Typicality of Heisenberg scaling precision in multimode quantum metrology

Giovanni Gramegna [1,2,*] Danilo Triggiani [3,†] Paolo Facchi [1,2] Frank A. Narducci [4] and Vincenzo Tamma [3,5,‡]

[1]*Dipartimento di Fisica and MECENAS, Università di Bari, I-70126 Bari, Italy*
[2]*INFN, Sezione di Bari, I-70126 Bari, Italy*
[3]*School of Mathematics and Physics, University of Portsmouth, Portsmouth PO1 3QL, United Kingdom*
[4]*Department of Physics, Naval Postgraduate School, Monterey, California 93943, USA*
[5]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, United Kingdom*

We propose a measurement setup reaching Heisenberg scaling precision for the estimation of any distributed parameter φ (not necessarily a phase) encoded into a generic M-port linear network composed only of passive elements. The scheme proposed can be easily implemented from an experimental point of view since it employs only Gaussian states and Gaussian measurements. Due to the complete generality of the estimation problem considered, it was predicted that one would need to carry out an adaptive procedure which involves both the input states employed and the measurement performed at the output; we show that this is not necessary: Heisenberg scaling precision is still achievable by only adapting a single stage. The nonadapted stage only affects the value of a prefactor multiplying the Heisenberg scaling precision: We show that, for large values of M and a random (unbiased) choice of the nonadapted stage, this prefactor takes a typical value which can be controlled through the encoding of the parameter φ into the linear network.

I. INTRODUCTION

The precision achievable in a measurement when all experimental noise sources are minimized is ultimately determined by the discreteness of all physical phenomena: Electronic devices will suffer the discreteness of the electric charge, whereas the quantum nature of light will affect optical devices. Due to this quantum noise, the error in the estimation of a physical parameter φ through a measurement employing N probes (e.g., photons, electrons) is strongly limited by the so-called "shot noise" factor of $1/\sqrt{N}$. However, it has been proven that quantum features such as entanglement and squeezing can be exploited to go beyond the shot-noise limit and reach a precision of order $1/N$, the so-called Heisenberg limit (HL) [1–8].

Several quantum metrological problems have been largely studied and a few approaches have been proposed to reach a HL sensitivity. However, these protocols are usually difficult to implement experimentally due to the convoluted and challenging measurement procedures [9–11] and the fragile quantum coherence needed in the input states [2,12–14]. Gaussian states, on the other hand, provide a promising avenue for quantum optical technologies [15,16], since they are easier to create and manipulate experimentally compared to non-Gaussian ones, such as Fock states. Moreover, they allow a complete analytical treatment from a theoretical point of view [15–17]. In particular, the *squeezing* of a Gaussian state, which allows for highly reduced-noise signals, appears to be a valuable tool to reach quantum super-sensitive precision [8]. From a metrological perspective, squeezed states are often used along with Gaussian measurements [18–20], defined as measurement schemes producing a Gaussian probability distribution of the outcomes for any Gaussian state [15]. Homodyne and heterodyne detection represent paradigmatic examples of Gaussian measurements. It has been shown, both theoretically [21] and experimentally [22], that an adaptive homodyne phase estimation performs better than heterodyne detection, and approaches closer to the intrinsic quantum uncertainty than any previous technique when no prior knowledge of the phase is given. The importance of feedback and adaptivity in quantum estimation protocols has been underlined also in subsequent works [18,23]. Adaptiveness can be avoided in an optimal protocol (or near optimal) only if some constraint in the range of variation of the parameter is given [24–26].

Within the domain of quantum optics, photons are sent as probes through an interferometer where a parameter φ to be estimated is encoded. The information about the parameter is imprinted then in the output state of the photons, and it can be extracted by a suitable measurement. The situation which has been often considered is the case where φ is an optical phase [1,3,6,18,20] or a phaselike parameter [2,4]. These results clearly apply also to situations in which other quantities of interest (e.g., a distance) can be converted into an optical phase [3], but they fail to cover more general situations, e.g.,

*giovanni.gramegna@ba.infn.it
†danilo.triggiani@port.ac.uk
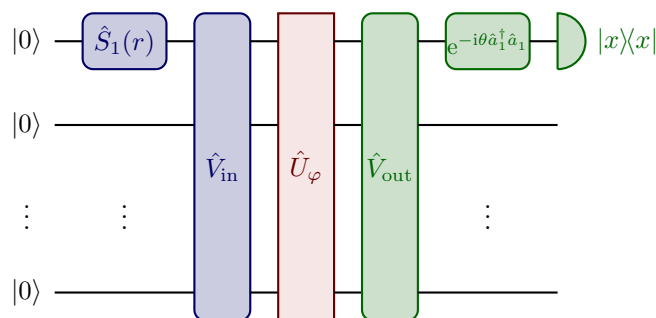‡vincenzo.tamma@port.ac.uk

FIG. 1. Block diagram of the investigated setup [34]. A single-mode squeezed vacuum state with real squeezing parameter r is injected into the first preparation stage $\hat{V}_{\text{in}}$, which inputs the linear network $\hat{U}_\varphi$ encoding the parameter φ to be estimated. After the network, there is a second stage $\hat{V}_{\text{out}}$ before the measurement. Finally, homodyne detection on the first output port of $\hat{V}_{\text{out}}$ is performed, and the quadrature field $\hat{x}_\theta$ is measured. In order to reach Heisenberg scaling sensitivity in the estimation of φ, it suffices to optimize only one of the two auxiliary stages, in such a way that one of the conditions (22) or (23) holds.

the unknown parameter is distributed among several components of the interferometer. Recently, some progress has been made along this direction concerning the estimation of particular functions of multiple parameters distributed in a specific manner within a particular network [27–33]. It has been also shown in a recent work [19] that the presence of a single unknown parameter distributed in multiple nodes of an arbitrary network introduces nontrivial complications if no constraints are given on the range of values the parameter is allowed to assume: In fact, it appears that a simultaneous adaptive procedure both in the input probe and in the measurement is needed in order to reach the HL, making the whole scheme quite unfeasible from a practical point of view. Furthermore, the proposed scheme requires an unquantified precision and number of resources in the adaptive procedure.

In this work we demonstrate the typicality of the Heisenberg limited sensitivity with a simple metrological technique which overcomes at the same time all these serious drawbacks [34]. In particular, we consider a general scenario in which φ can be any parameter embedded into an arbitrary linear passive M-mode interferometer: It can be a parameter characterizing any specific component of the interferometer, or arbitrarily distributed among different components of the circuit. We will show that an experimentally feasible scheme achieving Heisenberg scaling is typically possible in such a general scenario. In our scheme (see Fig. 1), a single-mode squeezed vacuum state is sent through a linear, passive preliminary stage which scatters the input photons among all the M channels of the interferometer, in order to extract the information on φ which is distributed among all the modes. A second auxiliary stage at the output of the interferometer refocuses the photons in the only observed output port. By employing a single-mode homodyne detection, we present two broad conditions which together suffice to reach the HL: the first being the requirement that most of the injected photons are successfully refocused on the observed output mode; the second simply being a minimal-resolution requirement

on the homodyne measurement. Remarkably, these conditions allow for imperfections both in the refocusing and in the measurement. Heisenberg scaling is thus achievable by choosing two additional passive and linear stages, whose roles are to conveniently scatter the input probe to all the M modes, and then to refocus the photons. Despite the fact that the choice of these unitary stages can be in general φ dependent, we show that it is always sufficient to adapt only one of the two stages, which we will thus call optimized, while the other stage can be chosen arbitrarily and independent of the parameter. Moreover we show that the optimized stage can be prepared with a precision which is achievable using only classical resources, or by means of a preliminary classical estimation. This is also consistent with the result obtained in [34]; namely, that a preliminary classical estimation of φ yields enough information to correctly prepare the optimized stage and thus to achieve Heisenberg scaling in the estimation protocol. Finally we show that the nonoptimized stage affects the precision simply by a constant prefactor. Using typicality and results of measure concentration in high-dimensional vector spaces, we show that distributing the unknown parameter among a high number of modes M allows this prefactor to typically take nonvanishing values.

The rest of the paper is organized as follows. In Sec. II we describe the proposed optical interferometer and the relative Fisher information. In Sec. III we use the Fisher information to prove that the Heisenberg scaling can be achieved under suitable physical conditions; we then show how, even in the most general case, all the adaptivity can be confined within one of the auxiliary stages. In Sec. IV we discuss the typicality of our results for interferometers with a large number of channels. Finally, in Sec. V we draw some conclusions and discuss the outlook.

II. THE PROPOSED SETUP

Let us consider a metrological scheme where the parameter to be estimated is encoded into an M-port passive linear network described by the unitary $\hat{U}_\varphi$ acting on M bosonic modes $\hat{a}_j$ ($j = 1, \ldots, M$) obeying the canonical commutation relations $[\hat{a}_j, \hat{a}_k^\dagger] = \delta_{jk}$, and $[\hat{a}_j, \hat{a}_k] = [\hat{a}_j^\dagger, \hat{a}_k^\dagger] = 0$. For a passive linear network, the action of $\hat{U}_\varphi$ on the annihilation operators is associated with an $M \times M$ unitary matrix via

$$\hat{U}_\varphi^\dagger \hat{a}_j \hat{U}_\varphi = \sum_{k=1}^{M} (U_\varphi)_{jk} \hat{a}_k. \tag{1}$$

The unitarity of the matrix U_φ is strictly related to the conservation of the number of photons injected. By definition, U_φ is the matrix of the single-photon transition amplitudes, i.e., $|(U_\varphi)_{jk}|^2$ is the probability that a single photon injected into the kth input channel ends up in the jth output channel due to the action of the network.

