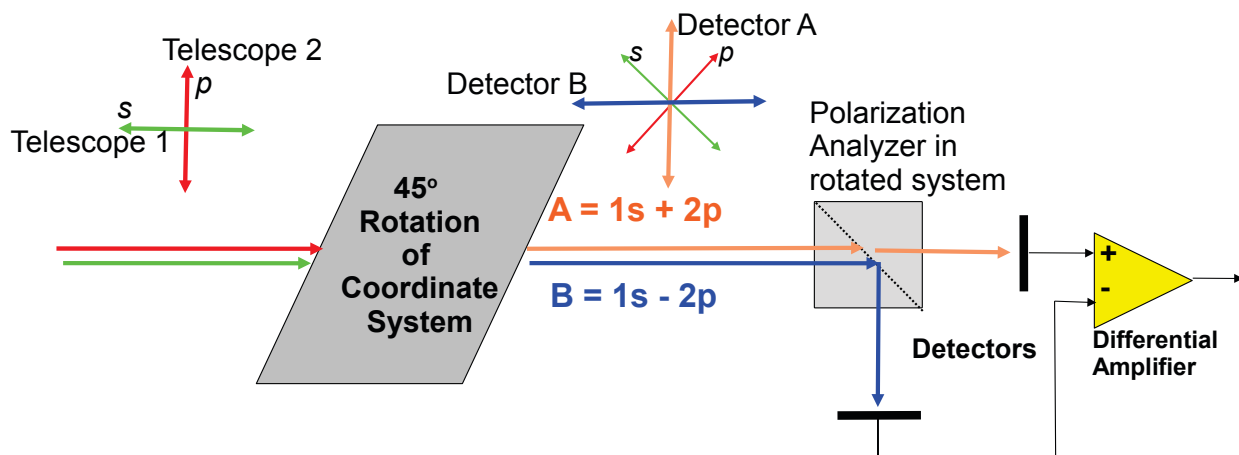


Figure 7. Interference is produced and detected when a *polarization recombined beam* (sect. 3.4) is separated not into its original constituents (telescopes 1 and 2) in the *s* and *p* polarizations as diagrammed at the left, but rather at 45°. The “*A*” and “*B*” polarizations defined by that rotated system will each contain half of the power of each incoming beam, but with their electric fields added in opposite phase. Differencing those detector readings obtains an interferometric signal in which the individual “photometric” levels of the original beams has been cancelled

3.4.2 End-around combiner

The PRS as explained so far, combines one half of the radiation between each adjacent beam $m = 1$ to M . That leaves the *p* wave of beam 1 and the *s* wave of beam M unused. As long as the number of telescopes being interfered, M , is even, then it will be possible to have an additional path which similarly combines those two waves in an extra *end-around* polarization recombined beam. This provides one more baseline on which interference is measured. Although redundant, this will lower the fringe tracking error, especially on the baseline it directly monitors.

One implementation of the end-around channel as used with the proposed NFT is shown in fig. 6A where the *s* wave of telescope 4 is paired with the *p* wave of telescope 1 in the lower-right polarizing beamsplitter. When only 2 telescopes are in use (numbered 3 and 4 in this figure) the end-around mirror is shifted to the other position indicated so that it deflects the *p* wave of telescope 3 instead. The end-around mirror also includes an additional alignment capability and (short range) OPD adjustment using a piezo actuator.

One interesting use for the end-around channel is for “baseline bootstrapping.” In this application,⁶ the phases between telescopes 1 and 2, 2 and 3, etc. are locked (or measured) and used to phase up a particularly long baseline between telescopes M and telescope 1 in the end-around channel. Even if the latter baseline is very over-resolved, its visibility can be now be measured anyway using coherent integration⁷ with the other $M - 1$ baselines providing a phase reference. This could include scanning the end-around fringe with respect to delay using the OPD actuator on the end-around mirror. Another special use for the end-around channel is in measuring the *closure phase* over the baseline polygon comprising the M telescope locations. In this case the end-around channel’s OPD actuator is first locked during a calibrator observation, and then a possible shift in the closure phase is observed after the telescopes are repointed toward an interesting science target.

3.5 Polarization Analyzer

The final essential stage of the polarization-based collimated beam combiner is the *polarization analyzer* which takes a polarization recombined beam that had been produced in the PRS, and actually generates interference between its two components (originating from the two telescopes). In the basic two-phase detection configuration, this consists of nothing more than a polarizing beam splitter with its polarization axes rotated exactly 45° with respect to the polarization directions defined by the PRS. The principle of operation is depicted in fig. 7.

The two incoming waves from telescopes 1 and 2, combined in the same spatial mode by the PRS, have their electric fields polarized in directions that we will take to be vertical and horizontal respectively. Now the reader is simply requested

to look at these polarized components following a 45° rotation about the beam axis. In this rotated system we shall call the new horizontal and vertical polarizations (actually at $+45^\circ$ and -45°) the “ A ” and “ B ” polarizations[¶] as depicted in fig. 7. The polarization analyzer, implemented as a polarizing beam splitter cube in that diagram, is oriented along the A and B axes. It is easy to see that the A polarization component (the component that would be transmitted by such an analyzer) consists of one half the power ($\sqrt{2}/2$ of the amplitude) of each of the telescopes’ beams added in some particular phase. The B polarization would likewise be composed of half the power of each telescope beam but added in the *opposite* phase.

Without loss of generality let us take the reference planes for E_1 and E_2 , the electric fields of the beams from telescopes 1 and 2, to be in the same phase (at some optical frequency) as the reference plane of E_A and E_B . Thus:

$$E_A = \frac{\sqrt{2}}{2}(E_1 + E_2) \quad , \quad E_B = \frac{\sqrt{2}}{2}(E_1 - E_2) \quad . \quad (12)$$

The A and B components resolved by the polarization analyzer each have optical intensities (powers) given by $I = |E|^2$. When those detector readings are subtracted, as depicted in fig. 7, we obtain:

$$\Delta I = I_A - I_B = \frac{1}{2}(|E_1|^2 + |E_2|^2 + 2\mathcal{R}e\{E_1 E_2^*\}) - \frac{1}{2}(|E_1|^2 + |E_2|^2 - 2\mathcal{R}e\{E_1 E_2^*\}) = 2\mathcal{R}e\{E_1 E_2^*\} \quad . \quad (13)$$

So this exactly implements the balanced beam combiner we had been seeking, with detector outputs which, when subtracted, yield a result only having to do with the correlation between the telescopes’ electric fields, and no sensitivity whatsoever to the telescopes’ individual photometric contributions $I_1 = |E_1|^2$ or $I_2 = |E_2|^2$. In other words, we have achieved $\gamma = 0$ (as defined in sect. 2.2.1).

It should be noted that the polarizing beam splitter implementing the analyzer in fig. 7 need not be ideal in its splitting of the two polarizations. A practical analyzer with some crosstalk between the polarizations in the outputs can still achieve photometric symmetry with a slight reduction of efficiency. However with respect to wavelength it would be important for its characteristics to remain unchanged, which would be difficult in a wideband application. A superior hardware solution (used in the NFT, see sect. 4.2.7) is to use a Wollaston prism to separate the polarizations. With such a device there is essentially zero crosstalk between the polarizations, putting such worries to rest.

3.5.1 Adjusting the analyzer for photometric symmetry

Note that this success has nothing to do with the precision of any coatings or optical components, but *only* to orienting the analyzer angle at exactly 45° . In practice it also requires that the gain of the two detectors be equalized through the application of a scale factor. We know that the *net* scale factor should be unity but because of gain discrepancies will actually be k before the balance is properly adjusted. We know that the rotation angle ψ should be 45° but cannot measure that with sufficient precision just using a protractor. These must be adjusted empirically.

