Moment-based superresolution: Formalism and applications

Giacomo Sorelli, Manuel Gessner, Mattia Walschaers, Nicolas Treps

Moment-based superresolution: Formalism and applications

Giacomo Sorelli, Manuel Gessner, Mattia Walschaers, and Nicolas Treps
Laboratoire Kastler Brossel, Sorbonne Université, ENS-Université PSL, CNRS, Collège de France, 4 Place Jussieu, 75252 Paris, France

(Received 26 May 2021; accepted 12 August 2021; published 15 September 2021)

Sensitivity limits are usually determined using the Cramér-Rao bound. Recently, this approach has been used to obtain the ultimate resolution limit for the estimation of the separation between two incoherent point sources. However, methods that saturate these resolution limits usually require the full measurement statistics, which can be challenging to access. In this work we introduce a simple superresolution protocol to estimate the separation between two thermal sources which relies only on the average value of a single accessible observable. We show how optimal observables for this technique may be constructed for arbitrary thermal sources and we study their sensitivities when one has access to spatially resolved intensity measurements (direct imaging) and photon counting after spatial-mode demultiplexing. For demultiplexing, our method is optimal, i.e., it saturates the quantum Cramér-Rao bound. We also investigate the impact of noise on the optimal observables, their measurement sensitivity, and the scaling with the number of detected photons of the smallest resolvable separation. For low signals in the image plane, we demonstrate that our method saturates the Cramér-Rao bound even in the presence of noise.

DOI: 10.1103/PhysRevA.104.033515

I. INTRODUCTION

Resolving small angular separations through an optical imaging system is an important problem both in microscopy and in astronomy. The most traditional imaging technique is a spatially resolved intensity measurement, also known as direct imaging. The resolution of this approach, as pointed out by Abbe [1] and Rayleigh [2] at the end of the nineteenth century, is limited by diffraction and noise. However, for a large enough signal-to-noise ratio, the diffraction limit is not fundamental [3] and can be overcome by superresolution techniques based on fluorescence microscopy [4–6], homodyne measurements [7–9], or intensity measurements in an appropriate spatial-mode basis [10,11].

Superresolution imaging has been analyzed from the point of view of quantum metrology [12–17] and the Fisher information was used to determine how well the separation between two incoherent sources can be resolved. In this framework, the diffraction limit manifests itself through the vanishing of the direct-imaging Fisher information for small source separations [18]. On the other hand, the quantum Fisher information, i.e., the Fisher information optimized over all possible measurements in the image plane, stays constant for small distances [18], leaving room for superresolution. Furthermore, it was shown [18] that the ultimate resolution limit given by the quantum Fisher information can be approached by photon counting after spatial-mode demultiplexing. Several experiments used spatial light modulators to implement a simplified version of this demultiplexing technique, which distinguishes only between the fundamental and the first excited modes [19–22]. On the other hand, distance estimation from the demultiplexing of multiple spatial modes, realized using multiplane light conversion [23], was only recently reported [24].

In general, reaching the Cramér-Rao bound, e.g., via maximum-likelihood estimation, requires the measurement of the full photon-counting statistics, which can be practically challenging. Here we demonstrate that this is not necessary in superresolution imaging, where the ultimate resolution can be obtained using a moment-based estimation technique, which requires one to measure only the average value of a single measurement observable. For this estimation technique, we identify the optimal observables when different measurements, such as spatially resolved intensity measurements (direct imaging) or photon counting after spatial-mode demultiplexing, are available.

In particular, we focus on the estimation of the transverse separation between two thermal sources of arbitrary and different brightnesses and we determine the sensitivity that can be reached with our optimized observables and consequently the minimal resolvable distance between the sources. For demultiplexing measurements, we construct the optimal observables also in the presence of experimental imperfections such as misalignment, measurement crosstalk, and detector noise. Therefore, our results are directly relevant for practical applications. Even in the presence of noise, for low brightnesses of the sources, we prove that our approach is sufficient to saturate the Cramér-Rao bound. Finally, for arbitrary brightnesses of the sources, we demonstrate that our optimized demultiplexing measurement allows us to reach the quantum Cramér-Rao bound [25,26], in the noiseless scenario, if sufficiently many modes are measured.

The paper is structured as follows. First, in Sec. II, we present our model for thermal sources in the image plane of an imaging system. After recalling the method of moments

for parameter estimation in Sec. III, in Sec. IV we use it to construct the optimal observable to estimate the source separation via spatial-mode demultiplexing, while in Sec. V we employ it to bound the sensitivity of ideal direct imaging. Detailed studies of the performances of our moment-based approach for ideal and noisy demultiplexing are presented in Secs. VI and VII, respectively. In Sec. VIII we discuss the smallest source separation which is resolvable by the different measurement techniques. Section IX summarizes our work.

II. THERMAL STATES IN THE IMAGE PLANE

We want to estimate the transverse distance between two thermal point sources located at positions $\pm\mathbf{r}_0$, with $\mathbf{r}_0 = (d\cos\theta/2, d\sin\theta/2)$. The sources emit a total mean photon number equal to $2N$ in the spatial modes associated with the field operators $\hat{s}_1$ and $\hat{s}_2$. We denote by $\hat{\rho}_a(N)$ a thermal state with mean photon number N in the mode associated with the field operator $\hat{a}$. Accordingly, the sources are described by the quantum state $\hat{\rho}_0 = \hat{\rho}_{s_1}[(1-\gamma)N] \otimes \hat{\rho}_{s_2}[(1+\gamma)N]$, where $-1 < \gamma < 1$ takes into account the possibly different (but finite) temperatures of the two sources. Such a state is described by the density matrix

$$\hat{\rho}_0 = \int d^2\alpha_1 d^2\alpha_2 P_{s_1,s_2}(\alpha_1,\alpha_2)|\alpha_1,\alpha_2\rangle\langle\alpha_1,\alpha_2|, \tag{1}$$

where $|\alpha_{1\,(2)}\rangle$ are coherent states of the field operators $\hat{s}_{1\,(2)}$ and $P_{s_1,s_2}(\alpha_1,\alpha_2) = P_{s_1}(\alpha_1)P_{s_2}(\alpha_2)$ is the Glauber-Sudarshan P function, with

$$P_{s_{1\,(2)}}(\alpha_{1\,(2)}) = \frac{1}{\pi N(1\mp\gamma)} e^{-|\alpha_{1\,(2)}|^2/(1\mp\gamma)N}. \tag{2}$$

The evolution of the field operators through a diffraction-limited imaging system, with transmissivity κ, is given by [26,27]

$$\hat{c}_1 = \sqrt{\kappa}\hat{s}_1 + \sqrt{1-\kappa}\hat{v}_1, \tag{3a}$$

$$\hat{c}_2 = \sqrt{\kappa}\hat{s}_2 + \sqrt{1-\kappa}\hat{v}_2, \tag{3b}$$

where $\hat{c}_{1\,(2)}$ are the field operators associated with the images $u_0(\mathbf{r}\pm\mathbf{r}_0)$ of the two sources, with $u_0(\mathbf{r})$ the point spread function (PSF) of the imaging system, which we assume to be real up to a global phase. The field operators $\hat{v}_{1\,(2)}$ are associated with auxiliary modes, which are in the vacuum state. This beam-splitter model for the propagation through an imaging system is illustrated in Fig. 1(a).

