Optimal Detection Scheme for Shot-Noise-Limited Phase Estimation in Passive Classical-Light Interferometry

Vincent Michaud-Belleau,[*] Jérôme Genest, and Jean-Daniel Deschênes

Centre d'optique, photonique et laser, Université Laval, Québec, QC G1V 0A6, Canada

Driven in part by the extreme requirements of gravitational-wave detection, recent research about interferometric phase estimation has focused on the use of nonclassical states of light in order to beat the shot-noise limit, allowing further improvements in sensitivity once the detected optical power has been maximized. Still, for a large class of interferometric applications, the efficient use of these states remains prohibitively complicated. Seeking simple improvements to instruments illuminated by classical laser light, we revisit the passive or unmodulated Mach-Zehnder interferometer by structuring the estimation problem in the form of a linear model for which an unbiased estimator attaining the Cramér-Rao bound is easily computed. This paper compares the performance of such linear and efficient estimators for several detection schemes used with two-output and four-output (in-phase and quadrature) interferometers. We find that the independent monitoring of all available output ports leads to an overall sensitivity that is, in most cases, better than that of balanced detection and single-output detection. In addition, it allows the cancellation of the technical amplitude noise contribution at all operating phases without resorting to modulation techniques. This seldom used but simple detection scheme should therefore be considered when designing instruments for operation close to and perhaps below the shot-noise limit.

I. INTRODUCTION

The precise measurement of phase fluctuations from an already known or measurable operating phase is one of the most common challenges in optical interferometry. For instance, it is relevant in the fields of laser stabilization and characterization [1,2], gravitational-wave detection [3,4], rotational measurements using optical gyroscopes [5,6], and, more generally, in the large family of applications in which a physical effect can be transduced to an optical path-length variation [7]. Optimizing the performance of a phase estimator, that is, minimizing its variance or power spectral density (PSD) in a given spectral band without the introduction of a bias, is important as it often constitutes the most straightforward way to improve the sensitivity of a given instrument.

The ultimate performance of interferometers illuminated by classical continuous-wave laser light is generally determined by the photon-counting error, commonly referred to as shot noise [8]. In the semiclassical framework, the shot-noise limit (SNL) for phase sensitivity can be expressed through the variance $\Lambda = 2h\nu_0 B/(\eta P_0)$ rad^2 in the ideal case, with h being Planck's constant, ν_0 the laser frequency, B the equivalent noise bandwidth, η the detection quantum efficiency, and P_0 the input laser power [9]. The

use of nonclassical light at the interferometer's inputs is required to beat the SNL [10]. Although impressive demonstrations have been made in this direction [11–13], the efficient use of nonclassical light, even for modest sensitivity improvements, still constitutes a considerable technical challenge [13,14] that might not be reasonable to take in a number of applications, especially when the SNL itself is difficult to attain.

In search for easy-to-implement methods to optimize the performance of standard instruments, we reinvestigate common detection schemes by structuring the estimation problem in the form of a linearized model. Even though this constitutes an approximation that is only valid for small phase excursions around the operating point, most interferometers naturally work in this regime when the SNL constitutes a concern. For each detection scheme, the linearization allows direct identification of an efficient minimum variance unbiased (MVU) estimator whose statistical behavior is fully defined [15]. Using this simple process, we are able to confirm well-known facts about the optimal operation of interferometers. We also find that using as many photodetectors as there are output ports is associated with significant benefits. This paper focuses on the elements supporting these findings.

Section II is dedicated to the theoretical analysis of the interferometric phase estimation problem based on the semiclassical model of photodetection. Although internal modulation [16] (similar to heterodyne readout [17]) and

———————
[*]vincent.michaud-belleau.1@ulaval.ca

external modulation [18] (similar to homodyne readout [19]) are often used to reduce the contribution of low-frequency technical amplitude noise and avoid additive flicker noise, they introduce other complications [20,21] and require tailored modulation and demodulation waveforms to efficiently approach the SNL [22]. For this last reason, a distinct theoretical framework would be required to properly treat these cases. The analysis made here is consequently restricted to fully passive interferometers (this format could also be called "dc readout" or "direct detection"). Several detection schemes which respect this restriction are compared using the introduced theory. Section III covers an experimental demonstration supporting the original conclusions of the theoretical analysis and providing additional insights about its application.

II. THEORY

In this section, a series of approximations and assumptions is introduced, allowing the interferometric signals to be put in the form of a linear model. The associated MVU estimator is then presented along with its statistical performance. This provides the theoretical tools required for a comparison of the phase sensitivity of several detection schemes used with passive two-output and four-output interferometers. The section concludes with a short comment about the effect of typical imperfections found in interferometric instruments.

A. Linearized form of the estimation problem

When illuminated by laser light of average power P_0, the lossless 50:50 interferometer illustrated in Fig. 1(a) outputs two time-varying power signals that take the following forms:

$$P'_1(t) = \frac{P_0}{2}\left[1 + \alpha(t)\right]\left\{1 - V\cos\left[\theta_0 + \theta(t)\right]\right\} + n_1(t),$$
(1a)

$$P'_2(t) = \frac{P_0}{2}\left[1 + \alpha(t)\right]\left\{1 + V\cos\left[\theta_0 + \theta(t)\right]\right\} + n_2(t),$$
(1b)

where $\alpha(t)$ represents a zero-mean relative power signal that can be related to amplitude modulation of the laser field [23] and V is the fringe visibility, a power ratio that takes value 1 only in an ideal interferometer. Here, the visibility V is modeled so that it can naturally account for a polarization mismatch in the interferometer; other imperfections such as losses and imperfectly balanced coupling have to be modeled in a slightly different manner, as we later discuss. The total interferometric phase is the sum of two terms: θ_0, the mean phase or operating point, which is determined by the laser's carrier frequency ν_0 and the
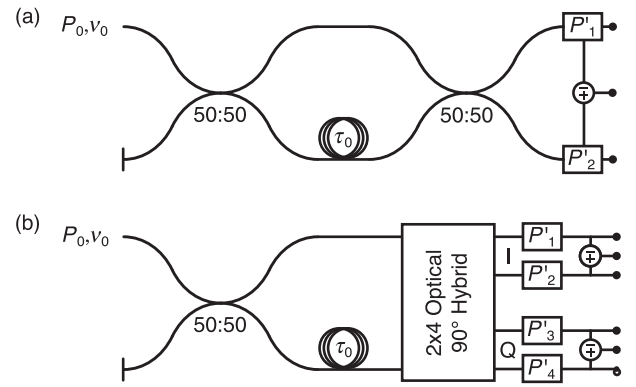

FIG. 1. (a) Passive two-output interferometer. (b) Passive four-output interferometer based on a 2×4 optical 90° hybrid coupler [25]. Only shot noise is considered in the displayed power signals; additive noise is introduced at a later stage, represented by the black dots. Although fiber interferometers are illustrated here, the treatment of this paper is also valid for free-space interferometers.

average interferometric delay τ_0 ($\theta_0 = 2\pi\nu_0\tau_0$), and $\theta(t)$, the zero-mean phase signal of interest that is typically produced through delay modulation of the interferometer or phase modulation of the input laser field [24]. The term "modulation" is used here in its broadest sense: it can represent a deterministic or stochastic signal without loss of generality.