That can be done by considering the effect of departures from $k = 1$ and $\psi = 45^\circ$ in which case it can be shown that the response after subtraction of the detector readings, ΔI in (13) becomes:

$$\Delta I = I_A - kI_B = I_1(\sin^2(\psi) - k\cos^2(\psi)) + I_2(\cos^2(\psi) - k\sin^2(\psi)) + (1 + k)\sin(2\psi)\mathcal{R}e\{E_1 E_2^*\} \quad . \quad (14)$$

Now we could, for instance, alternately open only one shutter and then the other shutter blocking the telescope beams to obtain $\Delta I_{(1)}$ and $\Delta I_{(2)}$. If $\Delta I_{(1)} > 0$ while $\Delta I_{(2)} < 0$ (or visa-versa) it would indicate that ψ must be greater (or less) than 45° and we could adjust the mechanical angle in the right direction. On the other hand, if both $\Delta I_{(1)}$ and $\Delta I_{(2)}$ were greater (or less) than 0, it would indicate that the net gain k was less (or greater) than unity. By iterating between adjustments of k and ψ on this basis we quickly converge toward the precise achievement of $k = 1$ and $\psi = 45^\circ$, and consequently $\gamma = 0$ as has been accomplished in the laboratory setting (sect. 5.1).

[¶]Note that the terminology used in this discussion is totally divorced from that of sect. 2. In particular, A and B have nothing to do with those designations in sect. 2 where they referred to two telescopes. In this section the telescope beams are rather numbered 1 and 2.

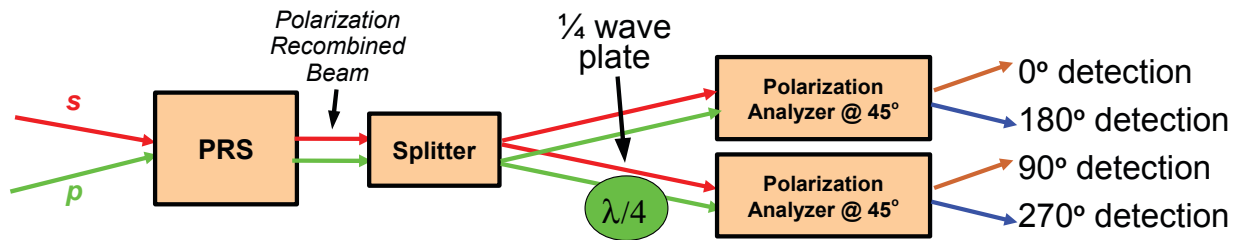


Figure 8. For quadrature detection, the polarization recombined beam is split and the two are run into separate polarization analyzers in the direction of the A and B polarizations as before. En route to the analyzer, one of the beams passes through a quarter wave plate oriented horizontally (delaying the p wave) so that the interference from its two outputs are in phase quadrature with respect to the other two outputs.

3.6 Implementations for Quadrature and Three-Phase Detection

While the 2-phase detection depicted in fig. 5 has been proposed for the NFT (sect. 4.3), a visibility measuring interferometer would require detection of both components of the complex visibility. This could of course be accomplished through OPD modulation as in sect. 2.4, but now only requiring sampling at 0° and 90° (rather than 4 quadrature phases) taking advantage of the hardware's photometric symmetry but the result would still be subject to time skew (though to a lesser degree). A more elegant solution is optical quadrature detection.⁸ Using a polarization recombined beam, optical quadrature (or 3-phase) detection can be effected by using slightly different optics in place of the simple polarization analyzer shown in fig. 7.

One example is shown in fig. 8. The configuration is almost the same as that of fig. 7 with a polarizing beam splitter (for instance) used to obtain the polarization in the A and B directions. But on its way to the analyzer, half (or a different proportion) of the light is split off in the horizontal direction using a beamsplitter which is nominally non-polarizing (however if the reflection coefficients for the horizontal and vertical polarizations are different there will be no serious consequences). That light split off from the beam then goes through a quarter wave plate oriented horizontally so that it delays the p wave from one telescope with respect to the s wave from the other telescope. This *altered* polarization recombined beam then enters a similar polarization analyzer which detects interference at a 90° phase with respect to the other analyzer. The four detected intensities then correspond to quadrature detection of the interference between the two telescopes, and track each other in every other respect. Note that if the earlier beamsplitter introduces a phase difference between the s and p polarization components in its reflection (or, less likely, transmission) this must be taken into account by altering the retardance of the quarter wave plate by a similar amount.

Obtaining slightly improved performance when limited by detector noise, it is also possible to implement (not diagrammed) 3-phase detection of interference by a different modification to the analyzer section of the polarization-based collimated beam combiner. A custom beam splitting surface is introduced into the polarization recombined beam at a rotational tilt of 45° (in other words, in the direction of the A polarization) which is detected. This beam splitter passes its p polarization (what we have called the A polarization) with no reflection, whereas the power in its s polarization (what we have called the B polarization) is $2/3$ reflected and $1/3$ transmitted. The transmitted wave then passes through a quarter wave plate oriented horizontally, as in the quadrature configuration, before being split into its A and B polarizations as before and detected. The three intensities measure interference at phases of 180° , $+60^\circ$ and -60° respectively. It is possible to use a beamsplitter which has a reflection coefficient of its s polarization other than $2/3$; such a partially polarizing beamsplitter can be easily implemented with a dielectric surface at the brewster angle, for instance. Then the detected phase angles will no longer be spaced by 60° , but will still provide the real and imaginary parts of the visibility using appropriate linear combinations of the detections (and retaining photometric symmetry). However there will be a SNR disparity between the real and imaginary parts (which may in some cases be desired, or might be a nuisance). One example based on this concept is the nulling configuration mentioned in the following section.

3.7 Additional Functions Applied to the Polarization Recombined Beams

As has been pointed out, a polarization recombined beam has rather unique properties in that it contains both interfering waves before they have actually interfered, but in a way that both waves can be manipulated together and identically by

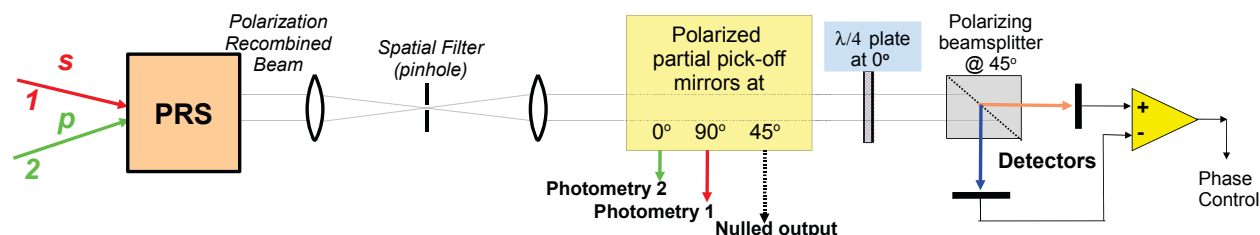


Figure 9. A diagram of the polarization-based collimated beam combiner implementation of a nulling interferometer⁹ with outputs to drive a control system. Among the elements illustrated in this example are common spatial filtering using a single optical element for both waves (in the *s* and *p* polarizations), photometric monitoring of the individual waves *after* spatial filtering, and 3-phase interferometric detection (sect. 3.6). The “polarized partial pick-off mirrors” can be implemented (for instance) as plates at the brewster angle which transmit all of the *p* wave but reflect some of the *s* wave, and are at rotation angles of 0° and 90° for the photometric monitoring (probably using a lower reflection) and 45° for the nulled output (using a high reflection but less than unity so that the following phase detector can still function).

optical elements (with the exception of some birefringent components which would cause undesired crosstalk between those waves). This may be desired either as a convenience or sometimes to insure identical treatment between the two waves. For instance, in the proposed NFT design, a common prism (sect. 4.2.6) is used by the polarization recombined beams (with all 4 or 6 of them passing through it at slightly different angles) in order to disperse the waves spectrally. Since the *s* and *p* polarizations are treated identically by such a prism, it has no effect on the interference at each wavelength and acts no differently than if it had been situated after the analyzer.