We now propose an estimation scheme reaching Heisenberg scaling if suitable conditions are satisfied. As shown in Fig. 1, the preparation of the input probe consists in two steps: First, we inject a single-mode squeezed vacuum in the first port of $\hat{V}_{\text{in}}$. Then, the unitary stage $\hat{V}_{\text{in}}$ is used to scatter the photons injected among all the modes. The input state of the network $\hat{U}_\varphi$ in our protocol is therefore given

by $|\psi_0\rangle = \hat{V}_{\rm in}\hat{S}_1(r)|{\rm vac}\rangle$, where $\hat{S}_1(r) = e^{\frac{r}{2}(\hat{a}_1^2 - \hat{a}_1^{\dagger 2})}$ is a single-mode squeezing operator with squeezing parameter $r > 0$, and $|{\rm vac}\rangle = |0\rangle^{\otimes M}$ is the M-mode vacuum state. The average number of photons injected in the apparatus is thus $N = \sinh^2 r$. The state $\hat{U}_\varphi|\psi_0\rangle$ at the output of the network $\hat{U}_\varphi$ undergoes the unitary $\hat{V}_{\rm out}$ which refocuses all the photons into a single mode, namely the first one, where a homodyne measurement of the field quadrature $\hat{x}_\theta$ is performed. If the refocusing procedure is not perfect there will be some photons scattered into other channels with probability $1 - P_\varphi$, where

$$P_\varphi = |(V_{\rm out}U_\varphi V_{\rm in})_{11}|^2 \qquad (2)$$

is defined by the probability amplitude $(u_\varphi)_{11} = (V_{\rm out}U_\varphi V_{\rm in})_{11}$ for the transition from the first input to the first output port in the overall interferometer $u_\varphi = V_{\rm out}U_\varphi V_{\rm in}$, with $V_{\rm in}$ and $V_{\rm out}$ being the single-photon unitary matrix representatives of $\hat{V}_{\rm in}$ and $\hat{V}_{\rm out}$, respectively, obtained analogously to (1).

The homodyne measurement is described by a positive operator valued measure (POVM) $\mathcal{M} = \{\hat{\Pi}_x\}$, whose elements are defined by

$$\hat{\Pi}_x = e^{i\theta\hat{a}_1^\dagger\hat{a}_1}|x\rangle_{11}\langle x|e^{-i\theta\hat{a}_1^\dagger\hat{a}_1}. \qquad (3)$$

The probability of obtaining a value x from a measurement of the quadrature $\hat{x}_\theta = e^{i\theta\hat{a}_1^\dagger\hat{a}_1}\hat{x}_1 e^{-i\theta\hat{a}_1^\dagger\hat{a}_1}$ is then given by Born's rule

$$p(x|\varphi) = {\rm Tr}(\hat{\Pi}_x\hat{u}_\varphi\hat{S}_1(r)|{\rm vac}\rangle\langle{\rm vac}|\hat{S}_1^\dagger(r)\hat{u}_\varphi^\dagger), \qquad (4)$$

over the output state $\hat{u}_\varphi\hat{S}_1(r)|{\rm vac}\rangle$ after the overall interferometric evolution $\hat{u}_\varphi = \hat{V}_{\rm out}\hat{U}_\varphi\hat{V}_{\rm in}$, which yields (see Appendix A)

$$p(x|\varphi) = \frac{1}{\sqrt{2\pi\Delta_\varphi}}\exp\left(-\frac{x^2}{2\Delta_\varphi}\right), \qquad (5)$$

where the variance of the Gaussian distribution,

$$\begin{aligned}\Delta_\varphi = &\tfrac{1}{2}\big(1 + |(u_\varphi)_{11}|^2(\cosh 2r - 1) \\ &+ {\rm Re}\big[e^{-2i\theta}(u_\varphi)_{11}^2\big]\sinh 2r\big),\end{aligned} \qquad (6)$$

encodes the parameter φ through the interferometric transition amplitude $(u_\varphi)_{11} = (V_{\rm out}U_\varphi V_{\rm in})_{11}$.

It is known from classical estimation theory that the maximum precision attainable when inferring the value of the unknown parameter φ, is given by the so-called Cramér-Rao bound [35,36]:

$$\delta\varphi \geqslant \frac{1}{\sqrt{\nu F(\varphi)}} = \delta\varphi_{\min}, \qquad (7)$$

where ν is the number of measurements performed, while $F(\varphi)$ is the Fisher information (FI):

$$F(\varphi) = \int p(x|\varphi)\left(\frac{\partial\log p(x|\varphi)}{\partial\varphi}\right)^2 dx. \qquad (8)$$

The bound (7) can be asymptotically saturated through post-processing Bayesian data analysis [37–39].

The Fisher information related to a Gaussian probability distribution with variance Δ_φ can be evaluated by inserting

(5) into (8), and it reads

$$F(\varphi) = \frac{1}{2}\left(\frac{\partial_\varphi\Delta_\varphi}{\Delta_\varphi}\right)^2. \qquad (9)$$

Plugging (6) into (9), it is possible to explicitly evaluate the FI (see Appendix B) obtaining

$$F(\varphi) = 2\left(\frac{(\partial_\varphi P_\varphi)f(N) - 2P_\varphi(\partial_\varphi\gamma_\varphi)h(N)}{1 + 2P_\varphi f(N)}\right)^2, \qquad (10)$$

where

$$f(N) := N(1 + \cos[2(\gamma_\varphi - \theta)]\sqrt{1 + 1/N}), \qquad (11)$$

$$h(N) := N\sin[2(\gamma_\varphi - \theta)]\sqrt{1 + 1/N}, \qquad (12)$$

with

$$\gamma_\varphi = \arg(u_\varphi)_{11} = \arg(V_{\rm out}U_\varphi V_{\rm in})_{11} \qquad (13)$$

being the accumulated phase through the interferometric evolution.

III. HEISENBERG SCALING

We now demonstrate that Heisenberg scaling sensitivity can be achieved in the proposed metrological setup shown in Fig. 1, if two conditions are met. The first condition is the constraint that the average number of photons scattered into channels which are not measured is a finite quantity ℓ_φ, independent of N, which translates into the condition,

$$P_\varphi = 1 - \frac{\ell_\varphi}{N} + O\left(\frac{1}{N^2}\right), \quad \ell_\varphi \geqslant 0, \qquad (14)$$

on the probability P_φ in (2). Here, ℓ_φ depends in general on the linear network U_φ in which the parameter is embedded, and on the auxiliary stages $V_{\rm in}$ and $V_{\rm out}$ in Fig. 1. In Sec. III A we will show how, for any given arbitrary U_φ, it is possible to optimize only a single stage so that the probability distribution in (2) can be expressed as (14). The second condition relates the accumulated phase $\gamma_\varphi = \arg(u_\varphi)_{11}$ through the whole setup and the phase $\theta = \theta_\varphi$ of the measured quadrature field $\hat{x}_\theta$, according to

$$\theta_\varphi = \gamma_\varphi \pm \frac{\pi}{2} + \frac{k_\varphi}{N} + O\left(\frac{1}{N^2}\right), \quad k_\varphi \neq 0, \qquad (15)$$

where k_φ can depend on φ, but is assumed to be independent of N. In practice, one can even fix k_φ to a constant value without using additional resources.

A heuristic explanation behind this condition can be found in Fig. 2: In order to maximize the ratio in (9) while keeping constant $N = \sinh^2 r$, the choice of the quadrature $\hat{x}_\theta$ to be measured is a trade-off between two opposite behaviors. One consists in minimizing the variance Δ_φ in the denominator of (9), while the other consists in maximizing the sensitivity of the variance with respect to the variations of φ, namely choosing θ such that $\partial_\varphi\Delta_\varphi$ in the numerator is maximal. The former is met for θ as close as possible to $\gamma \pm \pi/2$, since $\hat{x}_{\gamma\pm\pi/2}$ are the squeezed quadratures after the rotation in phase space by the phase γ_φ accumulated through the interferometer; the latter instead requires a choice of $\theta - \gamma$ far enough from the stationary points of the variance at $\gamma \pm \pi/2$, where
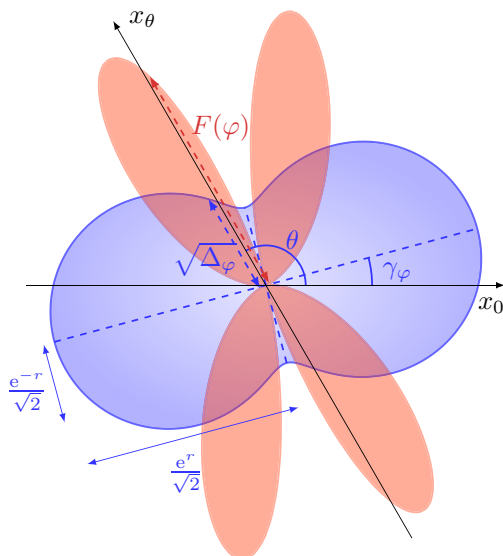
FIG. 2. Phase space representation of the squeezed vacuum state (with squeezing parameter $re^{2i\gamma_\varphi}$) at the first output channel of the whole setup shown in Fig. 1 (blue), and of the Fisher information (10) (four red lobes). We have considered for simplicity the case where all the photons are refocused in the first output channel (when condition (14) reduces to $P_\varphi = 1$). Given any axis at an angle θ with respect to the horizontal axis, corresponding to the $x_0 = x$ quadrature, the distance between its intersections with the ellipse and the origin represents the standard deviation $\sqrt{\Delta_\varphi}$ of the quadrature $\hat{x}_\theta$. In other words, the blue graph is the polar plot of $\sqrt{\Delta_\varphi}$ shown in (6) as a function of θ. A polar plot of the Fisher information is overlaid in red. The Fisher information takes vanishing values if the minimum-variance quadratures are measured, namely for $\theta_{\min} = \gamma_\varphi \pm \pi/2$. This happens because the variance Δ_φ of the quadrature along $\theta_{\min}$ is locally insensitive to the variations of φ. Thus, one needs to get far enough from $\theta_{\min}$ to achieve a suitable high value of the Fisher information. In particular, we have shown that for a large number N of photons it is enough to move from $\theta_{\min}$ of an additional angle of the order $1/N$ as in (15) to reach the Heisenberg scaling in the measure of the parameter φ.

Δ_φ and therefore the overall probability distribution $p(x|\varphi)$ are insensitive to variation in the parameter φ. Noticeably, the larger is N, and thus the squeezing parameter, the closer to the squeezed direction the quadrature field should be measured, as can be seen in (15).