The terms $n_1(t)$ and $n_2(t)$ account for all measurement noises expressed in units of optical power. This includes shot noise, which is the most fundamental contributor, but also all forms of additive noise, for instance, thermal noise and amplifier noise in the photodetector's electronics. These measurement noises are assumed to be zero mean and Gaussian, an accurate model even for Poissonian shot noise as long as more than approximately ten photons are measured during the system's effective integration time [26,27], a condition that is most often respected in practical contexts. Quantization noise is notably non-Gaussian, but it is rarely dominant in typical analog-to-digital converter (ADC) designs. We thus consider its small contribution to the overall noise to be Gaussian, when it is not simply negligible.

The use of the Gaussian model allows the zero-mean measurement noises to be fully defined using a single covariance matrix. Each element of this matrix can include a contribution from shot noise and a contribution from additive noise. Since our treatment is based on the semi-classical model of photodetection, we assume that shot noise at one output of the interferometer is statistically independent from shot noise at the other output, with a variance that is proportional to the average output power. The exact variance can be computed from the mean photocurrent and the detector's quantum efficiency. The

additive noise in each detection channel is also assumed to be independent from that in the other channel, although in this case we consider the two variances to be equal and known. In practice, this information is obtained by measuring the noise level in the absence of illumination. To simplify the theoretical results, we introduce the dimensionless parameter γ, which we define as the ratio between the additive-noise variance and the shot-noise variance that would be measured at the input of the interferometer using a detector with the same quantum efficiency η. In units of optical power, this "total" shot-noise variance is given as $2h\nu_0 BP_0/\eta = \Lambda P_0^2$. Using this definition, it is understood that only shot noise contributes to the measurement noise when $\gamma = 0$, while additive noise has the potential to be the dominant contributor when $\gamma > 0$, depending on the average power that is measured at the output under consideration.

When $\theta(t) \ll 1$, that is, when the unknown phase fluctuations can be assumed to be sufficiently small, it is possible to linearize Eq. (1) around the known operating point θ_0. This linearization leads to the following approximate expressions:

$$P_1'(t) \approx \frac{P_0}{2}(1 - V\cos\theta_0)[1 + \alpha(t)]$$
$$+ \left(\frac{P_0}{2}V\sin\theta_0\right)\theta(t) + n_1(t), \quad (2a)$$

$$P_2'(t) \approx \frac{P_0}{2}(1 + V\cos\theta_0)[1 + \alpha(t)]$$
$$- \left(\frac{P_0}{2}V\sin\theta_0\right)\theta(t) + n_2(t), \quad (2b)$$

where the cross terms are neglected under the additional assumption that the relative power signal is small [$\alpha(t) \ll 1$]. Since all time-varying signals are modeled to be zero mean [28], the mean part of each power signal in Eq. (2) can be removed or filtered out without loss of information. This leads to a natural description of the observed data in matrix form:

$$P = G\psi + n, \quad (3)$$

in which P is the mean-subtracted measurement vector (or data vector), G is the observation matrix (or sensitivity matrix or Jacobian matrix), ψ is the parameter vector, and n is the noise vector having statistical properties defined by C_n, which can represent a variance or covariance depending on context. Note that optical power signals without the prime symbol implicitly have their mean subtracted and that the explicit time dependence is dropped from this point on. Equation (3) constitutes the linear model; the exact structure of the variables it contains depends on the detection scheme used to conduct the

interferometric experiment and on assumptions about the relative contribution of θ versus α.

For a two-output interferometer [Fig. 1(a)], the three detection schemes we consider are as follows: "single detection," where the power is measured at only one of the two output ports (P_1 or P_2), "balanced detection," for which the difference between the powers incident on both detectors is recorded ($P_2 - P_1$), and "double detection," where both output ports are monitored independently (P_1 and P_2). For single detection and balanced detection, the linear model is set to estimate θ assuming α has negligible contribution, regardless of the validity of this assumption. We call this method "individual estimation" since the available information is insufficient to separately estimate both parameters. However, double detection *does* provide sufficient information for this to be achieved, in which case the method is called "simultaneous estimation." Still, if α is considered to be negligible, double detection can be set for individual estimation through a proper organization of the linear model. The structure of the model for each of these detection schemes is detailed in Table I.

Under certain experimental conditions, it can be beneficial to replace the interferometer's 50:50 output coupler with an optical 90° hybrid coupler [Fig. 1(b)]. This has an effect analogous to splitting the laser signal in two parts of equal power $P_0/2$, sending one part to the interferometer, and sending the other part to a second interferometer that is identical to the first one except for an extra differential delay that ideally corresponds to a $\pi/2$ phase shift at the laser's carrier frequency [2]. An interferometer based on such a coupler outputs four signals instead of two:

$$P_1' = \frac{P_0}{4}(1 + \alpha)[1 - V\cos(\theta_0 + \theta)] + n_1, \quad (4a)$$

$$P_2' = \frac{P_0}{4}(1 + \alpha)[1 + V\cos(\theta_0 + \theta)] + n_2, \quad (4b)$$

$$P_3' = \frac{P_0}{4}(1 + \alpha)[1 - V\sin(\theta_0 + \theta)] + n_3, \quad (4c)$$

$$P_4' = \frac{P_0}{4}(1 + \alpha)[1 + V\sin(\theta_0 + \theta)] + n_4, \quad (4d)$$

where P_0 still represents the average laser power at the overall system's input. For such a four-output interferometer, the previous approximations can also be applied in order to obtain a linear model. The main difference with the case of the two-output interferometer resides in the available detection schemes. Here, we consider two of them: "double balanced detection," for which the in-phase I signals (P_1 and P_2) and quadrature Q signals (P_3 and P_4) are each treated as in balanced detection, and "quadruple

TABLE I. Elements of the linear model and MVU phase estimator variance for the three detection schemes studied in a passive 50:50 two-output interferometer. Λ is the SNL, γ is the ratio between the additive-noise variance in one detection channel and the total shot-noise variance measured at the input, and V is the fringe visibility. In double detection with simultaneous estimation, α can be estimated concurrently with θ and the SNL for relative power, also expressed as Λ, can be reached when $\gamma \to 0$ and $V \to 1$. In all cases of individual estimation, a non-negligible α instead contaminates the phase estimate $\hat{\theta}$.

Detection scheme Estimation method	Single Individual	Balanced Individual	Double Individual	Double Simultaneous
P	P_1	$P_2 - P_1$	$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$	$\begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$
G	$\frac{VP_0}{2}\sin\theta_0$	$-VP_0\sin\theta_0$	$\frac{VP_0}{2}\begin{bmatrix}\sin\theta_0 \\ -\sin\theta_0\end{bmatrix}$	$\frac{P_0}{2}\begin{bmatrix}(1-V\cos\theta_0) & (V\sin\theta_0) \\ (1+V\cos\theta_0) & (-V\sin\theta_0)\end{bmatrix}$
ψ	θ	θ	θ	$\begin{bmatrix}\alpha\\\theta\end{bmatrix}$
$C_n/(\Lambda P_0^2)$	$\frac{1}{2}(2\gamma + 1 - V\cos\theta_0)$	$\gamma + 1$	$\frac{1}{2}\begin{bmatrix}2\gamma+1-V\cos\theta_0 & 0 \\ 0 & 2\gamma+1+V\cos\theta_0\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}2\gamma+1-V\cos\theta_0 & 0 \\ 0 & 2\gamma+1+V\cos\theta_0\end{bmatrix}$
$\sigma_{\hat{\theta}}^2/\Lambda$ (any γ, V)	$\frac{2(2\gamma+1-V\cos\theta_0)}{V^2\sin^2\theta_0}$	$\frac{\gamma+1}{V^2\sin^2\theta_0}$	$\frac{(2\gamma+1)^2-V^2\cos^2\theta_0}{(2\gamma+1)V^2\sin^2\theta_0}$	$\frac{2\gamma+1+(2\gamma-1)V^2\cos^2\theta_0}{V^2\sin^2\theta_0}$
$\sigma_{\hat{\theta}}^2/\Lambda$ ($\gamma=0$, $V=1$)	$\frac{2}{1+\cos\theta_0}$	$\frac{1}{\sin^2\theta_0}$	1	1

detection," where the four output ports are monitored independently. The structure of the linear model for each of these cases is presented in Table II.