One interesting configuration is the use of a spatial filter which is common mode to both input beams. This solves a dilemma concerning the placement of a spatial filter. In a bulk optics configuration, individual spatial filters before the beam combiner need to be precisely adjusted to match their output wavefronts in order to interfere efficiently, and any beam degradation between the spatial filter and the beam combination also causes the wavefront match to suffer. On the other hand, spatial filtering *after* beam combination (as in PRIMA⁸) makes it difficult for complementary outputs to be looking at the same spatial mode, thus compromising photometric symmetry even with a beam combiner having $\gamma = 0$. What’s more, the photometric contribution of the individual input beams in the selected spatial mode cannot be monitored since they have already been interfered. However the polarization-based collimated beam combiner can overcome all of these difficulties and accommodate a single spatial filter (using bulk optics) which operates identically on both telescope beams at a point within the polarization recombined beam.

One problem where this solution was of particular value was in the implementation of a nulling interferometer for ground-based use with sufficient outputs for control of the relative phase and amplitudes of the input waves, while producing a nulled output employing sufficient spatial filtering for the rejection of unwanted modes that wouldn’t be simultaneously nulled. While the limited value of nulling in a ground-based instrument was pointed out by the author, the proposed configuration⁹ illustrates the utility of the polarization-based collimated beam combiner in implementing such a spatial filter as well as an example of 3-phase interferometric detection (with one of the phases being the nulled output) as described in sect. 3.6. The polarization combined beam in fig. 9 passes through a pinhole spatial filter, for instance, followed by three partially reflecting polarizing reflectors, two of which pick off samples of the original telescope beams in order to drive variable attenuators in order to match the amplitudes of those waves *following* spatial filtering. The third is the actual nulled output which combines the two polarizations at 180°. Additional radiation from both input beams is analyzed at 45°, as before, after passing through a quarter wave plate, so that the difference between those two detector readings is a measure of the phase offset between the beams which is driven to zero in order to maximize the nulling. Not shown, but of importance for accommodating an atmospheric path in a wideband system, is spectral dispersion (with a resolution of at least 2) of the phase monitoring detections. A difference between the phase balance at two different wavelengths is used for realtime control (see sect. 4.3.4) of an upstream dispersion compensator so that fluctuations in atmospheric water vapor are tracked, enabling nulling over a wide bandwidth.

4. THE NOVA FRINGE TRACKER (NFT)

The NOVA Fringe Tracker (NFT) is a proposed instrument for the VLTI based on the polarization-based collimated beam combiner concept. It is the result of one of three phase A studies conducted under contract to ESO towards the realization of a second-generation fringe tracking facility for the VLTI. The study was conducted by a consortium led by NOVA^{||} which included Sterrewacht Leiden (Leiden Observatory), Technische Universiteit Delft (Delft Institute of Technology), and TNO Science and Industry.

ESO requested a concept for fringe tracking 4 or (probably in a future upgrade) 6 telescopes using either the science star or an off-axis reference star with interferometric detection in the K and/or H band. Such a fringe tracker would track the VLTI delay lines with a residual OPD error of no more than 100nm (with a sufficiently bright reference star) enabling other VLTI instruments to observe dimmer targets with the benefit of atmospheric OPD stabilization enabling long coherent exposures, and also for the purpose of phase referenced imaging. In addition to stabilizing the telescopes' relative phases, it would need to measure changes in longitudinal dispersion (or "group delay"). The fringe tracker must accommodate baselines on a possibly resolved reference star having unequal visibilities. It was desired for it to be able to cophase a UT (8 meter telescope) with an AT (1.8 meter telescope). It would require an update rate of up to 2KHz in order to track high frequency vibrations currently affecting the VLTI optical train in addition to the atmosphere.

4.1 Design Choices

Early in the study it was determined that the design goals could be better met using a bulk optics design rather than employing integrated optics technology. After examining various possibilities it was decided to perform pairwise interference (rather than schemes combining all 4 or 6 telescopes' light together) using collimated (on-axis) beam combination thus illuminating as few detector pixels as possible in order to reduce the effect of detection noise. We will not detail the numerous trade-offs considered in reaching these decisions.

The concept of the polarization-based collimated beam combiner soon emerged and was further elaborated as an efficient implementation of these early decisions in order to meet the performance goals as set forth by ESO. The ability of this topology to realize a high degree of photometric symmetry was further appreciated as a solution to the design challenges involved in the detection of small visibilities or when employing an AT with a UT beam (20 times brighter) while avoiding the problems encountered during the commissioning of PRIMA (the first-generation fringe tracker) that could be attributed to photometric asymmetry in its beam combination**.

4.1.1 Wavelength coverage

Use of bulk optics technology allowed us to address both desires of ESO for fringe tracking in the K and H bands without requiring additional hardware (beyond specifying the optical components over a wider bandwidth), and for these to be operational concurrently and employing a common beam combining (actually PRS) surface so that their relative phase relationship is absolutely fixed. The final design specifies operation from the K band edge (2.45 μm) all the way to 1.2 μm , although it is not certain whether the VLTI will be capable of performing useful interferometry at such short wavelengths (J band).

4.1.2 Spectral resolution

It was recognized that the topology enabled separation of the wavelength components using a common prism for all beams as they were approaching the camera lens. This could be replaced with one having a greater or lesser degree of spectral dispersion, and the spectral resolution could also be varied somewhat by translating the axial position of that prism using a zero-deviation design. It would be possible to allow for spectral resolutions which project the entire spectrum (1.2 - 2.45 μm) on between 4 and 7 pixels of the detector, and, using our prism design which is more dispersive in the K band than

^{||}NOVA, the Nederlandse Onderzoekschool Voor Astronomie, is a federation of the leading astronomical institutes in The Netherlands.

^{**}Although not a matter of attention at the time of design, the polarization-based collimated beam combiner also avoids the problems affecting PRIMA due to interaction between birefringence encountered in the VLTI optical train and the intended polarization characteristics of the incoming beam. Since for each interferometric channel (baseline combination) the PRS collapses one polarization from the incoming telescope beam, that wave following the PRS has no memory of its previous polarization characteristics. The net effect of any polarization transforms up to the PRS is limited to a reduction (expected to be less than 1%) in the correlated power, or phase shifts whose effects are absorbed into the routine OPD calibration (referencing to the science instrument).

at short wavelengths, the K band will be divided between at least two pixels even at the lowest resolution. That enables the detection of dispersion fluctuations (and fringe jumps) using K band radiation alone. No point was seen in a greater spectral resolution (than 7 pixels for the entire bandwidth) although the introduction of a higher (or lower) resolution prism in that position is not ruled out.

4.1.3 Spatial filtering

In order to achieve the best limiting sensitivity, use of spatial filtering was not adopted in the proposed design. Rather there is simply a mask at the intermediate focus with holes whose radii are about $3\lambda/D$ (at the longest wavelength) in order to reduce noise from the thermal background while accepting over 95% of the starlight. While spatial filtering has become essential in the design of interferometers for achieving a high accuracy in visibility measurement, its use in the present fringe tracking application is of limited value and acts to reduce the throughput and thus limiting sensitivity in the case of detector noise limited operation (on the dimmest sources). Indeed there may be cases where the introduction of a spatial filter will increase the SNR, namely when the strehl ratio is very poor (so that most of the light would be incoherent and rejected by the filter) and the detection is photon noise limited (so that rejection of that incoherent light is more beneficial to the SNR than is the inadvertent reduction of signal power). It is not excluded that the implementation of the NFT design *could* include spatial filtering capabilities through shifting the intermediate focus mask between two (or more) sets of holes which may be smaller. However the primary design goal of maximizing the performance at low light levels was best achieved with the oversized mask holes.