In order to prove the claim of HL scaling, we will evaluate the asymptotics of the Fisher information (10) as $N \to \infty$. Substituting the value $\theta = \theta_\varphi$ in (15) into (11) and (12), we get

$$f(N) = -\frac{1}{2} + \frac{2k_\varphi^2}{N} + \frac{1}{8N} + O\left(\frac{1}{N^2}\right), \qquad (16)$$

$$h(N) = 2k_\varphi\left(1 + \frac{1}{2N}\right) + O\left(\frac{1}{N^2}\right). \qquad (17)$$

Hence, substituting (16) and (17) in (10), and neglecting higher order terms, the asymptotic behavior of the Fisher information reads

$$F(\varphi) \sim 8\varrho(k_\varphi, \ell_\varphi)(\partial_\varphi \gamma_\varphi)^2 N^2, \qquad (18)$$

with

$$\varrho(k, \ell) = \left(\frac{8k}{1 + 16k^2 + 4\ell}\right)^2. \qquad (19)$$

The quadratic scaling in the mean number photons N in (18) finally proves that conditions (14) and (15) suffice to reach the Heisenberg scaling.

The asymptotics for the Fisher information carries two prefactors, $\varrho(k_\varphi, \ell_\varphi)$ and $(\partial_\varphi \gamma_\varphi)^2$. We easily notice that the prefactor $\varrho(k, \ell)$ vanishes only at $k = 0$, and attains its maximum at $\ell = 0$, $k = \pm 1/4$:

$$\varrho(k, \ell) \leqslant \varrho(\pm 1/4, 0) = 1, \qquad (20)$$

so that, with this choice of the constants k and ℓ, the Fisher Information asymptotically reads

$$F(\varphi)|_{\substack{k = \pm\frac{1}{4} \\ \ell = 0}} \sim 8(\partial_\varphi \gamma_\varphi)^2 N^2. \qquad (21)$$

Moreover, $\varrho(k, \ell)$ is a decreasing function of ℓ independent of k, so that $\ell = 0$ is always the best case, meaning that the less photons that are scattered in different channels, the higher the sensitivity in the estimation. Instead, for a fixed arbitrary positive value of ℓ, the maximum of $\varrho(k, \ell)$ is reached for $k = \pm\sqrt{4\ell + 1}/4$.

One-sided adaptivity

Since $P_\varphi = |(V_{\text{out}} U_\varphi V_{\text{in}})_{11}|^2$, condition (14) may appear to require a simultaneous optimization of the input V_{in} and the output V_{out} in a parameter-dependent way. This two-sided adaptation can be quite difficult to realize in practice.

However, we are going to show that in fact conditions (14) and (15) can always be satisfied with just a one-sided parameter-dependent adaptation, which can be performed either at the input or at the output of the network equivalently. And remarkably, this adaptation can be accomplished by performing a preliminary classical, shot-noise limited, estimation of φ.

In particular, one can choose to adaptively optimize only V_{out} and fix V_{in} to an arbitrary parameter-independent unitary stage: In this case, one can set the parameter-dependent condition,

$$(V_{\text{out}})_{1i} = (V_{\text{in}}^\dagger U_\varphi^\dagger)_{1i} + O\left(\frac{1}{\sqrt{N}}\right). \qquad (22)$$

Alternatively, it is possible to adaptively optimize only V_{in} with the condition,

$$(V_{\text{in}})_{i1} = (U_\varphi^\dagger V_{\text{out}}^\dagger)_{i1} + O\left(\frac{1}{\sqrt{N}}\right), \qquad (23)$$

while V_{out} can be arbitrarily chosen.

Remarkably both Eqs. (22) and (23) imply that an error of the order of $O(\frac{1}{\sqrt{N}})$ is allowed to prepare the optimized stage to reach condition (14). To show that both Eqs. (22) and (23) satisfy condition (14), we notice that P_φ can be expressed as a transition probability,

$$P_\varphi = |(V_{\text{out}} U_\varphi V_{\text{in}})_{11}|^2 = |\langle v_{\text{out}} | v_{\text{in}} \rangle|^2, \qquad (24)$$

between the two normalized vectors $|v_{\text{in}}\rangle = U_\varphi V_{\text{in}} |e_1\rangle$ and $|v_{\text{out}}\rangle = V_{\text{out}}^\dagger |e_1\rangle$, with $|e_1\rangle = (1, 0, \ldots, 0)^T$. Then, Eq. (22)

translates into

$$|v_{\text{out}}\rangle = |v_{\text{in}}\rangle + |\delta v\rangle = e^{i\varepsilon H}|v_{\text{in}}\rangle, \quad \varepsilon = O\left(\frac{1}{\sqrt{N}}\right), \quad (25)$$

with some $H = H^\dagger$, by unitarity. Therefore, we can see that

$$
\begin{aligned}
P &= |\langle v_{\text{out}}|v_{\text{in}}\rangle|^2 = |\langle v_{\text{in}}|e^{-i\varepsilon H}v_{\text{in}}\rangle|^2 \\
&= |1 - i\varepsilon\langle v_{\text{in}}|Hv_{\text{in}}\rangle - O(\varepsilon^2)|^2 = 1 - O(\varepsilon^2) \\
&= 1 - O\left(\frac{1}{N}\right).
\end{aligned}
\quad (26)
$$

We can notice that in both Eqs. (22) and (23), no assumption on the nonoptimized stage is made, so that its choice is completely arbitrary. This freedom affects the precision of the estimation of φ through the N-independent prefactor $(\partial_\varphi\gamma_\varphi)^2$ which appears in the Fisher information (18). At this point, one may argue that this prefactor may be vanishing if a poor choice for the nonadapted unitary is made. Remarkably, in the next section we will show that the prefactor is typically nonvanishing for random choices of the nonadapted stage and suitably well-behaved given linear networks $\hat{U}_\varphi$.

IV. TYPICAL SENSITIVITY

In this section we will address in more detail the study of the prefactor $(\partial_\varphi\gamma_\varphi)^2$ in the Fisher information (18), clarifying under what circumstances it can be safely considered nonvanishing and characterizing its magnitude for random choices of the nonoptimized stage. First of all, we can link $(\partial_\varphi\gamma_\varphi)^2$ to the derivative of the matrix element $(u_\varphi)_{11} = (V_{\text{out}}U_\varphi V_{\text{in}})_{11} = \sqrt{P_\varphi}e^{i\gamma_\varphi}$:

$$
\begin{aligned}
|(\partial_\varphi u_\varphi)_{11}|^2 &= |(\partial_\varphi\sqrt{P_\varphi} + i(\partial_\varphi\gamma_\varphi)\sqrt{P_\varphi})e^{i\gamma_\varphi}|^2 \\
&= (\partial_\varphi\sqrt{P_\varphi})^2 + (\partial_\varphi\gamma_\varphi)^2 P_\varphi.
\end{aligned}
\quad (27)
$$

If condition (14) is satisfied, Eq. (27) simplifies to

$$(\partial_\varphi\gamma_\varphi)^2 = |(\partial_\varphi u_\varphi)_{11}|^2 + O\left(\frac{1}{N}\right), \quad (28)$$

so that the two quantities are equal up to order $1/N$.

Now, if the adaptation is performed in the output, i.e., we choose an arbitrary V_{in} and adapt V_{out} according to Eq. (22), we see that

$$|(\partial_\varphi u_\varphi)_{11}|^2 = (V_{\text{in}}^\dagger G_\varphi V_{\text{in}})_{11}^2 + O\left(\frac{1}{N}\right), \quad (29)$$

where the Hermitian operator,

$$G_\varphi := iU_\varphi^\dagger\frac{\partial_\varphi U_\varphi}{\partial\varphi}, \quad (30)$$

is the (φ-dependent) generator of U_φ. If, on the other hand, condition (14) is realized through an adaptation on the input while taking an arbitrary V_{out}, then Eq. (23) implies that

$$|(\partial_\varphi u_\varphi)_{11}|^2 = (V_{\text{out}}U_\varphi G_\varphi U_\varphi^\dagger V_{\text{out}}^\dagger)_{11}^2 + O\left(\frac{1}{N}\right). \quad (31)$$

Using Eqs. (28)–(31), we can finally rewrite the asymptotic expression of the Fisher information (18) as

$$F(\varphi) \sim \varrho(k_\varphi, \ell_\varphi)f(U, G_\varphi)N^2, \quad (32)$$

as $N \to \infty$, where

$$f(U, G_\varphi) = (U^\dagger G_\varphi U)_{11}^2, \quad (33)$$

with $U = V_{\text{in}}$ if the optimization is performed on the output, while $U = U_\varphi^\dagger V_{\text{out}}^\dagger$ if the optimization is carried out on the input. We emphasize that the prefactor $f(U, G_\varphi)$ is completely independent of the choice of the optimized stage.

The maximization of the prefactor (33) can be realized, for example, if $U = V_\varphi$ is some unitary diagonalizing G_φ, i.e., satisfying equation $V_\varphi^\dagger G_\varphi V_\varphi = D_\varphi$ with $D_\varphi = \text{diag}(g_1, g_2, \ldots, g_M)$ being the diagonal matrix of the eigenvalues of G_φ, ordered in such a way that $|g_1| = \|G_\varphi\|$ is the maximum eigenvalue in absolute value [19]. Actually, it is not necessary to take a diagonalizing unitary to maximize (33), since only the first column of U enters in the definition of $f(U, G_\varphi)$; hence, to maximize $f(U, G_\varphi)$ it is sufficient to require this column to be the eigenvector of G_φ corresponding to the maximum eigenvalue $\|G_\varphi\|$. However, even that requirement would necessitate the complete knowledge of G_φ, which in general depends on the unknown parameter φ. Therefore, it is more relevant to consider arbitrary choices of the nonadapted network (the unitary U) independently of φ in order to determine the practical advantages of the obtained Heisenberg scaling precision for finite values of N and only one (classically) adapted stage.

For this reason, we will perform now a statistical analysis on the typical values which can be assumed by the prefactor $f(U, G_\varphi)$ for random choices of the unitary U. Assuming no prior knowledge of the unitary U, we sample it from the unitary group U(M) according to the unbiased uniform distribution probability, i.e., the unitarily invariant Haar measure $\mathcal{P}$.

