

Review

Estimation with Heisenberg-Scaling Sensitivity of a Single Parameter Distributed in an Arbitrary Linear Optical Network

Danilo Triggiani ¹  and Vincenzo Tamma ^{1,2,*} 

¹ School of Mathematics and Physics, University of Portsmouth, Portsmouth PO1 3QL, UK; danilo.triggiani@port.ac.uk

² Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK

* Correspondence: vincenzo.tamma@port.ac.uk

Abstract: Quantum sensing and quantum metrology propose schemes for the estimation of physical properties, such as lengths, time intervals, and temperatures, achieving enhanced levels of precision beyond the possibilities of classical strategies. However, such an enhanced sensitivity usually comes at a price: the use of probes in highly fragile states, the need to adaptively optimise the estimation schemes to the value of the unknown property we want to estimate, and the limited working range, are some examples of challenges which prevent quantum sensing protocols to be practical for applications. This work reviews two feasible estimation schemes which address these challenges, employing easily realisable resources, i.e., squeezed light, and achieve the desired quantum enhancement of the precision, namely the Heisenberg-scaling sensitivity. In more detail, it is here shown how to overcome, in the estimation of any parameter affecting in a distributed manner multiple components of an arbitrary M -channel linear optical network, the need to iteratively optimise the network. In particular, we show that this is possible with a single-step adaptation of the network based only on a prior knowledge of the parameter achievable through a “classical” shot-noise limited estimation strategy. Furthermore, homodyne measurements with only one detector allow us to achieve Heisenberg-limited estimation of the parameter. We further demonstrate that one can avoid the use of any auxiliary network at the price of simultaneously employing multiple detectors.

Keywords: quantum metrology; quantum sensing; distributed parameter; heisenberg limit; gaussian metrology; squeezing



Citation: Triggiani, D.; Tamma, V. Estimation with Heisenberg-Scaling Sensitivity of a Single Parameter Distributed in an Arbitrary Linear Optical Network. *Sensors* **2022**, *22*, 2657. <https://doi.org/10.3390/s22072657>

Academic Editors: Ivano Ruo-Berchera, Paolo Traina and Sylvain Girard

Received: 28 January 2022

Accepted: 27 March 2022

Published: 30 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Due to the discrete nature of physical phenomena, the error in the estimation of physical properties, such as lengths, delays, temperatures, or refractive indexes, when employing N probes (e.g., photons, electrons) is strongly limited by the so-called shot-noise scaling factor of $1/\sqrt{N}$ when a classical estimation strategy, i.e., in which the probe and the measurement employed can be fully described classically, is performed. However, it has been proven that it is possible, by exploiting quantum features such as entanglement and the squeezing of light, to overcome this classical limitation, and to reach an enhanced scaling in the precision of order $1/N$, called the Heisenberg limit [1–17].

A promising path towards this quantum-enhanced sensitivity is the one enabled by the squeezing of light [18–21]. Squeezed states are particular states of the electromagnetic field characterised by a quadrature field with smaller fluctuations than the field quadratures of the vacuum itself. This useful property, together with the Gaussian nature of these states which makes them relatively easy to produce, resilient to noise, and mathematically simple to manipulate [20,22], makes these states evident candidates for metrological purposes. The working principle of these protocols is rather straightforward, and can be outlined as follows. The probe, a pure squeezed state, undergoes an optical phase delay of a magnitude which depends on the unknown parameter we are interested to measure. The phase delay

causes a transformation, or more precisely a rotation, of the state of the probe, that needs to be observed to infer the unknown parameter. In order to obtain the quantum enhancement, the measured quadrature field must be ‘sufficiently’ squeezed, namely it must possess a variance reduced below the vacuum fluctuations.

It is possible to highlight two major different approaches undertaken in literature, according to what type of squeezed state is employed and, ultimately, to how the information about the parameter is encoded on, and then retrieved from, the probe. In ‘displacement-encoding’ approaches [2,23], the initial overall state possesses a non-vanishing displacement—i.e., the average of the quadrature fields—so that the value of the parameter is encoded into the variation of the displacement. This approach presents the advantage that the unknown parameter can be retrieved through the measurement of the displacement of the probe, namely the mean value of the signal experimentally observed in a laboratory when performing homodyne detection. In ‘squeezing-encoding’ approaches [24–26], all the resources are accumulated in the squeezing of the probes, which are thus squeezed vacuum states, so that the information on the parameter is encoded into the parameter-dependent rotation of the covariance matrix. This is the optimal approach for squeezing-based estimation protocols in the sense that, for a fixed total average number of photons in a Gaussian state, the strategy that maximises the precision is to concentrate all the photons in the squeezing [24,26,27]. Since the information about the parameter is encoded in the covariance matrix of the probe, with this approach the estimation consists of retrieving the value of the parameter from the modulation of the noise. Another interesting approach involves the use of an active component for the estimation of phases, i.e., anti-squeezing the signal before the detection [28–32]. In this approach, the goal is generally to estimate the value of the unknown phase detecting whether the final state of the probe is different from the one injected in the network, namely, performing a projective measurement on the initial state.

Another interesting aspect of quantum metrology that has been frequently investigated is the possibility to estimate a parameter which is not localised into a single node of the network which encodes it, but distributed in an arbitrary manner among multiple components. Although this framework entails a further complication, given by the generally non-trivial encoding of the value of the parameter onto the probe, it allows for the study of estimation schemes with arbitrary structures of the networks. Since, in this approach, the network is generally described as a ‘black-box’, namely the structure of the network is not specified, the applicability of the estimation technique is universal—i.e., guaranteed for any passive and unitary evolution of the probe—covering in this way experimental situations rarely treated in literature. Moreover, the absence of a defined network structure allows to analyse general properties, such as the average performance of the estimation over some set of local transformations [33], the best possible Gaussian strategy for a multi-channel network [27], or the typicality of the Heisenberg scaling [34]. Applications range from high-precision biomedical phase imaging, quantum enhanced pattern recognition, and gravitational-waves sensing, to the mapping of external fields (magnetic and electric fields or temperatures).

However, standard Gaussian estimation protocols still present some challenges. In particular, adaptivity—i.e., the fact that a scheme for the estimation of the parameter depends on the parameter itself—is known to be a typical feature of *ab initio* Gaussian metrology [24,35–37], namely schemes for the estimation of a parameter of which no prior information is known. Intuitively, the cause of the adaptivity stems from the fact that the phase of the squeezed-quadrature which needs to be measured to achieve the quantum-enhanced precision depends on the phase acquired by the probe through the interferometric evolution in arbitrary parameter-encoding networks and, ultimately, on the unknown parameter. Moreover, in distributed quantum metrology, a further experimental challenge arises from the need to adaptively optimize the preparation of the probe and of the measurement [27]. A common approach to avoid adaptivity is limiting the range of values that the unknown parameter is allowed to take, for example, requiring that the parameter is small [38–41].

In fact, in the regime of small phases, the transformation the probe undergoes is small enough to make the rotation of the squeezed quadrature negligible, which in turn remains practically unchanged. Nevertheless, in some practical scenarios the experimenters have no control on the value of the phase—e.g., the system under investigation is unreachable or cannot be manipulated—and in such cases, this approach ceases to be feasible.

We will show here how the challenge of adaptivity, and any restriction in the range of possible values of the parameter to be estimated, can be overcome thanks to some recent results in the field of distributed quantum metrology, for the estimation at the Heisenberg scaling sensitivity of a single parameter, distributed throughout an arbitrary, multi-channel passive and linear network (see Figure 1). In particular, the unknown parameter can be thought of as a physical property of an external field—such as the temperature of the environment, or the magnitude of an electromagnetic field—which globally affects some or all of the components of the network. This review is organised in two parts, in which we introduce two distinct estimation schemes, discussing the features that differentiate the two approaches. In Section 2, we will discuss a squeezing-encoding scheme achieving Heisenberg-scaling sensitivity by employing a single squeezed vacuum state and homodyne detection at a single output channel [42]. We will assume that no prior knowledge on the value of the parameter is known, and that the structure of the network, as well as the nature of the parameter it encodes, are completely arbitrary. We will present the conditions that need to be satisfied in order to reach the Heisenberg scaling in such a generic model, and we will show that, in general, only a classical prior estimation of the unknown parameter suffices to prepare a single auxiliary linear network required to optimize the setup. These results show that it is possible to conceive feasible two-step estimation strategies, composed of a first classical estimation of the unknown parameter required to engineer the auxiliary stage, and then the actual quantum estimation. In Section 3, we describe a different scheme achieving the Heisenberg scaling, which makes use of a single-mode squeezed-coherent state as a probe, and homodyne detection in every output port of the linear network [43]. Once again, the model will assume no prior knowledge on the parameter, nor on the structure of the network. We will see that, in this case, no auxiliary stage is required to achieve the Heisenberg scaling, so that the estimation can be carried out in a single step, at the cost of employing a multiple homodyne detector. Moreover, we will show that, due to the introduction of a non-vanishing displacement in the probe, the overall precision becomes the sum of two contributions, one deriving from the information encoded in the sample mean of the outcomes of the homodyne measurements, the other in the sample covariance matrix.

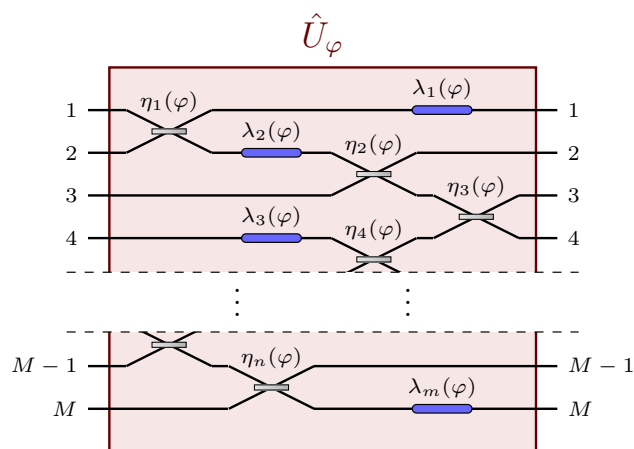


Figure 1. Example of a passive and linear network \hat{U}_φ which depends on a single global parameter φ . The parameter can be thought of as a physical property of an external agent (e.g., temperature, electromagnetic field) which affects multiple components, possibly of different natures, of the network [42,43]. Reprinted with permission from ref. [42], © 2021 The Author(s).

2. Quantum Estimation Based on Single-Homodyne Measurements

In this section, we will describe a generic model of a M -channel linear network \hat{U}_φ that allows the Heisenberg scaling to be reached in the estimation of a parameter φ distributed arbitrarily in \hat{U}_φ , regardless of the structure of the network. Our model makes use of two auxiliary stages, namely two other linear networks \hat{V}_{in} and \hat{V}_{out} , whose purposes are to distribute the probe, a single-mode squeezed vacuum injected in the first channel, in all the optical modes of \hat{U}_φ , and then to refocus it in the only channel observed through homodyne detection. We will show that only one of the two auxiliary networks must be optimized, and although the optimal choice of such network depends on the value of φ , the required precision for this optimization can be achieved with a classical estimation strategy. This allows us to conceive two-step estimation protocols, where a prior classical estimation to engineer the optimal auxiliary stage is followed by the quantum Heisenberg scaling-achieving estimation.

2.1. Setup

Let us consider a M -channel passive linear network, whose action on the state of the probe is described by the unitary operator \hat{U}_φ , φ being a single, generally distributed, unknown parameter we are interested to estimate. Due to its passive and linear nature, this network can be represented by a unitary matrix. Let then, U_φ be the $M \times M$ unitary matrix representing the action of \hat{U}_φ on the annihilation operators \hat{a}_i , $i = 1, \dots, M$, associated with each channel of the network, satisfying the commutation relations $[\hat{a}_i, \hat{a}_j] = 0$, $[\hat{a}_i, \hat{a}_j^\dagger] = \delta_{ij}$, where we denote with δ_{ij} the Kronecker delta. The matrix U_φ is thus defined by the transformation

$$\hat{U}_\varphi^\dagger \hat{a}_i^\dagger \hat{U}_\varphi = \sum_j (U_\varphi)_{ij} \hat{a}_j^\dagger. \quad (1)$$

We will consider a single-mode squeezed state

$$|\psi_0\rangle = |\mathbf{r}\rangle \equiv \hat{S}(\mathbf{r})|\text{vac}\rangle, \quad \mathbf{r} = (r, 0, \dots, 0), \quad (2)$$

as a probe, with $\hat{S}(\mathbf{r}) = e^{\frac{1}{2}r(\hat{a}_1^2 - \hat{a}_1^\dagger^2)}$ squeezing operator and $N = \sinh^2 r$ average number of photons, all injected in a single channel of the apparatus, i.e., the first one with the choice of squeezing parameters \mathbf{r} in Equation (2). In other words, the state $|\psi_0\rangle$ presents a non-vanishing number of photons only in the first mode. As discussed in Section 1, the approach of squeezing-based estimation strategies is to infer the value of φ from the transformation of the covariance matrix $\Gamma = \text{diag}(e^{2r}, 1, \dots, e^{-2r}, 1, \dots)/2$ of the state $|\psi\rangle$ after the interferometric evolution \hat{U}_φ . To do so, we will consider the model where a single output channel, say the first, is measured through homodyne detection. We will denote with θ the phase of the local oscillator, which coincides with the phase of the measured quadrature \hat{x}_θ . We will assume that $r > 0$ without loss of generality. We notice that, with this assumption, the squeezed quadrature of the state $|\psi_0\rangle$ is $\hat{p}_1 \equiv -i(\hat{a}_1 - \hat{a}_1^\dagger)/\sqrt{2}$, with $\text{Var}[\hat{p}_1] = e^{-2r}/2$, while the anti-squeezed quadrature field is $\hat{x}_1 \equiv (\hat{a}_1 + \hat{a}_1^\dagger)/\sqrt{2}$, with $\text{Var}[\hat{x}_1] = e^{2r}/2$. In terms of the creation and annihilation operators, the measured quadrature field can be expressed as $\hat{x}_\theta = (e^{-i\theta}\hat{a}_1 + e^{i\theta}\hat{a}_1^\dagger)/\sqrt{2}$.

Since the linear network \hat{U}_φ is arbitrary, the average number of photons that can be actually detected after the interferometric evolution ranges between 0 and N . Naturally, if this number were to be small, or far from N , we would expect a sub-optimal performance of the estimation scheme, since most of the photons would come out of the network \hat{U}_φ from channels that are not observed, and information on φ would in this way be lost. Moreover, it may happen that the transformation that it imposes on the probe in the transition to the first output port is trivial, namely that the element $(U_\varphi)_{11}$ of the transition amplitude matrix U_φ does not depend on φ . This occurrence would preclude the probe from acquiring any observable information on the parameter. In order to prevent these conditions from happening, this model includes the presence of two auxiliary linear and passive networks

acting on the probe, \hat{V}_{in} before and \hat{V}_{out} after the linear network \hat{U}_φ (see Figure 2). The first auxiliary stage \hat{V}_{in} can be understood as network scattering, which distributes the probe through multiple input channels of the parameter-dependent network \hat{U}_φ . The purpose of the second stage \hat{V}_{out} is instead to refocus the probe, after the interaction with \hat{U}_φ , into the only observed channel. The unitary matrix describing the overall network is thus

$$u_\varphi = V_{out}U_\varphi V_{in}, \tag{3}$$

given by the matrix product of the three single matrix representations V_{out} , U_φ and V_{in} .

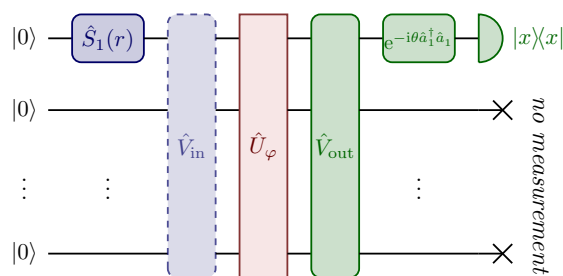


Figure 2. Schematic diagram of the setup described in Section 2.1. The squeezed vacuum state in Equation (2) is injected in the first channel of a network composed of a first auxiliary stage \hat{V}_{in} , a network \hat{U}_φ which depends on a generally distributed parameter φ we want to estimate, and a second auxiliary stage \hat{V}_{out} , before being detected through homodyne measurements in the first output port. The role of the two auxiliary stages \hat{V}_{in} and \hat{V}_{out} is to respectively distribute the photons of the probe through multiple channels, and then to refocus them into the only observed channel. We will show that only one auxiliary network needs to be optimized to reach the Heisenberg scaling, while, for networks with a large number of channels, the effect of the non-optimized network is typically irrelevant on the overall precision of the estimation [42]. Reprinted with permission from ref. [42], © 2021 The Author(s).

Since in this model all the photons are injected in the first input channel, which is also the only channel observed at the output, the only relevant transition amplitude is the element $(u_\varphi)_{11}$ of the overall unitary matrix in Equation (3). We can then rewrite

$$(u_\varphi)_{11} = (V_{out}U_\varphi V_{in})_{11} = \sqrt{P_\varphi}e^{i\gamma_\varphi}, \tag{4}$$

where $P_\varphi = |(V_{out}U_\varphi V_{in})_{11}|^2$ is the probability that a single photon injected in the first port is detected at the first output port of the overall network, and $\gamma_\varphi = \arg((V_{out}U_\varphi V_{in})_{11})$ is the phase acquired by the probe during the interferometric evolution. The Gaussian nature of the probe and of the homodyne measurements yield a Gaussian probability density function

$$p_\varphi(x) = \frac{1}{\sqrt{2\pi\sigma_\varphi^2}}e^{-\frac{x^2}{2\sigma_\varphi^2}} \tag{5}$$

which governs the outcomes of the homodyne detection [18–21]. The univariate Gaussian probability density function in Equation (5) is centred at zero due to the absence of a displacement in the probe, while its variance is given by (see Appendix A)

$$\sigma_\varphi^2 = \frac{1 - P_\varphi}{2} + \frac{P_\varphi}{2} [\cosh(2r) + \cos(2\gamma_\varphi - 2\theta) \sinh(2r)], \tag{6}$$

where θ is the phase of the local oscillator.