4.1.4 Method of detecting interference

As has been emphasized, the elimination of photometric asymmetry in the beam combining optics relaxes the usual requirement of OPD dithering (sect. 2.4) for the purpose of interferometric detection utilized in *most* astronomical interferometers employing collimated beam combination (to which PRIMA, the first-generation VLTI fringe tracker also was an exception). This allows for an estimate of the visibility phase to be based on a single detector readout with no further latency or crosstalk concerns.

A second, again unusual, design decision was to totally dispense with the detection of the complex visibility either through quadrature optical detection (sect. 3.6) or OPD dithering of 90° . Since only the imaginary part of the visibility responds to a phase offset when near zero phase, we elected not to waste any of the precious starlight on measuring the real part of the visibility and thus gain up to a doubling of the sensitivity. This means that the gain coefficient of the imaginary part of the visibility in the feedback system isn't based on a frame-by-frame measurement of the (fluctuating) visibility magnitude, but only an off-line estimate of what that visibility magnitude is on the average. In a closed loop system, this approximation is not a serious impediment to fringe tracking especially in situations approaching the limiting sensitivity where the SNR in an individual detector frame is marginal.

4.2 Detailed Design and Proposed Layout

Figures 10 and 11 show the proposed implementation of the NFT concept as it has been submitted to ESO. We will briefly point out the elements of the system, whose design choices and principles of operation have mostly been discussed.

4.2.1 Dispersion compensator

Not shown in fig. 10 is a set of adjustable elements which introduce longitudinal dispersion into the incoming beams so as to mostly cancel the large dispersion due to the unbalanced air columns in the VLTI optical paths and delay lines,¹⁰ and secondly to track fluctuations (on the order of seconds) in the atmospheric paths through differing amounts of water vapor.¹¹ Although it is a requirement for the operation of the NFT in particular, it has been proposed that this compensation would best be placed *upstream* in the VLTI laboratory so that its correction is applied to the light reaching *all* near infrared instruments, thus including the science instrument whose OPD would be controlled by the fringe tracker. Because of the demanding requirements of the NFT design, to simultaneously perform interference near 0° across the near IR spectrum, it will require the movement of wedges from two materials having different dispersion characteristics. Simply removing the higher order component of dispersion (or *group velocity dispersion*), as might suffice for a quadrature measuring interferometer, would not be sufficient in this more demanding configuration.

It should also be mentioned that a certain amount of dispersion is required for the operation of the polarization-based collimated beam combiner itself across a wide bandwidth, since the detection of interference at what we call "zero phase"

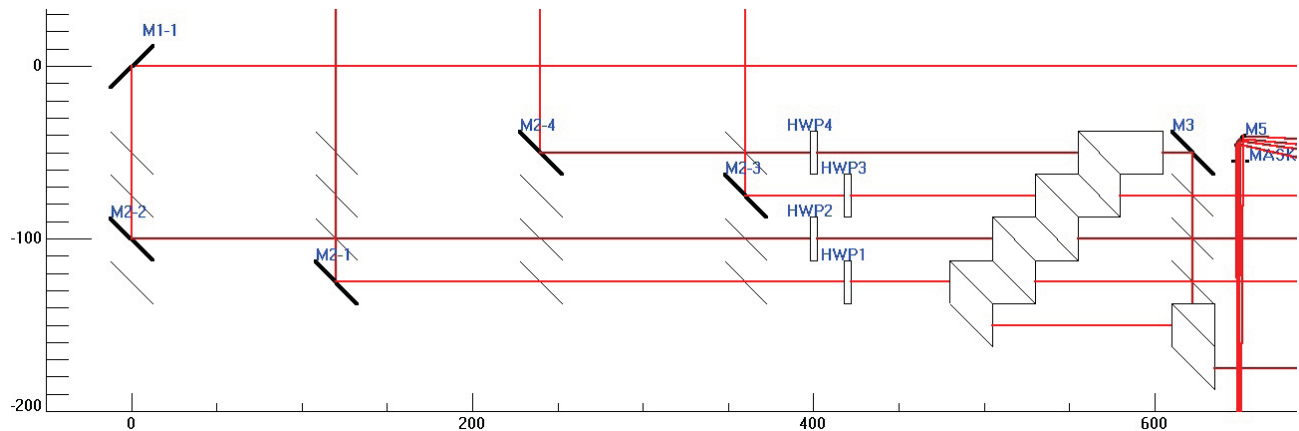


Figure 10. One proposed layout of the NFT (for 4 telescope beams) showing the front end up to the PRS, indicating the center axes of the beam paths. Light from up to 4 VLT beams (one shown) is picked off by M1. Those nominally 18mm diameter beams are selected by the M2 (switchyard mirrors) sending them towards the *polarization recombining stage* (PRS), consisting mainly of a giant prism block with internal polarizing beam splitting surfaces, and an *end-around combiner* using the beam picked off by M3 (on a translation stage). Alternate positions for M2 and M3 are indicated. Drawing is to scale with axes dimensions in millimeters.

actually requires phase quadrature between the incoming beams. This 90° of first-order dispersion would be required on every second beam and would be achieved without being noticed in the process of adjusting the dispersion compensator to cancel the (much larger) atmospheric dispersion. Also, depending on the design of the coating used, there may be dispersion introduced in the s waves reflected twice by the polarizing beam combiner surfaces in the PRS. Although this could (and would) be corrected using the dispersion compensator, that is not a proper solution because the corrections would accumulate and confound the end-around channel. Such dispersion would better be corrected at the PRS itself.

4.2.2 Front end

The front end of the NFT consists of beam pick-off mirrors (or dichroic mirrors) labelled M1 (one of which is shown in fig. 10) followed by a local *switchyard* capable of routing any VLT beam to any interferometric channel. This is accomplished through multiple flip-up M2 mirrors, or one M2 mirror for each interferometric channel each on a translation stage so it can be moved into one of the positions indicated in fig. 10). Both the M1 and M2 mirrors have fine alignment motors so that the beam directions, and, less importantly, any lateral pupil shift, can be equalized in between the incoming beams to maximize the interferometric contrast. The M2's also include a short range ($\sim 2\text{mm}$) OPD equalization capability as well as a very short range ($\sim 50\mu\text{m}$) but very fast ($\sim .1\text{ms}$) piezo actuator with internal metrology for precise OPD compensation, notably in control mode 2 and in scanning mode 3 (sect. 4.3). For the end-around channel, these functions are duplicated on the M3 mirror (fig. 10).

The proposed NFT design uses an almost achromatic half wave plate in each telescope beam, which in the even channels are oriented at 45° in order to effect the required polarization reversing function discussed in sect. 3.3.

4.2.3 PRS

The polarization recombining stage (PRS) of the NFT follows the configuration already presented in sect. 3.4.1 closely following the layout of fig. 6A. The hardware is sized for a beam separation of at least 25mm (as shown in fig. 10) in order to easily accommodate a nominal beam diameter of 18mm (insuring a clear aperture of 20mm at all points). The end-around channel (sect. 3.4.2) is also implemented as in fig. 6A with M3 in fig. 10 on a translation stage which can shift it to positions for interfering 2, 4, or 6 telescopes (in a future upgrade), or in other positions to implement certain utility functions.