For a random unitary U, sampled according to this distribution, the average value of the prefactor $f(U, G_\varphi)$ can be computed using techniques from random matrix theory (see Appendix D):

$$\mathbb{E}[f(U, G_\varphi)] = \frac{\text{Tr}(G_\varphi^2) + \text{Tr}(G_\varphi)^2}{M(M+1)}, \quad (34)$$

where $\mathbb{E}[\cdot]$ denotes the expectation value over U(M) with respect to the Haar measure.

In the trivial case of a generator proportional to the identity, $G_\varphi = \|G_\varphi\|\mathbb{1}$, which corresponds to the case of a network $U_\varphi = e^{i\|G_\varphi\|}\mathbb{1}$ acting as a φ-dependent global phase shifter, we have $\text{Tr}(G_\varphi)^2 = M^2\|G_\varphi\|^2$ and $\text{Tr}(G_\varphi^2) = M\|G_\varphi\|^2$, so that the average value of the prefactor equals the maximum one, $f_{\text{max}} = \|G_\varphi\|^2$, in accordance to the fact that in this particular case *every* unitary in U(M) diagonalizes G_φ.

In general, we are interested in determining the conditions which make this average value in (34) as large as possible. First of all, we can note directly from expression (34) that eigenvalues of opposite signs can have a detrimental effect on this average, since they lower the value of $\text{Tr}(G_\varphi)$. In general, we can find a lower bound on the average value (34) using Jensen's inequality $\mathbb{E}[X^2] \geqslant \mathbb{E}[X]^2$ to obtain

$$\mathbb{E}[f(U, G_\varphi)] \geqslant \mathbb{E}[(U^\dagger G_\varphi U)_{11}]^2 = \left[\frac{\text{Tr}(G_\varphi)}{M}\right]^2, \quad (35)$$

where again, the average $\mathbb{E}[(U^\dagger G_\varphi U)_{11}]$ has been computed using standard techniques (see Appendix D).

Notice that the right-hand side of this inequality is nothing but the square of the average between G_φ's eigenvalues. Hence, if we have some degree of control on the eigenvalues, we can achieve a result which is a certain fraction α of the maximum value $f_{\max}$ if the average of G_φ's eigenvalues is at least a fraction $\sqrt{\alpha}$ of the maximum eigenvalue, namely,

$$\left[\frac{\mathrm{Tr}(G_\varphi)}{M} \right]^2 \geqslant \alpha \|G_\varphi\|^2 \quad \Rightarrow \quad \mathbb{E}[f(U, G_\varphi)] \geqslant \alpha f_{\max}. \tag{36}$$

However, this may be not sufficient for our purposes, since the average of a random variable alone does not determine its *typical* behavior: A paradigmatic elementary example is that of a real random variable taking only the values 0 or 1 with equal probabilities, thus having an average of $1/2$ even if it *never* takes values close to $1/2$.

We will show now that this is not the case for the prefactor $f(U, G_\varphi)$, thanks to the fact that it is a sufficiently well-behaved function with respect to the random unitary U. In fact, by using results on concentration of measure in high-dimensional probability spaces, we prove in Appendix E that for a network with a large number M of ports the prefactor $f(U, G_\varphi)$ becomes *typical*, meaning that it becomes almost constant with respect to random choices of $U \in \mathrm{U}(M)$ (according to the unitarily invariant measure), hence concentrating around its average value (34), bounded below by (36).

In formulas, we have

$$\mathcal{P}(|f - \mathbb{E}[f]| \geqslant \varepsilon) \leqslant 2 \exp\left(-\frac{AM}{\|G_\varphi\|^4} \varepsilon^2 \right), \tag{37}$$

where $A = (72\pi^3)^{-1}$. This result tells us that for large interferometers it is extremely unlikely to obtain a prefactor sensibly different from its average, since for large values of M the probability of $f(U, G_\varphi)$ being different from its average is exponentially suppressed.

This result can be also seen from the exact distribution of the prefactor computed for some particular cases, which is shown in Fig. 3. It can be seen from these figures that as M is increased the distribution of the prefactor concentrates around its average: in particular, for the chosen configuration, a value of $M = 20$ is already sufficient to get this concentration. Thus, for any well-behaved linear network $\hat{U}_\varphi$ such that the expectation value in (34) is far enough from zero, any random choice of the nonadapted stage in the proposed interferometric setup typically yields an Heisenberg-scaling precision for the estimation of φ if the number M of interferometric channels is large enough.

V. CONCLUSIONS

We demonstrated by using a simple metrological technique the typicality of Heisenberg scaling precision for the estimation of a generic parameter φ encoded into an arbitrary M-mode network. Our scheme can be applied regardless of the nature of the parameter, which can even be distributed among several components of the network. In particular, the
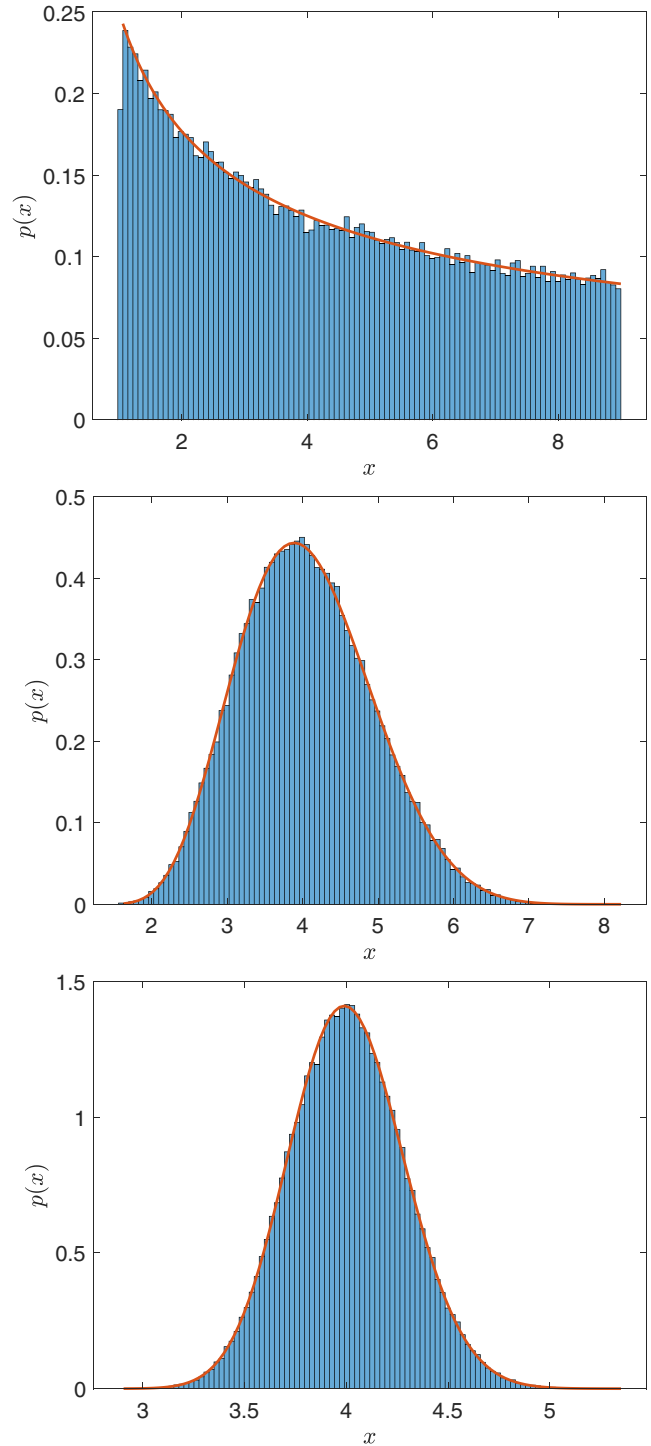

FIG. 3. Histograms of the random variable $f(U, G_\varphi) = (U^\dagger G_\varphi U)_{11}^2$ numerically obtained with 10^5 samplings of U from the unitary group with Haar measure and choosing G_φ as a diagonal matrix with half entries 3s and half entries made of 1s, with $M = 2$ (top), $M = 20$ (middle), and $M = 200$ (bottom). The normalization is chosen in such a way that the total area under the curve equals 1. For these particular cases an explicit analytic expression of the distribution is achievable, and is given by the orange curves. The derivation is shown in Appendix C.

proposed scheme makes use of a single-mode squeezed state as a probe, scattered throughout all the modes by means of an auxiliary passive linear stage. Once the information on the parameter is gathered by the probe, this gets refocused on a single output channel by a second auxiliary stage, and then detected with homodyne measurement. The analysis of the Fisher information associated with such scheme reveals that, if a constant average number of photons (not scaling with the total number of photons injected) is scattered into channels different from the one measured, due to an imperfect refocusing procedure, the Heisenberg limit can be asymptotically reached, provided that the homodyne detection is performed with a sufficient resolution. For a distributed parameter, the refocusing is generally parameter dependent, implying some sort of adaptive procedure in order to correctly refocus the probe. However, we have shown that all the dependence on the parameter can be entirely put in only one of the two auxiliary stages, while the other only affects the estimation through a multiplicative pre-factor. Moreover, we have also discussed how all the information on the parameter needed to sufficiently refocus the probe can be obtained with a classical shot-noise precision, meaning that the number of resources required to adaptively optimize the auxiliary stages is not detrimental for the Heisenberg scaling precision. Finally, we have shown that, for a large number of modes, Heisenberg scaling is typically obtained by an arbitrary nonadapted stage, with an overwhelming probability, i.e., an exponentially suppressed probability of failure.