The modes $u_0(\mathbf{r}\pm\mathbf{r}_0)$ are nonorthogonal and therefore the operators $\hat{c}_1^\dagger$ and $\hat{c}_2$ do not commute [see Fig. 1(a)]. As a consequence, these modes cannot be used to properly represent the quantum state of the sources in the image plane. To obviate this problem, we introduce the orthonormal image modes

$$u_\pm(\mathbf{r}) = \frac{u_0(\mathbf{r}+\mathbf{r}_0) \pm u_0(\mathbf{r}-\mathbf{r}_0)}{\sqrt{2(1\pm\delta)}}, \tag{4}$$

where δ represents the overlap between the source images

$$\delta = \int d^2\mathbf{r}\, u_0^*(\mathbf{r}+\mathbf{r}_0)u_0(\mathbf{r}-\mathbf{r}_0). \tag{5}$$

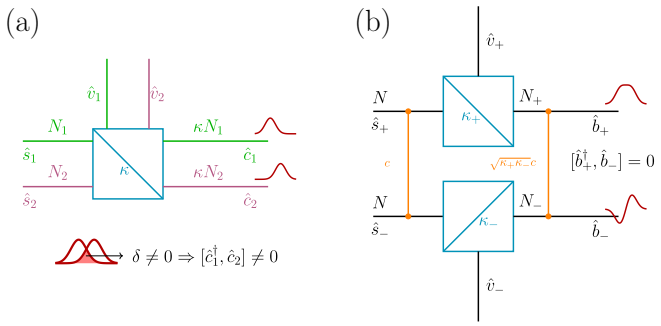

FIG. 1. Equivalent beam-splitter models for the propagation of thermal states through a diffraction-limited imaging system. (a) The source's modes $\hat{s}_{1/2}$ are populated with two thermal states with photon numbers $N(1\pm\gamma)$ and are mixed with the nonorthogonal vacuum modes $\hat{v}_{1/2}$ on a beam splitter with transmissivity κ, resulting in the nonorthogonal image modes $\hat{c}_{1/2}$. (b) The symmetric and antisymmetric modes $\hat{s}_\pm$ have equal mean photon number N, but are classically correlated with phase-insensitive correlations $\langle\hat{s}_\pm^\dagger\hat{s}_\mp\rangle = \gamma N$. The modes $\hat{s}_\pm$ are mixed with the orthogonal vacuum modes $\hat{v}_\pm$ on two beam splitters with transmissivities $\kappa_\pm$, resulting in the orthogonal image modes $\hat{b}_\pm$

The relation between the field operators $\hat{b}_\pm$ associated with the modes $u_\pm(\mathbf{r})$ and the field operators $\hat{s}_\pm = (\hat{s}_1 \pm \hat{s}_2)/\sqrt{2}$ in the object plane can be obtained from the sum and difference of Eqs. (3) [26],

$$\hat{b}_\pm = \sqrt{\kappa_\pm}\hat{s}_\pm + \sqrt{1-\kappa_\pm}\hat{v}_\pm, \tag{6}$$

with $\hat{v}_\pm$ associated with auxiliary modes, which are in the vacuum state, and $\kappa_\pm = \kappa(1\pm\delta)$.

We now use Eq. (6) to propagate the quantum state $\hat{\rho}_0$ of the sources to the image plane. First, we note that the transformation to the modes $\hat{s}_\pm$ can be interpreted as a 50:50 beam splitter

$$\begin{pmatrix}\hat{s}_+\\ \hat{s}_-\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}\hat{s}_1\\ \hat{s}_2\end{pmatrix}. \tag{7}$$

According to Eq. (7), the coherent states $|\alpha_1\rangle|\alpha_2\rangle$ correspond to the coherent states $|\frac{\alpha_1+\alpha_2}{\sqrt{2}}\rangle|\frac{\alpha_1-\alpha_2}{\sqrt{2}}\rangle \equiv |\alpha_+\rangle|\alpha_-\rangle$ of the field operators $\hat{s}_\pm$ [see Fig. 1(b) for a schematic illustration]. The quantum state of the sources can therefore be rewritten in terms of the latter coherent states as

$$\hat{\rho}_0 = \int d^2\alpha_+ d^2\alpha_- P_{s_+,s_-}(\alpha_+,\alpha_-)|\alpha_+,\alpha_-\rangle\langle\alpha_+,\alpha_-|, \tag{8}$$

where $P_{s_+,s_-}(\alpha_+,\alpha_-) = P_{s_1,s_2}(\frac{\alpha_++\alpha_-}{\sqrt{2}}, \frac{\alpha_+-\alpha_-}{\sqrt{2}})$. Accordingly, the modes $u_\pm(\mathbf{r})$ have both mean photon numbers $\langle\hat{s}_\pm^\dagger\hat{s}_\pm\rangle = N$ and the photon-number imbalance γ appears in the classical phase-insensitive correlations $\langle\hat{s}_\pm^\dagger\hat{s}_\mp\rangle = N\gamma$ [see orange lines in Fig. 1(b)]. Going through the imaging system according to Eq. (6), we have $|\alpha_+,\alpha_-\rangle \to |\sqrt{\kappa_+}\alpha_+, \sqrt{\kappa_-}\alpha_-\rangle \equiv |\beta_+,\beta_-\rangle$, with $|\beta_\pm\rangle$ coherent states of the field operators $\hat{b}_\pm$. We can therefore write the quantum state in the image plane as

$$\hat{\rho}(d,\theta) = \int d^2\beta_+ d^2\beta_- P_{b_+,b_-}(\beta_+,\beta_-)|\beta_+,\beta_-\rangle\langle\beta_+,\beta_-|, \tag{9}$$

with

$$P_{b_+,b_-}(\beta_+,\beta_-)=\frac{1}{\kappa_+\kappa_-}P_{s_+,s_-}\left(\frac{\beta_+}{\sqrt{\kappa_+}},\frac{\beta_-}{\sqrt{\kappa_-}}\right). \tag{10}$$

Combining Eqs. (10) and (2), we can write the P function as

$$P_{b_+,b_-}(\beta_+,\beta_-)=\frac{1}{\pi^2\det V}e^{-\boldsymbol{\beta}^\dagger V\boldsymbol{\beta}}, \tag{11}$$

where we have defined $\boldsymbol{\beta}=(\beta_+,\beta_-)^T$ and

$$V=\begin{pmatrix} N_+ & \gamma\sqrt{N_+N_-}\\ \gamma\sqrt{N_+N_-} & N_-\end{pmatrix}, \tag{12}$$

with $N_\pm=N\kappa_\pm$. For equally bright sources ($\gamma=0$), the off-diagonal elements of the matrix V vanish and Eq. (11) reduces to the product of two Gaussian functions corresponding to $\hat{\rho}(\theta,d)=\hat{\rho}_{b_+}(N_+)\otimes\hat{\rho}_{b_-}(N_-)$, as reported in Ref. [28]. Finally, the sources in the image plane are described by Eqs. (9) and (11), with the information on the parameter d contained in the shape of the modes $u_\pm(\mathbf{r})$ and the mean photon numbers $N_\pm$.

III. METHOD OF MOMENTS

We estimate the distance d between the two sources with the method of moments [17,29]. Following this approach, given an observable $\hat{X}$, an estimator $\tilde{d}$ for the parameter d is extracted from the sample mean $\bar{x}_\mu=\sum_{i=1}^{\mu}x_i/\mu$ of μ independent measurements of $\hat{X}$. The distance estimator is obtained by comparing the sample mean $\bar{x}_\mu$ with a calibration curve given by the expectation value $\langle\hat{X}\rangle=\mathrm{tr}[\hat{X}\hat{\rho}(d,\theta)]$ of the measurement operator $\hat{X}$ as a function of the source separation d, which could be known either from theory or from a previous calibration experiment.

When sufficiently many measurements are performed ($\mu\gg1$), it follows from the central limit theorem that $\bar{x}_\mu$ is normally distributed with mean value $\langle\hat{X}\rangle$ and variance $(\Delta\hat{X})^2/\mu$, with $(\Delta\hat{X})^2=\langle\hat{X}^2\rangle-\langle\hat{X}\rangle^2$. Accordingly, we choose as a separation estimator the parameter value $\tilde{d}$ at which $\langle\hat{X}\rangle$ equates the measurement mean $\bar{x}_\mu$. The associated estimation error is given by $(\Delta d)^2=\chi^2[d,\theta,\hat{X}]/\mu$, with

$$\chi^2[d,\theta,\hat{X}]=\frac{(\Delta\hat{X})^2}{\left(\frac{\partial\langle\hat{X}\rangle}{\partial d}\right)^2}. \tag{13}$$

The quantity $\chi^2[d,\theta,\hat{X}]$ determines the sensitivity of the method-of-moments estimation strategy for the quantum state $\rho(d,\theta)$ and the observable $\hat{X}$. It obeys the chain of inequalities

$$\chi^{-2}[d,\theta,\hat{X}]\leqslant\mathcal{F}[d,\theta,\hat{X}]\leqslant\mathcal{F}_Q[d,\theta], \tag{14}$$

where the second is saturable by an optimal operator $\hat{X}$ [13,17,29–31]. Here $\mathcal{F}[d,\theta,\hat{X}]$ is the Fisher information that bounds the achievable sensitivity when estimating d from measurements of $\hat{X}$ according to the Cramér-Rao lower bound $(\Delta d)^2\geqslant(\mu\mathcal{F}[d,\theta,\hat{X}])^{-1}$ [12,13]. Finally, $\mathcal{F}_Q[d,\theta]=\max_{\hat{X}}(\mathcal{F}[d,\theta,\hat{X}])$ is the quantum Fisher information which defines the ultimate metrological sensitivity [13].