4.2.4 Intermediate focus

The 18mm collimated polarization recombined beams emerging from the PRS are folded and focussed using the M4 concave mirrors towards an *intermediate focus* after 600mm, after which they travel another 300mm where they have diverged to 9mm diameter where they are accepted by the 10mm camera lens. In order to straighten their paths toward

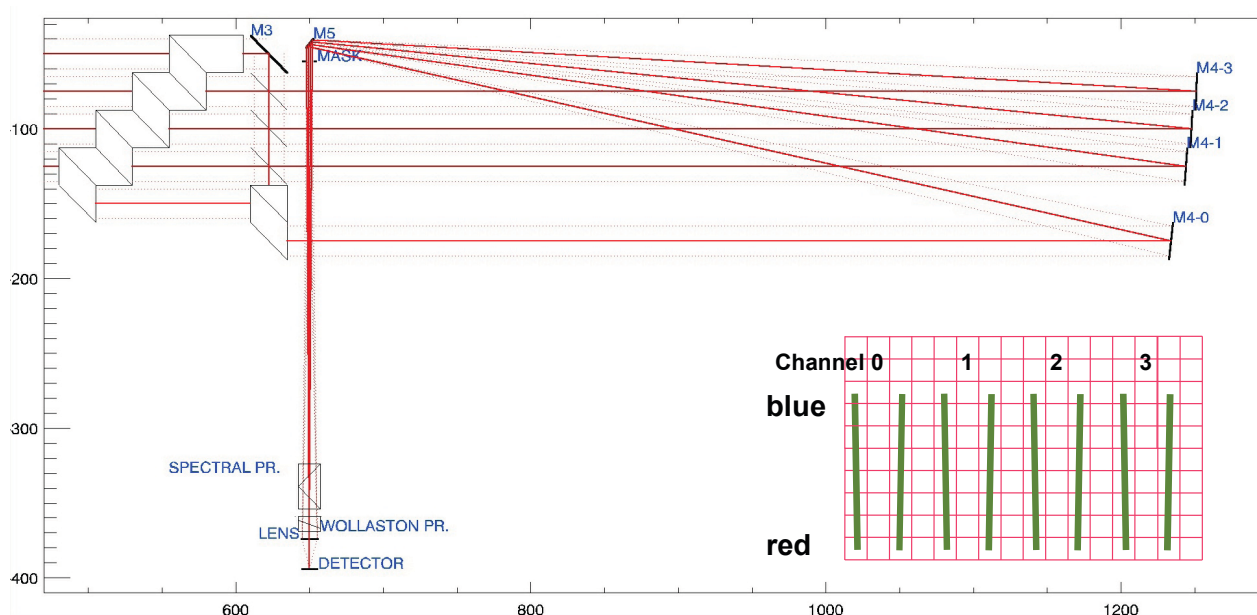


Figure 11. Continuation of fig. 10 including the PRS to show the back end of one proposed layout of the NFT. The collimated beams from the PRS are focussed by the M4 slow ($f/30$) off-axis paraboloids toward an intermediate focus where each passes through the hole of a refrigerated mask for the purpose of reducing thermal background radiation. Right before the mask, the still converging beams are individually redirected so that they eventually overlap at the camera lens. Prior to focussing, the beam directions are dispersed by a spectral prism in Y and the two interferometric phases separated by a Wollaston prism in X, to be focussed onto the PICNIC detector inside a cryostat. The resulting image on the detector consists of 8 short spectra centered on pixels in a pattern such as shown in the lower right inset.

the camera, there are tiny mirrors (or a mirror block) labelled M5 in fig. 11. The M4 and M5 alignments insure that the intermediate focus locations are centered in the mask holes and are focussed on the detector in a linear pattern matching the pixel pitch ($40\mu\text{m}$) of the detector. A pointing mechanism at the camera centers these spots within their intended pixels.

4.2.5 Mask

At the point of the intermediate focus, each polarization recombined beam passes through the middle of a hole in a mask whose purpose is *not* spatial filtering (as discussed in sect. 4.1.3) but only to reduce the effect of the warm background which radiates significantly around the longest wavelengths that the detector is sensitive to ($2.5\mu\text{m}$). Due to the spectral prism even pixels which receive the star's light at a shorter wavelength would also be sensitive to $2.4\mu\text{m}$ radiation from a different point on the sky, which is thus blocked by the mask. The mask holes (about $.5\text{mm}$ in diameter according to the proposed design) correspond to just under 1 pixel in size when referred to the detector. The mask itself is blackened and refrigerated in order to prevent being a radiation source itself.

4.2.6 Spectral prism

The light from all 4 polarization combined beams, before being split by the analyzer (sect. 4.2.7) into its interferometric components, are spectrally dispersed in the Y direction using a single zero-deviation chromatic prism composed of calcium fluoride (low dispersion) and fused quartz (higher dispersion) sections. In order to reconfigure the spectral resolution it is foreseen for there to be two prisms of this sort with different internal angles available and also for the resolution to be continuously varied over a certain range by shifting the prism longitudinally (sect. 4.1.2). In conjunction with fine pointing adjustments by the camera, the spectral resolution can then be reconfigured with respect to pixel boundaries with between 4 and 7 pixels illuminated by radiation from the $1.2 - 2.45\mu\text{m}$ full bandwidth of the instrument.

4.2.7 Wollaston prism (polarization analyzer)

Just before (or possibly inside) the cryostat is placed a low deviation Wollaston prism composed of two sections of quartz with its optic axis at + and -45° with respect to the horizontal (thus in the “A” and “B” directions in fig. 7). The surface between the two sections is vertically oriented so that the deviation between the two polarizations is expressed horizontally. Thus each polarization combined beam produces two short spectra side by side centered on detector pixels representing opposite interferometric phases. One plausible layout of the 8 spectra from 4 interferometric channels is diagrammed in fig. 11. There is a very small tilt (about 1°) in the spectra which is due to spectral dependence of the birefringence of the quartz; this changes the centering of the spectra within a pixel by only a small amount and presents no concern.

4.2.8 Camera and detector

The beams enter a cryostat and are focussed using a 2 (or more) element achromat having (in the specific design proposed) a focal length of 20mm and a diameter of 10mm (thus f/2) for a nominal beam diameter at that point of 9mm. The PICNIC detector chip was one of two offered by ESO who will implement that portion of the instrument, and was preferred especially due to its large pixel size: 40μm square. The layout of the spectra with respect to the pixel grid is depicted in fig. 11. The lens’ PSF is close to the diffraction limit at the longest wavelengths, and the width of the spectra in that diagram correctly represents the width (10 μm, about 1/4 of a pixel) within which 90% of the optical power is confined. The detector is required to be read out at speeds up to 2KHz in order to accommodate vibrations affecting the VLTI optical train, however operation at much lower frame rates, if possible, would be desired in the case of the dimmest targets which otherwise would be overwhelmed by readout noise. The PICNIC chip has obtained readout noise levels of about 20 detected photons rms per pixel per exposure, and is sensitive to wavelengths shorter than 2.5μm, thus including the entire K band and excluding most thermal radiation.

4.3 Operating Modes and Feedback Loop

The primary operating modes as a fringe tracker involve measuring interference using 2-phase detection which corresponds to the imaginary part of the visibility with respect to the intended operating point. In other words the net effect of the dispersion compensator along with all other dispersive elements (notably the unbalanced air paths in the VLTI) places all wavelengths from the two telescopes simultaneously in phase quadrature in the polarization recombined beam, so that the result for all spectral channels, after subtracting the two phases, is near zero. Those measurements from all spectral channels are summed together in an optimum weighting to obtain an error signal which we will call $\Delta\tau$ thus given by:

$$\Delta\tau = \sum_i w_i \frac{1}{2\pi\nu_i} \frac{v_i}{V_i} \quad (15)$$

where v_i is the reading from the differencing of the detectors of spectral channel i and V_i is a calibration for that spectral channel such that $\frac{v_i}{V_i}$ is an estimate of the phase angle in radians. ν_i is the center optical frequency of that channel and the w_i are a set of weights such that $\sum w_i = 1$. The scale of $\Delta\tau$ with respect to the actual OPD offset will fluctuate due to wavefront fluctuations, but can nevertheless be used in a feedback system.