B. Minimum variance unbiased estimator

When the observed signals can be put in the linear form of Eq. (3), it can be shown [15] that the MVU estimator is given by

$$\hat{\psi} = \left(G^T C_n^{-1} G\right)^{-1} G^T C_n^{-1} P, \tag{5}$$

provided that $G^T C_n^{-1} G$ is invertible. Moreover, this MVU estimator is efficient in the sense that it attains the Cramér-Rao bound (CRB). It is thus the best possible unbiased

estimator and it makes full use of the recorded information. Since the last equation represents a linear transformation of a Gaussian noise vector, the statistical performance of the estimator $\hat{\psi}$ is also completely specified, i.e.,

$$\hat{\psi} \sim \mathcal{N}(\psi, C_{\hat{\psi}}), \tag{6}$$

where

$$C_{\hat{\psi}} = \left(G^T C_n^{-1} G\right)^{-1} \tag{7}$$

can either represent the estimation variance $\sigma_{\hat{\theta}}^2$ if ψ is modeled to contain only θ or the estimation covariance under simultaneous estimation of α and θ. In the latter case, the

TABLE II. Elements of the linear model and MVU phase estimator variance for the two detection schemes studied in a passive 50:50 four-output interferometer. Λ is the SNL, γ is the ratio between the additive-noise variance in one channel and the total shot-noise variance measured at the input, and V is the fringe visibility.

Detection scheme Estimation method	Double balanced Individual	Double balanced Simultaneous	Quadruple Individual	Quadruple Simultaneous
P	$\begin{bmatrix}P_2-P_1\\P_4-P_3\end{bmatrix}$	$\begin{bmatrix}P_2-P_1\\P_4-P_3\end{bmatrix}$	$\begin{bmatrix}P_1 & P_2 & P_3 & P_4\end{bmatrix}^T$	$\begin{bmatrix}P_1 & P_2 & P_3 & P_4\end{bmatrix}^T$
G	$\frac{VP_0}{2}\begin{bmatrix}-\sin\theta_0\\\cos\theta_0\end{bmatrix}$	$\frac{VP_0}{2}\begin{bmatrix}(\cos\theta_0) & (-\sin\theta_0)\\(\sin\theta_0) & (\cos\theta_0)\end{bmatrix}$	$\frac{VP_0}{4}\begin{bmatrix}\sin\theta_0\\-\sin\theta_0\\-\cos\theta_0\\\cos\theta_0\end{bmatrix}$	$\frac{P_0}{4}\begin{bmatrix}(1-V\cos\theta_0) & (V\sin\theta_0)\\(1+V\cos\theta_0) & (-V\sin\theta_0)\\(1-V\sin\theta_0) & (-V\cos\theta_0)\\(1+V\sin\theta_0) & (V\cos\theta_0)\end{bmatrix}$
ψ	θ	$\begin{bmatrix}\alpha\\\theta\end{bmatrix}$	θ	$\begin{bmatrix}\alpha\\\theta\end{bmatrix}$
$C_n/(\Lambda P_0^2)$	$\frac{1}{2}\begin{bmatrix}2\gamma+1 & 0\\0 & 2\gamma+1\end{bmatrix}$	$\frac{1}{2}\begin{bmatrix}2\gamma+1 & 0\\0 & 2\gamma+1\end{bmatrix}$	$\frac{1}{4}\begin{bmatrix}w_{1,1} & 0 & 0 & 0\\0 & w_{2,2} & 0 & 0\\0 & 0 & w_{3,3} & 0\\0 & 0 & 0 & w_{4,4}\end{bmatrix}^{\text{a}}$	$\frac{1}{4}\begin{bmatrix}w_{1,1} & 0 & 0 & 0\\0 & w_{2,2} & 0 & 0\\0 & 0 & w_{3,3} & 0\\0 & 0 & 0 & w_{4,4}\end{bmatrix}^{\text{a}}$
$\sigma_{\hat{\theta}}^2/\Lambda$ (any γ, V)	$\frac{2(2\gamma+1)}{V^2}$	$\frac{2(2\gamma+1)}{V^2}$	$\frac{2[(4\gamma+1)^2-V^2\cos^2\theta_0][(4\gamma+1)^2-V^2\sin^2\theta_0]}{V^2(4\gamma+1)[(4\gamma+1)^2-V^2(\sin^4\theta_0+\cos^4\theta_0)]}$ [b]	$\frac{2[(4\gamma+1)^2-V^2\cos^2\theta_0][(4\gamma+1)^2-V^2\sin^2\theta_0]}{V^2(4\gamma+1)[(4\gamma+1)^2-V^2(\sin^4\theta_0+\cos^4\theta_0)]}$ [b]
$\sigma_{\hat{\theta}}^2/\Lambda$ ($\gamma=0$, $V=1$)	2	2	1	1

[a] $w_{1,1}=4\gamma+1-V\cos\theta_0$, $w_{2,2}=4\gamma+1+V\cos\theta_0$, $w_{3,3}=4\gamma+1-V\sin\theta_0$, $w_{4,4}=4\gamma+1+V\sin\theta_0$.

[b] This expression is only valid for individual estimation, although it constitutes a good approximation in simultaneous estimation for which the true expression is significantly more elaborate. At worst, the displayed and nondisplayed expressions differ by 0.5% when $\gamma \approx 0.1$; they become identical for $\gamma=0$ and $\gamma \to \infty$. The displayed expression is thus sufficiently accurate in most practical cases.

individual estimation variances are found on the diagonal of the matrix $C_{\hat\psi}$. These variances, $\sigma_{\hat\alpha}^2$ and $\sigma_{\hat\theta}^2$, are typically used to define the sensitivity of an interferometric measurement in relative power and in phase, respectively, although relative power estimation is seldom done this way since it can easily be performed without an interferometer.

The advantage of simultaneous estimation rather lies in the absence of crosstalk between the two parameters. This property can be recognized by switching the mean-subtracted measurement matrix containing the power signals in Eq. (5) for $G\psi$, which is the same minus the measurement noise contribution (i.e., $P - n = G\psi$). This produces the following expression:

$$\left(G^T C_n^{-1} G\right)^{-1} G^T C_n^{-1} G\psi, \qquad (8)$$

in which the premultiplier for ψ can be identified as the identity matrix of size 2 if simultaneous estimation is performed. The presence of null off-diagonal elements confirms there is no crosstalk on average between the two estimates. For most applications, the important implication is that simultaneous estimation allows phase to be estimated without contamination from technical relative intensity noise (RIN) regardless of the level of the phase signal relative to RIN and the point of operation, but provided the linear approximation remains valid.