In practical situations, one does not have access to all possible measurement observables. Therefore, we now assume that we can measure only a finite number K of observables

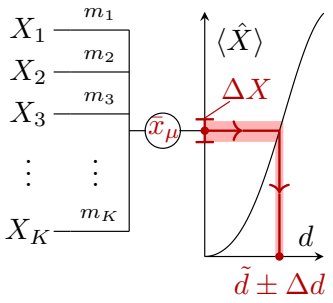

FIG. 2. Moment-based estimation of the separation between the sources. The measured mean values of the available observables are linearly combined with optimal coefficients and compared with a calibration curve.

$\hat{X}_k$, as well as their linear combinations $\hat{X}_{\tilde{\mathbf{m}}}=\tilde{\mathbf{m}}\cdot\hat{\mathbf{X}}$, with $\hat{\mathbf{X}}=(\hat{X}_1,\ldots,\hat{X}_K)^T$ and $\tilde{\mathbf{m}}$ the measurement coefficients vector. Under these assumptions, it is possible to perform an analytical optimization over all possible linear combinations. Such an optimization yields the measurement sensitivity [29]

$$\begin{aligned}M[d,\theta,\hat{\mathbf{X}}]&=\max_{\tilde{m}}(\chi^{-2}[d,\theta,\hat{X}_{\tilde{\mathbf{m}}}])\\&=\mathbf{D}[d,\theta,\hat{\mathbf{X}}]^T\Gamma[d,\theta,\hat{\mathbf{X}}]^{-1}\mathbf{D}[d,\theta,\hat{\mathbf{X}}],\end{aligned} \tag{15}$$

where $\Gamma[d,\theta,\hat{\mathbf{X}}]$ is the covariance matrix, whose elements are given by $\Gamma_{k,l}[d,\theta,\hat{\mathbf{X}}]=\langle\hat{X}_k\hat{X}_l\rangle-X_kX_l$, and $D[d,\theta,\hat{\mathbf{X}}]=\frac{\partial X_k}{\partial d}$ is the derivative vector, where we defined as $X_k=\langle\hat{X}_k\rangle$ the expectation value of $\hat{X}_k$. The sensitivity given by Eq. (15) is achieved by the measurement coefficients $\tilde{\mathbf{m}}=\mathbf{m}$ [29], with

$$\mathbf{m}=\eta\Gamma^{-1}[d,\theta,\hat{\mathbf{X}}]D[d,\theta,\hat{\mathbf{X}}], \tag{16}$$

where η is a normalization constant. The procedure discussed in this section to obtain the separation estimation $\tilde{d}$ from the optimal coefficients $\mathbf{m}$ in Eq. (16) is illustrated in Fig. 2.

IV. CONSTRUCTION OF THE OPTIMAL OBSERVABLE FOR DEMULTIPLEXING

Fisher-information-based studies [18,25] suggested that photon counting after spatial-mode demultiplexing allows one to approach the ultimate limit for the separation estimation. Therefore, in this section we assume that we have access to K spatial modes $\{v_k(\mathbf{r})\}$ with associated field operators $\hat{a}_k$ and that we can measure the photon number in each mode $\hat{N}_k=\hat{a}_k^\dagger\hat{a}_k$. We use the formalism presented in Sec. III to derive their optimal linear combination

$$\hat{N}_{\mathbf{m}}=\mathbf{m}\cdot\hat{\mathbf{N}}=\sum_{i=1}^{K}m_i\hat{N}_i, \tag{17}$$

with $\hat{\mathbf{m}}$ determined by Eq. (16). Once this optimal observable is known, we only need to access its average value. The latter can be derived by measuring the mean photon number in each of the modes $\{v_k(\mathbf{r})\}$ and then combining them according to the coefficients $\mathbf{m}$.

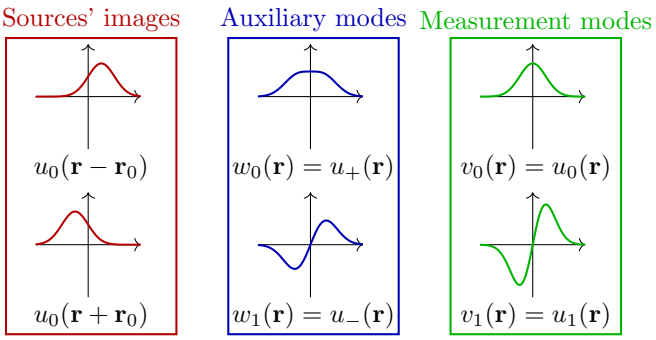

FIG. 3. The images of the two sources $u_0(\mathbf{r}\pm\mathbf{r}_0)$ [red (left)] are compared with the first two modes of the auxiliary mode basis $\{w_i(\mathbf{r})\}$ [blue (middle)] and the ideal HG measurement basis $\{v_i(\mathbf{r})=u_i(\mathbf{r})\}$ [green (right)]. The modes $\{v_i(\mathbf{r})\}$ are modified by misalignment and crosstalk.

A. Covariance matrix and derivative vector

We now use the P function (11) to calculate the photon-number covariance matrix and derivative vector which are needed to compute the measurement sensitivity (15) and coefficients (16). For this goal, it is convenient to define an auxiliary mode basis $\{w_i(\mathbf{r})\}$ obtained by extending $w_0(\mathbf{r})=u_+(\mathbf{r})$ and $w_1(\mathbf{r})=u_-(\mathbf{r})$ to a complete orthonormal basis. Accordingly, for the field operators $\hat{b}_i$ associated with this basis, we have $\hat{b}_0=\hat{b}_+$ and $\hat{b}_1=\hat{b}_-$. The field operators $\hat{a}_k$ in the measurement basis $\{v_k(\mathbf{r})\}$ can be expanded in terms of the field operators $\hat{b}_k$ as $\hat{a}_k=\sum_l g_{kl}\hat{b}_l$, with $g_{kl}=\int d^2\mathbf{r}\, v_k^*(\mathbf{r})w_l(\mathbf{r})$. A one-dimensional comparison between the sources' images and the first two modes of the basis $\{w_k(\mathbf{r})\}$ and $\{v_k(\mathbf{r})\}$ is presented in Fig. 3. Using the definition of the mean photon number in the measurement modes, we have

$$N_k=\langle\hat{N}_k\rangle=\langle\hat{a}_k^\dagger\hat{a}_k\rangle=\sum_{ij=\pm}g_{ki}^*g_{kj}\langle\hat{b}_i^\dagger\hat{b}_j\rangle, \tag{18}$$

where we used that all $\hat{b}_{i\geqslant 2}$ are in the vacuum. Analogously, the covariance matrix is given by $\Gamma_{kl}[d,\theta,\hat{\mathbf{N}}]=\langle\hat{N}_k\hat{N}_l\rangle-N_kN_l$, with

$$\begin{aligned}\langle\hat{N}_k\hat{N}_l\rangle&=\langle\hat{a}_k^\dagger\hat{a}_l^\dagger\hat{a}_k\hat{a}_l\rangle+\delta_{kl}\langle\hat{a}_k^\dagger\hat{a}_k\rangle\\&=\sum_{mnpq=\pm}g_{mk}^*g_{nl}^*g_{pk}g_{ql}\langle\hat{b}_m^\dagger\hat{b}_n^\dagger\hat{b}_p\hat{b}_q\rangle+\delta_{kl}N_k.\end{aligned} \tag{19}$$