4.3.1 Modes 1 and 2

In “mode 1” $\Delta\tau$ (after appropriate filtering) is applied to the main delay line of the VLTI in order to drive the OPD back towards the equilibrium tracking point. In operational “mode 2” the feedback is used to generate a rapid correction using the *local* piezo OPD actuators (sect. 4.2.2) to counter the detected OPD offset. Those actuators include internal metrology and it is the reading from that metrology which represents the best estimate of the *external* OPD offset (that is, the portion of the atmospheric OPD which has *not* been corrected by the VLTI delay lines) and it is that value which is communicated to the VLTI system, presumably to generate corrections to the main delay lines. Mode 2 is more robust because it tracks the residual OPD using a much faster actuator (a local piezo) so that the offset stays well within a radian (the 2-phase interferometry loses sensitivity when the offset is not much less than a radian). On the other hand mode 1 sends more direct feedback to the VLTI delay lines and can thus potentially achieve a lower residual OPD affecting the science instrument (which after all, is the main point of fringe tracking!).

4.3.2 Mode 3: OPD scanning

Operational mode 3 refers to a number of possibilities involving OPD scanning, which in the primary modes we have avoided by using 2-phase interference (and taking advantage of the hardware's photometric symmetry) for increased sensitivity. But for utility purposes at least, it will be useful to scan the OPD across the entire fringe packet. The result can be of use at least in the following ways:

- For optimum scaling of the interference signal and optimum weighting of the spectral channels, we wish to estimate the visibility *magnitude* of the tracking star in each spectral channel. While this might be predictable in many cases, we would at least want to check our assumptions with such a measurement.
- In fringe *acquisition* it will be useful to scan across a possibly large OPD range in order to examine the interferogram for a region of correlated power (an interference signal) after which the OPD can be adjusted to home in on the zero phase near the group delay (peak coherence) and switch to mode 1 or 2.
- Scanning the fringe across a wide OPD range can retrieve spectral detail concerning the source's correlated flux by taking the fourier transform of the autocorrelation thus obtained, including spectral phase (usually of instrumental origin, thus sensitive to residuals from the NFT's dispersion compensation). This works particularly well if the residual OPD of the channel performing this function is stabilized by tracking on the other baselines (or, when using only two telescopes, the same baseline in the other interference channel).

4.3.3 Offline estimation of atmospheric OPD

The estimates of atmospheric OPD applied to the local and/or the main VLT delay line in modes 1 and 2, are based on data from the interferometer as filtered on the basis of the detection noise spectrum (usually white) and the spectrum of atmospheric (and possibly instrumental) OPD fluctuations. Of course the realtime estimate can only utilize data that has already been measured, whereas a superior estimate of what the atmospheric OPD *had been* will be available following the complete observation using a two-sided data set and optimal smoothing filter. Thus the raw data from the instrument is saved in order to later produce an *off-line* estimate of the atmospheric OPD which, in conjunction with the record of the actual delay line settings, supplies a superior phase reference for the offline coherent integration of data that had been concurrently obtained by the science instrument.¹²

4.3.4 Detection of dispersion and fringe jumps

As mentioned above, the net detection of the OPD offset from the intended tracking point is based on a (weighted) sum of the detections from several spectral channels. At the same time, a parallel computation examines the *difference* between the detected interference in those wavelengths, and smoothes those results. Let us call $\Delta\tau_{ij}$ the difference in the apparent phase delays between channels i and j , computed as:

$$\Delta\tau_{ij} = \frac{1}{2\pi\nu_i} \frac{v_i}{V_i} - \frac{1}{2\pi\nu_j} \frac{v_j}{V_j} \quad (16)$$

Fluctuations in $\Delta\tau_{ij}$ (for some i and j) on a time scale of seconds will result from water vapor seeing,¹¹ and that detection may be used in order to servo the dispersion compensator (sect. 4.2.1) in order to cancel that dispersion. Although it would be possible for the dispersion compensator to ignore atmospheric water vapor fluctuations without badly hurting interferometry (and fringe tracking) in the K band, the phases at much shorter wavelengths could very well exceed a radian and incapacitate those channels' contribution to the detection of atmospheric OPD.

On the other hand, there may be *shift* in $\Delta\tau_{ij}$, although such a shift might not be noticed as such for several seconds due to the smoothing, but this would more likely have been caused by a fringe jump, generally (but not always) due to a momentary loss of correlated flux during which an atmospheric OPD shift in excess of $\lambda/2$ failed to be tracked. The main estimator of OPD tracking error $\Delta\tau$ in (15) is blind to jumps of about 1 wavelength, but a fringe jump will lead to a significant $\Delta\tau_{ij}$. For instance, if the effective tracking wavelength is $2.2\mu m$ and spectral channels 1 and 2 are centered at $2.3\mu m$ and $2.1\mu m$, then a fringe jump of $2.2\mu m$, which would be undetected by (15), would cause a systematic bias in $\Delta\tau_{12}$ of $.2\mu m$. While this detection might be below the noise floor, it would probably be noticed after several seconds of smoothing.

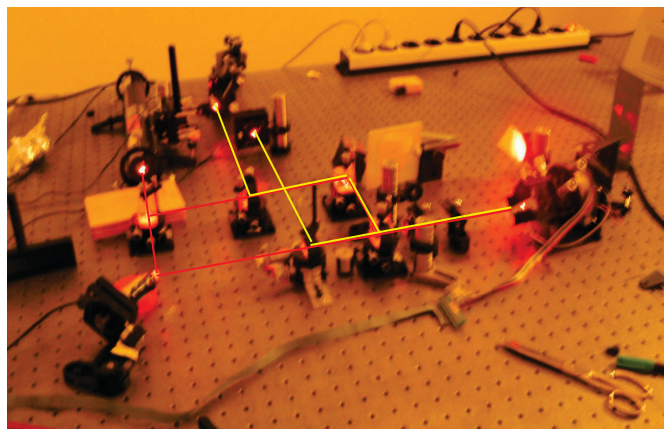
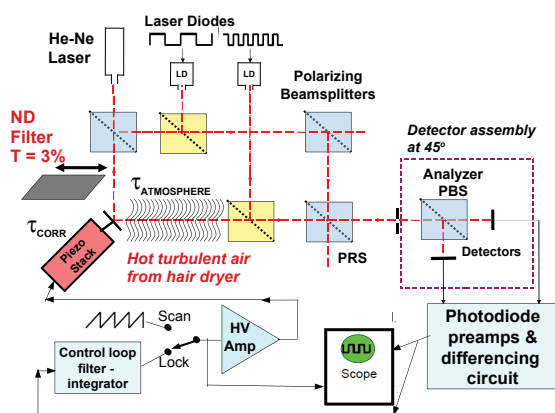


Figure 12. **Left:** Diagram of the laboratory setup. **Right:** Photograph of the apparatus with the beam paths of the He-Ne laser drawn in red, and the overlapping paths containing (incoherent) light from two laser diodes in yellow.

An average reading $\Delta\tau_{12}$ of $.2\mu m$, however, would also result from a change in the differential atmospheric water vapor column density by about $4.5\text{ moles}/m^2$, which is within possible limits.¹⁰ How would we distinguish between these two possibilities? One answer is that the fringe jump would occur abruptly while such a great change in the differential atmospheric water vapor column density would normally take more than a few seconds. However in the case of a very poor SNR, near the sensitivity limit of tracking by the NFT, $\Delta\tau_{12}$ may well be so noisy that it has to be smoothed over a much longer period than a few seconds so that the abruptness cannot be distinguished. However this degeneracy can be broken through simultaneous interferometric measurement at shorter wavelengths. For instance, with a third channel centered at $1.55\mu m$ such a water vapor jump would cause a shift in the differential phase delay at that wavelength (with respect to the tracking wavelength $2.2\mu m$) of about 1.2 fs or $.37\mu m$, and would result in an average $\Delta\tau_{32} = .16\mu m$ using (16). On the other hand a 1λ fringe jump at $2.2\mu m$ would amount to $511^\circ (=151^\circ)$ of phase at $1.55\mu m$, thus changing the average detected interference in that channel from $v_3 = 0$ to $v_3 = V_3 \sin(151^\circ) = .48V_3$. This would be (mis-) interpreted as an OPD offset of $\frac{.48 \cdot 1.7\mu m}{2\pi} = .12\mu m$ and generate a $\Delta\tau_{32} = .02\mu m$ (probably indistinguishable from 0). Thus even though both occurrences will lead to the same $\Delta\tau_{12} = .2\mu m$, they could be distinguished using the channel at $1.55\mu m$. This is an additional benefit of the wideband capability of the NFT design even when the shorter wavelength channel is of little value in phase tracking.