C. Performance comparison in a two-output interferometer

The variances of the phase estimates obtained by linear MVU analysis of a two-output interferometer are compiled in Table I and illustrated in Fig. 2 for a few insightful combinations of γ and V. One apparent conclusion is that all detection schemes can theoretically reach and practically approach the SNL at a minimum of one operating point θ_0 per 2π range (one complete fringe cycle) in the regime of small additive noise and high visibility. For single detection, this occurs close to the dark fringe ($\theta_0 = 0$ for P_1), at a phase offset which depends sharply on γ and V. This property was recognized in [16] and it is considered in the design of LIGO [20]. For a nonzero offset from the dark fringe, the contribution from the relative power α is not canceled and can be dominant at low frequencies. In balanced detection, the SNL is reached halfway between the bottom and the top of the balanced fringe ($\theta_0 = \{\pi/2, 3\pi/2\}$), a fact that is also well known in the interferometry community. This is a convenient property since it is also where the relative power α is naturally canceled when balancing, or significantly attenuated in the presence of imperfections. This remains true for all values of γ and V. Finally, in double detection, whether it is based on individual or simultaneous estimation, all operating points are optimal in the asymptotical sense ($\gamma \to 0$, $V \to 1$). Far from one asymptote, however, the minimum
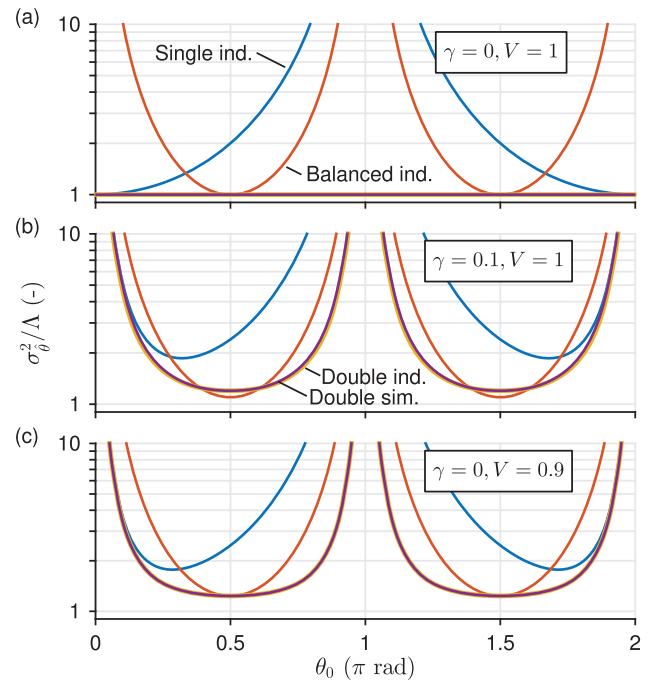

FIG. 2. (a) Theoretical ratio between the phase estimator variance and the SNL in the absence of additive noise for a two-output interferometer displaying perfect visibility. (b) Same for perfect visibility with additive noise. (c) Same for imperfect visibility without additive noise.

variance is reached at the same operating points as balanced detection, that is, $\theta_0 = \{\pi/2, 3\pi/2\}$. Just like in balanced detection, the relative power α is canceled at those points (individual estimation). This scheme is almost never used in interferometry and we believe that its interesting properties are not well known. Wider recognition of these properties would likely lead to broader usage.

The optimal detection scheme ultimately depends on various application-dependent factors, but we can draw several general conclusions from the linear MVU estimation analysis. Single detection is the simplest in terms of hardware and implementation, but its performance suffers from the largest penalty for higher γ and/or lower V. Balanced detection is the most robust under all experimental conditions with a variance that reaches $(\gamma + 1)\Lambda/V^2$ at the optimal point of operation. This corresponds to the absolute minimum that is achievable for a two-output interferometer fed with a classical state of light, a condition that occurs for this scheme because we assume that additive noise is introduced after balancing [Fig. 1(a)] as is typically the case in practice. Double detection requires more hardware than balanced detection for the same total shot-noise variance, which implies a larger additive noise contribution and thus higher estimation variance at the optimal operating point when γ is increased [see Fig. 2(b)]. However, it offers a better overall performance when shot noise is dominant ($\gamma \ll 1$), as can be recognized by the

fact that its estimation variance remains closer to the SNL over a large range of operating phases. This feature can be explained by the fact that the two signals contain more information in total than what remains in their difference. This can provide some advantages for nonstationary phase measurements (phase drift) [29,30], when there is no prior knowledge about θ_0 and the overall estimation fidelity has to be maximized [31,32], or when active stabilization of θ_0 through actuation of the laser or the interferometer is impractical or undesired [33,34]. Double detection is also the only passive scheme that permits independent and simultaneous measurement of α and θ in a two-output interferometer, a convenient property when both parameters convey useful information or when crosstalk is problematic. However, simultaneous estimation implies a penalty with respect to individual estimation [see Fig. 2(b)], a phenomenon that becomes more apparent for higher γ, so it should only be used if the advantages outweigh this penalty.

It is worth noting that double detection has already been recognized by Pezzé *et al.* to perform better than balanced detection for operation close to the SNL in classical-light interferometry [35]. Besides the general form of the mathematical model, notable differences between our works include the initial assumptions (linearity and Gaussian noise, mainly), which are more restrictive here but lead to simple analytical expressions even under explicit consideration of additive noise and various interferometric imperfections, the capacity of our model to properly take technical RIN into account if required, and our following treatment of four-output interferometers. Still, our conclusions about the sensitivity of a two-output interferometer are similar, suggesting that the formalism described here could be successfully translated to interferometers fed by squeezed-vacuum light, as was done by Pezzé *et al.* in a following paper [36].

D. Performance comparison in a four-output interferometer

The results of the linear MVU estimation analysis for a four-output interferometer are compiled in Table II. Once again, the phase estimate variances for three selected combinations of γ and V are illustrated in Fig. 3. By inspection of Table II, it is apparent that double-balanced detection produces measurements that are in quadrature. This allows independent and simultaneous extraction of amplitude and phase, even when the phase signal cannot be linearized [2]. Here, however, our interest still lies in cases where $\theta \ll 1$ and for which the estimation problem can be put in the linear form of Eq. (3). The quadrature relation implies there are no optimal operating points even when $\gamma > 0$ and $V < 1$. In other words, the variance of the phase estimate is always constant, which means that active stabilization is not required for optimal estimation. Moreover, individual
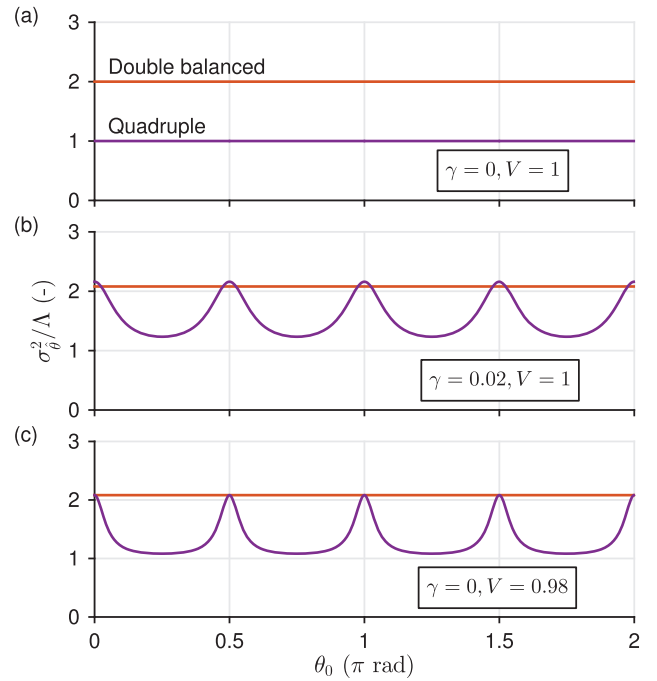