APPENDIX A: DERIVATION OF THE PROBABILITY DISTRIBUTION IN (5)

In order to evaluate the probability distribution associated with the measure of the quadrature field x_θ, it is natural to proceed with the phase-space formalism for Gaussian quantum optics. For a deep and systematic overview in this topic, several works and reviews can be found in literature [15,40,41]. In this Appendix, we will briefly introduce only the concepts and tools needed to obtain the expression (5) for the probability distribution $p(x|\varphi)$ of measuring the quadrature value x at the output of the proposed interferometer in Fig. 1. We recall first the definition (4),

$$p(x|\varphi) = \langle \text{vac}|\hat{S}_1^\dagger(r)\hat{u}_\varphi^\dagger e^{i\theta\hat{a}_1^\dagger\hat{a}_1}|x\rangle_{11}\langle x|e^{-i\theta\hat{a}_1^\dagger\hat{a}_1}\hat{u}_\varphi\hat{S}_1(r)|\text{vac}\rangle, \tag{A1}$$

where $\hat{u}_\varphi$ describes the overall interferometric evolution of the single mode squeezed state $\hat{S}_1(r)|\text{vac}\rangle$. To evaluate (A1), it is

useful to first recover its Fourier transform,

$$\chi(\xi|\varphi) = \int dx \, p(x|\varphi)e^{-ix\xi}$$

$$= \langle \text{vac}|\hat{S}_1^\dagger(r)\hat{u}_\varphi^\dagger e^{i\theta\hat{a}_1^\dagger\hat{a}_1}e^{-i\hat{x}_1\xi}e^{-i\theta\hat{a}_1^\dagger\hat{a}_1}\hat{u}_\varphi\hat{S}_1(r)|\text{vac}\rangle. \tag{A2}$$

It is possible to write this characteristic function in a more canonical way. Indeed, first notice that we can write (the derivation is given in Appendix A 1 for completeness)

$$e^{i\theta\hat{a}_1^\dagger\hat{a}_1}e^{-i\hat{x}_1\xi}e^{-i\theta\hat{a}_1^\dagger\hat{a}_1} = \hat{D}(\vec{\xi}_\theta), \tag{A3}$$

where

$$\hat{D}(\vec{\xi}_\theta) = e^{-i\vec{\xi}_\theta\cdot\vec{z}} \tag{A4}$$

is the *displacement operator*, with

$$\vec{\xi}_\theta = \begin{pmatrix} \xi\cos\theta \\ \xi\sin\theta \end{pmatrix}, \quad \vec{z} = \begin{pmatrix} \hat{x}_1 \\ \hat{p}_1 \end{pmatrix}. \tag{A5}$$

Then, using Eq. (A3) we can write the characteristic function (A2) as

$$\chi(\xi|\varphi) = \text{Tr}[\hat{D}(\vec{\xi}_\theta)\hat{u}_\varphi\hat{S}_1(r)|\text{vac}\rangle\langle\text{vac}|S_1^\dagger(r)\hat{u}_\varphi^\dagger]. \tag{A6}$$

Due to the Gaussian nature of the squeezed vacuum state and the linearity of the interferometric setup, the characteristic function (A6) is a Gaussian bivariate function centered in zero, of the form [15],

$$\chi(\xi|\varphi) = e^{-\frac{1}{2}\vec{\xi}_\theta^T\sigma_\varphi\vec{\xi}_\theta}, \tag{A7}$$

where σ_φ is the 2×2 covariance matrix of the whole interferometer output state $\hat{u}_\varphi\hat{S}_1(r)|\text{vac}\rangle$, reduced to the first mode.

In order to evaluate this matrix, we first recover the covariance matrix Γ_0 of the input state $\hat{S}_1(r)|\text{vac}\rangle$, which reads

$$\Gamma_0 = \frac{1}{2}\begin{pmatrix} e^{2\mathcal{R}} & 0 \\ 0 & e^{-2\mathcal{R}} \end{pmatrix}, \tag{A8}$$

where $\mathcal{R}$ is the $M \times M$ diagonal matrix with a single nonzero entry $\mathcal{R}_{11} \equiv r$. After the action of the interferometer, the covariance matrix transforms into

$$\Gamma_\varphi = R_\varphi\Gamma_0 R_\varphi^T, \tag{A9}$$

where R_φ is the orthogonal and symplectic matrix associated with the interferometer unitary matrix u_φ,

$$R_\varphi = W^\dagger\begin{pmatrix} u_\varphi & 0 \\ 0 & u_\varphi^* \end{pmatrix}W, \tag{A10}$$

and

$$W = \frac{1}{\sqrt{2}}\begin{pmatrix} \mathbb{1} & i\mathbb{1} \\ \mathbb{1} & -i\mathbb{1} \end{pmatrix}, \tag{A11}$$

where $\mathbb{1}$ is the $M \times M$ identity matrix. R_φ can be easily evaluated to be

$$R_\varphi = \begin{pmatrix} \text{Re}[u_\varphi] & -\text{Im}u_\varphi \\ \text{Im}u_\varphi & \text{Re}[u_\varphi] \end{pmatrix}, \tag{A12}$$

so that Γ_φ in (A9) reads

$$\Gamma_\varphi = \begin{pmatrix} \Delta X_\varphi^2 & \Delta XP_\varphi \\ (\Delta XP_\varphi)^T & \Delta P_\varphi^2 \end{pmatrix}, \tag{A13}$$

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

$$\Delta X_\varphi^2 \equiv \frac{1}{2}[\text{Re}[u_\varphi]e^{2\mathcal{R}}\text{Re}[u_\varphi^\dagger] - \text{Im}u_\varphi e^{-2\mathcal{R}}\text{Im}u_\varphi^\dagger]$$
$$= \frac{1}{2}[\text{Re}[u_\varphi \cosh(2\mathcal{R})u_\varphi^\dagger] + \text{Re}[u_\varphi \sinh(2\mathcal{R})u_\varphi^T]], \quad \text{(A14)}$$

$$\Delta P_\varphi^2 \equiv \frac{1}{2}[-\text{Im}u_\varphi e^{2\mathcal{R}}\text{Im}u_\varphi^\dagger + \text{Re}[u_\varphi]e^{-2\mathcal{R}}\text{Re}[u_\varphi^\dagger]]$$
$$= \frac{1}{2}[\text{Re}[u_\varphi \cosh(2\mathcal{R})u_\varphi^\dagger] - \text{Re}[u_\varphi \sinh(2\mathcal{R})u_\varphi^T]], \quad \text{(A15)}$$

$$\Delta XP_\varphi \equiv \frac{1}{2}[-\text{Re}[u_\varphi]e^{2\mathcal{R}}\text{Im}u_\varphi^\dagger - \text{Im}u_\varphi e^{-2\mathcal{R}}\text{Re}[u_\varphi^\dagger]]$$
$$= \frac{1}{2}[-\text{Im}[u_\varphi \cosh(2\mathcal{R})u_\varphi^\dagger] + \text{Im}[u_\varphi \sinh(2\mathcal{R})u_\varphi^T]]. \quad \text{(A16)}$$

In the second lines of each of the previous expression, we have exploited the fact that $\mathcal{R}$ is real. We are interested to evaluate σ_φ, the covariance matrix reduced to the first mode, which we can now readily write

$$\sigma_\varphi = \begin{pmatrix} (\Delta X_\varphi^2)_{11} & (\Delta XP_\varphi)_{11} \\ (\Delta XP_\varphi)_{11} & (\Delta P_\varphi^2)_{11} \end{pmatrix}, \quad \text{(A17)}$$

and insert in (A7). Our final step is to invert the Fourier transform to finally get the expression of the probability distribution $p(x|\varphi)$ given in (5). In order to do that, we introduce the 2×2 orthogonal matrix,

$$O_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad \text{(A18)}$$

such that $\vec{\xi}_\theta = O_\theta \vec{\xi}_0$, with $\vec{\xi}_0 = (\xi, 0)^T$. Then, the characteristic function (A7) can be written in a more convenient way, namely

$$\chi(\xi|\varphi) = e^{-\frac{1}{2}\vec{\xi}_0^T O_\theta^T \sigma_\varphi O_\theta \vec{\xi}_0} = e^{-\frac{1}{2}(O_\theta^T \sigma_\varphi O_\theta)_{11}\xi^2}. \quad \text{(A19)}$$

Exploiting, by the definition of $\mathcal{R}$, the identities,

$$(u_\varphi \cosh(2\mathcal{R})u_\varphi^\dagger)_{11} = \cosh(2r)|(u_\varphi)_{11}|^2 + \sum_{i=2}^{M}|(u_\varphi)_{1i}|^2$$
$$= \cosh(2r)|(u_\varphi)_{11}|^2 + (1 - |(u_\varphi)_{11}|^2), \quad \text{(A20)}$$

$$(u_\varphi \sinh(2\mathcal{R})u_\varphi^T)_{11} = \sinh(2r)(u_\varphi)_{11}^2, \quad \text{(A21)}$$

and some elementary trigonometry, the term $(O_\theta^T \sigma_\varphi O_\theta)_{11}$ can be further manipulated to match the expression of Δ_φ given in (6). In fact,

$$(O_\theta^T \sigma_\varphi O_\theta)_{11}$$
$$= \sum_{i,j=1,2}(O_\theta)_{i1}(O_\theta)_{j1}(\sigma_\varphi)_{ij}$$
$$= \cos^2\theta(\Delta X_\varphi^2)_{11} + \sin^2\theta(\Delta P_\varphi^2)_{11}$$
$$+ 2\cos\theta\sin\theta(\Delta XP_\varphi)_{11}$$
$$= \frac{1}{2}\cos^2\theta(\cosh(2r)|(u_\varphi)_{11}|^2 + \sinh(2r)\text{Re}((u_\varphi)_{11}^2))$$
$$+ \frac{1}{2}\sin^2\theta(\cosh(2r)|(u_\varphi)_{11}|^2 - \sinh(2r)\text{Re}((u_\varphi)_{11}^2))$$

$$+ \frac{1}{2}(\cos^2\theta + \sin^2\theta)(1 - |(u_\varphi)_{11}|^2)$$
$$+ \cos\theta\sin\theta\sinh(2r)\text{Im}[(u_\varphi)_{11}^2]$$
$$= \frac{1}{2}(1 + |(u_\varphi)_{11}|^2(\cosh(2r) - 1)$$
$$+ \sinh(2r)(\cos(2\theta)\text{Re}((u_\varphi)_{11}^2) + \sin(2\theta)\text{Im}[(u_\varphi)_{11}^2]))$$
$$= \frac{1}{2}(1 + |(u_\varphi)_{11}|^2(\cosh(2r) - 1)$$
$$+ \text{Re}[e^{-2i\theta}(u_\varphi)_{11}^2]\sinh(2r)) \equiv \Delta_\varphi. \quad \text{(A22)}$$

After applying the inverse Fourier transformation on the Gaussian characteristic function (A7), the probability distribution reads

$$p(x|\varphi) = \frac{1}{2\pi}\int d\xi \, \chi(\xi|\varphi)e^{ix\xi} = \frac{1}{\sqrt{2\pi\Delta_\varphi}}\exp\left(-\frac{x^2}{2\Delta_\varphi}\right), \quad \text{(A23)}$$

as displayed in (5).