The use of the $\Delta\tau_{ij}$ for the detection of fringe jumps will be used in realtime to generate an intentional fringe jump in the opposite direction and to flag the duration of the error as such so that data from the science instrument during that period can be rejected. Detection of dispersion fluctuations using the $\Delta\tau_{ij}$ will be interpreted as atmospheric water vapor fluctuations (since changes in the dispersion of the air columns inside the VLTI will be much more gradual, and mainly predictable) invoking a correction of the dispersion compensator to return the simultaneous detection of each interference channel at zero phase. A similar analysis of the multispectral data from the NFT will be performed offline (sect. 4.3.3), again with improved precision employing data from the past and future for these determinations. The actual setting of the dispersion compensator, possibly along with residual dispersion detected in post processing, will aid in the interpretation of the scientific data, especially if the scientific instrument does not take advantage of the same dispersion compensator in the light supplied to it.

5. LABORATORY DEMONSTRATION

A demonstration of the polarization-based collimated beam combiner was performed at a laboratory of TU Delft where the essential elements of the optical system and tracking system were implemented for one channel (one baseline) along with a simulation of starlight perturbed by atmospheric turbulence and large photometric fluctuations in each “telescope” beam. Figure 12 diagrams the optical setup employed.

The helium-neon laser (with its polarization at about 45°) is split into two paths, one of which has its OPD adjusted using a mirror on a piezo actuator. There is also a portion of the path which can simulate “atmospheric turbulence” through

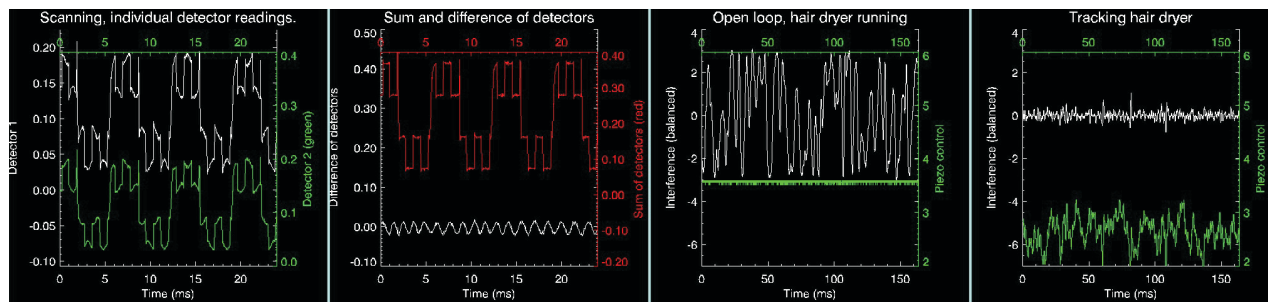


Figure 13. Waveforms downloaded from a digital oscilloscope monitoring the raw detectors (left) or the balanced interferometric output and the piezo control voltage (right 2 plots). From left to right: The raw detector readings I_A and I_B while scanning the fringe showing the interferometric signal dominated by the injected “photometric fluctuations” from the pulsing laser diodes. The sum of those readings cancels the interference leaving only the “photometry” (upper trace in red), while the difference signal (white) retrieves the interference while rejecting the photometric noise. With the piezo control fixed at 3.6 volts, running the hair dryer to simulate “atmospheric turbulence” over several wavelengths causes the interference to fluctuate across its maximum range ($\pm 2.8\text{v}$ in this case). With the piezo tracking the OPD, the residual interference is reduced to a few tenths of a volt, about .05 radians rms or 5nm of red light, while the piezo voltage (lower trace in green) tracks the fluctuations over about 5 wavelengths. When the hair dryer is turned off (not shown) the piezo tracks only very small vibrations and the residuals are too small to quantify.

the injection of warm air from a hair dryer. “Photometric fluctuations” are simulated in each of the two beams through the addition of pulsing light from two red laser diodes using beamsplitters. Since each only enters one arm of the system, this light does not contribute to the interference but would leak into the interferometric determination in a beam combiner lacking photometric symmetry (sect. 2.1.1). The two laser diodes are driven by square waves at 250Hz and 1000Hz respectively and their powers are similar to that of the He-Ne laser.

The PRS (sect. 3.4) is implemented with a polarizing beam splitter to generate a polarization recombined beam which travels to the polarization analyzer consisting of another polarizing beam splitter inside of a rotating assembly. Silicon photodiodes detect its two outputs which are run into preamps and a differential amplifier to obtain the interference signal, essentially free of photometric crosstalk from the pulsing laser diodes.

5.1 Basic Results

The setup demonstrated the principle of rejection of photometric crosstalk through implementation of balanced beam combination as has been described. The two interfering laser diodes were pulsed at distinct frequencies which could of course be seen in the individual detectors (fig. 13, left), and one goal of the experiment was to measure the maximum achieved rejection ratio. However it would not be meaningful to quote a number, since it was possible to adjust the analyzer angle and detector balance (sect. 3.5.1) to the point of being limited by measurement noise in the detection of residual photometric crosstalk. The success of the rejection of photometric noise cannot be properly appreciated from the plot in fig. 13 in which these digitized signals were numerically subtracted, but involved (analog) electronic subtraction followed by substantial amplification, and required eliminating the interference signal by blocking one beam from the He-Ne laser (rather than just reducing it through misalignment as was done in fig. 13).

Of more significance is that following a proper adjustment, the rejection ratio would consistently remain far greater than 100 following the passage of time (weeks) or after any changes to portions of the system not involving those two critical components (the angle of the analyzer and the balancing of the photodiode preamps and differential amplifier). Such stability and repeatability was only really achieved after placing a small ($\sim 1.5\text{mm}$) aperture stop right in front of the analyzer (the detector surfaces are about 3mm wide) and insuring that the detectors were placed in approximately the same optical positions with respect to the beam as it was split by the polarizing beam splitter implementing the analyzer. The rotational alignment of the analyzer could not possibly have been adjusted with sufficient precision had it not been for a finely geared angular adjustment mechanism on which that assembly was mounted.

5.2 Fringe Tracking

With the switch in fig. 12 in the lower position, a control loop was implemented to servo the piezo actuator in response to OPD variations detected interferometrically. In other words it was phase tracking in the exact manner that the NFT

proposes to do, by measuring interference in a balanced 2-phase detector with rapid feedback to an OPD actuator. By adding a 30Hz low amplitude square wave to the error signal from the interferometer, shifting the equilibrium point of the tracking by about 20nm, it was possible to observe the step response of the control loop. At best, the filter (including a gain and a damping adjustment) was adjusted to obtain near critical damping with a response time of about .5ms. This control loop faithfully tracked intentional vibrations and pressure applied to mirrors in the interferometric paths. Turning on a hair dryer simulated atmospheric turbulence as might be observed by a stellar interferometer, with a smaller peak amplitude (about 5 wavelengths) but on shorter time scales. The full scale of the interferometers's response, $V \sin(2\pi\nu\tau)$ with the correlated power V being about 2.8 volts, is seen in the 3rd plot of fig. 13 with the piezo voltage being held fixed. Tracking in the rightmost plot shows a residual tracking error of about .05 radians rms.