FIG. 3. (a) Theoretical ratio between the phase estimator variance and the SNL in the absence of additive noise for a four-output interferometer displaying perfect visibility. (b) Same for perfect visibility with additive noise. (c) Same for imperfect visibility without additive noise. Individual estimation and simultaneous estimation give almost identical results, so only one case is displayed.

and simultaneous estimation always yield the same performance since crosstalk is directly eliminated. The only disadvantage of the double-balanced detection scheme is that the SNL cannot be reached: a 3-dB penalty arises even in the ideal case because the total input power is effectively split between two interferometers (I and Q) that both rely on balanced detection, an operation that leads to a loss of useful information as recognized in the previous section.

In contrast, the SNL can be reached in theory if the quadruple detection scheme is used. This means that the 3-dB penalty of double-balanced detection can be eliminated at most but not all operating points. To benefit from this improvement, it might thus be necessary to reintroduce active stabilization. Of course, this constitutes a net advantage only when the increased hardware complexity is tolerable, when additive noise is small, and when fringe visibility is high. As γ increases, the performance of quadruple detection becomes worse than that of double-balanced detection, first at certain operating points and eventually over the full 2π range, because the extra detection channels add their own noise. As was the case for a two-output interferometer, there are advantages and disadvantages to the independent measurement of all output signals. The trade-offs are, however, different in the

two cases since crosstalk is rarely problematic in a four-output interferometer, regardless of the detection scheme that is used, whereas the 3-dB penalty is not an issue for the two-output interferometer based on single or balanced detection as long as the operating point is controlled. Still, for best overall performance close to the SNL, monitoring all output ports is preferable in all cases.

E. Impact of imperfections in the interferometer

At this point, only the imperfect fringe visibility due to polarization mismatch has been considered. It is, however, easy to model other imperfections such as losses and imperfectly balanced input and output couplers by adjusting the matrix G and the shot-noise level in C_n. It should be stressed that any imperfection or departure from the ideal case implies some form of penalty with respect to the estimation variances given in Tables I and II, although the final effect may be neutral in some occurrences. Since it can be difficult to accurately model all interferometric imperfections, it is preferable to characterize the instrument in a prior measurement. This calibration can be easily performed through controlled modulation of the laser phase, laser frequency, or the interferometric delay, provided that a full 2π range of θ_0 is explored over the spectral range of interest.

III. EXPERIMENTAL DEMONSTRATION

This section starts with a description of the essential elements of the laboratory setup. We then provide an overview of the signal processing steps before presenting the sensitivities measured experimentally for all detection schemes studied in the previous section, first for a two-output interferometer and then for a four-output interferometer. Instead of working with the phase variance, we consider the integrated PSD between 350 kHz and 1 MHz since it is difficult to reach a satisfying value of γ (for the purpose of this demonstration) when integrating over lower frequencies. With trivial adaptations, the conclusions of the theoretical section remain valid when integrating quantities over a restricted spectral range. Note that the SNL is simply $2h\nu_0/(\eta P_0)$ rad^2/Hz when expressed in PSD.

A. Setup and equipment

The laser used in the experiment (OEWaves, WGM Gen3) outputs continuous-wave light at 194.7 THz (1540 nm) and provides enough optical power to simultaneously saturate all the detectors. A RIN PSD level that is well approximated by the expression $[(4 \times 10^{-11})/f + 3 \times 10^{-17}]$ 1/Hz is measured above 1 kHz, where f is the frequency (shot noise is subtracted for this PSD measurement). The same is performed for the frequency-noise PSD, with a measured level of roughly $(4 \times 10^6)/f$ Hz^2/Hz. An electro-optic modulator (EOM,

from EOspace) is used at the interferometer's input, allowing the generation of a sinusoidal phase-modulation signal at 300 kHz for verification purposes.

We assemble a fiber interferometer similar to the one illustrated in Fig. 1(b) and set a small differential delay $\tau_0 \approx 3$ ps (corresponding to approximately 600 μm in silica) in order to minimize losses, phase drift, and unwanted contribution from the laser's frequency noise [37]. However, a small gain for frequency noise also implies a small gain for laser phase modulation (PM). This explains why we observe a significant residual amplitude modulation (AM) at 300 kHz when using the EOM, as the interferometric gain for AM is rather close to unity. It is necessary to consider the contribution of this residual AM, displaying a phase shift of 2.6 radians and a −15-dB level with respect to the phase component (units of α versus θ), in order to explain the results presented later. Since our goal is to expose the ultimate noise floor that lies under the verification signal at 300 kHz, the laser's frequency noise is undesired in our experiment; however, it would constitute the signal of interest in many applications.

Instead of implementing all detection schemes in succession, we simply perform quadruple detection, record the data, and emulate the action of other schemes through signal processing; to study the two-output interferometer, we discard the two signals of the interferometer's Q branch. This shortcut is associated with one important difference with respect to our previous theoretical treatment: it introduces more additive noise than strictly possible in balanced detection, since a typical balanced detector would add negligible additive noise prior to the subtraction step, as assumed in the theoretical analysis (see Fig. 1). However, this has a small impact in our chosen regime of operation and our conclusions remain the same (the theoretical model could, however, be adapted in cases for which γ is closer to 1).

Amplified InGaAs detectors are designed to minimize γ, the ratio between additive-noise variance and shot-noise variance. The first step in this direction usually consists in maximizing the output voltage, given by the product of the transimpedance gain R and the photocurrent I_0. Here, a maximum value of $RI_0 \approx 18$ V is possible. This figure determines the ratio between the thermal-noise variance associated to the feedback resistor and the shot-noise variance. With RI_0 fixed, what remains is the adjustment of R so that other forms of additive noise, namely ADC noise, are not contributing too much to the overall noise. With this objective in mind, we chose a value $R = 39$ kΩ. Considering the properties of the photodiode (Thorlabs, FGA01FC), circuit board, and operational amplifier, this value of R sets the detection bandwidth to roughly 1 MHz (resonance at 1.4 MHz with an OP-27 operational amplifier) and the maximum photocurrent to $I_0 = 450$ μA.