1. Derivation of (A3)

Exploiting the unitarity of $e^{i\theta\hat{a}_1^\dagger\hat{a}_1}$, we can write

$$e^{i\theta\hat{a}_1^\dagger\hat{a}_1}e^{-i\xi\hat{x}_1}e^{-i\theta\hat{a}_1^\dagger\hat{a}_1} = \exp(-i\xi e^{i\theta\hat{a}_1^\dagger\hat{a}_1}\hat{x}_1 e^{-i\theta\hat{a}_1^\dagger\hat{a}_1}). \quad \text{(A24)}$$

By using the definition $\hat{x}_1 = (\hat{a}_1 + \hat{a}_1^\dagger)/\sqrt{2}$, the first-mode quadrature operator along θ reads

$$e^{i\theta\hat{a}_1^\dagger\hat{a}_1}\hat{x}_1 e^{-i\theta\hat{a}_1^\dagger\hat{a}_1} = \frac{1}{\sqrt{2}}(\hat{a}_1(\theta) + \hat{a}_1(\theta)^\dagger), \quad \text{(A25)}$$

where $\hat{a}_1(\theta) = e^{i\theta\hat{a}_1^\dagger\hat{a}_1}\hat{a}_1 e^{-i\theta\hat{a}_1^\dagger\hat{a}_1}$ is the first-mode annihilation operator at time θ in the Heisenberg picture. The Heisenberg equation is obtained by taking the derivative of $\hat{a}_1(\theta)$ with respect to θ, and reads

$$\frac{d\hat{a}_1(\theta)}{d\theta} = i[\hat{a}_1(\theta)^\dagger\hat{a}_1(\theta), \hat{a}_1(\theta)] = -i\hat{a}_1(\theta), \quad \text{(A26)}$$

since $[\hat{a}_1(\theta), \hat{a}_1(\theta)^\dagger] = [\hat{a}_1, \hat{a}_1^\dagger] = 1$ and $[\hat{a}_1(\theta), \hat{a}_1(\theta)] = [\hat{a}_1, \hat{a}_1] = 0$. Therefore,

$$\hat{a}_1(\theta) = e^{-i\theta}\hat{a}_1, \qquad \hat{a}_1(\theta)^\dagger = e^{-i\theta}\hat{a}_1^\dagger, \quad \text{(A27)}$$

the second equality being obtained by taking the adjoint of the first. By plugging (A27) into (A25) one gets

$$xie^{i\theta\hat{a}_1^\dagger\hat{a}_1}\hat{x}_1 e^{-i\theta\hat{a}_1^\dagger\hat{a}_1} = \xi\cos\theta\,\hat{x}_1 + \xi\,\sin\theta\,\hat{p}_1 = \vec{\xi}_\theta \cdot \vec{z}, \quad \text{(A28)}$$

where the vectors $\vec{\xi}_\theta$ and $\vec{z}$ are given in (A5). Inserting this expression into (A24), we finally obtain (A3).

APPENDIX B: DERIVATION OF THE FISHER INFORMATION IN (10)

In this Appendix we will evaluate the FI in (10) from the expression in (9). Let us recall that the variance Δ_φ of the Gaussian probability density function (5) reads

$$\Delta_\varphi = \frac{1}{2}(1 + |(u_\varphi)_{11}|^2(\cosh 2r - 1) + \text{Re}[e^{-2i\theta}(u_\varphi)_{11}^2]\sinh 2r). \quad \text{(B1)}$$

The derivative of Δ_φ is written as a sum of two contributions,

$$\partial_\varphi \Delta_\varphi = \tfrac{1}{2}(\cosh 2r - 1)\partial_\varphi |(u_\varphi)_{11}|^2$$
$$+ \tfrac{1}{2} \sinh 2r \operatorname{Re}\big[e^{-2i\theta} \partial\varphi(u_\varphi)_{11}^2\big]. \tag{B2}$$

The derivative in the first contribution is thus evaluated:

$$\partial_\varphi |(u_\varphi)_{11}|^2 = (u_\varphi)_{11}^* \partial_\varphi(u_\varphi)_{11} + (u_\varphi)_{11} \partial_\varphi(u_\varphi)_{11}^*$$
$$= 2\operatorname{Re}[(u_\varphi)_{11}^* \partial_\varphi(u_\varphi)_{11}], \tag{B3}$$

while the derivative in the second contribution reads

$$\operatorname{Re}\big[e^{-2i\theta} \partial_\varphi(u_\varphi)_{11}^2\big] = 2\operatorname{Re}[e^{-2i\theta}(u_\varphi)_{11} \partial_\varphi(u_\varphi)_{11}]. \tag{B4}$$

Then, defining γ_φ as the phase of $(u_\varphi)_{11}$, and recalling that $\sinh^2 r = N$, (B2) reads

$$\partial_\varphi \Delta_\varphi$$
$$= \operatorname{Re}[(u_\varphi)_{11}^* \partial_\varphi(u_\varphi)_{11}((\cosh 2r - 1) + e^{2i(\gamma_\varphi - \theta)} \sinh 2r)]$$
$$= 2\operatorname{Re}[(u_\varphi)_{11}^* \partial_\varphi(u_\varphi)_{11}(N + e^{2i(\gamma_\varphi - \theta)}\sqrt{N^2 + N})], \tag{B5}$$

while (B1) can be written as

$$\Delta_\varphi = \tfrac{1}{2}(1 + 2|(u_\varphi)_{11}|^2)$$
$$\times (N + \cos(2\gamma_\varphi - 2\theta)\sqrt{N^2 + N})). \tag{B6}$$

Inserting the expressions (B5) and (B6) into (9), we get the FI,

$$F(\varphi) = 8\left(\frac{\operatorname{Re}[(u_\varphi)_{11}^* \partial_\varphi(u_\varphi)_{11}(N + e^{2i(\gamma_\varphi - \theta)}\sqrt{N^2 + N})]}{1 + 2|(u_\varphi)_{11}|^2(N + \cos(2\gamma_\varphi - 2\theta)\sqrt{N^2 + N})}\right)^2. \tag{B7}$$

Moreover we can write $(u_\varphi)_{11} = e^{i\gamma_\varphi}\sqrt{P_\varphi}$, so that

$$(u_\varphi)_{11}^* \partial_\varphi(u_\varphi)_{11} = \frac{\partial_\varphi P_\varphi}{2} + iP_\varphi \partial_\varphi \gamma_\varphi. \tag{B8}$$

Since $P_\varphi, \gamma_\varphi$, and their derivatives are real, once we define the quantities,

$$f(N) := N\big(1 + \cos[2(\gamma_\varphi - \theta)]\sqrt{1 + 1/N}\big), \tag{B9}$$

$$h(N) := N \sin[2(\gamma_\varphi - \theta)]\sqrt{1 + 1/N}, \tag{B10}$$

we easily obtain

$$F(\varphi) = 2\left(\frac{(\partial_\varphi P_\varphi)f(N) - 2P_\varphi(\partial_\varphi \gamma_\varphi)h(N)}{1 + 2P_\varphi f(N)}\right)^2, \tag{B11}$$

as displayed in (10).

APPENDIX C: ANALYTIC DISTRIBUTION OF THE PREFACTOR (33) IN THE FISHER INFORMATION (32) FOR GENERATORS WITH ONLY TWO DISTINCT EIGENVALUES

We will derive here the explicit form of the probability density function for the prefactor $f(U, G_\varphi) = (U^\dagger G_\varphi U)_{11}^2$ in (33) for a fixed generator G_φ as U is sampled from $U(M)$ with the Haar measure. First of all, note that this distribution depends only on the eigenvalues of G_φ, which we denote with $g = (g_1, \ldots, g_M)$, dropping the φ subscript for notation

simplicity. This can be seen using the spectral decomposition of $G_\varphi = V_\varphi^\dagger D_\varphi V_\varphi$, where $D_\varphi = \operatorname{diag}(g)$, yielding

$$U^\dagger G_\varphi U = (V_\varphi U)^\dagger D_\varphi(V_\varphi U) \stackrel{d}{=} U^\dagger D_\varphi U, \tag{C1}$$

where in the last step we used the invariance property of the Haar measure and we used the notation $\stackrel{d}{=}$ to say that the two random variables have the same distribution. In light of this remark, we have

$$f(U, G_\varphi) \stackrel{d}{=} (U, D_\varphi)$$
$$= (U^\dagger D_\varphi U)_{11}^2$$
$$= \left(\sum_{j=1}^M |u_j|^2 g_j\right)^2, \tag{C2}$$

having defined the random vector $u = Ue_1$ obtained by the application of the random matrix $U \in U(M)$ to the fixed basis vector $e_1 = (1, \ldots, 0)^T \in \mathbb{C}^M$, where $\mathbb{C}^M$ denotes the set of M-tuples of complex numbers. We see that $f(U, D_\varphi)$ can be interpreted as a weighted average of the eigenvalues of G_φ with random weights; these weights are given by the square modulus of the components of a random vector drawn from the unit sphere in $\mathbb{C}^M$ with the Haar measure. The distribution of this random variable can be quite complicated for a generic choice of the G_φ's eigenvalues $g = (g_1, \ldots, g_M)$. We will consider here the situation in which there are at most two distinct eigenvalues $g_1 \geqslant g_2 \geqslant 0$, i.e.,

$$g = (\underbrace{g_1, \ldots, g_1}_{k}, \underbrace{g_2, \ldots, g_2}_{M-k}), \tag{C3}$$

so that

$$f(U, D_\varphi) = \left(g_1 \sum_{j=1}^k |u_j|^2 + g_2 \sum_{j=k+1}^M |u_j|^2\right)^2$$
$$= \left[(g_1 - g_2)\sum_{j=1}^k |u_j|^2 + g_2\right]^2, \tag{C4}$$

where we used the normalization constraint,

$$\sum_{j=1}^M |u_j|^2 = 1. \tag{C5}$$

In order to get the distribution of (C4), let us first consider the random quantity $\tau(U)$ defined by the sum inside the brackets, namely,

$$\tau(U) = \sum_{j=1}^k |u_j|^2. \tag{C6}$$

We start from the distribution $q(t)$ of $\tau(U)$, defined in such a way that $q(t)dt$ is the probability to have

$$t \leqslant \tau(U) = \sum_{j=1}^k |u_j|^2 \leqslant t + dt, \tag{C7}$$

or, defining $x_{2j-1} := \mathrm{Re}\, u_j$ and $x_{2j} := \mathrm{Im}\, u_j$, the probability to have

$$t \leqslant \sum_{j=1}^{k} \left(x_{2j-1}^2 + x_{2j}^2 \right) \leqslant t + dt. \tag{C8}$$