Once the probability density function $p_\varphi(x)$ is known, it is possible to evaluate the Fisher information [44,45]

$$\mathcal{F}(\varphi) = \int_{\mathbb{R}} dx p_\varphi(x) \frac{d}{d\varphi} \ln p_\varphi(x) \quad (7)$$

of the estimation scheme, which in turn fixes the ultimate precision $\delta\varphi_{\min}$ achievable in the estimation of φ through ν iteration of the measurement, given by the Cramer-Rao bound [44,45]

$$\delta\varphi_{\min} = \frac{1}{\sqrt{\nu\mathcal{F}(\varphi)}}. \quad (8)$$

For a Gaussian distribution centred on zero, the Fisher information reads (see Appendix B)

$$\mathcal{F}(\varphi) = \frac{1}{2} \left(\frac{\partial_\varphi \sigma_\varphi^2}{\sigma_\varphi^2} \right)^2, \quad (9)$$

where $\partial_\varphi := d/d\varphi$. We notice from Equation (6) that all the information on the parameter φ is encoded in the variance σ_φ^2 of the measured quadrature \hat{x}_θ through the two quantities P_φ and γ_φ . Thus, we can split $\partial_\varphi \sigma_\varphi^2$ into two contributions, one containing the derivative of P_φ , the other the derivative of γ_φ , namely

$$\partial_\varphi \sigma_\varphi^2 = (\partial_\varphi P_\varphi) \partial_P \sigma_\varphi^2 + (\partial_\varphi \gamma_\varphi) \partial_\gamma \sigma_\varphi^2, \quad (10)$$

where ∂_P and ∂_γ are derivatives with respect to P_φ and γ_φ , respectively, so that

$$\begin{aligned} \partial_P \sigma_\varphi^2 &= \frac{1}{2} (-1 + \cosh(2r) + \cos(2\gamma_\varphi - 2\theta) \sinh(2r)) \\ \partial_\gamma \sigma_\varphi^2 &= -P_\varphi \sin(2\gamma_\varphi - 2\theta) \sinh(2r). \end{aligned} \quad (11)$$

2.2. Heisenberg Scaling

Generally, without imposing any condition on the setup, this model does not achieve the Heisenberg scaling in the precision for the estimation of φ . In fact, we can explicitly rewrite the variance σ_φ^2 in Equation (6) in terms of the average number of photons $N = \sinh^2 r$ in the probe

$$\begin{aligned} \sigma_\varphi^2 &= \frac{1 - P_\varphi}{2} + \frac{P_\varphi}{2} \left[1 + 2N + 2 \cos(2\gamma_\varphi - 2\theta) \sqrt{N(N+1)} \right] \\ &= NP_\varphi (1 + \cos(2\gamma_\varphi - 2\theta)) + \mathcal{O}(1), \end{aligned} \quad (12)$$

where $\mathcal{O}(1)$ is a term of order equal to or smaller than 1, negligible in the asymptotic regime of N large (We will say that given two functions $f(N)$ and $g(N)$, $f(N) = \mathcal{O}(g(N))$ when $\lim_{N \rightarrow \infty} |f(N)/g(N)| < +\infty$). We can also rewrite the derivatives $\partial_P \sigma_\varphi^2$ and $\partial_\gamma \sigma_\varphi^2$ in terms of N

$$\begin{aligned} \partial_P \sigma_\varphi^2 &= (N + \cos(2\gamma_\varphi - 2\theta) \sqrt{N(N+1)}) \\ &= N(1 + \cos(2\gamma_\varphi - 2\theta)) + \mathcal{O}(1) \\ \partial_\gamma \sigma_\varphi^2 &= -2P_\varphi \sin(2\gamma_\varphi - 2\theta) \sqrt{N(N+1)} \\ &= -2NP_\varphi \sin(2\gamma_\varphi - 2\theta) + \mathcal{O}(1). \end{aligned} \quad (13)$$

Plugging the asymptotics shown in Equations (12) and (13) into the expression of the Fisher information in Equation (9), we notice that the numerator of $\mathcal{F}(\varphi)$ can be of the order N^2 at most. Since the denominator is in general of the order N^2 as well, it yields an overall general scaling of the Fisher information of $\mathcal{O}(1)$ —i.e., even lower than the SQL.

In order for this setup to reach the Heisenberg scaling, we thus need to impose some constraints which prevent the denominator of the Fisher information, i.e., the variance σ_φ^2 , to grow with N . We show in Appendix C that the asymptotic conditions (Given a function $f(N)$ and a finite sum $p(N)$ of powers of N , we will say that $f(N) \sim p(N)$ when they show the same asymptotic behaviour. In formulas, $f(N) \sim p(N)$ when $f(N) = p(N) + \mathcal{O}(N^{s-\epsilon})$, $\forall \epsilon > 0$, with s exponent of the smallest power of N appearing in the sum $p(N)$)

$$\gamma_\varphi - \theta \sim \pm \frac{\pi}{2} + \frac{k}{N} \quad k \neq 0 \quad (14)$$

$$P_\varphi \sim 1 - \frac{\ell}{N} \quad 0 \leq \ell < N \quad (15)$$

need to be satisfied for large N , with $0 \leq \ell < N$ and $k \neq 0$ arbitrary constant independent of N . We will discuss more in detail the physical meaning of these conditions in Section 2.3, and we will see that Equation (14) is a minimum-resolution requirement on the tuning of the local oscillator, while Equation (15) is the condition on the refocusing of the probe. Intuitively, Equations (14) and (15) assure that σ_φ^2 at the denominator of $\mathcal{F}(\varphi)$ does not grow as fast as $\partial_\varphi \sigma_\varphi^2$ at the numerator: instead, we can see in Appendix C that, when these conditions hold, the variance σ_φ^2 becomes of the order $1/N$, while its derivative $\partial_\varphi \sigma_\varphi^2$ remains constant for large N . In particular, we show in Appendix C that the Fisher information in Equation (9) asymptotically reads

$$\mathcal{F}(\varphi) \sim 8\varrho(k, \ell)(\partial_\varphi \gamma_\varphi)^2 N^2, \quad (16)$$

proving the achievement of the Heisenberg scaling, with

$$\varrho(k, \ell) = \left(\frac{8k}{1 + 16k^2 + 4\ell} \right)^2 \quad (17)$$

positive factor reaching its maximum value for $k = \pm 1/4$ and $\ell = 0$, namely $\varrho(1/4, 0) = 1$.

Compared with the conditions found in the literature for single-parameter Gaussian estimation schemes based on squeezed-vacuum probes which, adjusted to the notation employed so far, can be translated into $\gamma_\varphi - \theta = \tan^{-1} e^{2r} \sim \pi/2 + (4N)^{-1}$ and $P_\varphi = 1$ [26,27], we see that Equations (14) and (15) achieve two important further results

- It is possible to loosen the optimal conditions found in literature, which still allow us to reach the Heisenberg scaling, at the price of a multiplying factor $0 < \varrho(k, \ell) \leq 1$ which does not depend on N and hence does not ruin the scaling of the precision;
- These conditions are explicitly expressed in terms of the average number N of photons in the probe and, therefore, in terms of the precision we want to achieve. In Section 2.3, we will discuss how this allows us to assess the precision needed to engineer suitable auxiliary stages V_{in} and V_{out} to reach the Heisenberg scaling, showing that it is possible to avoid an iterative adaptation of the optical network.

Lastly, we recall that it is always possible to asymptotically saturate the Cramér-Rao bound in Equation (8) in the limit $\nu \rightarrow +\infty$ of samples with a large number ν of observations. In particular, the maximum-likelihood estimator $\tilde{\varphi}_{\text{MLE}}$ is an asymptotically efficient and Gaussian estimator which can be obtained through the maximisation of the Likelihood function

$$\mathcal{L}(\varphi; \mathbf{x}) = \prod_{i=1}^{\nu} p_\varphi(x_i) = \frac{1}{(2\pi\sigma_\varphi^2)^{\nu/2}} \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma_\varphi^2}\right), \quad (18)$$

associated with the set $\{x_i\}_{i=1, \dots, \nu}$ of the ν measurement outcomes of the quadrature field \hat{x}_θ [44,45]. In Appendix D we see that the non-trivial solution which maximises

the Likelihood function $\mathcal{L}(\varphi; \mathbf{x})$ in Equation (18) is simply given by the estimator $\tilde{\varphi}_{\text{MLE}}$ satisfying

$$\sigma(\tilde{\varphi}_{\text{MLE}})^2 := \sigma_{\tilde{\varphi}_{\text{MLE}}}^2 = S(\mathbf{x})^2, \quad (19)$$

where $\sigma^2(\varphi)$ is the variance σ_{φ}^2 in Equation (6) as a function of φ , and $S(\mathbf{x})^2$ is the usual sample variance

$$S(\mathbf{x})^2 = \frac{1}{\nu} \sum_{i=1}^{\nu} x_i^2. \quad (20)$$

Generally, Equation (19) cannot be solved analytically, so that numerical methods need to be employed to find non-trivial solutions. Nevertheless, it is possible to find some exceptions, particularly for elementary functional dependencies of P_{φ} and γ_{φ} on the unknown parameter φ . For example, in the case for which $P_{\varphi} \equiv P$ is independent of φ , and the functional dependence of $\gamma(\varphi) := \gamma_{\varphi}$ on φ of the phase acquired by the probe is invertible, the function $\sigma(\varphi)^2$ in Equation (19) can be easily inverted as well, and the maximum-likelihood estimator reads

$$\tilde{\varphi}_{\text{MLE}}(\mathbf{x}) = \gamma^{-1} \left(\theta + \frac{1}{2} \left(2n\pi \pm \arccos \left(\frac{(2S(\mathbf{x})^2 - 1) - 2P_{\varphi} \sinh^2 r}{2P_{\varphi} \sinh r \cosh r} \right) \right) \right). \quad (21)$$

We can notice how, due to the presence of the cosine in σ_{φ}^2 in Equation (6), some prior knowledge on the parameter φ is required in order to correctly choose the invertibility interval for $\cos(2\gamma(\varphi) - 2\theta)$ —i.e., to choose the correct value of $n \in \mathbb{N}$ and the sign of the arccos function in Equation (21). In the next section, we will see how a classical prior knowledge of the parameter φ is required to satisfy condition (15), achievable with a prior coarse estimation reaching an uncertainty of the order of $1/\sqrt{N}$. Such prior knowledge on the parameter, for a large enough N , can be employed to choose the correct invertibility interval.

2.3. Conditions for the Heisenberg Scaling

We can see that both conditions in Equations (14) and (15) are φ -dependent, suggesting that an adaptive procedure must take place in order to employ the estimation scheme described in Section 2.1, as it is customary for ab-initio Gaussian estimation strategies [24,27,35–37]. However, some considerations can be made in this regard.

Condition (14) fixes the phase of the quadrature \hat{x}_{θ} which needs to be measured. The quantity γ_{φ} is in fact the phase acquired by the squeezed vacuum during the interferometric evolution from the first input port to the first output port, and for $\gamma_{\varphi} \equiv \varphi$ Equation (14) resembles the condition $\theta = \varphi + \tan^{-1} e^{2r}$ found in the literature for single-phase estimation [24]. On the other hand, condition (14) is a looser condition to reach the Heisenberg scaling, and it puts in relation the precision with which we are able to choose the phase θ of the local oscillator—given by the resolution of the homodyne detection apparatus—with the precision achievable in the estimation of φ . In particular, it is evident how the minimum resolution for the homodyne detector required to reach an uncertainty $\delta\varphi$ of order $1/N$ must be, in turn, of order $1/N$. This is in agreement with the common notion in metrology for which a sensor cannot detect changes in the quantity that is being measured which are smaller than its resolution.

Interestingly, we notice from Equation (14) that the constant k cannot be equal to zero. Counterintuitively, the value $k = 0$ coincides with the choice of measuring the quadrature $\hat{x}_{\gamma_{\varphi} + \pi/2}$, namely the minimum-variance quadrature after the squeezed vacuum undergoes a phase-shift of magnitude γ_{φ} , i.e., after the interferometric evolution given by \hat{u}_{φ} in Equation (3). This apparent incongruity can be explained by observing the expression of $\partial_{\varphi} \sigma_{\varphi}^2$ in Equation (10). Differently from displacement-encoding approaches—in which the information on the parameter is obtained from the transformation of the displacement of the probe, and thus minimizing the noise of the signal is always the optimal choice—here the value of φ is encoded in the variance of the quadrature itself. For us to be able to extract information on the parameter, the variance of the signal needs to be sensitive

to small variations in φ —i.e., the derivative $\partial_\varphi \sigma_\varphi^2$ must be non-vanishing. Of the two contributions of $\partial_\varphi \sigma_\varphi^2$ in Equation (10), the one originating from the variations in P_φ is identically vanishing when condition (15) is satisfied, since $\partial_\varphi P_\varphi \simeq 0$ for P_φ close to its maximum. The remaining contribution derives from the variations in the overall phase γ_φ , and the variance of the maximally squeezed quadrature $\hat{x}_{\gamma_\varphi + \pi/2}$ is a stationary point with respect to variations in the phase γ_φ and is thus insensitive to φ , namely $\partial_\gamma \sigma_\varphi^2 = 0$ for $\gamma_\varphi - \theta = \pi/2$ in Equation (11) (See Figure 3).

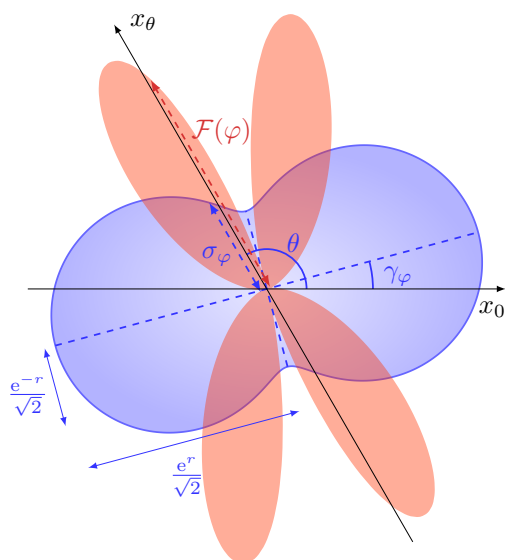


Figure 3. Polar plot of the standard deviation σ_φ (see Equation (6)) in blue, and of the Fisher information $\mathcal{F}(\varphi)$ in Equation (16) in orange as functions of the phase θ of the quadrature \hat{x}_θ measured, for $P_\varphi = 1$. The large values of $\mathcal{F}(\varphi)$ are reached for θ , satisfying condition (14). Interestingly, for $\theta = \gamma_\varphi \pm \pi/2$, namely when measuring the quadrature with minimum variance, σ_φ reaches its minimum, but the Fisher information drops to zero: as a squeezing-encoding estimation scheme, this model relies on the information about φ inscribed in the variance of the quadrature measured. On the other hand, the minimum variance is a stationary point as a function of φ , and thus is locally insensitive to the variations of the parameter. Reprinted with permission from ref. [34], © 2021 The Author(s).

Condition (15) is the requirement that most of the photons injected into the network end up in the observed output channel. In fact, this condition can be rewritten in terms of the average number of photons that are not correctly refocused $N(1 - P_\varphi) \sim \ell$. Thus, condition (15) tells us that the number of photons which are not observed must be a constant ℓ , not growing with N . In other words, this condition assures that most of the information on φ encoded in the probe is not lost in channels that are not observed. As a matter of fact, we can see from Equation (6) that the variance of the observed quadrature \hat{x}_θ after the interferometric evolution is the convex combination of the variances of a squeezed vacuum and the pure vacuum, with coefficients P_φ and $1 - P_\varphi$, respectively. In order for this variance to be ‘squeezed’, in the sense that it is of order $1/N$, the contribution from the pure vacuum must be of order $1/N$, namely $1 - P_\varphi = \mathcal{O}(1/N)$.

This condition can also be seen as a requirement of the performance of the refocusing stage \hat{V}_{out} . In fact, in order to satisfy condition (15) for a given choice of \hat{V}_{in} , the auxiliary stage \hat{V}_{out} must be chosen so that $|V_{\text{out}} U_\varphi V_{\text{in}}|^2 \sim 1 - \ell/N$. As discussed earlier, this implies that, in general, the auxiliary stage \hat{V}_{out} which satisfies this condition depends on the value of the parameter itself, requiring an adaptive approach to find an optimal refocusing stage to reach the Heisenberg scaling. We show now that the information on φ required to engineer an adequate refocusing stage to reach the Heisenberg scaling can be obtained through a classical estimation strategy, namely that which achieves the scaling $1/\sqrt{N}$

typical of the shot-noise limit. This result is due to the structure of $P_\varphi = |(V_{\text{out}}U_\varphi V_{\text{in}})_{11}|^2$, which is essentially a transition probability $P = |\mathbf{v}_{\text{out}} \cdot \mathbf{v}_{\text{in}}|^2$ between the unitary vectors $\mathbf{v}_{\text{in}} = U_\varphi V_{\text{in}} \mathbf{e}_1$ and $\mathbf{v}_{\text{out}} = V_{\text{out}}^\dagger \mathbf{e}_1$, with $\mathbf{e}_1 = (1, 0, \dots, 0)^T$. Hence, a small tilt of order $O(1/\sqrt{N})$ between the unit vectors \mathbf{v}_{in} and \mathbf{v}_{out} yields a quadratic reduction in their transition probability. To prove this, we will call φ_{cl} the rough guess of the value of φ that is sufficiently precise to engineer a refocusing stage \hat{V}_{out} which satisfy condition (15), and we will show that the estimation strategy to obtain this rough estimate of φ is classical, namely that the error $\delta\varphi = \varphi - \varphi_{\text{cl}}$ associated with the prior rough estimation is allowed to be of order $1/\sqrt{N}$. For a given choice of \hat{V}_{in} and ℓ , we will call $\hat{V}_{\text{out}}(\varphi_{\text{cl}})$ a solution of Equation (15). The single-photon transition probability P_φ appearing in this condition can be written as the squared complex modulus of the scalar product of two M -dimensional complex vectors $U_\varphi V_{\text{in}} \mathbf{e}_1$ and $V_{\text{out}}(\varphi_{\text{cl}})^\dagger \mathbf{e}_1$, with $\mathbf{e}_1 = (1, 0, \dots, 0)^T$. We can then write

$$P_\varphi = \left| \mathbf{e}_1^T V_{\text{out}}(\varphi_{\text{cl}}) U_\varphi V_{\text{in}} \mathbf{e}_1 \right|^2 \equiv \eta(\varphi, \varphi_{\text{cl}}), \quad (22)$$

where the transition probability $\eta(\varphi, \varphi_{\text{cl}})$ is a smooth function of φ and φ_{cl} , with a locus of points of maxima along the condition $\varphi_{\text{cl}} = \varphi$, since, for a perfect knowledge of the parameter the auxiliary stage, $\hat{V}_{\text{out}}(\varphi_{\text{cl}} = \varphi)$ would satisfy $\eta(\varphi, \varphi) = 1$. If the prior estimation φ_{cl} slightly deviates from the real value of the parameter $\varphi = \varphi_{\text{cl}} + \delta\varphi$, we can write the expansion

$$\begin{aligned} P_\varphi &= \eta(\varphi, \varphi - \delta\varphi) \\ &= 1 - \underbrace{\frac{\partial \eta(\varphi, x)}{\partial x} \Big|_{x=\varphi}}_0 \delta\varphi + \frac{1}{2} \frac{\partial^2 \eta(\varphi, x)}{\partial x^2} \Big|_{x=\varphi} \delta\varphi^2 + \mathcal{O}(\delta\varphi^3), \end{aligned} \quad (23)$$

where the derivative of $\eta(\varphi, \varphi_{\text{cl}})$ is zero along the condition $\varphi = \varphi_{\text{cl}}$. We can see, comparing Equations (23) and (15), that an error $\delta\varphi$ of order $1/\sqrt{N}$ suffices to correctly engineer a refocusing stage \hat{V}_{out} that allows for the Heisenberg scaling. It is then possible to conceive two-step ab initio protocols exploiting the model presented in this section: a first, coarse, classical estimation of the parameter φ is performed and the rough estimate φ_{cl} is obtained, with an error $\delta\varphi = \varphi - \varphi_{\text{cl}} = O(1/\sqrt{N})$ of the same order of the shot-noise limit. Then, the classical information obtained on φ can be employed to engineer the refocusing stage \hat{V}_{out} , once \hat{V}_{in} is fixed, so that the overall network satisfies condition (15), allowing us to reach the Heisenberg scaling through the quantum strategy described in Section 2.1.