The maximum value of P_0 is also determined by these design choices since the power that can reach a single
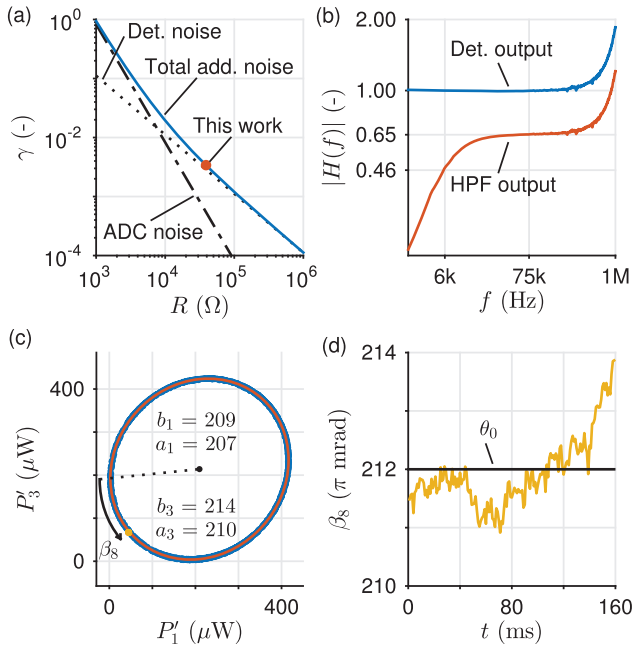
FIG. 4. (a) Theoretical relation between γ and the transimpedance gain R for $\eta = 0.79$, $P_0 = 865$ μW, $T = 300$ K (the effective temperature is $T \approx 550$ K considering the operational-amplifier current noise) and $\nu_0 = 194.7$ THz, and for an acquisition noise of density 15 nV/Hz$^{1/2}$. (b) Measured frequency response's modulus normalized by the transimpedance gain at the detector's output and at the HP filter's output. (c) Samples of the calibration data (blue points) and constrained least-squares elliptical fit (orange line). Four of the five extracted parameters are displayed in units of μW. The fifth parameter is an angle which has to be combined with the angles provided by the two other fits (not displayed) to extract the four χ_i. Samples of the LP signals associated to the eighth sensitivity measurement are superimposed in yellow. (d) Time-resolved values of β_8 extracted using Eq. (9). The value of θ_0 is taken as the mean of this signal.

detector in a four-output interferometer is $P_0/2$ at most in the absence of imperfections. For one of the four detectors to output its maximum voltage at one operating point, $P_0 \approx 1$ mW would be needed at the interferometer's input. In the experiment, the laser signal is instead attenuated down to 865 μW in order to maintain an acceptable level of nonlinearity. Taking all design parameters into account, we calculate and measure $\gamma \approx 0.003$ [Fig. 4(a)], a value that could have been reduced further at the expense of a lower detection bandwidth. However, in spite of these design efforts, the final performance of the interferometer is in large part determined by the fringe visibility, measured at $V = 0.986$ (the geometric mean of the four slightly different visibilities).

The dynamic range of our digitizer (GaGe CSE8389) is insufficient to reach this low value of γ when measuring small fluctuations on top of a dc signal that can reach more than 15 V. Consequently, each signal is split using a crossover network into a low-pass-filtered part

(LP) and high-pass-filtered part (HP) with cutoff at 6 kHz; both parts are recorded independently. With their higher signal-to-noise ratio, the HP signals are used for the final noise analysis. However, the LP signals are required to conduct the calibration and extract the average phase, two operations that are based on the processing of dc or low-frequency signals. The ADC's voltage-noise PSD is measured to be 120 nV/Hz$^{1/2}$ at 40 kHz, a value associated to $\gamma(40 \text{ kHz}) \approx 0.05$, and decreasing towards a white floor of level 15 nV/Hz$^{1/2}$ from 300 kHz and up, or $\gamma(f > 300 \text{ kHz}) \approx 0.003$. With the useful detection bandwidth limited at 1 MHz, this explains why the analysis is performed over a restricted spectral range. For a total duration of 160 ms, all signals are sampled at 125×10^6 samples/s, the available digitizer's highest possible rate, in order to minimize noise aliasing

B. Signal processing

Several offline measurements are performed in order to obtain results as free of the effects of experimental imperfections as possible. First, the frequency response of each detection channel is characterized through laser-power modulation over the full detection bandwidth [Fig. 4(b)]. This allows the construction of an equalizing digital filter, effective from the 6 kHz cutoff frequency up to 1 MHz (the gain is also brought back to 1). This filter is applied to all HP signals prior to further processing. Second, the mean of the additive-noise PSD between 350 kHz and 1 MHz is computed; the single figure obtained for each detection channel is used as the first part of the measurement noise covariance (or variance) C_n, the other part being shot noise, which is modeled based on the results of the calibration procedure outlined in the theoretical section. As stated before, this calibration is also required to obtain an accurate estimate of the observation matrix G.

Since the calibration is performed using the same laser and interferometer as in the final measurement, it is affected by the same types of noise, leading to a limited sensitivity. However, this calibration is required to characterize the *average* behavior of the interferometer; therefore, aggressive averaging can be performed so as to minimize the uncertainty. We thus start by modeling each LP calibration signal in the following modified form:

$$P'_k = b_k + a_k \cos(\beta + \chi_k) + n_k \tag{9}$$

for $k = 1, 2, 3,$ and 4, where b_k is the fringe bias, a_k is the fringe amplitude, and χ_k is a phase offset. For a perfect interferometer, the parameters of the last expression take values $b_k = a_k = P_0/2$ and $\chi_k = \{\pi, 0, -\pi/2, \pi/2\}$. Of course, this is never exactly the case in practice, namely because an optical 90° hybrid coupler does not produce a perfect $\pi/2$ phase shift and because it can introduce some crosstalk between its I and Q branches, but also

because of the typical imperfections which are mentioned in Sec. II. To estimate the value of each parameter, we perform a constrained least-squares elliptical fit [38] on the three data sets $P_3'(P_1')$, $P_4'(P_2')$, and $P_2'(P_1')$, which is possible since β [analogous to the full phase $\theta_0 + \theta(t)$ outside of calibration, but more affected by measurement noise] is slowly varied over a full 2π range during the calibration run [Fig. 4(c)]. Three such fits provide a total of 15 ellipse parameters, of which four are redundant, in the context of IQ interferometry; this is enough to estimate the 12 unknowns of Eq. (9), of which 11 are independent.