The normally ordered expectation values of the field operators $\hat{b}_i$ in Eqs. (18) and (19) can be calculated from the P function (11) according to

$$\langle\hat{b}_m^\dagger\hat{b}_n\rangle=\int\beta_m^*\beta_nP(\beta_+,\beta_-)d^2\beta_+d^2\beta_-, \tag{20a}$$

$$\langle\hat{b}_m^\dagger\hat{b}_n^\dagger\hat{b}_p\hat{b}_q\rangle=\int\beta_m^*\beta_n^*\beta_p\beta_qP(\beta_+,\beta_-)d^2\beta_+d^2\beta_-. \tag{20b}$$

Given that $P(\beta_+,\beta_-)$ [see Eq. (11)] is a Gaussian involving only the modes $u_\pm(\mathbf{r})$, the only nonzero expectation values are

$$\langle\hat{b}_\pm^\dagger\hat{b}_\pm\rangle=N_\pm, \tag{21a}$$

$$\langle\hat{b}_\pm^\dagger\hat{b}_\mp\rangle=-\gamma\sqrt{N_+N_-}, \tag{21b}$$

$$\langle\hat{b}_\pm^\dagger\hat{b}_\pm^\dagger\hat{b}_\pm\hat{b}_\pm\rangle=2N_\pm^2, \tag{21c}$$

$$\langle\hat{b}_\pm^\dagger\hat{b}_\mp^\dagger\hat{b}_\pm\hat{b}_\mp\rangle=N_+N_-(1+\gamma^2), \tag{21d}$$

$$\langle\hat{b}_\pm^\dagger\hat{b}_\pm^\dagger\hat{b}_\mp\hat{b}_\mp\rangle=2\gamma^2N_+N_-, \tag{21e}$$

$$\langle\hat{b}_\pm^\dagger\hat{b}_\mp^\dagger\hat{b}_\pm\hat{b}_\mp\rangle=-2\gamma N_\pm\sqrt{N_+N_-}. \tag{21f}$$

It is useful to write the mean photon number and the covariance matrices in terms of the sources images. For this purpose, we note that the coefficients $g_{k\pm}$ can be expressed in terms of the overlap functions with the source images $f_{\pm,k}=\int d\mathbf{r}\, v_k^*(\mathbf{r})u(\mathbf{r}\pm\mathbf{r}_0)$ as $g_{k\pm}=(f_{+,k}\pm f_{-,k})/\sqrt{2(1\pm\delta)}$ and using $N_\pm=N\kappa(1\pm\delta)$. Accordingly, inserting Eqs. (21) into Eqs. (18) and (19), using $N_\pm=N\kappa(1\pm\delta)$, we get the mean photon numbers

$$N_k=N\kappa(|f_{+,k}|^2+|f_{-,k}|^2)-\gamma N\kappa(|f_{+,k}|^2-|f_{-,k}|^2) \tag{22}$$

and the covariance matrix

$$\Gamma_{kl}[d,\theta,\hat{\mathbf{N}}]=\Gamma_{kl}^0+\gamma\Gamma_{kl}^1+\gamma^2\Gamma_{kl}^2, \tag{23}$$

with

$$\begin{aligned}\Gamma_{kl}^0=&(N\kappa)^2(|f_{-,k}|^2|f_{-,l}|^2+|f_{+,k}|^2|f_{+,l}|^2\\&+2\,\mathrm{Re}[f_{-,k}f_{-,l}^*f_{+,l}f_{+,k}^*])\\&+\delta_{kl}N\kappa(|f_{+,k}|^2+|f_{-,k}|^2),\end{aligned} \tag{24a}$$

$$\Gamma_{kl}^1=-\delta_{kl}N\kappa(|f_{+,k}|^2-|f_{-,k}|^2), \tag{24b}$$

$$\begin{aligned}\Gamma_{kl}^2=&(N\kappa)^2(|f_{-,k}|^2|f_{-,l}|^2+|f_{+,k}|^2|f_{+,l}|^2\\&-2\,\mathrm{Re}[f_{-,k}f_{-,l}^*f_{+,k}f_{+,l}^*]).\end{aligned} \tag{24c}$$

The derivative vector is obtained by differentiating Eq. (22) with respect to the parameter d, which gives

$$\begin{aligned}D_k[d,\theta,\hat{\mathbf{N}}]=2N\kappa\Bigg[&\mathrm{Re}\left(f_{+,k}^*\frac{\partial f_{+,k}}{\partial d}+f_{-,k}^*\frac{\partial f_{-,k}}{\partial d}\right)\\&-\gamma\,\mathrm{Re}\left(f_{+,k}^*\frac{\partial f_{+,k}}{\partial d}-f_{-,k}^*\frac{\partial f_{-,k}}{\partial d}\right)\Bigg].\end{aligned} \tag{25}$$

In the case of two equally bright sources ($\gamma=0$), we recover the expressions given in Ref. [28]. Analogously, setting $\gamma=0$ in Eq. (23), we have $\Gamma_{kl}[d,\theta,\hat{\mathbf{N}}]=\Gamma_{kl}^0$, which coincides with the covariance matrix reported in Ref. [28].

B. Noise sources

For the rest of this work, we focus on the case of a Gaussian PSF $u_0(\mathbf{r})=\sqrt{2/\pi w^2}\exp(-|\mathbf{r}|^2/w^2)$. For this PSF, the quantum Cramér-Rao bound can be approached by demultiplexing Hermite-Gauss (HG) modes with width matching that of the PSF [18,25], i.e., $v_k(\mathbf{r})=u_k(\mathbf{r})\equiv u_{nm}(\mathbf{r})$, with $k=(n,m)$, such that $u_{00}(\mathbf{r})=u_0(\mathbf{r})$. The HG modes are defined, for $\mathbf{r}=(x,y)$, as

$$u_{nm}(x,y)=\mathcal{N}_{nm}H_n\left(\frac{\sqrt{2}x}{w}\right)H_m\left(\frac{\sqrt{2}y}{w}\right)e^{-(x^2+y^2)/w^2}, \tag{26}$$

where $H_n(x)$ are the Hermite polynomials and $\mathcal{N}_{nm}=[(\pi/2)w^2 2^{n+m}n!m!]^{-1/2}$ is the normalization constant. For ideal measurements in the HG mode basis, we have the

overlap functions $f_{\pm,k} = \beta_{nm}(\pm\mathbf{r}_0)$, with

$$\beta_{nm}(\mathbf{a}) \equiv \int d^2\mathbf{r}\, u^*_{nm}(\mathbf{r}) u_{00}(\mathbf{r}-\mathbf{a}), \tag{27}$$

which fully determine the covariance matrix (23) and the derivative vector (25) and consequently the measurement sensitivity (15) and coefficients (16). A detailed discussion of the performances of our estimation strategy for ideal measurements is reported in Sec. VI.

We now study the impact of different noise sources on the covariance matrix (23) and the derivative vector (25). In particular, we consider misalignment between the demultiplexing basis and the source centroid, as well as crosstalk between the detection modes. These two imperfections affect Eqs. (23) and (25) only through modifications of the overlap functions $f_{\pm,k}$. Finally, we consider dark counts at the detection stage, which modify the diagonal of the covariance matrix (23). A schematic illustration of how the different noise sources enter a demultiplexing measurement is presented in Fig. 4.

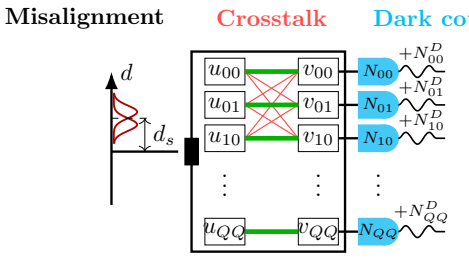

FIG. 4. Graphical illustration of a noisy demultiplexing measurement. The image of the two sources enters a demultiplexer, which performs mode sorting affected by crosstalk. Photon-counting measurements affected by electronic noise are performed at each demultiplexer's output.