This probability can be interpreted as the geometrical surface of a $2k$-dimensional hyperspherical cap of a $(2M-1)$-dimensional hypersphere sitting in $\mathbb{R}^{2M}$. Using this interpretation, one then finds that [42,43]

$$q(t) = \frac{(M-1)!}{(k-1)!(M-k-1)!} t^{k-1}(1-t)^{M-k-1} \chi_{[0,1]}(t), \tag{C9}$$

where

$$\chi_{[0,1]}(t) = \begin{cases} 1 & 0 \leqslant t \leqslant 1, \\ 0 & \text{otherwise.} \end{cases} \tag{C10}$$

Starting from the distribution (C9) of $\tau(U)$, the probability density function $p(x)$ of

$$f(U, \mathrm{diag}(g)) = [(g_1 - g_2)\tau(U) + g_2]^2, \tag{C11}$$

can be found with a change of variables to be

$$p(x) = \frac{1}{2\sqrt{x}\,\Delta g}\, q\!\left(\frac{\sqrt{x} - g_2}{\Delta g} \right), \tag{C12}$$

where $\Delta g := g_1 - g_2$. We then have explicitly,

$$p(x) = C \frac{(g_1 - \sqrt{x})^{M-k-1}(\sqrt{x} - g_2)^{k-1}}{\sqrt{x}} \chi_{[g_2,g_1]}(\sqrt{x}), \tag{C13}$$

where C is a normalization constant given by

$$C = \frac{1}{2(\Delta g)^{M-1}} \frac{(M-1)!}{(k-1)!(M-k-1)!}. \tag{C14}$$

This distribution is valid whenever G_φ has only two distinct positive eigenvalues $g_1 \geqslant g_2 \geqslant 0$. Numerical results are compared with the probability density function (C13) in Fig. 3.

APPENDIX D: DERIVATION OF THE AVERAGE VALUE (34) OF THE PREFACTOR IN (33) FOR RANDOM UNITARY MATRICES U

We collect here some results needed for the computation of averages over the unitary group which have been used in the main text. Denoting with $\mathcal{P}$ the Haar probability measure, the average of a function $f : \mathrm{U}(M) \to \mathbb{C}$ is defined as

$$\mathbb{E}[f(U)] = \int f(U)\, d\mathcal{P}(U), \tag{D1}$$

whenever this integral is defined. In order to derive the results of this work we are interested only in the moments of the matrix elements U_{ij}, i.e., the averages of some powers of the matrix elements and their complex conjugates, which we give here in the following lemma.

Lemma 1 ([44], Proposition 4.2.3). Denoting with U_{ij} the matrix elements of a Haar distributed random unitary matrix $U \in \mathrm{U}(M)$, the only nonvanishing moments up to the fourth order are given by

$$\mathbb{E}[|U_{ij}|^2] = \frac{1}{M} \quad (1 \leqslant i, j \leqslant M), \tag{D2}$$

$$\mathbb{E}[|U_{ij}|^4] = \frac{2}{M(M+1)} \quad (1 \leqslant i, j \leqslant M), \tag{D3}$$

$$\mathbb{E}[|U_{ij}|^2 |U_{kj}|^2] = \frac{1}{M(M+1)} \quad (i \neq k), \tag{D4}$$

$$\mathbb{E}[|U_{ij}|^2 |U_{il}|^2] = \frac{1}{M(M+1)} \quad (j \neq l), \tag{D5}$$

$$\mathbb{E}[|U_{ij}|^2 |U_{kl}|^2] = \frac{1}{M^2 - 1} \quad (i \neq k,\ j \neq l), \tag{D6}$$

$$\mathbb{E}[U_{ij} U_{kl} U_{il}^* U_{kj}^*] = -\frac{1}{M(M^2 - 1)} \quad (i \neq k,\ j \neq l). \tag{D7}$$

The results gathered in the previous lemma can be expressed with two compact formulas [45] as

$$\mathbb{E}[U_{ij} U_{kl}^*] = \frac{\delta_{ik}\delta_{jl}}{M}, \tag{D8}$$

$$\mathbb{E}[U_{ij} U_{kl} U_{mn}^* U_{pq}^*] = \frac{\delta_{im}\delta_{jn}\delta_{kp}\delta_{lq} + \delta_{ip}\delta_{jq}\delta_{km}\delta_{ln}}{M^2 - 1}$$
$$- \frac{\delta_{im}\delta_{jq}\delta_{kp}\delta_{ln} + \delta_{ip}\delta_{jn}\delta_{km}\delta_{lq}}{M(M^2 - 1)}. \tag{D9}$$

These formulas allow us to prove the following lemma, which has been used to derive the averages (34) and (35) of the main text.

Lemma 2. Given a generic $M \times M$ complex matrix A, the following results hold:

$$\mathbb{E}[(U^\dagger A U)_{ij}] = \frac{\mathrm{Tr}(A)}{M} \delta_{ij}, \tag{D10}$$

$$\mathbb{E}\big[(U^\dagger A U)_{ij}^2\big] = \frac{\mathrm{Tr}(A)^2 + \mathrm{Tr}(A^2)}{M(M+1)} \delta_{ij}. \tag{D11}$$

Proof. Equation (D10) can be proved using (D8):

$$\mathbb{E}[(U^\dagger A U)_{ij}] = \sum_{k,l} \mathbb{E}[U_{ik}^\dagger A_{kl} U_{lj}]$$
$$= \sum_{k,l} A_{kl} \mathbb{E}[U_{ki}^* U_{lj}]$$
$$= \sum_{k,l} A_{kl} \frac{\delta_{kl}\delta_{ij}}{M} = \frac{\mathrm{Tr}(A)}{M} \delta_{ij}, \tag{D12}$$

while Eq. (D11) can be proved using (D9):

$$\mathbb{E}\big[(U^\dagger A U)_{ij}^2\big]$$
$$= \sum_{k,l,m,n} \mathbb{E}[U_{ik}^\dagger A_{kl} U_{lj} U_{im}^\dagger A_{mn} U_{nj}]$$
$$= \sum_{k,l,m,n} A_{kl} A_{mn} \mathbb{E}[U_{lj} U_{nj} U_{mi}^* U_{ki}^*]$$
$$= [\mathrm{Tr}(A^2) + \mathrm{Tr}(A)^2]\left(\frac{1}{M^2 - 1} - \frac{1}{M(M^2 - 1)} \right)\delta_{ij}$$
$$= \frac{\mathrm{Tr}(A^2) + \mathrm{Tr}(A)^2}{M(M+1)} \delta_{ij}. \tag{D13}$$

∎

For $A = G_\varphi$ and $i = j = 1$, the expressions (D10) and (D11) reduce to the equalities in (35) and (34), respectively.

APPENDIX E: DERIVATION OF THE TYPICALITY RESULT IN (37)

In this Appendix we will show how to derive Eq. (37) starting from a standard result on concentration of measure in high-dimensional spaces known as Levy's Lemma, which we report in the following theorem for the sake of completeness.

Theorem 1. Let $f : \mathbb{S}^{n-1} \to \mathbb{R}$ be a function defined over the unit euclidean sphere,

$$\mathbb{S}^{n-1} = \left\{ x \in \mathbb{R}^n \,\middle|\, \sum_{k=1}^{n} x_k^2 = 1 \right\}, \qquad \text{(E1)}$$

endowed with the invariant Haar probability measure $\mathcal{P}$. Denote with L the Lipschitz constant of the function, i.e., the minimum L such that

$$|f(x) - f(y)| \leqslant L\|x - y\|_2, \qquad \text{(E2)}$$

for all $x, y \in \mathbb{S}^{n-1}$, where $\|x\|_2 = \sqrt{\sum_{k=1}^{n} x_k^2}$ is the Euclidean norm. Then

$$\mathcal{P}(|f - \mathbb{E}[f]| \geqslant \varepsilon) \leqslant 2e^{-\frac{n\varepsilon^2}{CL^2}}, \qquad \text{(E3)}$$

where C is some positive constant which can be taken to be $C = 9\pi^3$ [46,47].