Lastly, we notice that, in order to satisfy condition (15), it is also possible to optimize the input auxiliary stage \hat{V}_{in} while arbitrarily fixing the refocusing stage \hat{V}_{out} . In such a case, identical considerations can be made regarding the possibility of a two-step protocol, since the optimization \hat{V}_{in} still requires only a classical coarse estimation φ_{cl} of the parameter. Interestingly, only one of the auxiliary stages needs to be optimized, and thus depends on φ_{cl} , whether it is \hat{V}_{in} or \hat{V}_{out} . This leaves the choice of the second auxiliary stage completely arbitrary, notwithstanding that the pre-factor $(\partial_\varphi \gamma_\varphi)^2$ appearing in the Fisher information in Equation (16) is not vanishing. Indeed, the condition $(\partial_\varphi \gamma_\varphi)^2 = 0$ corresponds to the situation in which the optimized network $\hat{u}_\varphi = \hat{V}_{\text{out}}(\varphi_{\text{cl}}) \hat{U}_\varphi \hat{V}_{\text{in}}$ acts trivially, namely without imprinting any information about φ , on the probe. Remarkably, it has been shown that, for a random choice of the non-optimized auxiliary network, sampled uniformly within the set of all the possible linear networks, the pre-factor $(\partial_\varphi \gamma_\varphi)^2$ multiplying the scaling N^2 in the Fisher information is typically different from zero [34]. In other words, within certain non-restrictive regularity conditions and for linear networks with a large enough number of channels, the value of the pre-factor becomes essentially unaffected by the choice of the non-optimised auxiliary network. This important feature can be exploited for experimental applications, for example, employing the arbitrary non-optimised stage to manipulate the information encoded into the probe regarding the structure of a linear

network with multiple unknown parameters, ultimately allowing the choice of functions of such parameters to be estimated at the Heisenberg scaling sensitivity [46].

2.4. A Two-Channel Network

In this section, we will apply our model for the estimation of distributed parameters to a particular example of a 2-channel network, in which the unknown parameter φ influences the reflectivity η_φ of a beam-splitter and the magnitudes λ_φ and λ'_φ of two phase-shifts (see Figure 4). We can think of the global parameter φ as an external physical property, such as the temperature or the magnitude of the electromagnetic field, affecting the components of the network \hat{U}_φ . We will suppose that the functional dependence of the phase-shifts λ_φ , λ'_φ and of the reflectivity η_φ on the true value of the parameter φ are known and smooth, whether given by some law of nature or opportunely engineered. With reference to Figure 4, we can write the matrices representing the action of the beam-splitter and the phase-shifts as

$$U_{BS}(\eta_\varphi) = \exp(i\eta_\varphi\sigma_2) = \begin{pmatrix} \cos \eta_\varphi & \sin \eta_\varphi \\ -\sin \eta_\varphi & \cos \eta_\varphi \end{pmatrix}$$

$$U_{PS}(\lambda_\varphi, \lambda'_\varphi) = \exp\left(i\frac{\lambda_\varphi + \lambda'_\varphi}{2}\mathbb{I}_2 + i\frac{\lambda_\varphi - \lambda'_\varphi}{2}\sigma_3\right) = \begin{pmatrix} \exp(i\lambda_\varphi) & 0 \\ 0 & \exp(i\lambda'_\varphi) \end{pmatrix} \quad (24)$$

respectively, where σ_i , $i = 1, 2, 3$, is the i -th Pauli matrix and \mathbb{I}_2 is the 2×2 identity matrix, so that the network \hat{U}_φ is represented by the matrix

$$U_\varphi = U_{PS}(\lambda_\varphi, \lambda'_\varphi)U_{BS}(\eta_\varphi) = \begin{pmatrix} \cos \eta_\varphi \exp(i\lambda_\varphi) & \sin \eta_\varphi \exp(i\lambda_\varphi) \\ -\sin \eta_\varphi \exp(i\lambda'_\varphi) & \cos \eta_\varphi \exp(i\lambda'_\varphi) \end{pmatrix}. \quad (25)$$

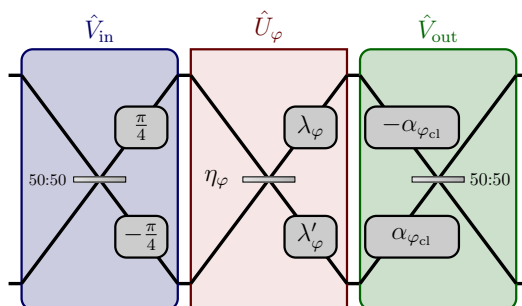


Figure 4. Schematic diagram of the two-channel network described in Section 2.4. The linear network \hat{U}_φ is composed of a beam splitter with coefficient η_φ and two phase-shifts of magnitudes λ_φ and λ'_φ . The auxiliary stage \hat{V}_{in} at the input is φ -independent, while the output stage \hat{V}_{out} is optimized after a classical prior estimation φ_{cl} of the parameter. In particular, the quantity $\alpha_{\varphi_{cl}} = (\lambda_{\varphi_{cl}} - \lambda'_{\varphi_{cl}})/2 - \pi/4$ depends on φ_{cl} only through the phase-shifts $\lambda_{\varphi_{cl}}$ and $\lambda'_{\varphi_{cl}}$. Reprinted with permission from ref. [42], © 2021 The Author(s).

We easily notice that $|(U_\varphi)_{11}|^2 = \cos^2 \eta_\varphi$, which, in general, is different from one and thus does not satisfy the condition (15), with the exception of the two values $\eta_\varphi = 0, \pi$ which correspond to the absence of the mixing between the two modes. As described in the model earlier, we then add two auxiliary stages V_{in} and $V_{out} \equiv V_{out}(\varphi_{cl})$, of which only one depends on a prior coarse estimation φ_{cl} of φ realised with a classical strategy, so that $\varphi - \varphi_{cl} = \delta\varphi = \mathcal{O}(1/\sqrt{N})$. In particular we choose as input stage

$$V_{in} = U_{PS}\left(\frac{\pi}{4}, -\frac{\pi}{4}\right)U_{BS}\left(\frac{\pi}{4}\right) = \frac{1}{\sqrt{2}} \begin{pmatrix} \exp(i\frac{\pi}{4}) & \exp(i\frac{\pi}{4}) \\ -\exp(-i\frac{\pi}{4}) & \exp(-i\frac{\pi}{4}) \end{pmatrix}, \quad (26)$$

and as output stage

$$V_{\text{out}} = U_{\text{BS}}\left(\frac{\pi}{4}\right)U_{\text{PS}}(-\alpha_{\varphi_{\text{cl}}}, \alpha_{\varphi_{\text{cl}}}) = \frac{1}{\sqrt{2}} \begin{pmatrix} \exp(-i\alpha_{\varphi_{\text{cl}}}) & \exp(i\alpha_{\varphi_{\text{cl}}}) \\ -\exp(-i\alpha_{\varphi_{\text{cl}}}) & \exp(i\alpha_{\varphi_{\text{cl}}}) \end{pmatrix}, \quad (27)$$

where $\alpha_{\varphi_{\text{cl}}} = (\lambda_{\varphi_{\text{cl}}} - \lambda'_{\varphi_{\text{cl}}})/2 - \pi/4$ is a quantity which can be obtained through a classical estimation φ_{cl} of φ . A straightforward calculation of $P_{\varphi} = |(V_{\text{out}}(\varphi_{\text{cl}})U_{\varphi}V_{\text{in}})_{11}|^2$ yields

$$P_{\varphi} = \frac{1}{2} \left(1 + \sin(\lambda_{\varphi} - \lambda'_{\varphi} - 2\alpha_{\varphi_{\text{cl}}})\right) = \frac{1}{2} (1 + \cos(\delta\lambda - \delta\lambda')), \quad (28)$$

where $\delta\lambda = \lambda_{\varphi} - \lambda_{\varphi_{\text{cl}}}$ and $\delta\lambda' = \lambda'_{\varphi} - \lambda'_{\varphi_{\text{cl}}}$ are the error in the estimates of λ_{φ} and λ'_{φ} due to the imprecision of the classical estimation φ_{cl} . We can then easily see that P_{φ} in Equation (28) satisfies condition (15), since both the errors $\delta\lambda$ and $\delta\lambda'$ are of order $1/\sqrt{N}$,—i.e., $\delta\lambda = (\partial_{\varphi}\lambda_{\varphi})\delta\varphi = \mathcal{O}(1/\sqrt{N})$, and similarly for $\delta\lambda'$ —and thus we obtain from Equation (28)

$$P_{\varphi} \sim 1 - \frac{1}{4}(\partial_{\varphi}\lambda_{\varphi} - \partial_{\varphi}\lambda'_{\varphi})^2\delta\varphi^2. \quad (29)$$

In order to evaluate the Fisher information in Equation (16), we need to calculate both the phase acquired by the probe throughout the whole interferometric evolution γ_{φ} , and the coefficient ℓ . The phase γ_{φ} is easily obtained as the complex phase of $(u_{\varphi})_{11} \equiv (V_{\text{out}}(\varphi_{\text{cl}})U_{\varphi}V_{\text{in}})_{11}$

$$\gamma_{\varphi} = \frac{\lambda_{\varphi} + \lambda'_{\varphi}}{2} + \eta_{\varphi} + \frac{\pi}{2}. \quad (30)$$

Since $\delta\varphi = \mathcal{O}(1/\sqrt{N})$, we call h the finite N -independent constant such that $\delta\varphi \sim h/\sqrt{N}$. The transition probability P_{φ} can then be written as

$$P_{\varphi} \sim 1 - \frac{h^2}{4}(\partial_{\varphi}\lambda_{\varphi} - \partial_{\varphi}\lambda'_{\varphi})^2 \frac{1}{N}, \quad (31)$$

so that the factor $\ell = h^2(\partial_{\varphi}\lambda_{\varphi} - \partial_{\varphi}\lambda'_{\varphi})^2/4$ appearing in the Fisher information in Equation (16) is easily evaluated comparing Equations (15) and (31). The Fisher information can be obtained from Equation (16), with $\partial_{\varphi}\gamma_{\varphi}$ given by Equation (30), and $\varrho(k, \ell)$ given by Equation (17), with k given by the condition on the local oscillator phase and $\ell = h^2(\partial_{\varphi}\lambda_{\varphi} - \partial_{\varphi}\lambda'_{\varphi})^2/4$.

We notice from the expression of $\alpha_{\varphi_{\text{cl}}} = (\lambda_{\varphi_{\text{cl}}} - \lambda'_{\varphi_{\text{cl}}})/2 - \pi/4$ that the unknown reflectivity of the beam splitter η_{φ} does not influence the refocusing stage $\hat{V}_{\text{out}}(\varphi_{\text{cl}})$, but it appears in the phase γ_{φ} acquired by the probe in Equation (30). In particular, if the two phases λ_{φ} and λ'_{φ} are vanishing, the dependence of $\alpha_{\varphi_{\text{cl}}}$, and thus of the refocusing stage $\hat{V}_{\text{out}}(\varphi_{\text{cl}})$, on the classical estimation φ_{cl} of the parameter disappears completely. In other words, this network for $\lambda_{\varphi} = \lambda'_{\varphi} = 0$ transforms the reflectivity η_{φ} of a beam splitter into the magnitude of a phase shift, independently from η_{φ} .

3. Quantum Estimation Based on Multi-Homodyne Measurements

In Section 2 we have presented a scheme for the estimation of a distributed parameter encoded in a multi-channel network, reaching the Heisenberg scaling employing a squeezed vacuum state and homodyne detection performed at a single output channel. In particular, in Section 2.2, we have discussed in depth about the conditions in Equations (14) and (15) which need to be satisfied to reach the Heisenberg scaling: Equation (14) imposes a minimum resolution in tuning the local oscillator phase, in order to infer the value of the parameter from the noise of a sufficiently squeezed quadrature. Equation (15) is a requirement on the refocusing of the probe into the only observed channel and, in order to be satisfied, a classical knowledge of the parameter is generally required to engineer the

optimal refocusing network. A natural question that arises is whether it is possible to ease these conditions by carrying out some changes on our scheme. In particular, what would happen if homodyne measurements were performed, not only at a single channel, but at all the output ports of the network instead? Would condition (15) become looser, allowing us to engineer the optimal auxiliary stages with even less information on the unknown parameter? Moreover, as we have already discussed in Section 1, a non-vanishing displacement in the probe would make it possible to infer the value of the parameter directly from the average of the quadrature with minimum variance, and not from its noise, which may be a more feasible approach in particular experimental scenarios. In this section, we will investigate a model that implements these changes (see Figure 5). We will see that, with these assumptions, not only is there no need to perform a prior estimation of the parameter to optimize the network, but the Heisenberg scaling can be achieved without employing an auxiliary stage in the first place. Moreover, the presence of displacement in the probe will allow us to perform the estimation directly measuring the minimum-variance quadrature, relying on the information about the parameter encoded in the average signal of the homodyne, and not in its noise.

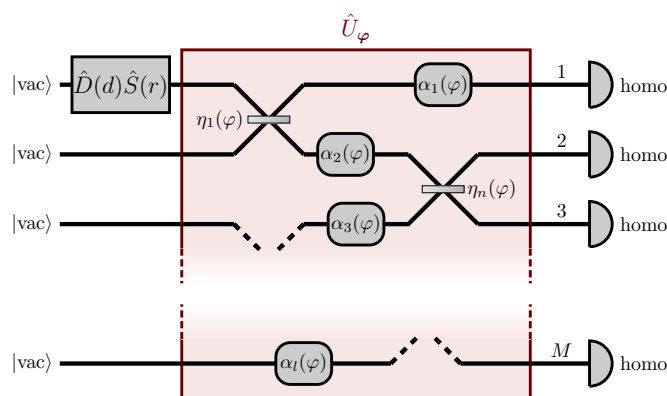


Figure 5. Scheme of the setup described in Section 3. A squeezed coherent state is injected in the first input port of a network \hat{U}_φ which depends on a parameter φ that is generally distributed among multiple components of the network. Homodyne detection is performed at each of the output ports. Differently from the setup in Figure 2, no auxiliary stage is required to reach the Heisenberg scaling. Reprinted with permission from ref. [43], © 2022 The Author(s).

3.1. Setup

We will consider a linear and passive network \hat{U}_φ , which depends on a single and generally distributed parameter, for example affecting several components of the network, as shown in Figure 1. Once again, \hat{U}_φ admits a unitary matrix representation U_φ obtained through Equation (1). Differently from Section 2 though, we consider as a probe the single-mode squeezed state

$$|\psi_0\rangle = |\alpha, r\rangle \equiv \hat{D}(\alpha)\hat{S}(r)|\text{vac}\rangle, \quad \alpha = (\alpha, 0, \dots, 0), \quad r = (r, 0, \dots, 0), \quad (32)$$

with $\hat{D}(\alpha) = e^{\alpha(\hat{a}_1^\dagger - \hat{a}_1)}$, i.e., a squeezed coherent state with a mean number of photons $N = N_D + N_S = \alpha^2 + \sinh^2 r$, $\alpha, r > 0$, where we can introduce the displacement \mathbf{d} , given by $\alpha_i = (d_i + id_{i+M})/\sqrt{2}$, $i = 1, \dots, M$. We remark that the choice $\alpha, r > 0$ is a specific (φ -independent) condition which is required in displacement-encoding approaches: squeezing and displacement can, in general, have different complex phases, and the condition $\alpha, r > 0$ assures that the squeezed quadrature ($\hat{x}_{\pi/2}$ for $r > 0$) is orthogonal—i.e., conjugated—to the displaced quadrature (\hat{x}_0 for $\alpha > 0$), and hence it is the most sensible to the presence of phases. Moreover, in our model, we will consider homodyne detection in all M output channels of the linear network, so that M quadrature fields \hat{x}_{i,θ_i} are measured, where θ_i is the phase of the i -th local oscillator, $i = 1, \dots, M$.