With these 11 parameters and the average additive-noise level in hand, the matrices G and C_n are calculated over the full phase range and for all detection schemes, providing benchmarks to assess the performance of the MVU estimates measured experimentally. Final sensitivity measurements are carried out at 25 static operating points. In each case, it is necessary to extract an accurate value of θ_0 so as to compute the proper estimate from Eq. (5). This is done by inverting a noiseless version of Eq. (9) in order to isolate β_n using the LP signals (P_k') recorded for the nth data set. The mean of β_n is taken as the θ_0 value for the nth measurement [Fig. 4(d)]; its standard deviation is used to trace the x-axis error bar. Finally, each MVU estimate's PSD is computed and the sensitivity is calculated as its integral between 350 kHz and 1 MHz, with y-axis error bars given by the standard deviation of the associated spectral noise.
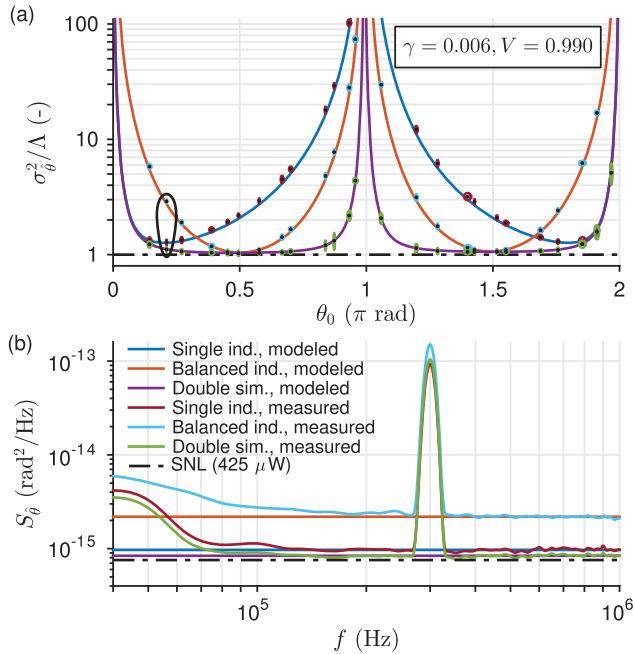

FIG. 5.   (a) Comparison between the modeled and measured phase estimate variances as a function of operating point in a two-output interferometer. (b) Same for the phase estimate PSD, but only at the operating point circled in the top panel.
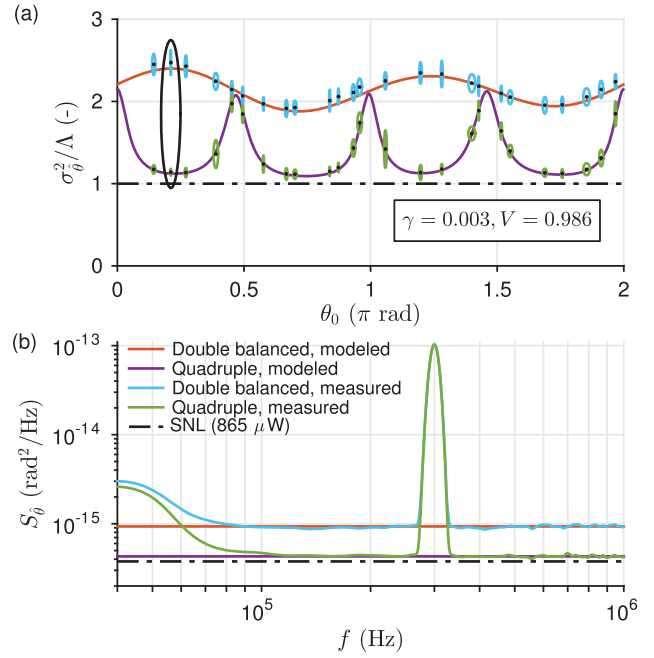

FIG. 6.   (a) Comparison between the modeled and measured phase estimate variances as a function of operating point in a four-output interferometer. (b) Same for the phase estimate PSD, but only at the operating point circled in the top panel.

C. Results

The outputs of the processing procedure are shown in Figs. 5 and 6. Since the results are nearly identical for all schemes in which both individual and simultaneous estimation is possible, we only display the latter to minimize clutter. By inspection of Fig. 6, it can be noticed that the model curves are slightly asymmetric with respect to $\theta_0 = \pi$. This is caused by the crosstalk between the I and Q branches of the 90° hybrid coupler, which produces a small phase offset measured at 0.02 rad. A similar asymmetry is found in Fig. 5, but in this case it is artificial since these results are obtained by discarding the two signals of the four-output interferometer's Q branch, as explained before. Therefore, this phenomenon could not occur in a real two-output interferometer. It is worth noting that the total power, the visibility, and the parameter γ are not perfectly identical in the I and Q branches, which justifies the different model parameters used in the two plots. Other asymmetries can be explained by the 0.1 rad error in the IQ phase shift at 194.7 THz, the definition of the reference angle ($\theta_0 = 0$), and other imperfections visible in Fig. 4(c).

For the two-output interferometer, the amplitude of the calibration signal at 300 kHz is visibly different for the three detection schemes. This arises since the input signal is not pure PM, but rather contains AM as stated previously. At the operating point $\theta_0 = 0.21\pi$, AM (and identically RIN) are fully canceled only in double detection

and simultaneous estimation: the theoretical interferometric gain for α-type signals instead is -8.8 dB for single detection and 2.1 dB (with a π phase shift) for balanced detection, taking the interferometer imperfections into account. These numbers are fully consistent with the observed signal levels considering the phase shift and relative contribution of AM versus PM in the verification tone. The AM-to-PM crosstalk also explains why the curve for balanced detection is significantly higher than the two others below 100 kHz: it is much more contaminated by technical RIN. Finally, we note that these effects are not visible in Fig. 6 primarily because a four-output interferometer is largely immune to such crosstalk.

Starting from 300 kHz and going down, the value of γ increases progressively by roughly one order of magnitude. Obviously, this increases the total measurement-noise variance, which in turn raises the phase-noise floor in this region. However, the main reason why the PSDs are seen to rise significantly below 70 kHz is because the estimator is built assuming a value of γ that is invalid below 300 kHz. Therefore, the estimator is not optimal since the model is inaccurate in this spectral range, which means the CRB cannot be reached. Nevertheless, in the range of interest for this demonstration between 350 kHz and 1 MHz, the experimental results are in excellent agreement with the model, a fact which confirms the conclusions of our theoretical treatment.

IV. CONCLUSION

For phase estimation in passive classical-light interferometry, that is, the common problem of precisely measuring a phase shift from a known operating point, our linearized analysis based on minimum variance unbiased estimation theory shows that the best direct-detection scheme when operating close to the shot-noise limit (SNL) is the one in which all output ports are independently monitored. Used in a two-output interferometer, this allows a cancellation of technical laser amplitude noise and broadens the phase range over which the SNL is approached with respect to other passive detection schemes. Used in a four-output interferometer, this almost fully removes the 3-dB sensitivity penalty of the more common double-balanced detection scheme. In both cases, an intuitive explanation for the improved performance is that the optimal detection scheme maximizes the ratio of recorded to available information, provided that the level of additive noise introduced by the supplemental detection hardware remains sufficiently low. Otherwise, more conventional detection schemes may be preferable, depending on the application. We have presented a conclusive experimental demonstration that supports these theoretical findings.

For a large class of instruments, the optimal detection scheme presented here can be implemented in a straightforward manner and its benefits can be reaped with modest signal processing that can be easily performed offline or online using modern digital electronics. This claim can be justified by recognizing that many instruments are based on relatively simple setups illuminated by unmodulated classical light and reach their best performance when operating close to the SNL. Therefore, the demonstration made in this paper should be relevant to a large number of scientists and engineers working on interferometric instruments such as fiber-optic sensors and laser gyroscopes. In the near future, the simple detection scheme may also be applied in order to beat the shot-noise-limited phase sensitivity using an interferometer fed with nonclassical light.

[1] F. Kéfélian, H. Jiang, P. Lemonde, and G. Santarelli, Ultralow-frequency-noise stabilization of a laser by locking to an optical fiber-delay line, Opt. Lett. **34**, 914 (2009).