1. Misalignment

The assumption that the demultiplexing basis is perfectly centered with respect to the centroid of the two sources is often not true in practice. In fact, the source centroid is in general *a priori* unknown and needs to be preestimated, possibly via direct imaging, to optimally align the demultiplexer. In the case of faint sources, it was observed [18] that an imperfect positioning of the demultiplexer makes the Fisher information go to zero for small source separations. Methods to mitigate this effect have been proposed by adaptively switching between demultiplexing and direct imaging [32] or by optimizing the detection basis [33]. An alternative approach, which we will not consider here, consists in estimating the source's centroid simultaneously with the separation [34].

Within our model, a two-dimensional shift $\mathbf{r}_s = (d_s\cos\theta_s, d_s\sin\theta_s)^T$ of the centroid of the sources with respect to the demultiplexer axis enters the measurement sensitivity (15) and the coefficients (16) through the overlap functions $f_{\pm,k} = \beta_{nm}(\pm\mathbf{r}_0 - \mathbf{r}_s)$. As a trick, to compute the overlap functions $\beta_{nm}(\mathbf{a})$, we can use an analogy with the quantum-mechanical harmonic oscillator. In particular, we can interpret the integral (27) as the overlap between the (n,m) excited state of a two-dimensional harmonic oscillator and a coherent state (displaced to the phase-space coordinate $\mathbf{a}$) of the same harmonic oscillator. We therefore obtain

$$\beta_{nm}(\pm\mathbf{r}_0 + \mathbf{r}_s) = \frac{1}{\sqrt{n!m!}} \exp\left(-\frac{\left(d_s\cos\theta_s \pm \frac{d}{2}\cos\theta\right)^2 + \left(d_s\sin\theta_s \pm \frac{d}{2}\sin\theta\right)^2}{2w^2}\right) \times \left(\frac{d_s}{w}\cos\theta_s \pm \frac{d}{2w}\cos\theta\right)^n \left(\frac{d_s}{w}\sin\theta_s \pm \frac{d}{2w}\sin\theta\right)^m. \tag{28}$$

2. Crosstalk

A recent experiment [24] identified imperfections in the mode decompositions as an important limitation for resolving the distance between incoherent point sources. For sources with low brightness, it was reported that, in the presence of crosstalk in the demultiplexing basis, the Fisher information drops to zero for small separations [35].

To include the impact of crosstalk between the detection modes in our model, we follow [35] and we describe it as a unitary matrix c_{kl} that maps the ideal HG modes $u_l(\mathbf{r})$ into the actual measurement basis $v_k = \sum_l c_{kl} u_l(\mathbf{r})$. Accordingly, the overlap functions $f_{\pm,k}$ are given by linear combinations of the functions (28),

$$f_{\pm,mn} = \sum_{pq} c_{nm,pq}\beta_{pq}(\pm\mathbf{r}_0 - \mathbf{r}_s). \tag{29}$$

To assess the impact of crosstalk on the sensitivity of our method, as well as on the shape of our optimal observable, we numerically generate $K \times K$ unitary matrices

$$C(\epsilon) = \exp\left(-i\epsilon \sum_{i=1}^{K^2-1} \lambda_i G_i\right) \tag{30}$$

by sampling the uniformly distributed random real coefficients λ_i ($0 \leqslant \lambda_i \leqslant 1$ and $\sum_i \lambda_i^2 = 1$), which multiply the generalized Gell-Mann matrices G_i [36]. The positive parameter ϵ allows us to tune the crosstalk strength. In experimentally relevant scenarios, crosstalk is generally weak [24,35]. Namely, the diagonal elements of the crosstalk matrix $C_{ii}(\epsilon) = c_{ii}$ are larger when compared to the off-diagonal ones $C_{i\neq j}(\epsilon) = c_{i\neq j}$. Accordingly, we select ϵ to ensure that

$$\overline{|c_{ij}|^2} = \frac{1}{K(K-1)} \sum_{k\neq l=1}^{K} |c_{kl}|^2 \ll 1. \tag{31}$$

In the following, when discussing results in the presence of crosstalk, we will consider multiple random realizations of the

matrix c_{ij} and quantify the average crosstalk probability via the ensemble average $\langle\overline{|c_{ij}|^2}\rangle$.

3. Dark counts

Electronic noise at the detection stage introduces additional photon counts, the dark counts, which contain no information on the parameter value d. Consequently, the signal-to-noise ratio in each detection mode is reduced by dark counts (dc). The impact of this noise source on distance estimation via spatial-mode demultiplexing has been investigated in the recent literature [37–39]. Despite the different approaches, all these works obtained the same qualitative result: Dark counts cause the Fisher information to drop to zero for small source separations.

To include dark counts in our model, we add to the quantum-mechanical photon-number operator in the measurement modes $\hat{N}_k$ a classical random variable ξ_k. In particular, ξ_k are thermally distributed with mean value

$$\langle\xi_k\rangle = N_k^{\mathrm{dc}}. \tag{32}$$

To quantify the strength of dark counts with respect to the number of photons received in the image plane, we define $\sigma_k = N_k^{\mathrm{dc}}/2N\kappa$. The measured mean photon number in each detection mode N_k' and the covariance matrix $\Gamma'[d,\theta,\hat{\mathbf{N}}]$ are now obtained by computing not only quantum-mechanical expectation values, but also classical averages over the probability distribution of ξ_k. Following this procedure, we obtain

$$N_k' = N_k + N_k^{\mathrm{dc}} \tag{33}$$

and, given that N_k^{dc} is independent of d, the derivative vector $D[d,\theta,\hat{\mathbf{N}}]$ is untouched by dark counts. On the other hand, the covariance matrix becomes

$$\Gamma'_{kl}[d,\theta,\hat{\mathbf{N}}] = \Gamma_{kl}[d,\theta,\hat{\mathbf{N}}] + \delta_{kl}N_k^{\mathrm{dc}}\left(N_k^{\mathrm{dc}}+1\right), \tag{34}$$

with the extra diagonal term describing the noise induced by dark counts.

V. DIRECT IMAGING

In Secs. VI and VII we will discuss the performances of ideal and noisy demultiplexing, respectively, and we will compare them with those of direct imaging. In order to perform such a comparison, we now consider an ideal direct-imaging system and we evaluate its optimized moment-based sensitivity (15) that bounds the Fisher information according to Eq. (14).

Direct imaging estimates the separation between the two sources from the intensity distribution in the image plane. The mean intensity is given by

$$I(\mathbf{r}) = \langle\hat{E}^{(+)\dagger}(\mathbf{r})E^{(+)}(\mathbf{r})\rangle = \sum_{i,j=\pm} u_i^*(\mathbf{r})u_j(\mathbf{r})\langle\hat{b}_i^\dagger\hat{b}_j\rangle = N\kappa[(1+\gamma)|u_0(\mathbf{r}+\mathbf{r}_0)|^2 + (1-\gamma)|u_0(\mathbf{r}-\mathbf{r}_0)|^2]. \tag{35}$$

In the first step, we used the expansion of the electric field in the basis $\{w_k(\mathbf{r})\}$ of the symmetric $w_0(\mathbf{r}) = u_+(\mathbf{r})$ and antisymmetric modes $w_1(\mathbf{r}) = u_-(\mathbf{r})$ (see Sec. IV) and the fact that, in the quantum state (8), all higher-order modes $w_{k\geqslant 2}(\mathbf{r})$ are in the vacuum state. In the second step, we used $\langle\hat{b}_\pm^\dagger\hat{b}_\pm\rangle = N_\pm$ and $\langle\hat{b}_\pm^\dagger\hat{b}_\mp\rangle = -\gamma\sqrt{N_+N_-}$ [see Eqs. (21)], as well as the relation (4) between $u_0(\mathbf{r}\pm\mathbf{r}_0)$ and $u_\pm(\mathbf{r})$.