Since we are observing all the output ports of the network, but a non-vanishing number of photons is injected only in the first channel, only the first column of the unitary matrix U_φ is relevant in our model, consisting of the transition amplitude of single photons from the first to every channel of the network (see Equation (1)). We can thus employ the parametrisation

$$(U_\varphi)_{i1} = \sqrt{P_i} e^{i\gamma_i}, \quad (33)$$

where P_i is the probability that a single photon is transmitted through the linear network from the first to the i -th channel, and γ_i is the phase that it would acquire (In order to keep the notation lighter, we are dropping the subscript φ in P_i , γ_i , μ and Σ . Nonetheless, it rests assured that these quantities, in general, depend on the unknown parameter). Once again, due to the Gaussian nature of the model, the (joint) probability distribution associated with the outcome x of the M homodyne detectors is also Gaussian and reads [18–21]

$$p_\varphi(x) = \frac{1}{\sqrt{(2\pi)^M \det[\Sigma]}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right). \quad (34)$$

In Equation (34), both the covariance matrix Σ and the mean μ depend on the parameter φ . The elements of the covariance matrix Σ are evaluated in Appendix A, and read

$$(\Sigma)_{ij} = \frac{\delta_{ij}}{2} + \sqrt{P_i P_j} (\cos(\tilde{\gamma}_i - \tilde{\gamma}_j) \sinh(r)^2 + \cos(\tilde{\gamma}_i + \tilde{\gamma}_j) \cosh(r) \sinh(r)), \quad (35)$$

where δ_{ij} is the Kronecker delta, $\tilde{\gamma}_i = \gamma_i - \theta_i$ is the phase acquired at the output of the i -th channel relative to the correspondent local oscillator, and μ is the mean vector (see Appendix A)

$$\mu_i = d \sqrt{P_i} \cos \tilde{\gamma}_i, \quad i = 1, \dots, M. \quad (36)$$

The determinant $\det[\Sigma]$ can also be written in compact form (see Appendix A)

$$\det[\Sigma] = \frac{1}{2^M} + \frac{\sinh(r)}{2^{M-1}} \sum_{i=1}^M P_i (\sinh(r) + \cos(2\tilde{\gamma}_i) \cosh(r)) - \frac{\sinh^2(r)}{2^{M-2}} \sum_{i=1}^M \sum_{j=i+1}^M P_i P_j \sin^2(\tilde{\gamma}_i - \tilde{\gamma}_j). \quad (37)$$

For a multivariate Gaussian distribution of the form shown in Equation (34), the Fisher information can be easily evaluated from its definition in Equation (7) and employ the expression of the probability distribution $p_\varphi(x)$ in Equation (34) as (see Appendix B)

$$\mathcal{F}(\varphi) = \underbrace{\frac{1}{\det[\Sigma]} \partial_\varphi \mu^T C \partial_\varphi \mu}_{\mathcal{F}_D(\varphi)} + \underbrace{\frac{1}{2} \left(\frac{\partial_\varphi \det[\Sigma]}{\det[\Sigma]} \right)^2 - \frac{1}{2 \det[\Sigma]} \text{Tr}[(\partial_\varphi \Sigma)(\partial_\varphi C)]}_{\mathcal{F}_S(\varphi)} \quad (38)$$

where $C = \det[\Sigma] \Sigma^{-1}$ is the cofactor matrix of Σ and $\text{Tr}[A] = \sum_i A_{ii}$ denotes the trace of the matrix A . Compared with the Fisher information shown in Equation (9) for the model described in Section 2, the Fisher information for this setup includes an additional term $\mathcal{F}_D(\varphi)$ which depends on the derivative of the average μ with respect to the parameter φ . Moreover, the contribution $\mathcal{F}_S(\varphi)$, representing the information on φ encoded in the covariance matrix Σ , can be split into two terms, of which the first resembles the Fisher information in Equation (9) once we perform the substitution $\sigma_\varphi^2 \rightarrow \det[\Sigma]$: we will show in the following that, in the asymptotic regime of large N_S , this is the only contribution of $\mathcal{F}_S(\varphi)$ which, besides $\mathcal{F}_D(\varphi)$, reaches the Heisenberg scaling.

3.2. Heisenberg Scaling

Similarly to what happens to the setup described in Section 2, the Fisher information in Equation (38) does not generally reach the Heisenberg scaling unless certain conditions are met. To show this, it is necessary to evaluate all the contributions of $\mathcal{F}(\varphi)$ in terms of the number of photons N_D and N_S in the asymptotic regime. First, it is convenient to express the cofactor matrix C explicitly in terms of the squeezing parameter r . In Appendix B, we see that a closed form for C in such terms exists

$$C_{ss} = \frac{1}{2^{M-1}} + \frac{1}{2^{M-2}} \sum_{\substack{i=1 \\ i \neq s}}^M (\Sigma_{ii} - \frac{1}{2}) - \frac{1}{2^{M-3}} \sum_{\substack{i=1 \\ i \neq s}}^M \sum_{\substack{j=i+1 \\ j \neq s}}^M S_{ijj}, \quad (39a)$$

$$C_{st} = -\frac{1}{2^{M-2}} \Sigma_{st} + \frac{1}{2^{M-3}} \sum_{\substack{i=1 \\ i \neq s,t}}^M S_{sti}, \quad s \neq t, \quad (39b)$$

where

$$S_{sti} = \sinh^2 r \sqrt{P_s P_t} P_i \sin(\gamma_s - \gamma_i) \sin(\gamma_t - \gamma_i), \quad (40)$$

and Σ_{ij} are shown in Equation (35). We then notice that, in order to analyse the generic asymptotic behaviour of the Fisher information in Equation (38), it suffices to separately examine the asymptotics of the terms $\boldsymbol{\mu}$, Σ_{ij} , S_{sti} , $\det[\Sigma]$, and of their derivatives.

From Equations (35)–(37), (39) and (40) we see that Σ , S and $\det[\Sigma]$ are all of order N_S in general, while $\boldsymbol{\mu}$ is of order $\sqrt{N_D}$. The same asymptotic behaviours are also kept for their respective derivatives, since both $\partial_\varphi \bar{\gamma}_i \equiv \partial_\varphi \gamma_i$ and $\partial_\varphi P_i$ are independent of the probe, and thus of N_D and N_S . We can thus see from the expression of $\mathcal{F}(\varphi)$ in Equation (38) that its numerator grows at most with N^2 , while the denominator—i.e., the denominator $\det[\Sigma]$ —in general grows with N_S . Therefore, in order for $\mathcal{F}(\varphi)$ to reach the Heisenberg scaling, namely a scaling of order N^2 , some conditions must be imposed so that the denominators in $\mathcal{F}_D(\varphi)$ and $\mathcal{F}_S(\varphi)$ do not grow for large N_S . In Appendix C, it can be seen that, to achieve the Heisenberg scaling, the determinant $\det[\Sigma]$ must be of the order N^{-1} , similarly to what happens to σ_φ^2 in the single-homodyne setup in Section 2.2, and the conditions for this to occur are

$$\bar{\gamma}_i = \pm \frac{\pi}{2} + \mathcal{O}(N_S^{-1}), \quad i = 1, \dots, M. \quad (41)$$

When these conditions hold, we can introduce the finite (possibly vanishing) quantities $k_i = \lim_{N_S \rightarrow \infty} N_S(\gamma_i \mp \pi/2)$, so that the determinant $\det[\Sigma]$ reduces to (see Appendix C)

$$\det[\Sigma] \sim \frac{1}{2^{M-2} N_S} \left(\left(\sum_{i=1}^M P_i k_i \right)^2 + \frac{1}{16} \right), \quad (42)$$

while $\partial_\varphi \det[\Sigma]$, $\partial_\varphi \Sigma$, $\partial_\varphi C$ and C tend to constant values, and $\partial_\varphi \boldsymbol{\mu}$ scales as $\sqrt{N_D}$. We can easily see that this makes only the first two terms of $\mathcal{F}(\varphi)$ in Equation (38) dominant for large N_D and N_S . When conditions (41) are met, we can thus neglect the last term in Equation (38), and the Fisher information

$$\begin{aligned} \mathcal{F}(\varphi) &\sim \frac{1}{\det[\Sigma]} \partial_\varphi \boldsymbol{\mu}^T C \partial_\varphi \boldsymbol{\mu} + \frac{1}{2} \left(\frac{\partial_\varphi \det[\Sigma]}{\det[\Sigma]} \right)^2 \\ &\sim 8(\partial\gamma)_{\text{avg}}^2 \left(\zeta(k_{\text{avg}}) 2N_D N_S + \varrho(k_{\text{avg}}) N_S^2 \right), \end{aligned} \quad (43)$$

asymptotically reaches the Heisenberg scaling in $N_D N_S$ and N_S^2 (see Appendix C), where we introduced the quantities

$$k_{\text{avg}} \equiv \sum_{i=1}^M P_i k_i, \quad (\partial\gamma)_{\text{avg}} \equiv \sum_{i=1}^M P_i \partial_\varphi \gamma_i, \quad (44)$$

while

$$\zeta(x) = (16x^2 + 1)^{-1}, \quad \varrho(x) = (8x)^2 / (16x^2 + 1)^2 \quad (45)$$

are positive, even functions which reach their maxima at $x = \pm 1/4$ and $x = 0$, respectively, namely $\varrho(1/4) = 1$ and $\zeta(0) = 1$.

It is now possible to compare the Fisher information in Equation (43) achieved with the present scheme, with the Fisher information in Equation (16) obtained with the setup for distributed metrology employing a squeezed vacuum state and homodyne detection on a single channel. Since this setup employs a squeezed probe with a non-vanishing displacement, it lends itself to both displacement-based and squeezing-based estimation approaches. This is reflected by the presence of two separate contributions to the Fisher information in Equation (43): the first is given by the variations in the displacement μ , the second by the variations in the determinant $\det[\Sigma]$, with respect to changes in the value of φ . Both terms present a pre-factor $(\partial\gamma)_{\text{avg}}^2$ shown in Equation (44) which resembles the pre-factor $(\partial_\varphi\gamma_\varphi)^2$ in Equation (16) for the single-homodyne counterpart: in particular, $(\partial\gamma)_{\text{avg}}^2$ is a weighted average of the sensitivities of the complex phases γ_i to changes in the parameter φ , each one weighted by the ‘fraction’ P_i of the probe undergoing the phase shift γ_i . Indeed, we can see that $(\partial\gamma)_{\text{avg}}^2$ reduces to the single-homodyne counterpart when the whole probe is refocused into a single channel—say the first—so that $P_1 = 1$ and $P_i = 0$ for $i = 2, \dots, M$, i.e., $(\partial\gamma)_{\text{avg}}^2 = (\partial_\varphi\gamma_1)^2$.

In the term of the Fisher information associated with the squeezing-encoding in Equation (43), the factor $\varrho(k_{\text{avg}})$ coincides in turn with the function $\varrho(k, \ell)$ in Equation (17) for $\ell = 0$ —i.e., the ideal case of perfect refocusing of the probe, and all photons being observed—and after the substitution $k \rightarrow k_{\text{avg}}$, namely the weighted average of the coefficients k_i with the same weights P_i . In other words, this term can be thought of as a generalisation of the Fisher information for a single homodyne due to the presence of multiple observed channels. Noticeably, this term can be set as equal to zero only for $N_S = 0$, in which case, the whole expression in Equation (43) vanishes, ruining the Heisenberg scaling, or for $k_{\text{avg}} = 0$. In fact, we can see from Equation (37) that the condition $k_{\text{avg}} = 0$ —i.e., $\bar{\gamma}_i = \pm\pi/2$, namely when the quadratures with minimum variances are measured at each channel minimises $\det[\Sigma]$, which becomes a stationary point with respect to the variations in φ , and $\partial_\varphi \det[\Sigma]$ in Equation (43) vanishes. On the other hand, the first contribution to Equation (43) is instead a new term not present in the Fisher information for the single homodyne, deriving from the information on the parameter encoded in the displacement μ . This term achieves the Heisenberg scaling in $N_D N_S$, in the sense that it reaches the Heisenberg scaling in $N = N_D + N_S$ if both N_D and N_S grow with N , i.e., if $N_D = \beta N$ and $N_S = (1 - \beta)N$ with $0 \leq \beta < 1$ independently of N . The function $\zeta(k_{\text{avg}})$ in Equation (45) does not have roots, hence the first contribution to \mathcal{F} vanishes only for $N_D = 0$, i.e., for a squeezed vacuum as a probe.

Finally, we will now write the Likelihood function $\mathcal{L}(\varphi; \mathbf{x})$ for the setup described here, and discuss the maximum-likelihood estimator $\hat{\varphi}_{\text{MLE}}$ saturating the Cramér-Rao bound shown in Equation (8). After performing ν measurements of the field quadratures \hat{x}_{i,θ_i} , the Likelihood of the outcomes (x_1, \dots, x_ν) is given by

$$\mathcal{L}(\varphi; x_1, \dots, x_\nu) = \prod_{j=1}^{\nu} p_\varphi(x_j), \quad (46)$$

with $p_\varphi(\mathbf{x}_j)$ found in Equation (34). In Appendix D, we show that the maximum-likelihood estimator, a non-trivial solution of the maximisation of the Likelihood function in Equation (46), is implicitly given by the estimator $\tilde{\varphi}_{\text{MLE}}$ which satisfies the equation

$$0 = \frac{1}{2} \text{Tr} \left[\partial_\varphi \Sigma^{-1} \left(\nu \Sigma - \sum_{j=1}^{\nu} (\mathbf{x}_j - \boldsymbol{\mu})(\mathbf{x}_j - \boldsymbol{\mu})^\top \right) \right]_{\varphi=\tilde{\varphi}_{\text{MLE}}} + \left[(\partial_\varphi \boldsymbol{\mu})^\top \Sigma^{-1} \left(\nu \boldsymbol{\mu} - \sum_{j=1}^{\nu} \mathbf{x}_j \right) \right]_{\varphi=\tilde{\varphi}_{\text{MLE}}}, \quad (47)$$

where Σ is the covariance matrix in Equation (35), and $\boldsymbol{\mu}$ the mean vector in Equation (36). This equation cannot, in general, be solved analytically, hence numerical methods typically need to be in place to find $\tilde{\varphi}_{\text{MLE}}$. On the other hand, Equation (47) simplifies for certain particular cases. For example, when measuring all the minimum-variance quadratures so that $\tilde{\gamma}_i = \pm\pi/2$ in Equation (41), and the probabilities P_i are independent of φ , we can see from Equation (35) that $\partial_\varphi \Sigma = 0$. In this case, the Likelihood Equation becomes

$$0 = \left[(\partial_\varphi \boldsymbol{\mu})^\top \Sigma^{-1} \left(\boldsymbol{\mu} - \frac{1}{\nu} \sum_{j=1}^{\nu} \mathbf{x}_j \right) \right]_{\varphi=\tilde{\varphi}_{\text{MLE}}}, \quad (48)$$

where the term in the right-hand side can be seen as a linear combination of the quantities $\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}$, where $\tilde{\boldsymbol{\mu}} = \sum_{j=1}^{\nu} \mathbf{x}_j / \nu$ are estimators of the mean $\boldsymbol{\mu}$. We see here how the displacement of the probe allows us to perform the estimation of φ through the sample mean $\boldsymbol{\mu}$, i.e., the average of the outcomes of the homodyne measurements. On the other hand, for a squeezed vacuum such as a probe, $\boldsymbol{\mu} = 0$, so that Equation (47) becomes

$$0 = \frac{1}{2} \text{Tr} \left[\partial_\varphi \Sigma^{-1} \left(\Sigma - \frac{1}{\nu} \sum_{j=1}^{\nu} \mathbf{x}_j \mathbf{x}_j^\top \right) \right]_{\varphi=\tilde{\varphi}_{\text{MLE}}}, \quad (49)$$

where it is possible to recognise the sample covariance matrix $\tilde{\Sigma} = \sum_{j=1}^{\nu} \mathbf{x}_j \mathbf{x}_j^\top / \nu$, estimator of the covariance matrix Σ .

3.3. Conditions for the Heisenberg Scaling

For the model introduced in this section, some considerations regarding the conditions in Equation (41) can also be drawn, especially in light of the feature of the single-homodyne, squeezed-vacuum scheme discussed in Section 2.3.

In fact, these conditions do nothing but fix the phases θ_i of the quadratures \hat{x}_{i,θ_i} that need to be measured to achieve the Heisenberg scaling. Regarding the minimum resolution of the homodyne sensors, it appears that there is no evident advantage in this setup compared to the single-homodyne scheme, since the resolution required to tune the local oscillators of each channel is still of order $1/N$. On the other hand, introducing displacement in the probe allows, as previously discussed, the value of the parameter to be inferred from the information inscribed in the average of the measured quadratures $\boldsymbol{\mu}$. Therefore, it is possible with this scheme to exactly measure the minimum-variance quadratures $\hat{x}_{i,\gamma_i \pm \pi/2}$, a possibility that was prevented in the previous scheme due to the requirement $k \neq 0$ in Equation (14). Although the situation for which $k_i = 0$ for $i = 1, \dots, M$ sets to zero the contribution $\mathcal{F}_S(\varphi)$ in Equation (43), associated with the information encoded in the covariance matrix Σ , the first contribution $\mathcal{F}_D(\varphi)$ of the Fisher information still reaches the Heisenberg scaling.

Another interesting feature of this protocol, which differentiates it from its single-homodyne counterpart, is that it does *not* require any adaptation of the network: since every output channel is observed, no condition on the transition probabilities P_i is required. Therefore, not only there is no requirement for a φ -dependent auxiliary stage, but there is no need for an auxiliary stage to reach the Heisenberg scaling in the first place. However,

the precision in the estimation is still affected by the network through the terms k_{avg} and $(\partial\gamma)_{\text{avg}}$, which appear in the constant factors multiplying the scaling in the Fisher information in Equation (38). In fact, we can see from Equation (45) that these two quantities depend on the transition probabilities P_i and on the derivatives of the complex phases $\partial_\varphi\gamma_i$. In particular, $\mathcal{F}(\varphi)$ can be vanishing for exceptionally poorly conceived networks, for which $(\partial\gamma)_{\text{avg}} = 0$: for example, a network for which γ_i is independent of φ for all values of i such that $P_i \neq 0$, is associated with a vanishing factor $(\partial\gamma)_{\text{avg}}$, as we can see from its definition in Equation (44). In this case, adding a φ -independent auxiliary network V , either at the input or at the output of U_φ , would modify the transition amplitude of the overall interferometric evolution, and ultimately yield a non-vanishing value of $(\partial\gamma)_{\text{avg}}$.