[2] V. Michaud-Belleau, H. Bergeron, P. S. Light, N. B. Hébert, J.-D. Deschênes, A. N. Luiten, and J. Genest, Passive coherent discriminator using phase diversity for the simultaneous measurement of frequency noise and intensity noise of a continuous-wave laser, Metrologia **53**, 1154 (2016).

[3] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GW150914: The Advanced LIGO Detectors in the Era of First Discoveries, Phys. Rev. Lett. **116**, 131103 (2016).

[4] G. M. Harry *et al.* (LIGO Scientific Collaboration), Advanced LIGO: the next generation of gravitational wave detectors, Class. Quantum Grav. **27**, 084006 (2010).

[5] H. C. Lefèvre, *The Fiber-Optic Gyroscope* (Artech House, Boston, 2014).

[6] W. W. Chow, J. Gea-Banacloche, L. M. Pedrotti, V. E. Sanders, W. Schleich, and M. O. Scully, The ring laser gyro, Rev. Mod. Phys. **57**, 61 (1985).

[7] P. Hariharan, *Basics of Interferometry* (Academic Press, Burlington, 2010).

[8] R. Demkowicz-Dobrzański, M. Jarzyna, and J. Kołodyński, Chapter four - quantum limits in optical interferometry, Prog. Opt. **60**, 345 (2015).

[9] P. Fritschel, G. González, B. Lantz, P. Saha, and M. Zucker, High Power Interferometric Phase Measurement Limited by Quantum Noise and Application to Detection of Gravitational Waves, Phys. Rev. Lett. **80**, 3181 (1998).

[10] C. M. Caves, Quantum-mechanical noise in an interferometer, Phys. Rev. D **23**, 1693 (1981).

[11] U. L. Andersen, T. Gehring, C. Marquardt, and G. Leuchs, 30 years of squeezed light generation, Phys. Scr. **91,** 053001 (2016).

[12] A. A. Berni, T. Gehring, B. M. Nielsen, V. Händchen, M. G. A. Paris, and U. L. Andersen, *Ab initio* quantum-enhanced optical phase estimation using real-time feedback control, Nat. Photonics **9,** 577 (2015).

[13] J. Aasi *et al.* (LIGO Scientific Collaboration), Enhanced sensitivity of the LIGO gravitational wave detector by using squeezed states of light, Nat. Photonics **7,** 613 (2013).

[14] J. H. Shapiro, Parity measurements versus dual-homodyne measurements in coherent-state Mach–Zehnder interferometry, J. Opt. Soc. Am. B **34,** 1699 (2017).

[15] S. M. Kay, *Fundamentals of Statistical Signal Processing* (Prentice Hall PTR, Upper Saddle River, 1993).

[16] A. J. Stevenson, M. B. Gray, H.-A. Bachor, and D. E. McClelland, Quantum-noise-limited interferometric phase measurements, Appl. Opt. **32,** 3481 (1993).

[17] B. P. Abbott *et al.* (LIGO Scientific Collaboration), LIGO: the laser interferometer gravitational-wave observatory, Rep. Prog. Phys. **72,** 076901 (2009).

[18] M. B. Gray, A. J. Stevenson, C. C. Harb, H.-A. Bachor, and D. E. McClelland, External phase-modulation interferometry, Appl. Opt. **35,** 1623 (1996).

[19] T. T. Fricke, Ph.D. thesis, Louisiana State University, 2011.

[20] S. Hild, H. Grote, J. Degallaix, S. Chelkowski, K. Danzmann, A. Freise, M. Hewitson, J. Hough, H. Lück, M. Prijatelj, K. A. Strain, J. R. Smith, and B. Willke, DC-readout of a signal-recycled gravitational wave detector, Class. Quantum Grav. **26,** 055012 (2009).

[21] A. J. Stevenson, M. B. Gray, C. C. Harb, D. E. McClelland, and H.-A. Bachor, Interferometers with internal and external phase modulation: experimental and analytical comparison, Aust. J. Phys. **48,** 971 (1995).

[22] T. M. Niebauer, R. Schilling, K. Danzmann, A. Rüdiger, and W. Winkler, Nonstationary shot noise and its effect on the sensitivity of interferometers, Phys. Rev. A **43,** 5022 (1991).

[23] $\alpha(t) \approx \epsilon(t) + \epsilon(t - \tau_0)$, where $\epsilon(t)$ is the laser's zero-mean relative amplitude modulation.

[24] $\theta(t) \approx \phi(t) - \phi(t - \tau_0) + 2\pi\nu_0\tau(t)$, where $\phi(t)$ is the laser's zero-mean phase modulation and $\tau(t)$ the interferometer's zero-mean delay modulation.

[25] R. Hui and M. O'Sullivan, *Fiber Optic Measurement Techniques* (Academic Press, Burlington, 2009).

[26] J.-D. Deschênes, Ph.D. thesis, Université Laval, 2014.

[27] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes* (Tata McGraw-Hill Education, New York, 2002).

[28] If they are not in practice, the means can be appropriately distributed by redefining P_0, V, and θ_0.

[29] B. T. Gard, C. You, D. K. Mishra, R. Singh, H. Lee, T. R. Corbitt, and J. P. Dowling, Nearly optimal measurement schemes in a noisy Mach-Zehnder interferometer with coherent and squeezed vacuum, Eur. Phys. J. Quant. Tech. **4,** 4 (2017).

[30] M. G. Genoni, S. Olivares, D. Brivio, S. Cialdi, D. Cipriani, A. Santamato, S. Vezzoli, and M. G. A. Paris, Optical interferometry in the presence of large phase diffusion, Phys. Rev. A **85,** 043817 (2012).

[31] T. B. Bahder, Phase estimation with nonunitary interferometers: Information as a metric, Phys. Rev. A **83,** 053601 (2011).

[32] T. B. Bahder and P. A. Lopata, Fidelity of quantum interferometers, Phys. Rev. A **74,** 051801 (2006).

[33] S. Chen, C. Li, and Y. Zhu, Sensitivity evaluation of quantitative phase imaging: a study of wavelength shifting interferometry, Opt. Lett. **42,** 1088 (2017).

[34] H. Yonezawa, D. Nakane, T. A. Wheatley, K. Iwasawa, S. Takeda, H. Arao, K. Ohki, K. Tsumura, D. W. Berry, T. C. Ralph, H. M. Wiseman, E. H. Huntington, and A. Furusawa, Quantum-enhanced optical-phase tracking, Science **337,** 1514 (2012).

[35] L. Pezzé, A. Smerzi, G. Khoury, J. F. Hodelin, and D. Bouwmeester, Phase Detection at the Quantum Limit with Multiphoton Mach-Zehnder Interferometry, Phys. Rev. Lett. **99,** 223602 (2007).

[36] L. Pezzé and A. Smerzi, Mach-Zehnder Interferometry at the Heisenberg Limit with Coherent and Squeezed-vacuum Light, Phys. Rev. Lett. **100,** 073601 (2008).

[37] With a low-frequency power gain of $(2\pi\tau_0)^2$ rad^2/Hz^2, the associated phase noise contribution is lower than that of shot noise above 10 Hz.

[38] A. Fitzgibbon, M. Pilu, and R. B. Fisher, Direct least square fitting of ellipses, IEEE Trans. Pattern Anal. Mach. Intell. **21,** 476 (1999).