Let us now assume that the intensity measurements are performed with an ideal pixelized detector, with unity quantum efficiency and noiseless number-resolved photon counting at each pixel. The pixels are defined by the area

$$\mathcal{A}_{ij} = \{(x,y) : |x-x_i| \leqslant x_p, |y-y_i| \leqslant y_p\}, \tag{36}$$

with $2x_p$ ($2y_p$) the horizontal (vertical) size of the pixels and (x_i, y_i) the coordinate of the center of the pixel. We further assume that the total area $\mathcal{A}$ of the detector is fixed and large enough to collect the full intensity in the image plane. Accordingly, the pixel area is fully determined by the number of segments on each axis N_p, i.e., $x_p \times y_p = \mathcal{A}/N_p^2$. The mean photon count per pixel is given by the mean intensity integrated over the pixel area

$$I_{ij} = N\kappa(1+\gamma)\Phi_{ij} + N\kappa(1-\gamma)\Psi_{ij}, \tag{37}$$

with

$$\Phi_{ij} = \int_{\mathcal{A}_{ij}} |u_0(\mathbf{r}+\mathbf{r}_0)|^2 d^2\mathbf{r}, \tag{38a}$$

$$\Psi_{ij} = \int_{\mathcal{A}_{ij}} |u_0(\mathbf{r}-\mathbf{r}_0)|^2 d^2\mathbf{r}. \tag{38b}$$

The intensity coherence function is defined as

$$\Gamma(\mathbf{r},\mathbf{r}') = \langle\hat{E}^{(+)\dagger}(\mathbf{r})E^{(+)}(\mathbf{r})\hat{E}^{(+)\dagger}(\mathbf{r}')E^{(+)}(\mathbf{r}')\rangle - I(\mathbf{r})I(\mathbf{r}'), \tag{39}$$

with

$$\langle \hat{E}^{(+)\dagger}(\mathbf{r})\hat{E}^{(+)}(\mathbf{r})\hat{E}^{(+)\dagger}(\mathbf{r}')\hat{E}^{(+)}(\mathbf{r}')\rangle = \sum_{ijkl=\pm} u_i^*(\mathbf{r})u_j(\mathbf{r})u_k^*(\mathbf{r}')u_l(\mathbf{r}')\langle \hat{b}_i^\dagger \hat{b}_j \hat{b}_k^\dagger \hat{b}_l\rangle$$
$$= \sum_{ijkl=\pm} u_i^*(\mathbf{r})u_j(\mathbf{r})u_k^*(\mathbf{r}')u_l(\mathbf{r}')\langle \hat{b}_i^\dagger \hat{b}_k^\dagger \hat{b}_j \hat{b}_l\rangle + \delta(\mathbf{r}-\mathbf{r}')I(\mathbf{r}), \tag{40}$$

where in the last step we used the commutation relation $[\hat{b}_i, \hat{b}_j^\dagger] = \delta_{ij}$ and the completeness relation $\sum_i w_i^*(\mathbf{r})w_i(\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}')$. Using the expectation values (21), we can finally write the intensity coherence function as

$$\Gamma(\mathbf{r},\mathbf{r}') = \Gamma^0(\mathbf{r},\mathbf{r}') + \gamma\Gamma^1(\mathbf{r},\mathbf{r}') + \gamma^2\Gamma^2(\mathbf{r},\mathbf{r}'), \tag{41}$$

with

$$\Gamma^0(\mathbf{r},\mathbf{r}') = N^2\kappa^2\{|u_0(\mathbf{r}+\mathbf{r}_0)|^2|u_0(\mathbf{r}'+\mathbf{r}_0)|^2 + |u_0(\mathbf{r}-\mathbf{r}_0)|^2|u_0(\mathbf{r}'-\mathbf{r}_0)|^2$$
$$+2\,\mathrm{Re}[u_0^*(\mathbf{r}+\mathbf{r}_0)u_0(\mathbf{r}-\mathbf{r}_0)u_0^*(\mathbf{r}'-\mathbf{r}_0)u_0(\mathbf{r}'+\mathbf{r}_0)]\}$$
$$+N\kappa\delta(\mathbf{r}-\mathbf{r}')[|u_0(\mathbf{r}+\mathbf{r}_0)|^2 + |u_0(\mathbf{r}-\mathbf{r}_0)|^2], \tag{42a}$$
$$\Gamma^1(\mathbf{r},\mathbf{r}') = N\kappa\delta(\mathbf{r}-\mathbf{r}')[|u_0(\mathbf{r}-\mathbf{r}_0)|^2 - |u_0(\mathbf{r}-\mathbf{r}_0)|^2], \tag{42b}$$
$$\Gamma^2(\mathbf{r},\mathbf{r}') = N^2\kappa^2\{|u_0(\mathbf{r}+\mathbf{r}_0)|^2|u_0(\mathbf{r}'+\mathbf{r}_0)|^2 + |u_0(\mathbf{r}-\mathbf{r}_0)|^2|u_0(\mathbf{r}'-\mathbf{r}_0)|^2$$
$$-2\,\mathrm{Re}[u_0^*(\mathbf{r}+\mathbf{r}_0)u_0(\mathbf{r}-\mathbf{r}_0)u_0^*(\mathbf{r}'-\mathbf{r}_0)u_0(\mathbf{r}'+\mathbf{r}_0)]\}. \tag{42c}$$

The covariance matrix for intensity measurements with the ideal pixelized detector described above is obtained by integrating the coherence function (41) over the pixel area according to

$$\Gamma_{ijkl}[d,\theta,N_p] = \int_{\mathcal{A}_{ij}} d^2\mathbf{r}\int_{\mathcal{A}_{kl}} d^2\mathbf{r}'\Gamma(\mathbf{r},\mathbf{r}') = \Gamma^0_{ijkl} + \gamma\Gamma^1_{ijkl} + \gamma^2\Gamma^2_{ijkl}, \tag{43}$$

with

$$\Gamma^0_{ijkl} = N^2\kappa^2(\Phi_{ij}\Phi_{kl} + \Psi_{ij}\Psi_{kl} + 2\,\mathrm{Re}\{\Xi^*_{ij}\Xi_{kl}\}) + \delta_{ik}\delta_{jl}N\kappa(\Phi_{ij}+\Psi_{ij}), \tag{44a}$$
$$\Gamma^1_{ijkl} = \delta_{ik}\delta_{jl}N\kappa(\Phi_{ij}-\Psi_{ij}), \tag{44b}$$
$$\Gamma^2_{ijkl} = N^2\kappa^2(\Phi_{ij}\Phi_{kl} + \Psi_{ij}\Psi_{kl} - 2\,\mathrm{Re}\{\Xi^*_{ij}\Xi_{kl}\}), \tag{44c}$$

where we have introduced

$$\Xi_{ij} = \int_{\mathcal{A}_{ij}} u_0^*(\mathbf{r}+\mathbf{r}_0)u_0(\mathbf{r}-\mathbf{r}_0)d^2\mathbf{r}. \tag{45}$$

The derivative vector is obtained from the mean intensity per pixel (37), where the dependence on the parameter d is only contained in the functions Φ_{ij} and Ψ_{ij} (38),

$$D_{ij}[d,\theta,N_p] = N\kappa(1+\gamma)\frac{\partial\Phi_{ij}}{\partial d} + N\kappa(1-\gamma)\frac{\partial\Psi_{ij}}{\partial d}. \tag{46}$$

For a Gaussian PSF, the integrals (38) and (45) can be evaluated analytically and result in

$$\Phi_{ij} = \tfrac{1}{4}\{\mathrm{erf}[\mathcal{C}_-(x_-)] - \mathrm{erf}[\mathcal{C}_-(x_i^+)]\}\{\mathrm{erf}[\mathcal{S}_-(y_i^-)] - \mathrm{erf}[\mathcal{S}_-(y_i^+)]\}, \tag{47a}$$
$$\Psi_{ij} = \tfrac{1}{4}\{\mathrm{erf}[\mathcal{C}_+(x_i^-)] - \mathrm{erf}[\mathcal{C}_+(x_i^+)]\}\{\mathrm{erf}[\mathcal{S}_+(y_i^-)] - \mathrm{erf}[\mathcal{S}_+(y_i^+)]\}, \tag{47b}$$
$$\Xi_{ij} = \tfrac{1}{4}e^{-d^2/2w^2}(\mathrm{erf}[x_i^-] - \mathrm{erf}[x_i^+])(\mathrm{erf}[y_i^-] - \mathrm{erf}[y_i^+]). \tag{47c}$$