4. Conclusions

The recent advances in quantum metrology made possible the realisation of protocols achieving super-sensitivity in the estimation of optical lengths, time delays and space-time distortions due to gravitational waves, refractive indices in given materials, density and thickness of biological samples up to the nanometer scale, temperatures, polarisations of light, magnitudes of fields and their gradients, and more. However, current quantum sensing technologies still present some limitations, whether caused by the requirement of adaptively optimising the estimation procedure to the unknown value of the parameter that needs to be measured, by the fragility of the metrological resources which yield super-sensitivity, or by the scarce operating range of the protocols, impracticalities arise when tackling the metrological schemes with quantum mechanics. Moreover, most standard approaches to distributed quantum metrology suffer from a lack of universal estimation schemes, namely those which can operate independently of the unknown parameter nature and value, and of the unitary evolution which encodes the value of the parameter into the probe employed for the estimation.

In this work, we have reviewed in detail two schemes which address these limitations [42,43]. By employing analyses based on the Cramér–Rao bound, i.e., the ultimate precision achievable for a given estimation scheme, and on the Fisher information, we were able to assess the super-sensitivity of various feasible metrological setups, always achievable in the regime of large statistical samples through the maximum-likelihood estimator. We showed that, without making any assumptions on the structure of the multi-channel passive and linear network encoding the unknown parameter, which can as well be distributed among multiple components of the interferometer (such as temperature affecting the network in a distributed manner) it is possible to relax the adaptivity of the network to a requirement of classical knowledge—i.e., not at the quantum limit—on the parameter. This prior knowledge serves to engineer an auxiliary stage needed to perform the super-sensitive estimation with a single squeezed-vacuum state and homodyne detection at a single output port. Moreover, the adaptivity on the network can actually be completely circumvented when all the output ports of the network are observed with homodyne detectors. We discussed how this allows us to conceive estimation schemes at the Heisenberg limit, which can be performed with a single optimization step, when a classical knowledge of the parameter is needed to engineer the auxiliary stage, or without any optimization at all, when no prior knowledge is required.

Author Contributions: Both authors contributed to the paper and the obtained results. The project was conceived and managed by V.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Office of Naval Research Global (N62909-18-1-2153).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We acknowledge our collaborators in this research P. Facchi, G. Gramegna and F. A. Narducci.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Probability Distributions from Homodyne Measurements

In this appendix, we will obtain the probability density functions which govern the outcomes of homodyne detections performed on a single-mode squeezed state $|\alpha, r\rangle = \hat{D}(\alpha)\hat{S}(r)|\text{vac}\rangle$ injected in the first port of a M -channel passive and linear network \hat{U} , with $\alpha = (\alpha, 0, \dots, 0) = (\sqrt{N_D}, 0, \dots, 0)$, and $r = (r, 0, \dots, 0)$, so that $N_S = \sinh^2 r$, with $\alpha, r > 0$, and $N = N_D + N_S$ sum of the average number of photons in the displacement and in the squeezing of the state $|\alpha, r\rangle$. We refer to refs. [18–21] for reviews on Gaussian states, and how to describe and manipulate them mathematically. The state $|\alpha, r\rangle$ is a Gaussian state which is completely described, through its Wigner distribution

$$W(z) = \frac{1}{\sqrt{(2\pi)^{2M} \det[\Gamma]}} \exp\left(-\frac{1}{2}(z-d)^T \Gamma^{-1}(z-d)\right), \quad (\text{A1})$$

by its vector of displacement $d = \sqrt{2}(\alpha, 0, \dots, 0)$, and its matrix of covariances $\Gamma = \text{diag}(e^{2r}, 1, \dots, 1, e^{-2r}, 1, \dots, 1)/2$. Moreover, passive and linear networks preserve the Gaussian nature of $|\alpha, r\rangle$. For this reason, we will first obtain in generality the final displacement d_U and covariance matrix Γ_U of the state $\hat{U}|\alpha, r\rangle$. Then, we will marginalise the Wigner distribution, rotated by the symplectic and orthogonal matrix

$$R(\theta) = \begin{pmatrix} \cos(\Theta) & \sin(\Theta) \\ -\sin(\Theta) & \cos(\Theta) \end{pmatrix}, \quad (\text{A2})$$

accordingly to the phases $\Theta = \text{diag}(\theta) = \text{diag}(\theta_1, \dots, \theta_M)$ of the local oscillators of the homodyne. In doing so, we will be able to obtain the expressions in Equations (6), and (35)–(37).

The action of a passive and linear unitary \hat{U} on a state is described by a symplectic rotation of the state Wigner distribution [18–21]. The Wigner distribution $W_U(z)$ of the state $\hat{U}|\alpha, r\rangle$ after the interferometric evolution \hat{U} is thus obtained, rotating the initial Wigner distribution $W(z)$ found in Equation (A1) with the orthogonal and symplectic matrix

$$R = \begin{pmatrix} \text{Re}(U) & -\text{Im}(U) \\ \text{Im}(U) & \text{Re}(U) \end{pmatrix}, \quad (\text{A3})$$

where U is the unitary matrix representing the transition amplitudes which can be found through Equation (1), thus obtaining the transformation

$$W_U(z) = W(R^T z). \quad (\text{A4})$$

Since the probe in our model is Gaussian, the transformation of the state is completely described by the rotations $d_U = Rd$ of the displacement and $\Gamma_U = R\Gamma R^T$ of the covariance matrix of the initial state, with $\Gamma = \text{diag}(e^{2\mathcal{R}}, e^{-2\mathcal{R}})/2$, $\mathcal{R} = \text{diag}(r, 0, \dots, 0)$, and $d = \sqrt{2} \text{diag}(\alpha, 0, \dots, 0)$. A straightforward calculation shows that

$$d_U = Rd = \sqrt{2}\alpha \begin{pmatrix} \text{Re}[U_{11}] \\ \vdots \\ \text{Re}[U_{M1}] \\ \text{Im}[U_{11}] \\ \vdots \\ \text{Im}[U_{M1}] \end{pmatrix} \quad (\text{A5})$$

and that

$$\Gamma_U = R\Gamma R^T = \begin{pmatrix} \Delta X^2 & \Delta XP \\ \Delta XP^T & \Delta P^2 \end{pmatrix}. \tag{A6}$$

where we have defined the $M \times M$ matrices

$$\begin{aligned} \Delta X^2 &\equiv \frac{1}{2} \left[\operatorname{Re}[U]e^{2\mathcal{R}} \operatorname{Re}[U^\dagger] - \operatorname{Im}[U]e^{-2\mathcal{R}} \operatorname{Im}[U^\dagger] \right] \\ &= \frac{1}{2} \left[\operatorname{Re}[U \cosh(2\mathcal{R})U^\dagger] + \operatorname{Re}[U \sinh(2\mathcal{R})U^\dagger] \right], \end{aligned} \tag{A7a}$$

$$\begin{aligned} \Delta P^2 &\equiv \frac{1}{2} \left[-\operatorname{Im}[U]e^{2\mathcal{R}} \operatorname{Im}[U^\dagger] + \operatorname{Re}[U]e^{-2\mathcal{R}} \operatorname{Re}[U^\dagger] \right] \\ &= \frac{1}{2} \left[\operatorname{Re}[U \cosh(2\mathcal{R})U^\dagger] - \operatorname{Re}[U \sinh(2\mathcal{R})U^\dagger] \right], \end{aligned} \tag{A7b}$$

$$\begin{aligned} \Delta XP &\equiv \frac{1}{2} \left[-\operatorname{Re}[U]e^{2\mathcal{R}} \operatorname{Im}[U^\dagger] - \operatorname{Im}[U]e^{-2\mathcal{R}} \operatorname{Re}[U^\dagger] \right] \\ &= \frac{1}{2} \left[-\operatorname{Im}[U \cosh(2\mathcal{R})U^\dagger] + \operatorname{Im}[U \sinh(2\mathcal{R})U^\dagger] \right]. \end{aligned} \tag{A7c}$$

We can see that the only elements of U entering in Equation (A5) are the ones in the first column U_{i1} , since these are all the transition amplitudes from the only populated input port (namely the first) to every output port. We can easily see that the same is true for the matrices in Equation (A7). In fact, exploiting the relations

$$\cosh(2\mathcal{R})_{ij} = \delta_{ij} + \delta_{i1}\delta_{j1}(\cosh 2r - 1) \tag{A8a}$$

$$\sinh(2\mathcal{R})_{ij} = \delta_{i1}\delta_{j1} \sinh 2r, \tag{A8b}$$

where δ_{ij} is the Kronecker delta, we can evaluate

$$(U \cosh(2\mathcal{R})U^\dagger)_{ij} = \sum_{k,l=1}^M U_{ik} \cosh(2\mathcal{R})_{kl} U_{lj}^\dagger = \delta_{ij} + U_{i1}U_{j1}^* (\cosh 2r - 1) \tag{A9a}$$

$$(U \sinh(2\mathcal{R})U^\dagger)_{ij} = \sum_{k,l=1}^M U_{ik} \sinh(2\mathcal{R})_{kl} U_{lj}^\dagger = U_{i1}U_{j1} \sinh 2r. \tag{A9b}$$

We can thus parametrise only the first column of U , introducing the quantities $P_i = |U_{i1}|^2$ and $\gamma_i = \arg U_{i1}$ for $i = 1, \dots, M$, which are respectively the transition probabilities from the first input to the i -th output port, and the complex phases of the relevant transition amplitudes, so that $U_{i1} = \sqrt{P_i}e^{i\gamma_i}$. We can then rewrite the displacement d' in Equation (A5) and the covariance matrix Γ' in Equation (A6) in terms of P_i and γ_i

$$d_U = Rd = \sqrt{2}\alpha \begin{pmatrix} \sqrt{P_1} \cos \gamma_1 \\ \vdots \\ \sqrt{P_M} \cos \gamma_M \\ \sqrt{P_1} \sin \gamma_1 \\ \vdots \\ \sqrt{P_M} \sin \gamma_M \end{pmatrix}, \tag{A10}$$

and

$$\Delta X_{ij}^2 = \frac{1}{2}\delta_{ij} + \sqrt{P_i P_j} \left(\cos(\gamma_i - \gamma_j) \sinh^2 r + \cos(\gamma_i + \gamma_j) \sinh r \cosh r \right) \tag{A11a}$$

$$\Delta P_{ij}^2 = \frac{1}{2}\delta_{ij} + \sqrt{P_i P_j} \left(\cos(\gamma_i - \gamma_j) \sinh^2 r - \cos(\gamma_i + \gamma_j) \sinh r \cosh r \right) \tag{A11b}$$

$$\Delta XP_{ij} = \sqrt{P_i P_j} \left[-\sin(\gamma_i - \gamma_j) \sinh^2 r + \sin(\gamma_i + \gamma_j) \sinh r \cosh r \right]. \tag{A11c}$$

Finally, we notice that if we are performing homodyne to measure the quadratures \hat{x}_{i,θ_i} , where $\theta = (\theta_1, \dots, \theta_M)$ are the phases of the local oscillators at each output channel, this corresponds to apply a further unitary rotation $U(\theta) = \text{diag}(e^{-i\theta_1}, \dots, e^{-i\theta_M})$ on the state $U|\alpha, r\rangle$. This corresponds to substituting $U \rightarrow U(\theta)U$ in the expression of R in Equation (A3), whose effect is to add a phase $e^{-i\theta_i}$ to every elements of the i -th row of U . Thus, if we define $\tilde{\gamma}_i = \gamma_i - \theta_i$, we can write the displacement d_θ and the covariance matrix Γ_θ of the (accessory) state $\hat{U}(\theta)\hat{U}|\alpha, d\rangle$

$$d_\theta = \sqrt{2\alpha} \begin{pmatrix} \sqrt{P_1} \cos \tilde{\gamma}_1 \\ \vdots \\ \sqrt{P_M} \cos \tilde{\gamma}_M \\ \sqrt{P_1} \sin \tilde{\gamma}_1 \\ \vdots \\ \sqrt{P_M} \sin \tilde{\gamma}_M \end{pmatrix}, \tag{A12}$$

and

$$\Gamma_\theta = \begin{pmatrix} \Delta X_\theta^2 & \Delta X P_\theta \\ \Delta X P_\theta^T & \Delta P_\theta^2 \end{pmatrix}, \tag{A13}$$

with

$$\Delta(X_\theta^2)_{ij} = \frac{1}{2}\delta_{ij} + \sqrt{P_i P_j} \left(\cos(\tilde{\gamma}_i - \tilde{\gamma}_j) \sinh^2 r + \cos(\tilde{\gamma}_i + \tilde{\gamma}_j) \sinh r \cosh r \right) \tag{A14a}$$

$$\Delta(P_\theta^2)_{ij} = \frac{1}{2}\delta_{ij} + \sqrt{P_i P_j} \left(\cos(\tilde{\gamma}_i - \tilde{\gamma}_j) \sinh^2 r - \cos(\tilde{\gamma}_i + \tilde{\gamma}_j) \sinh r \cosh r \right) \tag{A14b}$$

$$\Delta(X P_\theta)_{ij} = \sqrt{P_i P_j} \left[-\sin(\tilde{\gamma}_i - \tilde{\gamma}_j) \sinh^2 r + \sin(\tilde{\gamma}_i + \tilde{\gamma}_j) \sinh r \cosh r \right]. \tag{A14c}$$

If we are performing homodyne detection only on a single channel, say the first, as it is done, for example, in Section 2, we need to first trace away the $M - 1$ channels we are not observing. The probability density function $p(x)$ is thus given by the marginal of the Wigner distribution of the state $\hat{U}(\theta)\hat{U}|\alpha, r\rangle$ associated with the first symplectic variable x_1 and, equivalently, a Gaussian distribution with average μ given by the first element of d_θ in Equation (A12)

$$\mu = \sqrt{2\alpha} \sqrt{P_1} \cos \tilde{\gamma}_1 \tag{A15}$$

and variance σ^2 given by the first row, first column element of Γ_θ in Equation (A13)

$$\sigma^2 = \frac{1}{2} + P_1 (\sinh^2 r + \cos 2\tilde{\gamma}_1 \sinh r \cosh r). \tag{A16}$$

For a squeezed vacuum state, $\alpha = 0$ and we recognise the expression of the variance in Equation (6) from Equation (A16).

If, instead, we are observing all M output channels of the network U , the probability density function $p(x)$ will be a Gaussian distribution with average μ given by the first M elements of d_θ

$$\mu = \sqrt{2\alpha} \begin{pmatrix} \sqrt{P_1} \cos \tilde{\gamma}_1 \\ \vdots \\ \sqrt{P_M} \cos \tilde{\gamma}_M \end{pmatrix}, \tag{A17}$$

from which we recognise Equation (36), while the covariance matrix Σ is given by the first M rows and columns of Γ_θ , i.e., by the matrix in Equation (A14a)

$$\Sigma_{ij} = \frac{1}{2}\delta_{ij} + \sqrt{P_i P_j} \left(\cos(\tilde{\gamma}_i - \tilde{\gamma}_j) \sinh^2 r + \cos(\tilde{\gamma}_i + \tilde{\gamma}_j) \sinh r \cosh r \right), \tag{A18}$$

from which we obtain the expression in Equation (35).

To evaluate the determinant $|\Sigma|$ in Equation (37), we first notice that the covariance matrix Γ of the initial state $|\alpha, r\rangle$ can be written as the sum $\Gamma \equiv \mathbb{I}_{2M}/2 + K_0$, where \mathbb{I} is the $2M \times 2M$ identity matrix and $K_0 = \Gamma - \mathbb{I}_{2M}/2$ is a diagonal matrix of rank 2. Since the rank is invariant under orthogonal rotations, the same holds true for $\Gamma_\theta = R\Gamma R^T \equiv \mathbb{I}_{2M} + K$, with $\text{rank}(K) = 2$. By definition of rank, none of the sub-matrices of K can have rank greater than 2, hence we can write

$$\Sigma = \mathbb{I}_M/2 + (\Sigma - \mathbb{I}_M/2) \equiv \mathbb{I}_M/2 + A, \quad (\text{A19})$$

with $\text{rank}(A) \leq 2$, since A is a submatrix of K

$$A_{ij} = \sqrt{P_i P_j} \left(\cos(\tilde{\gamma}_i - \tilde{\gamma}_j) \sinh^2 r + \cos(\tilde{\gamma}_i + \tilde{\gamma}_j) \sinh r \cosh r \right) \quad (\text{A20})$$

We can thus apply the results in Appendix E to Σ and write $|\Sigma| = |\mathbb{I}_M/2 + A|$ as a sum of determinants of the matrices obtained replacing any number of columns of $\mathbb{I}_M/2$, with the respective columns of A

$$|\Sigma| = \frac{1}{2^M} + \frac{1}{2^{M-1}} \sum_{i=1}^M A_{ii} + \frac{1}{2^{M-2}} \sum_{i=1}^M \sum_{j=i+1}^M A_{ii} A_{jj} - A_{ij}^2, \quad (\text{A21})$$

where the first term is the determinant of $\mathbb{I}_M/2$, the terms in the first summation are determinants of the matrices obtained by substituting the i -th column of $\mathbb{I}_M/2$ with the i -th column of A , and the terms in the last summation from substituting the i -th and j -th columns, while also exploiting the symmetry of A . Noticeably, since $\text{rank } A \leq 2$, all the contributions involving the replacement of three or more columns of A are vanishing for the definition of rank of a matrix. Replacing the expression of A found in Equation (A20) into Equation (A21), the expression in Equation (37) can then be easily obtained after some elementary trigonometry.

Appendix B. Fisher Information for Gaussian probabilities

In this Appendix, we will evaluate the Fisher information $\mathcal{F}(\varphi)$ associated with the estimation of a parameter φ from a M -variate Gaussian distribution

$$p_\varphi(x) = \frac{1}{\sqrt{(2\pi)^M \det[\Sigma]}} \exp\left(-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu)\right), \quad (\text{A22})$$

where we assume that both the covariance matrix Σ and the average μ depend on the parameter φ .