In order to apply Theorem 1 to our case, we need to compute the Lipschitz constant associated with the prefactor (33). First, note that $f(U, G_\varphi)$ can be interpreted as a function defined on a real unit sphere. In fact, it can be written as

$$f(U, G_\varphi) = \left(\sum_{j=1}^{M} |u_j|^2 g_j \right)^2, \qquad \text{(E4)}$$

where u is a *complex* vector on the unit sphere, given by $u = Ue$ with $e = (1, 0, \dots, 0)^T \in \mathbb{C}^M$. Since only the squared moduli $|u_j|^2$ appear in this expression, we can recast the problem in terms of a real vector $x \in \mathbb{R}^{2M}$ whose components are defined by

$$x_{2j-1} = \text{Re}\,u_j, \quad x_{2j} = \text{Im}\,u_j, \quad j = 1, \dots, M. \qquad \text{(E5)}$$

The normalization constraint $\sum_{j=1}^{M} |u_j|^2 = 1$ becomes

$$\sum_{j=1}^{2M} x_j^2 = 1, \qquad \text{(E6)}$$

so that $x \in \mathbb{S}^{2M-1}$, the unit sphere sitting inside $\mathbb{R}^{2M}$. We see then that the random factor in Eq. (E4) can be envisioned as a function defined over the unit sphere $\mathbb{S}^{2M-1}$:

$$\left(\sum_{j=1}^{M} |u_j|^2 g_j \right)^2 = \left(\sum_{j=1}^{M} (x_{2j-1}^2 + x_{2j}^2) g_j \right)^2$$

$$= (x^T \widetilde{G} x)^2 =: f(x), \qquad \text{(E7)}$$

where we have defined the diagonal matrix $\widetilde{G} = \text{diag}(\widetilde{g})$ with $\widetilde{g} = (g_1, g_1, \dots, g_M, g_M) \in \mathbb{R}^{2M}$. In order to apply Theorem 1 we need to estimate the Lipschitz constant L of the function f; to this aim, we evaluate the gradient of f, which is given by

$$\nabla f(x) = 4(x^T \widetilde{G} x)\widetilde{G} x. \qquad \text{(E8)}$$

The Lipschitz constant for f can be then obtained as

$$L = \max_{x \in \mathbb{S}^{2M-1}} \|\nabla f(x)\|_2 = 4\|G_\varphi\|^2. \qquad \text{(E9)}$$

To see this, note simply that

$$\|\nabla f(x)\|_2 = \sqrt{[\nabla f(x)]^T [\nabla f(x)]}$$

$$= 4|x^T \widetilde{G} x|\sqrt{x^T \widetilde{G}^2 x}$$

$$\leqslant 4\|\widetilde{G}\|^2$$

$$= 4\|G_\varphi\|^2, \qquad \text{(E10)}$$

where in the inequality we used the fact that $|x^T \widetilde{G} x| \leqslant \|\widetilde{G}\|$ and $x^T \widetilde{G}^2 x = \|\widetilde{G} x\|_2^2 \leqslant \|\widetilde{G}\|^2$, while in the last equality we used the fact that $\|\widetilde{G}\| = \|G_\varphi\|$. The value $\|\nabla f(x)\|_2 = 4\|G_\varphi\|^2$ can be obtained with $x = (1, 0, \dots, 0)^T$, which together with (E10) proves (E9). Applying Theorem 1 to our case with $n = 2M$ and $L = 4\|G_\varphi\|^2$ finally yields Eq. (37).

[1] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum-enhanced measurements: Beating the standard quantum limit, Science **306**, 1330 (2004).

[2] V. Giovannetti, S. Lloyd, and L. Maccone, Quantum Metrology, Phys. Rev. Lett. **96**, 010401 (2006).

[3] J. P. Dowling, Quantum optical metrology - the lowdown on high-N00N states, Contemp. Phys. **49**, 125 (2008).

[4] V. Giovannetti, S. Lloyd, and L. Maccone, Advances in quantum metrology, Nat. Photonics **5**, 222 (2011).

[5] J. P. Dowling and K. P. Seshadreesan, Quantum optical technologies for metrology, sensing, and imaging, J. Lightwave Technol. **33**, 2359 (2015).

[6] R. S. Bondurant and J. H. Shapiro, Squeezed states in phase-sensing interferometers, Phys. Rev. D **30**, 2548 (1984).

[7] D. J. Wineland, J. J. Bollinger, W. M. Itano, F. L. Moore, and D. J. Heinzen, Spin squeezing and reduced quantum noise in spectroscopy, Phys. Rev. A **46**, R6797(R) (1992).

[8] L. Maccone and A. Riccardi, Squeezing metrology, Quantum **4**, 292 (2020).

[9] C. W. Helstrom, Quantum detection and estimation theory, J. Stat. Phys. **1**, 231 (1969).

[10] S. L. Braunstein and C. M. Caves, Statistical Distance and the Geometry of Quantum States, Phys. Rev. Lett. **72**, 3439 (1994).

[11] K. P. Seshadreesan, P. M. Anisimov, H. Lee, and J. P. Dowling, Parity detection achieves the Heisenberg limit in interferometry with coherent mixed with squeezed vacuum light, New J. Phys. **13**, 083026 (2011).

[12] A. De Pasquale, P. Facchi, G. Florio, V. Giovannetti, K. Matsuoka, and K. Yuasa, Two-mode bosonic quantum metrology with number fluctuations, Phys. Rev. A **92**, 042115 (2015).

[13] H. Uys and P. Meystre, Quantum states for Heisenberg-limited interferometry, Phys. Rev. A **76**, 013804 (2007).

[14] M. D. Lang and C. M. Caves, Optimal quantum-enhanced interferometry, Phys. Rev. A **90**, 025802 (2014).

[15] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Gaussian quantum information, Rev. Mod. Phys. **84**, 621 (2012).

[16] G. Adesso, S. Ragy, and A. R. Lee, Continuous variable quantum information: Gaussian states and beyond, Open Systems & Information Dynamics **21**, 1440001 (2014).

[17] A. Ferraro, S. Olivares, and M. G. A. Paris, *Gaussian States in Quantum Information*, Napoli Series on Physics and Astrophysics (Bibliopolis, Naples, 2005).

[18] A. Monras, Optimal phase measurements with pure gaussian states, Phys. Rev. A **73**, 033821 (2006).

[19] T. Matsubara, P. Facchi, V. Giovannetti, and K. Yuasa, Optimal Gaussian metrology for generic multimode interferometric circuit, New J. Phys. **21**, 033014 (2019).

[20] C. Oh, C. Lee, C. Rockstuhl, H. Jeong, J. Kim, H. Nha, and S.-Y. Lee, Optimal Gaussian measurements for phase estimation in single-mode Gaussian metrology, npj Quantum Inf. **5**, 10 (2019).

[21] H. M. Wiseman, Adaptive Phase Measurements of Optical Modes: Going Beyond the Marginal q Distribution, Phys. Rev. Lett. **75**, 4587 (1995).

[22] M. A. Armen, J. K. Au, J. K. Stockton, A. C. Doherty, and H. Mabuchi, Adaptive Homodyne Measurement of Optical Phase, Phys. Rev. Lett. **89**, 133602 (2002).

[23] M. Aspachs, J. Calsamiglia, R. Muñoz-Tapia, and E. Bagan, Phase estimation for thermal Gaussian states, Phys. Rev. A **79**, 033834 (2009).

[24] D. Gatto, P. Facchi, F. A. Narducci, and V. Tamma, Distributed quantum metrology with a single squeezed-vacuum source, Phys. Rev. Research **1**, 032024(R) (2019).

[25] D. Gatto, P. Facchi, and V. Tamma, Phase space Heisenberg-limited estimation of the average phase shift in a Mach-Zehnder interferometer, Int. J. Quantum. Inform. **18**, 1941019 (2020).

[26] M. Takeoka, K. P. Seshadreesan, C. You, S. Izumi, and J. P. Dowling, Fundamental precision limit of a Mach-Zehnder interferometric sensor when one of the inputs is the vacuum, Phys. Rev. A **96**, 052118 (2017).

[27] W. Ge, K. Jacobs, Z. Eldredge, A. V. Gorshkov, and M. Foss-Feig, Distributed Quantum Metrology with Linear Networks and Separable Inputs, Phys. Rev. Lett. **121**, 043604 (2018).

[28] Q. Zhuang, Z. Zhang, and J. H. Shapiro, Distributed quantum sensing using continuous-variable multipartite entanglement, Phys. Rev. A **97**, 032329 (2018).

[29] J. S Sidhu and P. Kok, Geometric perspective on quantum parameter estimation, AVS Quantum Science **2**, 014701 (2020).

[30] K. Qian, Z. Eldredge, W. Ge, G. Pagano, C. Monroe, J. V. Porto, and A. V. Gorshkov, Heisenberg-scaling measurement protocol for analytic functions with quantum sensor networks, Phys. Rev. A **100**, 042304 (2019).

[31] Y. Xia, W. Li, W. Clark, D. Hart, Q. Zhuang, and Z. Zhang, Demonstration of a Reconfigurable Entangled Radio-Frequency Photonic Sensor Network, Phys. Rev. Lett. **124**, 150502 (2020).

[32] X. Guo, C. R. Breum, J. Borregaard, S. Izumi, M. V. Larsen, T. Gehring, M. Christandl, J. S. Neergaard-Nielsen, and U. L. Andersen, Distributed quantum sensing in a continuous-variable entangled network, Nat. Phys. **16**, 281 (2020).

[33] R. Nair, Quantum-Limited Loss Sensing: Multiparameter Estimation and Bures Distance between Loss Channels, Phys. Rev. Lett. **121**, 230801 (2018).

[34] G. Gramegna, D. Triggiani, P. Facchi, F. A. Narducci, and V. Tamma, Heisenberg scaling precision in multi-mode distributed quantum metrology, arXiv:2003.12550.

[35] H. Cramér, *Mathematical Methods of Statistics*, Vol. 9 (Princeton University Press, Princeton, 1999).

[36] C. R. Rao, Information and the accuracy attainable in the estimation of statistical parameters, in *Breakthroughs in Statistics* (Springer, Berlin/Heidelberg, 1992), pp. 235–247.

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

[38] S. Olivares and M. G. A. Paris, Bayesian estimation in homodyne interferometry, J. Phys. B: At., Mol. Opt. Phys. **42**, 055506 (2009).

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

[40] M. O. Scully and M. S. Zubairy, *Quantum Optics* (Cambridge University Press, Cambridge, 1997).

[41] W. P. Schleich, *Quantum Optics in Phase Space* (John Wiley and Sons, New York, 2005).

[42] J. v. Neumann, Beweis des ergodensatzes und desh-theorems in der neuen mechanik, Z. Phys. **57**, 30 (1929).

[43] J. von Neumann, Proof of the ergodic theorem and the h-theorem in quantum mechanics, Eur. Phys. J. H **35**, 201 (2010).

[44] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*, Vol. 77 (American Mathematical Society, Providence, 2000).

[45] Z. Puchała and J. A. Miszczak, Symbolic integration with respect to the Haar measure on the unitary groups, Bull. Pol. Acad. Sci.: Tech. Sci. **65**, 21 (2017).

[46] P. Facchi and G. Garnero, Quantum thermodynamics and canonical typicality, Int. J. Geom. Methods Mod. Phys. **14**, 1740001 (2017).

[47] S. Popescu, A. J. Short, and A. Winter, Entanglement and the foundations of statistical mechanics, Nat. Phys. **2**, 754 (2006).