5.2.1 Fringe tracking with large photometric imbalance

One demonstration using the apparatus was performed for the sake of addressing concerns by ESO on the use of proposed fringe trackers when combining the light of a large UT (8 meter) telescope and an AT (1.8 meter) supplying 20 times less light. The increased effect of photometric variations by the brighter beam was immediately recognized as a challenge to beam combiners with significant photometric asymmetry, but one which the polarization-based collimated beam combiner was well suited to address. A neutral density filter was added to one of the He-Ne laser beams before reaching the PRS, which reduced that beam's power by a factor of 30. Tracking was again achieved, although this required increasing the gain of the loop filter since the loss of flux in one beam reduced the response of the interferometer by about $\sqrt{30}$ (and increasing the sensitivity to photometric fluctuations by the same factor). With the incoherent laser diodes still pulsing with a power similar to that of the unattenuated beam, there was still no crosstalk noted and tracking proceeded just as before.

5.2.2 Insensitivity to optical disturbances in the back end

Another demonstration with predictable results verified the principle of the polarization recombined beam being insensitive to perturbations in its optical paths (except ones involving polarization or birefringence). This demonstrated that the only interferometrically critical components of the NFT design were those up to and including the PRS (implemented as a solid prism block as shown in fig. 6). Thus one could be assured that the relatively long optical paths following the PRS would not be critical in regards to achieving the interferometric efficiency or photometric symmetry, and in particular that any path length or other mechanical fluctuations in those paths would not affect the interferometric OPD. The demonstration involved no change to the setup of fig. 12 except for the deliberate insertion of an obviously improper transmissive optical element in between the PRS and the polarization analyzer. Of course it would need to be something that still would pass most of the light toward the photodetectors (and which wasn't birefringent) but we were looking for the sort of thing that wouldn't be found in the optics storeroom. In the kitchen a wine glass was found, and it was inserted such that the beam from the PRS crossed both sides of it before reaching the analyzer. While fringe tracking as described above, the wine glass was moved and tilted with, as predicted, no introduction of photometric crosstalk from the pulsing laser diodes, and no noticeable effect on the tracking or on the OPD reported by the piezo voltage.

6. CONCLUSION

The polarization-based collimated beam combiner achieves higher performance in a number of astronomical interferometric applications through the attainment of *photometric symmetry* (sect. 2.1.1). That solves a problem without the performance reductions involved in photometric monitoring (sect. 2.3) or OPD modulation (sect. 2.4). The rejection of photometric crosstalk results from a single geometrical adjustment (sect. 3.5.1) to achieve robust performance independent of wavelength rather than relying on the precision of manufactured optical elements or coatings. Application of the polarization-based collimated beam combiner in a visibility measuring interferometer has been described with implementations of the polarization analyzer to achieve quadrature or 3-phase interferometric detection (sect. 3.6).

The immediate application for this topology in the proposed NFT (sect. 4) has been detailed to illustrate its virtues. The maximization of limiting sensitivity as a design goal for a fringe tracker enables the widest application of the scientific instruments which depend on it. In addition to avoiding the classical and unsatisfactory solutions to the problem of photometric crosstalk, the achievement of photometric symmetry will allow for tracking on low visibility targets which have been resolved beyond the first visibility null, and will allow for interferometric detection involving light from an AT (1.8 meter) telescope with a UT (8 meter telescope) whose beam is 20 times as bright. Low resolution spectral detection (sect.

4.2.6) enables the instrument to track dispersion fluctuations (sect. 4.3.4) which are removed by a dispersion compensator (sect. 4.2.1), allowing for 2-phase interference for maximum sensitivity (sect. 4.3).

Demonstration of the concept in the laboratory (sect. 5) confirmed its ability to reject photometric crosstalk with a stable $\gg 100:1$ rejection ratio (sect. 5.1) and to perform fringe tracking (sect. 5.2) using a hair dryer to simulate atmospheric OPD fluctuations, even with a 30:1 intensity ratio between the incident beams (sect. 5.2.1).

ACKNOWLEDGMENTS

We appreciated the use of the laboratory facilities at TU Delft and especially the assistance of optics technician Thim Zuidwijk in facilitating the experimental work described. In addition to funding by ESO and contributions from the partners of the NOVA consortium (see sect. 4), this work was supported by a grant from OPTICON. Some results presented were derived from VINCI data obtained courtesy of ESO through their science archive facility.

REFERENCES

- [1] Coude Du Foresto, V., Ridgway, S., and Mariotti, J.-M., “Deriving object visibilities from interferograms obtained with a fiber stellar interferometer,” *Astronomy and Astrophysics Supplement series* **121**, 379–392 (Feb. 1997).
- [2] Coude Du Foresto, V., Perrin, G., Ruilier, C., Mennesson, B. P., Traub, W. A., and Lacasse, M. G., “FLUOR fibered instrument at the IOTA interferometer,” in [*Proc. SPIE Vol. 3350, p. 856-863, Astronomical Interferometry, Robert D. Reasenberg; Ed.*], 856–863 (July 1998).
- [3] Kervella, P., Coude du Foresto, V., Glindemann, A., and Hofmann, R., “VINCI: the VLT Interferometer commissioning instrument,” in [*Proc. SPIE Vol. 4006, p. 31-42, Interferometry in Optical Astronomy, Pierre J. Lena; Andreas Quirrenbach; Eds.*], 31–42 (July 2000).
- [4] Colavita, M. M., Wallace, J. K., Hines, B. E., Gursel, Y., Malbet, F., Palmer, D. L., Pan, X. P., Shao, M., Yu, J. W., Boden, A. F., Dumont, P. J., Gubler, J., Koresko, C. D., Kulkarni, S. R., Lane, B. F., Mobley, D. W., and van Belle, G. T., “The Palomar Testbed Interferometer,” *Astrophysical Journal* **510**, 505–521 (Jan. 1999).
- [5] Meisner, J. A., “Scientific and technical results from VINCI using coherent estimation of fringe visibility,” *Astrophysics and Space Science* **286**, 119–127 (2003).
- [6] Armstrong, J. T., Mozurkewich, D., Pauls, T. A., and Hajian, A. R., “Bootstrapping the NPOI: keeping long baselines in phase by tracking fringes on short baselines,” in [*Proc. SPIE Vol. 3350, p. 461-466, Astronomical Interferometry, Robert D. Reasenberg; Ed.*], 461–466 (July 1998).
- [7] Hummel, C. A., Mozurkewich, D., Benson, J. A., and Wittkowski, M., “Coherent integration using phase bootstrapping,” in [*Interferometry for Optical Astronomy II. Edited by Wesley A. Traub. Proceedings of the SPIE, Volume 4838, pp. 1107-1114 (2003).*], 1107–1114 (Feb. 2003).
- [8] Sahlmann, J., Ménardi, S., Abuter, R., Accardo, M., Mottini, S., and Delplancke, F., “The PRIMA fringe sensor unit,” *Astronomy and Astrophysics* **507**, 1739–1757 (Dec. 2009).
- [9] Meisner, J., “Real-time Control Systems for a Ground-based Nulling Interferometer: Requirements and Configurations,” in [*GENIE - DARWIN Workshop - Hunting for Planets*], *ESA Special Publication* **522** (Mar. 2003).
- [10] Meisner, J. A. and Le Poole, R. S., “Dispersion affecting the VLTI and 10 micron interferometry using MIDI,” in [*Interferometry for Optical Astronomy II. Edited by Wesley A. Traub. Proceedings of the SPIE, Volume 4838, pp. 609-624 (2003).*], 609–624 (Feb. 2003).
- [11] Colavita, M. M., Swain, M. R., Akeson, R. L., Koresko, C. D., and Hill, R. J., “Effects of Atmospheric Water Vapor on Infrared Interferometry,” *Publications of the Astronomical Society of the Pacific* **116**, 876–885 (Sept. 2004).
- [12] Jorgensen, A. M. and Mozurkewich, D., “Coherent integration: To real time or not to real time? That is the question,” in [*Proc. SPIE Vol. 7734 (these proceedings)*], (July 2010).