Recalling the definition of the Fisher information in Equation (7), we first evaluate the logarithmic derivative

$$\begin{aligned} \frac{\partial}{\partial \varphi} \ln p_\varphi(x) &= -\frac{1}{2} \frac{\partial}{\partial \varphi} \ln \det[\Sigma] - \frac{1}{2} (x - \mu)^\top \frac{\partial \Sigma^{-1}}{\partial \varphi} (x - \mu) \\ &\quad + \frac{\partial \mu^\top}{\partial \varphi} \Sigma^{-1} (x - \mu). \end{aligned} \quad (\text{A23})$$

In order to evaluate the Fisher information in Equations (9) and (38), we first notice that the expression $\left(\frac{\partial}{\partial \varphi} \ln p_\varphi(x)\right)^2$ contains terms of order up to fourth in $(x - \mu)$. Obtaining the

Fisher information thus reduces to the evaluation of the expectation values of a polynomial in $(x - \boldsymbol{\mu})$, for which we can exploit standard results of Gaussian integrals:

$$\mathbb{E}_{p_\varphi}[x_i - \boldsymbol{\mu}_i] = 0 \quad (\text{A24a})$$

$$\mathbb{E}_{p_\varphi}[(x_i - \boldsymbol{\mu}_i)(x_j - \boldsymbol{\mu}_j)] = \Sigma_{ij} \quad (\text{A24b})$$

$$\mathbb{E}_{p_\varphi}[(x_i - \boldsymbol{\mu}_i)(x_j - \boldsymbol{\mu}_j)(x_k - \boldsymbol{\mu}_k)] = 0 \quad (\text{A24c})$$

$$\mathbb{E}_{p_\varphi}[(x_i - \boldsymbol{\mu}_i)(x_j - \boldsymbol{\mu}_j)(x_k - \boldsymbol{\mu}_k)(x_l - \boldsymbol{\mu}_l)] = \Sigma_{ij}\Sigma_{kl} + \Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk}, \quad (\text{A24d})$$

from which we see that only even powers of the term $(x - \boldsymbol{\mu})$ have a non-vanishing contribution. We can thus evaluate

$$\begin{aligned} \mathcal{F}(\varphi) &= \mathbb{E}_{p_\varphi} \left[\left(\frac{\partial}{\partial \varphi} \ln p_\varphi(\mathbf{x}) \right)^2 \right] \\ &= \frac{1}{4} \left(\frac{\partial}{\partial \varphi} \ln \det[\Sigma] \right)^2 \\ &\quad + \sum_{s,t=1}^M \left(\frac{\partial \boldsymbol{\mu}^T}{\partial \varphi} \Sigma^{-1} \right)_s \left(\frac{\partial \boldsymbol{\mu}^T}{\partial \varphi} \Sigma^{-1} \right)_t \mathbb{E}_{p_\varphi}[(x_s - \boldsymbol{\mu}_s)(x_t - \boldsymbol{\mu}_t)] \\ &\quad + \frac{1}{2} \left(\frac{\partial}{\partial \varphi} \ln \det[\Sigma] \right) \sum_{s,t=1}^M \frac{\partial \Sigma_{st}^{-1}}{\partial \varphi} \mathbb{E}_{p_\varphi}[(x_s - \boldsymbol{\mu}_s)(x_t - \boldsymbol{\mu}_t)] \\ &\quad + \frac{1}{4} \sum_{s,t,u,v=1}^M \frac{\partial \Sigma_{st}^{-1}}{\partial \varphi} \frac{\partial \Sigma_{uv}^{-1}}{\partial \varphi} \mathbb{E}_{p_\varphi}[(x_s - \boldsymbol{\mu}_s)(x_t - \boldsymbol{\mu}_t)(x_u - \boldsymbol{\mu}_u)(x_v - \boldsymbol{\mu}_v)]. \end{aligned} \quad (\text{A25})$$

By substituting the expressions in Equations (A24) into Equation (A25), and exploiting the Jacobi's formula for the derivative of the determinant of square matrices which assures that

$$\frac{1}{\det[\Sigma]} \frac{\partial \det[\Sigma]}{\partial \varphi} = \text{Tr} \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi} \right] \equiv - \text{Tr} \left[\frac{\partial \Sigma^{-1}}{\partial \varphi} \Sigma \right], \quad (\text{A26})$$

we see that most of the terms in Equation (A25) cancel out, yielding the Fisher information

$$\mathcal{F}(\varphi) = \frac{\partial \boldsymbol{\mu}^T}{\partial \varphi} \Sigma^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \varphi} + \frac{1}{2} \text{Tr} \left[\frac{\partial \Sigma^{-1}}{\partial \varphi} \Sigma \frac{\partial \Sigma^{-1}}{\partial \varphi} \Sigma \right]. \quad (\text{A27})$$

We can easily reduce Equation (A27) to the Fisher information in Equation (9) for the single-homodyne case with $\boldsymbol{\mu} = 0$.

To obtain the FIM in Equation (38), we first exploit the relation

$$\Sigma \frac{\partial \Sigma^{-1}}{\partial \varphi} \Sigma = - \frac{\partial \Sigma}{\partial \varphi}, \quad (\text{A28})$$

and thus we express the covariance matrix Σ^{-1} in terms of its cofactor matrix, i.e., $\Sigma^{-1} = C / \det[\Sigma]$, where we have exploited the symmetry of Σ so that $C^T = C$. We can thus rewrite, in the single-parameter case

$$\begin{aligned} \mathcal{F}(\varphi) &= \frac{\partial \boldsymbol{\mu}^T}{\partial \varphi} \Sigma^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \varphi} - \frac{1}{2} \text{Tr} \left[\frac{\partial \Sigma^{-1}}{\partial \varphi} \frac{\partial \Sigma}{\partial \varphi} \right] \\ &= \frac{\partial \boldsymbol{\mu}^T}{\partial \varphi} \Sigma^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \varphi} + \frac{1}{2 \det[\Sigma]^2} \frac{\partial \det[\Sigma]}{\partial \varphi} \text{Tr} \left[C \frac{\partial \Sigma}{\partial \varphi} \right] \\ &\quad - \frac{1}{2 \det[\Sigma]} \text{Tr} \left[\left(\frac{\partial C}{\partial \varphi} \right) \left(\frac{\partial \Sigma}{\partial \varphi} \right) \right] \end{aligned} \quad (\text{A29})$$

We can recognise in the second term of Equation (A29) the Jacobi’s formula for the derivative of the determinant in Equation (A26), which allows us to obtain the expression for the Fisher information shown in Equation (38)

$$\mathcal{F}(\varphi) = \frac{\partial \boldsymbol{\mu}^T}{\partial \varphi} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \varphi} + \frac{1}{2 \det[\boldsymbol{\Sigma}]^2} \left(\frac{\partial \det[\boldsymbol{\Sigma}]}{\partial \varphi} \right)^2 - \frac{1}{2 \det[\boldsymbol{\Sigma}]} \text{Tr} \left[\left(\frac{\partial \mathbf{C}}{\partial \varphi} \right) \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \varphi} \right) \right]. \tag{A30}$$

In order to elicit the cofactor matrix \mathbf{C} in terms of the elements of $\boldsymbol{\Sigma}$ and thus obtain Equation (39), we first need to make some observations. First, the (s, t) -cofactor C_{st} is defined as the determinant of the $L - 1 \times L - 1$ sub-matrix of $\boldsymbol{\Sigma}$ obtained by deleting its s -th row and t -th column, then multiplied by $(-1)^{s+t}$. This can be equivalently thought of as the determinant of the $L \times L$ matrix $\boldsymbol{\Sigma}^{[s,t],1}$, where we denote with $X^{[s,t],n}$ the matrix obtained from the matrix X replacing all its elements in the s -th row and in the p -th column with zeros, with the exception of the element (s, t) which is replaced by n , namely

$$C_{st} = \det \left[\begin{pmatrix} \Sigma_{11} & \dots & \Sigma_{1t-1} & 0 & \Sigma_{1t+1} & \dots & \Sigma_{1L} \\ & \vdots & & \vdots & & \vdots & \\ \Sigma_{s-11} & \dots & \Sigma_{s-1t-1} & 0 & \Sigma_{s-1t+1} & \dots & \Sigma_{s-1L} \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \Sigma_{s+11} & \dots & \Sigma_{s+1t-1} & 0 & \Sigma_{s+1t+1} & \dots & \Sigma_{s+1L} \\ & \vdots & & \vdots & & \vdots & \\ \Sigma_{L1} & \dots & \Sigma_{Lt-1} & 0 & \Sigma_{Lt+1} & \dots & \Sigma_{LL} \end{pmatrix} \right]. \tag{A31}$$

For the particular case of the covariance matrix $\boldsymbol{\Sigma}$ in Equation (35), as discussed in Appendix A, we can write $\boldsymbol{\Sigma} = \mathbb{I}_M/2 + A$, with A symmetric matrix such that $\text{rank}(A) = \rho \leq 2$ given in Equation (A20). Thus, we can write the (s, t) -cofactor as

$$C_{st} = \det \left[(\mathbb{I}_M/2)^{[s,t],0} + A^{[s,t],1} \right], \tag{A32}$$

and evaluate this determinant with the methods discussed in detail in Appendix E. In particular, C_{st} is obtained as a sum of determinants of matrices obtained by replacing any possible set of columns of $(\mathbb{I}_M/2)^{[s,t],0}$ with the columns in the same positions of $A^{[s,t],1}$. Noticeably, by substituting a row and a column of A with an arbitrary row or column, in general increases its rank by one, so that $\text{rank}(A^{[s,t],1}) \leq 3$. To evaluate the cofactor matrix \mathbf{C} , it is thus convenient to consider separately the simpler case $s = t$ first, and then $s \neq t$.

For $s = t$ in Equation (A32), the cofactor becomes $C_{ss} = \det \left[(\mathbb{I}_M/2)^{[s,s],0} + A^{[s,s],1} \right]$. We notice that the matrix $(\mathbb{I}_M/2)^{[s,s],0}$ is a diagonal matrix with a single zero eigenvalue—i.e., the s -th—and thus each contribution to C_{ss} is non-vanishing only if the s -th columns of $(\mathbb{I}_M/2)^{[s,s],0}$ —a column of only 0s—is replaced. Moreover, the s -th column of $A^{[s,s],1}$ is also a column of all 0s, with the exception of the s -th element, which is equal to 1. We thus obtain

$$C_{ss} = \frac{1}{2^{M-1}} + \frac{1}{2^{M-2}} \sum_{\substack{i=1 \\ i \neq s}}^M A_{ii} - \frac{1}{2^{M-3}} \sum_{\substack{i=1 \\ i \neq s}}^M \sum_{\substack{j=i+1 \\ j \neq s}}^M A_{ii} A_{jj} - A_{ij}^2, \tag{A33}$$

which is the sum of terms obtained substituting the s -th, the s -th and i -th, and the s -th, i -th and j -th columns, respectively. Noticeably, replacing more than 3 columns yields vanishing contributions, since $\text{rank}(A^{[s,t],1}) \leq 3$. When $s \neq t$, $(\mathbb{I}_M/2)^{[s,t],0}$ is still a diagonal matrix but with two zeros in the diagonal, more precisely the s -th and the t -th elements. Both the s -th and the t -th columns are thus columns of zeros, and all non-vanishing contributions

to C_{st} must replace these two columns. For example, the only contribution obtained by swapping the s -th and t -th columns is of the type

$$\frac{1}{2^{M-2}} \det \begin{bmatrix} 0 & 1 \\ A_{ts} & 0 \end{bmatrix} = -\frac{1}{2^{M-2}} A_{st}, \quad (\text{A34})$$

exploiting at the same time the symmetry of A . The contributions obtained by swapping the s -th, t -th and i -th columns, with $i \neq s, t$, are of the type

$$\frac{1}{2^{M-3}} \det \begin{bmatrix} 0 & 1 & 0 \\ A_{ts} & 0 & A_{ti} \\ A_{is} & 0 & A_{ii} \end{bmatrix} = \frac{1}{2^{M-3}} (A_{si}A_{ti} - A_{st}A_{ii}), \quad (\text{A35})$$

where once again, we are exploiting the symmetry of A . Once again, substituting more than 3 columns yields no contributions, since $\text{rank}(A^{[s,t],1}) \leq 3$. The final expression for C_{st} , $s \neq t$ thus reads

$$C_{st} = -\frac{1}{2^{M-2}} A_{st} + \frac{1}{2^{M-3}} \sum_{\substack{i=1 \\ i \neq s,t}}^M (A_{si}A_{ti} - A_{st}A_{ii}). \quad (\text{A36})$$

Replacing in (A33) and (A36) the definition of $A = \Sigma - \mathbb{I}_M/2$ in Equation (A32), it is straightforward to obtain Equation (39).

Appendix C. Asymptotic Analyses of Gaussian Metrology

In this appendix, we will perform all the asymptotic analyses for a large number of photons N in the setups presented in Sections 2 and 3.

Appendix C.1. Single Homodyne

Here, we evaluate the asymptotic expressions of the Fisher information in Equation (16), showing that the conditions in Equations (14) and (15) yield the Heisenberg scaling for the estimation of the respective quantities of interest.

As shown in Equations (10) and (11), the dependence of the variance

$$\sigma_\varphi^2 = \frac{1 - P_\varphi}{2} + \frac{P_\varphi}{2} [\cosh(2r) + \cos(2\gamma_\varphi - 2\theta) \sinh(2r)], \quad (\text{A37})$$

on the parameters φ only appears through the transition probability P_φ and the acquired complex phase γ_φ

$$\frac{\partial \sigma_\varphi^2}{\partial \varphi} = (\partial_P \sigma_\varphi^2) \frac{\partial P_\varphi}{\partial \varphi} + (\partial_\gamma \sigma_\varphi^2) \frac{\partial \gamma_\varphi}{\partial \varphi}, \quad (\text{A38})$$

where ∂_P and ∂_γ represent the differentiation with respect to P_φ and γ_φ , and

$$\partial_P \sigma_\varphi^2 = \frac{1}{2} (-1 + \cosh(2r) + \cos(2\gamma_\varphi - 2\theta) \sinh(2r)) \quad (\text{A39a})$$

$$\partial_\gamma \sigma_\varphi^2 = -P_\varphi \sin(2\gamma_\varphi - 2\theta) \sinh(2r). \quad (\text{A39b})$$

To achieve the Heisenberg scaling, some conditions must be imposed so that the variance in Equation (A37) does not grow with N . The only option to do that without ruining the sensitivity of the setup is requiring that $\gamma_\varphi - \theta \simeq \pi/2$, as we can see from Equation (A37). In particular, we impose condition (14), and evaluate the variance in (A37) and its gradient (A38) in the large N limit

$$\sigma_\varphi^2 = \frac{1}{2} + NP_\varphi \left(1 - \cos\left(\frac{2k}{N}\right) \sqrt{1 + \frac{1}{N}} \right)$$

$$\begin{aligned}
 &= \frac{1}{2} + NP_\varphi \left(1 - \left(1 - \frac{2k^2}{N^2} \right) \left(1 + \frac{1}{2N} - \frac{1}{8N^2} \right) \right) + \mathcal{O}\left(\frac{1}{N^2}\right) \\
 &= \frac{1 - P_\varphi}{2} + P_\varphi \left(\frac{2k^2}{N} + \frac{1}{8N} \right) + \mathcal{O}\left(\frac{1}{N^2}\right), \tag{A40}
 \end{aligned}$$

$$\begin{aligned}
 \frac{\partial}{\partial \varphi} \sigma_\varphi^2 &= N \frac{\partial}{\partial \varphi} P_\varphi \left(1 - \cos\left(\frac{2k}{N}\right) \sqrt{1 + \frac{1}{N}} \right) + 2NP_\varphi \frac{\partial}{\partial \varphi} \gamma_\varphi \sin\left(\frac{2k}{N}\right) \sqrt{1 + \frac{1}{N}} \\
 &= N \frac{\partial}{\partial \varphi} P_\varphi \left(1 - \left(1 + \frac{1}{2N} \right) \right) + 2NP_\varphi \frac{\partial}{\partial \varphi} \gamma_\varphi \frac{2k}{N} + \mathcal{O}\left(\frac{1}{N}\right) \\
 &= -\frac{1}{2} \frac{\partial}{\partial \varphi} P_\varphi + 4kP_\varphi \frac{\partial}{\partial \varphi} \gamma_\varphi + \mathcal{O}\left(\frac{1}{N}\right). \tag{A41}
 \end{aligned}$$

Then, by imposing condition (15) on P_φ , we get

$$\sigma_\varphi^2 = \left(2k^2 + \frac{1}{8} + \frac{\ell}{2} \right) \frac{1}{N} + \mathcal{O}\left(\frac{1}{N^2}\right), \tag{A42a}$$

$$\frac{\partial}{\partial \varphi} \sigma_\varphi^2 = 4k \frac{\partial}{\partial \varphi} \gamma_\varphi + \mathcal{O}\left(\frac{1}{N}\right). \tag{A42b}$$

By substituting the expressions in Equation (A42) into the Fisher information in Equation (9), we can finally evaluate its asymptotic expression

$$\mathcal{F}(\varphi) \sim 8\varrho(k, \ell) N^2 \left(\frac{\partial \gamma_\varphi}{\partial \varphi} \right) \left(\frac{\partial \gamma_\varphi}{\partial \varphi^T} \right) \tag{A43}$$

with

$$\varrho(k, \ell) = \left(\frac{8k}{16k^2 + 1 + 4\ell} \right)^2 \tag{A44}$$

a positive and N -independent pre-factor.

Appendix C.2. Multiple Homodyne

In this appendix, we will study the asymptotic regime of the Fisher information in Equation (38),

$$\mathcal{F}(\varphi) = \underbrace{\frac{1}{\det[\Sigma]} \partial_\varphi \boldsymbol{\mu}^T C \partial_\varphi \boldsymbol{\mu}}_{\mathcal{F}_D(\varphi)} + \underbrace{\frac{1}{2} \left(\frac{\partial_\varphi \det[\Sigma]}{\det[\Sigma]} \right)^2 - \frac{1}{2 \det[\Sigma]} \text{Tr}[(\partial_\varphi \Sigma)(\partial_\varphi C)]}_{\mathcal{F}_S(\varphi)} \tag{A45}$$

for large $N = N_S + N_D = \sinh^2 r + \alpha^2$. We will show that conditions (41) are necessary to reach the Heisenberg scaling, and that, when they hold, the asymptotic expression for the Fisher information is the one shown in Equation (43).

As discussed in Section 3.2, all terms in the numerators of the Fisher information in Equation (A45) are at most of the order N_S^2 or $N_S N_D$. We thus need to study the asymptotic behaviour of the determinant $\det[\Sigma]$ in Equation (37), which is at the denominators of the FI, and find what conditions prevent $\det[\Sigma]$ from growing with N_S as well, ruining the Heisenberg scaling. In particular, in order to reach the Heisenberg scaling, namely a scaling of the order of N^2 in the FI, the determinant $\det[\Sigma]$ must be at most of the order $\mathcal{O}(1)$.