In Eqs. (47) we have defined

$$\mathcal{C}_\pm(x) = \frac{d\cos\theta}{\sqrt{2}w} \pm x, \tag{48a}$$
$$\mathcal{S}_\pm(y) = \frac{d\sin\theta}{\sqrt{2}w} \pm y, \tag{48b}$$

and $x_i^{\pm} = \sqrt{2}(x_i \pm x_p)/w$ and $y_i^{\pm} = \sqrt{2}(y_i \pm y_p)/w$. Analogously, for the derivatives $\frac{\partial \Phi_{ij}}{\partial d}$ and $\frac{\partial \Psi_{ij}}{\partial d}$ we obtain

$$\frac{\partial \Phi_{ij}}{\partial d} = \frac{\cos\theta}{2\sqrt{2\pi}\,w}(e^{-\mathcal{C}_-^2(x_i^-)} - e^{-\mathcal{C}_-^2(x_i^+)})\{\mathrm{erf}[S_-(y_i^-)] - \mathrm{erf}[S_-(y_i^+)]\} + \frac{\sin\theta}{2\sqrt{2\pi}\,w}(e^{-\mathcal{S}_-^2(y_i^-)} - e^{-\mathcal{S}_-^2(y_i^+)})\{\mathrm{erf}[C_-(x_i^-)] - \mathrm{erf}[C_-(x_i^+)]\}, \tag{49a}$$

$$\frac{\partial \Phi_{ij}}{\partial d} = \frac{\cos\theta}{2\sqrt{2\pi}\,w}(e^{-\mathcal{C}_+^2(x_i^-)} - e^{-\mathcal{C}_+^2(x_i^+)})\{\mathrm{erf}[\mathcal{S}_+(y_i^-)] - \mathrm{erf}[\mathcal{S}_+(y_i^+)]\} + \frac{\sin\theta}{2\sqrt{2\pi}\,w}(e^{-\mathcal{S}_+^2(y_i^-)} - e^{-\mathcal{S}_+^2(y_i^+)})\{\mathrm{erf}[\mathcal{C}_+(x_i^-)] - \mathrm{erf}[\mathcal{C}_+(x_i^+)]\}. \tag{49b}$$

The covariance matrix (43) and the derivative vector (46), together with Eqs. (47) and (49), generalize the results in the Supplemental Material of [25] to arbitrary two-dimensional arrangements and unequal brightnesses of the two sources. Let us now note that the covariance matrix (43) can be rewritten in the form

$$\Gamma[d, \theta, N_P] = I + UU^T, \tag{50}$$

where I is a diagonal matrix whose elements are given by the mean intensities per pixels $I_{k=ij}$ as reported in Eq. (37) and $U = (v^{(1)}, v^{(2)}, v^{(3)})^T$ is $N_p^2 \times 3$ matrix whose columns are given by the vectors

$$v^{(1)}_{k=ij} = N\kappa\sqrt{1+\gamma^2}\,\Phi_{ij}, \tag{51a}$$

$$v^{(2)}_{k=ij} = N\kappa\sqrt{1+\gamma^2}\,\Psi_{ij}, \tag{51b}$$

$$v^{(3)}_{k=ij} = N\kappa\sqrt{2(1-\gamma^2)}\,\Xi_{ij}, \tag{51c}$$

where we used that, according to Eqs. (47), the functions Φ_{ij}, Ψ_{ij}, and Ξ_{ij} are all real. Then, using the Woodbury matrix identity [40], we can write the inverse of the covariance matrix as

$$\Gamma^{-1}_{ijkl}[d, \theta, N_P] = \frac{\delta_{ik}\delta_{jl}}{I_{ij}} - \Lambda_{ijkl}, \tag{52}$$

where the correction $\Lambda = U(\mathbb{1}_3 + U^TU)^{-1}U^T$, with $\mathbb{1}_3$ the 3×3 matrix identity, can be evaluated numerically. In our numerical calculations, independently of the mean photon number of the two thermal sources, we did not observe appreciable changes in the behavior of the measurement sensitivity $M[d, \theta, N_p]$ for $N_p \gtrsim 50 \times 50$. Accordingly, $N_p = 50 \times 50$ was used to obtain all direct-imaging curves reported in this work.

The representation (52) allows us to write the optimal direct-imaging sensitivity as

$$M[d, \theta, N_p] = \sum_{ij=1}^{N_p} \frac{1}{I_{ij}}\left(\frac{\partial I_{ij}}{\partial d}\right)^2 - \sum_{ijkl=1}^{N_p} D_{ij}[d, \theta, N_p]\Lambda_{ijkl}D_{kl}[d, \theta, N_p], \tag{53}$$

where the first term is of order $N\kappa$, while the second one is of order $(N\kappa)^2$ and negative. As a consequence, the latter tends to reduce the relative sensitivity when increasing the source brightness. This behavior can be observed in Fig. 5, where the direct-imaging sensitivity $M[d, \theta, N_p]$ is plotted for different mean photon numbers of the sources.

The limit of continuous direct imaging is obtained for $N_p \to \infty$. In this limit, the first term in Eq. (53) becomes

$$\mathcal{F}_{\mathrm{DI}}[d, \theta] = \int \frac{1}{I(\mathbf{r})}\left(\frac{\partial I(\mathbf{r})}{\partial d}\right)^2 d^2\mathbf{r}, \tag{54}$$

which coincides with the direct-imaging Fisher information calculated assuming Poissonian sources in the $N\kappa \ll 1$ regime [18]. Accordingly, when the number of photons in the image plane is low ($N\kappa \ll 1$), when the second term in Eq. (53) is negligible, the optimized moment-based sensitivity saturates the Cramér-Rao bound for continuous direct imaging.

In Fig. 5 we see that the reduction of the measurement sensitivity due the Λ term in Eq. (53) is relevant only for intermediate source separations. In fact, the behavior of the measurement sensitivity $M[d, \theta, N_p]$ [Eq. (53)] for small distances between the sources is dominated by the diagonal part of the covariance matrix (43). Accordingly, in this regime, we can approximate the direct-imaging sensitivity by considering an expansion of the integral (54) for $d/2w \ll 1$. Such an

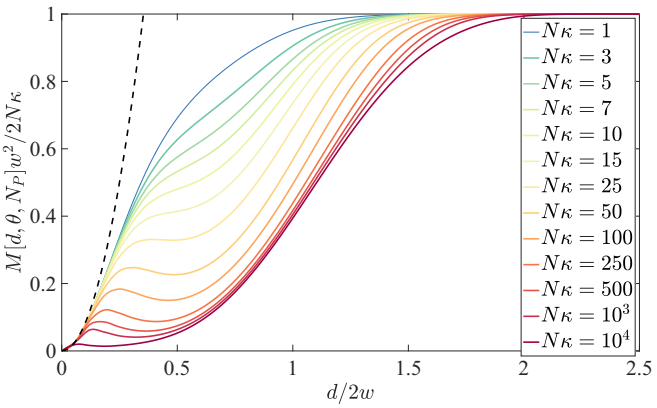

FIG. 5. Direct-imaging sensitivity $M[d, \theta, N_p]$ for two equally bright thermal sources ($\gamma = 0$) with a different mean number of received photons $N\kappa$. We considered an alignment angle $\theta = \pi/4$ and a square detector of side $6w$ divided into $N_p = 50 \times 50$ pixels. The black dashed line represents the analytical approximation for small source separation given by Eq. (55).