Thus, we first focus our attention on $\det[\Sigma]$ in Equation (37)

$$\begin{aligned}
 \det[\Sigma] &= \frac{1}{2^M} + \frac{\sinh(r)}{2^{M-1}} \sum_{i=1}^M P_i(\sinh(r) + \cos(2\tilde{\gamma}_i) \cosh(r)) \\
 &\quad - \frac{\sinh^2(r)}{2^{M-2}} \sum_{i=1}^M \sum_{j=i+1}^M P_i P_j \sin^2(\tilde{\gamma}_i - \tilde{\gamma}_j). \tag{A46}
 \end{aligned}$$

In particular, we will suppose that the asymptotic behaviour of the complex phases $\tilde{\gamma}_i$ for large N_S is given by $\tilde{\gamma}_i = \tilde{\gamma}_{0i} + k_i N_S^{-\alpha}$, with $k_i \in \mathbb{R}$ independent of N and $\alpha > 0$ (Notice that this condition is not restrictive, since if $\tilde{\gamma}_i$ were to grow with N (i.e., $\alpha < 0$), it would yield an oscillating asymptotic behaviour of $\det[\Sigma]$ and every other term appearing in the FI). By expanding in powers of N_S the terms in which the squeezing parameter r appears in (A46) we obtain

$$\det[\Sigma] = D_1 N_S + D_2 + D_3 \frac{1}{N_S} + \mathcal{O}\left(\frac{1}{N_S^2}\right), \tag{A47}$$

where

$$D_1 = \frac{1}{2^{M-1}} \left(1 + \sum_{i=1}^M P_i \cos(2\tilde{\gamma}_i)\right) - \frac{1}{2^{M-2}} \left(\sum_{i=1}^M \sum_{j=i+1}^M P_i P_j \sin(\tilde{\gamma}_i - \tilde{\gamma}_j)^2\right), \tag{A48a}$$

$$D_2 = \frac{1}{2^M} \left(1 + \sum_{i=1}^M P_i \cos(2\tilde{\gamma}_i)\right), \tag{A48b}$$

$$D_3 = -\frac{1}{2^{M+2}} \sum_{i=1}^M P_i \cos(2\tilde{\gamma}_i). \tag{A48c}$$

To prevent the scaling with N_S of the determinant $\det[\Sigma]$, D_1 must be set as equal to, or tending to, zero. After some trigonometry, and exploiting the fact that $\sum_i P_i = 1$ due to the unitarity of the network, we can rewrite

$$D_1 = \frac{1}{2^{M-2}} \left(\left(\sum_{i=1}^M P_i \cos^2 \tilde{\gamma}_i\right)^2 + \left(\sum_{i=1}^M P_i \sin \tilde{\gamma}_i \cos \tilde{\gamma}_i\right)^2 \right), \tag{A49}$$

which tends to zero only if both terms in the parenthesis tend to zero, i.e., when $\tilde{\gamma}_{0i} = \pi/2 + n\pi$, with $n \in \mathbb{Z}$. To analyse the order for large N_S of the determinant when the condition $\tilde{\gamma}_{0i} = \pi/2 + n\pi$ holds—and thus $\tilde{\gamma}_i = \pi/2 + n\pi + k_i N_S^{-\alpha}$ —we need to analyse the behaviour of each term in Equation (A48). Since $\cos(2\tilde{\gamma}_i) \sim -1 + 2k_i N_S^{-2\alpha}$, and $\sin(\tilde{\gamma}_i - \tilde{\gamma}_j)^2 \sim (k_i - k_j) N_S^{-2\alpha}$, both D_1 and D_2 scale with $N_S^{-2\alpha}$, while D_3 is asymptotically constant. Thus, we find the asymptotic behaviour of $\det[\Sigma]$ through Equation (A47): the term D_2 is always negligible with respect to $D_1 N_S$, while for $\alpha \leq 1$ the term $D_1 N_S$ dominates over D_3/N_S so that $\det[\Sigma]$ is of order $N_S^{1-2\alpha}$, and for $\alpha > 1$ instead D_3/N_S dominates $D_1 N_S$ and the determinant scales with N_S^{-1} . In equations:

$$\det[\Sigma] \sim \begin{cases} D_1 N_S \propto N_S^{1-2\alpha}, & \text{for } \alpha \leq 1 \\ D_3 N_S^{-1} \propto N_S^{-1}, & \text{for } \alpha > 1 \end{cases} \tag{A50}$$

Noticeably, the determinant stops growing with N_S only for $\alpha \geq 1/2$.

We thus now study the asymptotic behaviour of the numerators appearing in the Fisher information in Equation (A45), when the condition $\tilde{\gamma}_i = \pi/2 + k_i N_S^{-\alpha}$, with $\alpha \geq 1/2$, is true. We will perform the same analysis done for $\det[\Sigma]$ earlier, considering only the dominating term for every value of α . We first obtain the derivative of Σ from Equation (35), substituting $\tilde{\gamma}_i = \pi/2 + k_i / N_S^\alpha$, with $\alpha \geq 1/2$,

$$\begin{aligned} \partial_\varphi \Sigma_{ij} = & -\frac{1}{2} \partial_\varphi(\sqrt{P_i P_j}) + \sqrt{P_i P_j} \left(\partial_\varphi(\tilde{\gamma}_i + \tilde{\gamma}_j)(k_i + k_j) \right. \\ & \left. - \partial_\varphi(\tilde{\gamma}_i - \tilde{\gamma}_j)(k_i - k_j) \right) N_S^{1-\alpha} + \mathcal{O}(N_S^{1-2\alpha}) + \mathcal{O}(N_S^{-1}), \end{aligned} \tag{A51}$$

and we notice that each element of Σ scales at most with $N_S^{1-\alpha}$ for $1/2 \leq \alpha < 1$, while it becomes constant for $\alpha \geq 1$. We then analyse the auxiliary term (40)

$$S_{sti} = \sinh^2(r) \sqrt{P_s P_t P_i} \sin(\tilde{\gamma}_s - \tilde{\gamma}_i) \sin(\tilde{\gamma}_t - \tilde{\gamma}_i) = \mathcal{O}(N_S^{1-2\alpha}), \tag{A52}$$

of which we evaluate the derivative, always with the conditions $\tilde{\gamma}_i = \pi/2 + k_i/N_S^\alpha$, with $\alpha \geq 1/2$

$$\begin{aligned} \partial_\varphi S_{sti} &= \sqrt{P_s P_t P_i} \left((\partial_\varphi(\tilde{\gamma}_s - \tilde{\gamma}_i))(k_t - k_i) \right. \\ &\quad \left. + (\partial_\varphi(\tilde{\gamma}_t - \tilde{\gamma}_i))(k_s - k_i) \right) N_S^{1-\alpha} + \mathcal{O}(N^{1-2\alpha}), \end{aligned} \tag{A53}$$

which scales at most with $N_S^{1-\alpha}$ for large N_S . Moreover, the covariance matrix Σ in (35) asymptotically reads

$$\Sigma_{ij} = \frac{\delta_{ij} - \sqrt{P_i P_j}}{2} + \mathcal{O}(N_S^{1-2\alpha}) + \mathcal{O}(N_S^{-1}), \tag{A54}$$

which is in general constant asymptotically for large N_S . Inserting Equations (A52) and (A54) in the cofactor matrix (39), we obtain

$$C_{st} = \frac{1}{2^{M-1}} \sqrt{P_s P_t} + \mathcal{O}(N_S^{1-2\alpha}) + \mathcal{O}(N_S^{-1}). \tag{A55}$$

Exploiting the asymptotic behaviours of the derivatives found in Equations (A51) and (A53), we are able to analyse the remaining terms appearing in the FI

$$\partial_\varphi \det[\Sigma] = \frac{1}{2^{M-1}} \sum_{i=1}^M \partial_\varphi \Sigma_{ii} - \frac{1}{2^{M-2}} \sum_{i=1}^M \sum_{j=i+1}^M \partial_\varphi S_{ijj}, \tag{A56}$$

$$\partial_\varphi C_{ss} = \frac{1}{2^{M-2}} \sum_{\substack{i=1 \\ i \neq s}}^M \partial_\varphi \Sigma_{ii} - \frac{1}{2^{M-3}} \sum_{\substack{i=1 \\ i \neq s}}^M \sum_{\substack{j=i+1 \\ j \neq s}}^M \partial_\varphi S_{ijj}, \tag{A57}$$

$$\partial_\varphi C_{st} = -\frac{1}{2^{M-2}} \partial_\varphi \Sigma_{st} + \frac{1}{2^{M-3}} \sum_{\substack{j=1 \\ j \neq s,t}}^M \partial_\varphi S_{stj}. \tag{A58}$$

It is easy to see that $\partial_\varphi \det[\Sigma]$ scales at most with $N_S^{1-\alpha}$, since $\sum_i^M \partial_\varphi P_i = 0$, while the derivatives of the elements of C scale at most with $N_S^{1-\alpha}$ for $1/2 \leq \alpha < 1$, and become constant for $\alpha \geq 1$. The last step is to evaluate the asymptotic behaviour for the derivative $\partial_\varphi \mu$ of the average, which can be easily obtained by differentiating Equation (36)

$$\begin{aligned} \partial_\varphi \mu_i &= \sqrt{2N_D} \left(\frac{\partial_\varphi P_i}{2\sqrt{P_i}} \cos \tilde{\gamma}_i - \sqrt{P_i} \partial_\varphi \tilde{\gamma}_i \sin \tilde{\gamma}_i \right) \\ &= -\sqrt{2N_D P_i} \partial_\varphi \tilde{\gamma}_i + \mathcal{O}(\sqrt{N_D} N_S^{-\alpha}). \end{aligned} \tag{A59}$$

Now we can finally draw the conclusions, and obtain the scaling of the Fisher information shown in Equation (43) by putting together all the asymptotic regimes found. First, we notice that, independently of $\alpha \geq 1/2$, the third contribution to the Fisher information in Equation (A45) always scales with N_S , hence only reaching shot-noise precision. Thus, it is always dominated by the other two terms in the Heisenberg scaling regime. The only terms which allow the Heisenberg scaling are then the first two: in fact, the term $\mathcal{F}_D(\varphi)$ scales with $N_D N_S^{2\alpha-1}$ for $1/2 \leq \alpha \leq 1$, and with $N_D N_S$ for $\alpha > 1$, thus reaching the Heisenberg scaling for $\alpha \geq 1$, while the second term reaches sub-shot noise scaling $N_S^{2-2|\alpha-1|}$ for $1/2 < \alpha < 3/2$, achieving the Heisenberg scaling for $\alpha = 1$. It is interesting to notice the major difference between the two contributions to the Heisenberg scaling:

the term $\mathcal{F}_D(\varphi)$ reaches the Heisenberg scaling for every value of $\alpha \geq 1$, which means that the choice $\tilde{\gamma}_i = \pi/2$ —corresponding to $\alpha = +\infty$ —still yields the Heisenberg scaling. On the other hand, α must be equal to 1 in order for the first term of $\mathcal{F}_S(\varphi)$ to achieve the Heisenberg scaling. The consequences and the physical implications of this difference are discussed in depth in Section 3.2.

The conditions to reach the Heisenberg scaling can thus be condensed into the request on the complex phases

$$\tilde{\gamma}_i = \pi/2 + \mathcal{O}(N_S^{-1}), \tag{A60}$$

for $i = 1, \dots, M$, as shown in Equation (41), or equivalently that $\tilde{\gamma}_i \simeq \pi/2 + k_i/N_S$ asymptotically, with $k_i \in \mathbb{R}$ independent of N_S . Moreover, we have seen that the only relevant terms under this condition are the first two in Equation (A45), and that only for $\alpha = 1$ the first term of $\mathcal{F}_S(\varphi)$ is non-vanishing.

Finally, we can now prove (43). Substituting condition (A60) into Equations (A51) and (A53), we obtain from Equation (A56)

$$\begin{aligned} \partial_\varphi \det[\Sigma] &\sim \frac{1}{2^{M-1}} \sum_{i=1}^M 4P_i k_i (\partial_\varphi \tilde{\gamma}_i) - \frac{1}{2^{M-2}} \sum_{i=1}^M \sum_{j=i+1}^M 2P_i P_j (k_i - k_j) (\partial_\varphi \tilde{\gamma}_i - \partial_\varphi \tilde{\gamma}_j) \\ &\sim \frac{1}{2^{M-3}} \sum_{i=1}^M P_i k_i \sum_{j=1}^M P_j \partial_\varphi \tilde{\gamma}_j \end{aligned} \tag{A61}$$

where we exploited the passivity of the interferometer which assures $\sum_{i=1}^M P_i = 1$. We also obtain from (A46)

$$\det[\Sigma] \simeq \frac{1}{2^{M-2} N_S} \left(\left(\sum_{i=1}^M P_i k_i \right)^2 + \frac{1}{16} \right), \tag{A62}$$

which coincides with the expression in Equation (42). Lastly, from Equation (A59) when conditions (A60) hold, we obtain the asymptotic behaviour

$$\partial_\varphi \mu_i \sim -\sqrt{2N_D P_i} \partial_\varphi \tilde{\gamma}_i. \tag{A63}$$

Substituting Equations (A55) and (A61)–(A63) into the Fisher information in Equation (38) yields the asymptotic expression shown in Equation (43).

Appendix D. Maximum-Likelihood Estimators for Gaussian Distributions

In this Appendix, we will derive the Likelihood equation associated with the estimation schemes presented in Sections 2 and 3, i.e., the equation whose solution maximises the Likelihood functions in Equations (18) and (46). In doing so, we will demonstrate the expressions in Equations (19) and (47).

We first suppose that the M -variate Gaussian probability distribution

$$p_\varphi(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^M \det[\Sigma]}} \exp\left(-(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right) \tag{A64}$$

governing the outcomes $\mathbf{x} = (x_1, \dots, x_M)$ depends on a single parameter φ through its average $\boldsymbol{\mu}$ and its covariance matrix Σ . If we perform the observation ν times x_1, \dots, x_ν , the Likelihood function becomes

$$\mathcal{L}(\varphi; x_1, \dots, x_\nu) = \frac{1}{\sqrt{(2\pi)^{M\nu} \det[\Sigma]^\nu}} \exp\left(-\frac{1}{2} \sum_{i=1}^\nu (x_i - \boldsymbol{\mu})^T \Sigma^{-1} (x_i - \boldsymbol{\mu})\right). \tag{A65}$$

The maximum-likelihood estimator is obtained maximising the Likelihood function in Equation (A65)—or its logarithm. Applying the maximisation to the likelihood function in Equation (A65), we obtain the equation

$$\begin{aligned}
 0 &= \partial_\varphi \ln \mathcal{L}(\varphi; \mathbf{x}_1, \dots, \mathbf{x}_v) \Big|_{\varphi=\hat{\varphi}_{MLE}} \\
 &= \partial_\varphi \sum_{j=1}^v \ln p(\mathbf{x}_j; \varphi) \Big|_{\varphi=\hat{\varphi}_{MLE}} \\
 &= \left[-\frac{v}{2} \partial_\varphi \ln(\det[\Sigma]) - \frac{1}{2} \partial_\varphi \sum_{j=1}^v (\mathbf{x}_j - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_j - \boldsymbol{\mu}) \right]_{\varphi=\hat{\varphi}_{MLE}} \\
 &= \left[-\frac{v}{2} \text{Tr}[\Sigma^{-1} \partial_\varphi \Sigma] - \frac{1}{2} \partial_\varphi \sum_{j=1}^v \text{Tr}[\Sigma^{-1} (\mathbf{x}_j - \boldsymbol{\mu})(\mathbf{x}_j - \boldsymbol{\mu})^T] \right]_{\varphi=\hat{\varphi}_{MLE}} \\
 &= \frac{1}{2} \text{Tr} \left[\partial_\varphi \Sigma^{-1} \left(v \Sigma - \sum_{j=1}^v (\mathbf{x}_j - \boldsymbol{\mu})(\mathbf{x}_j - \boldsymbol{\mu})^T \right) \right]_{\varphi=\hat{\varphi}_{MLE}} \\
 &\quad + \left[(\partial_\varphi \boldsymbol{\mu})^T \Sigma^{-1} \left(v \boldsymbol{\mu} - \sum_{j=1}^v \mathbf{x}_j \right) \right]_{\varphi=\hat{\varphi}_{MLE}}, \tag{A66}
 \end{aligned}$$

where we exploited Jacobi’s formula for the derivative of the determinant of a matrix shown in Equation (A26), the identity in Equation (A28), and the symmetry of Σ . Equation (A66) is the expression of the Likelihood Equation (47). This can be easily simplified to the case of univariate Gaussian distribution for $M = 1$ and absence of average $\boldsymbol{\mu}$, for which $\Sigma \rightarrow \sigma^2$ and $\mathbf{x}_j \rightarrow x_j$, obtaining

$$0 = \Sigma - \frac{1}{v} \sum_{j=1}^v x_j^2 \tag{A67}$$

as shown in Equation (19).

Appendix E. Formulas for the Determinant of a Sum of Two Matrices

Let us consider an $L \times L$ matrix Z of the form $Z = D + W$, where $D = \text{diag}(d_1, \dots, d_L)$ is a diagonal matrix, and $\text{rank}(W) = \rho \leq L$. In this appendix, we will show a convenient method to express the determinant $\det[Z]$ in terms of the elements of W , for the evaluation of the determinant in Equation (37) and of the cofactor matrix in Equations (39).

We exploit the identity [47]

$$\det[Z] = \det[D + W] = \sum_{\alpha=0}^L \Theta_\alpha(D, W), \tag{A68}$$

where $\Theta_\alpha(X, Y)$ is the sum of the determinants of the matrices obtained by replacing any set of α columns (rows) of X with the α columns (rows) of Y at the same position. For example, for two 3×3 matrices

$$X = \begin{pmatrix} X_{11} & X_{12} & X_{13} \\ X_{21} & X_{22} & X_{23} \\ X_{31} & X_{32} & X_{33} \end{pmatrix} \quad Y = \begin{pmatrix} Y_{11} & Y_{12} & Y_{13} \\ Y_{21} & Y_{22} & Y_{23} \\ Y_{31} & Y_{32} & Y_{33} \end{pmatrix}, \tag{A69}$$

the quantity $\Theta_\alpha(D, W)$ is given, for $\alpha = 1$, by

$$\Theta_1(X, Y) = \det \begin{bmatrix} Y_{11} & X_{12} & X_{13} \\ Y_{21} & X_{22} & X_{23} \\ Y_{31} & X_{32} & X_{33} \end{bmatrix} + \det \begin{bmatrix} X_{11} & Y_{12} & X_{13} \\ X_{21} & Y_{22} & X_{23} \\ X_{31} & Y_{32} & X_{33} \end{bmatrix} + \det \begin{bmatrix} X_{11} & X_{12} & Y_{13} \\ X_{21} & X_{22} & Y_{23} \\ X_{31} & X_{32} & Y_{33} \end{bmatrix}, \tag{A70}$$

while for $\alpha = 2$

$$\Theta_2(X, Y) = \det \begin{bmatrix} Y_{11} & Y_{12} & X_{13} \\ Y_{21} & Y_{22} & X_{23} \\ Y_{31} & Y_{32} & X_{33} \end{bmatrix} + \det \begin{bmatrix} X_{11} & Y_{12} & Y_{13} \\ X_{21} & Y_{22} & Y_{23} \\ X_{31} & Y_{32} & Y_{33} \end{bmatrix} + \det \begin{bmatrix} Y_{11} & X_{12} & Y_{13} \\ Y_{21} & X_{22} & Y_{23} \\ Y_{31} & X_{32} & Y_{33} \end{bmatrix}. \tag{A71}$$

Since $\text{rank}(W) = \rho$, the determinant of any $\alpha \times \alpha$ sub-matrix of W is zero if $\alpha > \rho$, so we can write

$$\det[Z] = \sum_{\alpha=0}^{\rho} \Theta_{\alpha}(D, W). \tag{A72}$$

Let us now make explicit the first, easier terms of the summation. In particular, with the goal of applying this expression to the covariance matrix Σ in Equation (35) and to the cofactor matrix C in Equation (39), which can both be expressed as the sum of a diagonal matrix and a rank-two and rank-three matrix (see Equations (A19), (A20) and (A32)), we will explicitly report the first three terms $\Theta_0(X, Y)$, $\Theta_1(X, Y)$ and $\Theta_2(X, Y)$.

For $\alpha = 0$, no columns are replaced from D , hence $\Theta_0(D, W) = \det[D] = \prod_k d_k$. We also notice that, if at least one of the eigenvalues of D is zero, this term vanishes.

For $\alpha = 1$, $\Theta_1(D, W)$ is the sum of determinants of matrices of the form

$$\begin{pmatrix} d_1 & 0 & \dots & 0 & W_{1i} & 0 & \dots & 0 & 0 \\ 0 & d_2 & \dots & 0 & W_{2i} & 0 & \dots & 0 & 0 \\ & & \ddots & & \vdots & & \ddots & & \\ 0 & 0 & \dots & d_{i-1} & W_{i-1i} & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & W_{ii} & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & W_{i+1i} & d_{i+1} & \dots & 0 & 0 \\ & & \vdots & & \vdots & & \ddots & & \\ 0 & 0 & \dots & 0 & W_{L-1i} & 0 & \dots & d_{L-1} & 0 \\ 0 & 0 & \dots & 0 & W_{Li} & 0 & \dots & 0 & d_L \end{pmatrix} \tag{A73}$$

with $i = 1, \dots, L$. The determinant of a matrix of the type in Equation (A73) is straightforward, and reduces to $W_{ii} \times \prod_{k \neq i} d_k$. Thus, in general, $\Theta_1(D, W) = \sum_i W_{ii} \times \prod_{k \neq i} d_k$. A key observation to make, in view of the application to the covariance matrix Σ and cofactor matrix C , is that if two or more eigenvalues of D are zero, all these determinants are zero, and thus $\Theta_1(D, W) = 0$; if, instead, a single eigenvalue is zero, say $d_j = 0$, only one of these determinants of $\Theta_1(D, W)$ is non-vanishing, namely the determinant of the matrix obtained by replacing exactly the j -th column of D . In this case, for $d_j = 0$, then we have $\Theta_1(D, W) = W_{jj} \times \prod_{k \neq j} d_k$.

Let us now consider lastly the case $\alpha = 2$. The matrices contributing to $\Theta_2(D, W)$ are of the form

$$\begin{pmatrix} d_1 & \dots & 0 & W_{1i} & 0 & \dots & 0 & W_{1i'} & 0 & \dots & 0 \\ & & \ddots & \vdots & \vdots & & & \vdots & & \ddots & \\ 0 & \dots & d_{i-1} & W_{i-1i} & 0 & \dots & 0 & W_{i-1i'} & 0 & \dots & 0 \\ 0 & \dots & 0 & W_{ii} & 0 & \dots & 0 & W_{ii'} & 0 & \dots & 0 \\ 0 & \dots & 0 & W_{i+1i} & d_{i+1} & \dots & 0 & W_{i+1i'} & 0 & \dots & 0 \\ & & \vdots & \vdots & & \ddots & & \vdots & & \ddots & \\ 0 & \dots & 0 & W_{i'-1i} & 0 & \dots & d_{i'-1} & W_{i'-1i'} & 0 & \dots & 0 \\ 0 & \dots & 0 & W_{i'i} & 0 & \dots & 0 & W_{i'i'} & 0 & \dots & 0 \\ 0 & \dots & 0 & W_{i'+1i} & 0 & \dots & 0 & W_{i'+1i'} & d_{i'+1} & \dots & 0 \\ & & \vdots & \vdots & \vdots & & & \vdots & & \ddots & \\ 0 & \dots & 0 & W_{Li} & 0 & \dots & 0 & W_{Li'} & 0 & \dots & d_L \end{pmatrix}, \tag{A74}$$

with $i < i' = 1, \dots, L$. Once again, the determinants of the matrices of the type in Equation (A74) are easy to be evaluated, and read $\det[W^{(i,i')}] \prod_{k \notin \{i,i'\}} d_k$, where

$$W^{(i,i')} = \begin{pmatrix} W_{ii} & W_{ii'} \\ W_{i'i} & W_{i'i'} \end{pmatrix}. \quad (\text{A75})$$

We notice that $\det[W^{(i,i')}] = \det[W^{(i',i)}]$. Thus, we can rewrite the terms $\Theta_2(D, W)$ as the sum $\Theta_2(D, W) = \sum_i \sum_{i' > i} \det[W^{(i,i')}] \prod_{k \notin \{i,i'\}} d_k$. Once again, key observations can be made: if the diagonal matrix D has at least three null eigenvalues, then $\Theta_2(D, W)$ is vanishing. If only two eigenvalues are zero, e.g., $d_j = d_{j'} = 0$, with $j \neq j'$, then there is only one non-vanishing contribution to $\Theta_2(D, W)$, given by the matrix obtained by substituting the j -th and j' -th columns of D , and in this case we have $\Theta_2(D, W) = \det[W^{(j,j')}] \prod_{k \notin \{j,j'\}} d_k$. If only one eigenvalue is zero, namely $d_j = 0$, then $\Theta_2(D, W)$ is given by the sum of all the determinants of the matrices where the j -th column of D has been replaced, namely $\Theta_2(D, W) = \sum_{i \neq j} \det[W^{(i,j)}] \prod_{k \neq i,j} d_k$.

It is then possible to extend with similar considerations the same method for every value of α , eventually obtaining the compact expression for the determinant in Equation (A68)

$$\det[Z] = \sum_{\alpha=0}^{\rho} \sum_{\gamma \in C_{\alpha}^L} \det[W^{(\gamma)}] \prod_{k \notin \gamma} d_k, \quad (\text{A76})$$

where C_{α}^L is the set of all the combinations without repetitions of α columns out of L , $W^{(\gamma)}$ denotes the $\alpha \times \alpha$ sub-matrix of W obtained by selecting the rows and columns with indices $\gamma_1, \dots, \gamma_{\alpha}$.

References

1. Caves, C.M. Quantum-mechanical noise in an interferometer. *Phys. Rev. D* **1981**, *23*, 1693–1708. [CrossRef]
2. Bondurant, R.S.; Shapiro, J.H. Squeezed states in phase-sensing interferometers. *Phys. Rev. D* **1984**, *30*, 2548–2556. [CrossRef]
3. Wineland, D.J.; Bollinger, J.J.; Itano, W.M.; Moore, F.L.; Heinzen, D.J. Spin squeezing and reduced quantum noise in spectroscopy. *Phys. Rev. A* **1992**, *46*, R6797–R6800. [CrossRef]
4. Giovannetti, V.; Lloyd, S.; Maccone, L. Quantum-Enhanced Measurements: Beating the Standard Quantum Limit. *Science* **2004**, *306*, 1330–1336. Available online: <https://www.science.org/doi/pdf/10.1126/science.1104149> (accessed on 28 January 2022). [CrossRef] [PubMed]
5. Giovannetti, V.; Lloyd, S.; Maccone, L. Quantum Metrology. *Phys. Rev. Lett.* **2006**, *96*, 010401. [CrossRef]
6. Dowling, J.P. Quantum optical metrology—The lowdown on high-N00N states. *Contemp. Phys.* **2008**, *49*, 125–143. [CrossRef]
7. Paris, M.G. Quantum estimation for quantum technology. *Int. J. Quantum Inf.* **2009**, *7*, 125–137. [CrossRef]
8. Giovannetti, V.; Lloyd, S.; Maccone, L. Advances in quantum metrology. *Nat. Photonics* **2011**, *5*, 222–229. [CrossRef]
9. Lang, M.D.; Caves, C.M. Optimal Quantum-Enhanced Interferometry Using a Laser Power Source. *Phys. Rev. Lett.* **2013**, *111*, 173601. [CrossRef]
10. Tóth, G.; Apellaniz, I. Quantum metrology from a quantum information science perspective. *J. Phys. A Math. Theor.* **2014**, *47*, 424006. [CrossRef]
11. Dowling, J.P.; Seshadreesan, K.P. Quantum Optical Technologies for Metrology, Sensing, and Imaging. *J. Light. Technol.* **2015**, *33*, 2359–2370. [CrossRef]
12. Szczykulska, M.; Baumgratz, T.; Datta, A. Multi-parameter quantum metrology. *Adv. Phys. X* **2016**, *1*, 621–639. [CrossRef]
13. Schnabel, R. Squeezed states of light and their applications in laser interferometers. *Phys. Rep.* **2017**, *684*, 1–51. [CrossRef]
14. Braun, D.; Adesso, G.; Benatti, F.; Floreanini, R.; Marzolino, U.; Mitchell, M.W.; Pirandola, S. Quantum-enhanced measurements without entanglement. *Rev. Mod. Phys.* **2018**, *90*, 035006. [CrossRef]
15. Pirandola, S.; Bardhan, B.R.; Gehring, T.; Weedbrook, C.; Lloyd, S. Advances in photonic quantum sensing. *Nat. Photonics* **2018**, *12*, 724–733. [CrossRef]
16. Polino, E.; Valeri, M.; Spagnolo, N.; Sciarrino, F. Photonic quantum metrology. *AVS Quantum Sci.* **2020**, *2*, 024703. [CrossRef]
17. Maccone, L.; Riccardi, A. Squeezing metrology: A unified framework. *Quantum* **2020**, *4*, 292. [CrossRef]
18. Schleich, W. *Quantum Optics in Phase Space*; Wiley: Hoboken, NJ, USA, 2011. [CrossRef]
19. Weedbrook, C.; Pirandola, S.; García-Patrón, R.; Cerf, N.J.; Ralph, T.C.; Shapiro, J.H.; Lloyd, S. Gaussian quantum information. *Rev. Mod. Phys.* **2012**, *84*, 621–669. [CrossRef]

20. Adesso, G.; Ragy, S.; Lee, A.R. Continuous Variable Quantum Information: Gaussian States and Beyond. *Open Syst. Inf. Dyn.* **2014**, *21*, 1440001. [[CrossRef](#)]
21. Lvovsky, A.I. Squeezed Light. In *Photonics*; John Wiley and Sons, Ltd.: Hoboken, NJ, USA, 2015; Chapter 5, pp. 121–163. Available online: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119009719.ch5> (accessed on 28 January 2022). [[CrossRef](#)]
22. Cerf, N.J.; Leuchs, G.; Polzik, E.S. *Quantum Information with Continuous Variables of Atoms and Light*; Imperial College Press: London, UK, 2007. Available online: <https://www.worldscientific.com/doi/pdf/10.1142/p489> (accessed on 28 January 2022). [[CrossRef](#)]
23. Grangier, P.; Slusher, R.E.; Yurke, B.; LaPorta, A. Squeezed-light—Enhanced polarization interferometer. *Phys. Rev. Lett.* **1987**, *59*, 2153–2156. [[CrossRef](#)]
24. Monras, A. Optimal phase measurements with pure Gaussian states. *Phys. Rev. A* **2006**, *73*, 033821. [[CrossRef](#)]
25. Yonezawa, H.; Nakane, D.; Wheatley, T.A.; Iwasawa, K.; Takeda, S.; Arao, H.; Ohki, K.; Tsumura, K.; Berry, D.W.; Ralph, T.C.; et al. Quantum-Enhanced Optical-Phase Tracking. *Science* **2012**, *337*, 1514–1517. Available online: <https://www.science.org/doi/pdf/10.1126/science.1225258> (accessed on 28 January 2022). [[CrossRef](#)]
26. Oh, C.; Lee, C.; Rockstuhl, C.; Jeong, H.; Kim, J.; Nha, H.; Lee, S.Y. Optimal Gaussian measurements for phase estimation in single-mode Gaussian metrology. *npj Quantum Inf.* **2019**, *5*, 10. [[CrossRef](#)]
27. Matsubara, T.; Facchi, P.; Giovannetti, V.; Yuasa, K. Optimal Gaussian metrology for generic multimode interferometric circuit. *New J. Phys.* **2019**, *21*, 033014. [[CrossRef](#)]
28. Yurke, B.; McCall, S.L.; Klauder, J.R. SU(2) and SU(1,1) interferometers. *Phys. Rev. A* **1986**, *33*, 4033–4054. [[CrossRef](#)]
29. Gatto, D.; Facchi, P.; Narducci, F.A.; Tamma, V. Distributed quantum metrology with a single squeezed-vacuum source. *Phys. Rev. Res.* **2019**, *1*, 032024. [[CrossRef](#)]
30. Gatto, D.; Facchi, P.; Tamma, V. Phase space Heisenberg-limited estimation of the average phase shift in a Mach–Zehnder interferometer. *Int. J. Quantum Inf.* **2020**, *18*, 1941019. [[CrossRef](#)]
31. Ou, Z.Y.; Li, X. Quantum SU(1,1) interferometers: Basic principles and applications. *APL Photonics* **2020**, *5*, 080902. [[CrossRef](#)]
32. Gatto, D.; Facchi, P.; Tamma, V. Heisenberg-limited estimation robust to photon losses in a Mach-Zehnder network with squeezed light. *Phys. Rev. A* **2022**, *105*, 012607. [[CrossRef](#)]
33. Farace, A.; Pasquale, A.D.; Adesso, G.; Giovannetti, V. Building versatile bipartite probes for quantum metrology. *New J. Phys.* **2016**, *18*, 013049. [[CrossRef](#)]
34. Gramegna, G.; Triggiani, D.; Facchi, P.; Narducci, F.A.; Tamma, V. Typicality of Heisenberg scaling precision in multimode quantum metrology. *Phys. Rev. Res.* **2021**, *3*, 013152. [[CrossRef](#)]
35. Armen, M.A.; Au, J.K.; Stockton, J.K.; Doherty, A.C.; Mabuchi, H. Adaptive Homodyne Measurement of Optical Phase. *Phys. Rev. Lett.* **2002**, *89*, 133602. [[CrossRef](#)]
36. Aspachs, M.; Calsamiglia, J.; Muñoz Tapia, R.; Bagan, E. Phase estimation for thermal Gaussian states. *Phys. Rev. A* **2009**, *79*, 033834. [[CrossRef](#)]
37. Berni, A.A.; Gehring, T.; Nielsen, B.M.; Händchen, V.; Paris, M.G.A.; Andersen, U.L. Ab initio quantum-enhanced optical phase estimation using real-time feedback control. *Nat. Photonics* **2015**, *9*, 577–581. [[CrossRef](#)]
38. Zhuang, Q.; Zhang, Z.; Shapiro, J.H. Distributed quantum sensing using continuous-variable multipartite entanglement. *Phys. Rev. A* **2018**, *97*, 032329. [[CrossRef](#)]
39. Guo, X.; Breum, C.R.; Borregaard, J.; Izumi, S.; Larsen, M.V.; Gehring, T.; Christandl, M.; Neergaard-Nielsen, J.S.; Andersen, U.L. Distributed quantum sensing in a continuous-variable entangled network. *Nat. Phys.* **2020**, *16*, 281–284. [[CrossRef](#)]
40. Grace, M.R.; Gagatsos, C.N.; Zhuang, Q.; Guha, S. Quantum-Enhanced Fiber-Optic Gyroscopes Using Quadrature Squeezing and Continuous-Variable Entanglement. *Phys. Rev. Appl.* **2020**, *14*, 034065. [[CrossRef](#)]
41. Grace, M.R.; Gagatsos, C.N.; Guha, S. Entanglement-enhanced estimation of a parameter embedded in multiple phases. *Phys. Rev. Res.* **2021**, *3*, 033114. [[CrossRef](#)]
42. Gramegna, G.; Triggiani, D.; Facchi, P.; Narducci, F.A.; Tamma, V. Heisenberg scaling precision in multi-mode distributed quantum metrology. *New J. Phys.* **2021**, *23*, 053002. [[CrossRef](#)]
43. Triggiani, D.; Facchi, P.; Tamma, V. Non-adaptive Heisenberg-limited metrology with multi-channel homodyne measurements. *Eur. Phys. J. Plus* **2022**, *137*, 125. [[CrossRef](#)]
44. Cramér, H. *Mathematical Methods of Statistics (PMS-9)*; Princeton University Press: Princeton, NJ, USA, 1946. [[CrossRef](#)]
45. Rohatgi, V.K.; Saleh, A.M.E. *An Introduction to Probability and Statistics*; John Wiley and Sons: Hoboken, NJ, USA, 2015.
46. Triggiani, D.; Facchi, P.; Tamma, V. Heisenberg scaling precision in the estimation of functions of parameters in linear optical networks. *Phys. Rev. A* **2021**, *104*, 062603. [[CrossRef](#)]
47. Xu, S.; Darouach, M.; Schaeffers, J. Expansion of $\det(A+B)$ and robustness analysis of uncertain state space systems. *IEEE Trans. Autom. Control* **1993**, *38*, 1671–1675. [[CrossRef](#)]