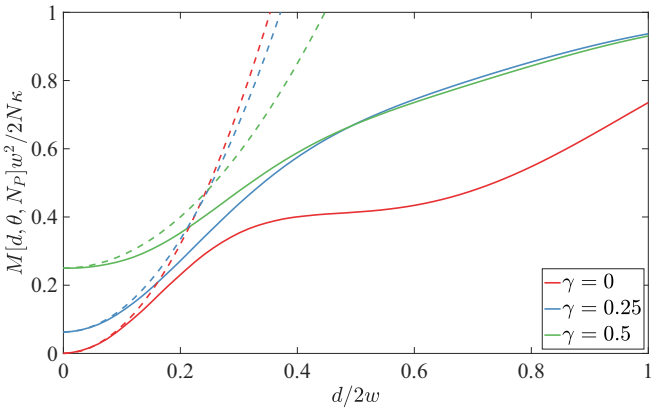

FIG. 6. Direct-imaging sensitivity $M[d,\theta,N_p]$ for thermal sources with fixed total mean photon number $N\kappa = 1.5$ and different brightness imbalances: $\gamma = 0$ [red (bottom) line], $\gamma = 0.25$ [blue (middle) line], and $\gamma = 0.5$ [green (top) line]. We considered an alignment angle $\theta = \pi/4$ and a square detector of length $6w$ divided into $N_p = 50 \times 50$ pixels. The dashed lines represent the analytical approximation for small source separation given by Eq. (55).

expansion can be performed analytically and results in

$$M_{\mathrm{DI}}[d,\theta] = \frac{2N\kappa}{w^2}[\gamma^2 + 4x^2(2 - 5\gamma^2 + 3\gamma^4)] + O(x^4). \tag{55}$$

The approximation (55) is compared (for $\gamma = 0$) with the exact numerical results for different mean photon numbers in Fig. 5. It is interesting to note that for $\gamma \neq 0$, Eq. (55) does not vanish for $d = 0$ (see Fig. 6). Accordingly, an asymmetry in the intensity distribution between the two sources eases the separation estimation. In particular, we see that for extreme intensity imbalances $\gamma \to 1$, the direct-imaging sensitivity for small separations tends to its maximal value $M_{\mathrm{DI}}[d,\theta] \to 2N\kappa/w^2$. We can understand this result by considering that in the $\gamma \to 1$ limit the estimation of the source separation becomes the localization of a single source.

VI. IDEAL DEMULTIPLEXING

In this section we present analytical expressions for the optimal moment-based sensitivity, according to Eq. (15), attainable with ideal demultiplexing measurements and for the optimal observable achieving it, as given by Eq. (16).

A. Analytical inversion of the covariance matrix

We now focus on the case of ideal intensity measurements of $K = (Q+1)^2$ HG modes $u_{nm}(\mathbf{r})$ with $0 \leqslant n, m \leqslant Q$. In this case, the overlap functions $\beta_{nm}(\pm\mathbf{r}_0)$ are obtained by setting $x_s = y_s = 0$ in Eq. (28), which implies the symmetry

$$\beta_{nm}(-\mathbf{r}_0) = (-1)^{n+m}\beta_{nm}(\mathbf{r}_0). \tag{56}$$

Equation (56) causes the mean photon number in the measurement modes N_k [see Eq. (22)] to be independent of the photon-number imbalance γ:

$$N_{nm} = 2N\kappa\beta_{nm}^2(\mathbf{r}_0). \tag{57}$$

Substituting Eq. (56) into Eq. (23) also simplifies the covariance matrix, which becomes

$$\Gamma = \mathrm{diag}(\boldsymbol{\nu}) + \boldsymbol{\xi}\boldsymbol{\xi}^T/2 + \boldsymbol{\zeta}\boldsymbol{\zeta}^T/2, \tag{58}$$

with $\mathrm{diag}(\boldsymbol{\nu})$ a diagonal $K \times K$ matrix whose elements are given by the vector $\boldsymbol{\nu}_{k=(n,m)} = N_{nm}$ [see Eq. (57)], while the other two terms are outer products of vectors with elements $\boldsymbol{\xi}_{k=(n,m)} = \sqrt{1+\gamma^2}N_{nm}$ and $\boldsymbol{\zeta}_{k=(n,m)} = (-1)^{n+m}\sqrt{1-\gamma^2}N_{nm}$. This particular form of the covariance matrix allows for its analytical inversion by two successive applications of the Sherman-Morrison formula [40]. Accordingly, we obtain

$$\Gamma^{-1}_{mnm'n'} = \frac{\delta_{mm'}\delta_{nn'}}{2N\kappa\beta_{mn}^2(\mathbf{r}_0)} - \frac{(-1)^{m+n+m'+n'}A_+ - B[(-1)^{m+n} + (-1)^{m'+n'}] + A_-}{A_+A_- - B^2}, \tag{59}$$

with

$$A_\pm = \frac{2}{1\pm\gamma^2} + 2N\kappa\sum_{mn=0}^{Q}\beta_{mn}^2(\mathbf{r}_0), \tag{60a}$$

$$B = 2N\kappa\sum_{mn=0}^{Q}(-1)^{m+n}\beta_{mn}^2(\mathbf{r}_0). \tag{60b}$$

Equation (59) is valid only under the assumption that the covariance matrix is invertible. According to the Sherman-Morrison formula, this is the case if and only if $\mathrm{diag}(\boldsymbol{\nu})$ is invertible, i.e., when $\beta_{nm}(\mathbf{r}_0) \neq 0$ for every n and m. Looking at Eq. (28), for $x_s = y_s = 0$, we see that this condition is not satisfied for $d = 0$ or $\theta = k\pi/2$, with $k = 0, 1, 2, \ldots$. Let us first focus on $d = 0$. In this case, $\beta_{nm}(\mathbf{r}_0) = 0$ for all n and m. However, this case corresponds to having only one source and the estimation problem under study is not well posed. On the other hand, for $\theta = k\pi/2$ with k even (odd), we have $\beta_{nm}(\mathbf{r}_0) = 0\,\forall\, n(m) \neq 0$. We can therefore make the covariance matrix (58) invertible by removing all modes with $n(m) = 0$ from our measurement basis and carry out the estimation within this smaller set of modes. This fact has a clear physical interpretation: The HG modes [see Eq. (26)] have the form $u_{nm}(x,y) = h_n(x)h_m(y)$; accordingly, when θ is an odd (even) multiple of $\pi/2$ the two sources are aligned with the x (y) axis and the only relevant modes are $u_{n0}(x,y) = h_n(x)$ [$u_{0m}(x,y) = h_m(y)$]. Equation (59) yields the correct results also for $\theta = k\pi/2$ by extracting only the relevant matrix elements.

B. Measurement sensitivity

The measurement sensitivity $M[d,\theta,\hat{\mathbf{N}}]$ [Eq. (15)], which is the maximal sensitivity achievable by any linear combination of the mean photon numbers in the measurement modes N_k [29], can be obtained by combining the inverse covariance matrix (59) with the derivative vector

$$D_{mn}[d,\theta,\hat{\mathbf{N}}] = \frac{2\kappa N}{wx}(m+n-x^2)\beta_{mn}^2(\mathbf{r}_0), \tag{61}$$

with $x = d/2w$, and is given by

$$M[d,\theta,\hat{\mathbf{N}}] = \frac{2N\kappa}{w^2}[F - (2N\kappa)(\delta_1+\delta_2+\delta_3)], \tag{62}$$

where

$$F = \sum_{mn=0}^{Q} \frac{(n+m-x^2)^2}{x^2}\beta^2_{mn}(\mathbf{r}_0), \tag{63a}$$

$$\delta_1 = \frac{A_+}{A_+A_- - B^2}\mathcal{S}_1^2, \tag{63b}$$

$$\delta_2 = -\frac{2B}{A_+A_- - B^2}\mathcal{S}_1\mathcal{S}_2, \tag{63c}$$

$$\delta_3 = \frac{A_-}{A_+A_- - B^2}\mathcal{S}_2^2, \tag{63d}$$

with

$$\mathcal{S}_1 = \sum_{p,q=0}^{Q} (-1)^{p+q}\frac{p+q-x^2}{x}\beta^2_{pq}(\mathbf{r}_0), \tag{64a}$$

$$\mathcal{S}_2 = \sum_{p,q=0}^{Q} \frac{p+q-x^2}{x}\beta^2_{pq}(\mathbf{r}_0). \tag{64b}$$

We immediately see that, when the number of received photons is low ($N\kappa \ll 1$), the sensitivity $M[d,\theta,\hat{\mathbf{N}}]$ is dominated by the F term. This term coincides with the Fisher information calculated by assuming Poissonian sources in the $N\kappa \ll 1$ regime [compare Eq. (63a) with Eq. (21) in the Supplemental Material of [35]]. Moreover, F does not depend on γ, i.e., the number of received photons is low, the sensitivity in the estimation of d does not depend on the relative brightness of the two sources.

For higher received photon numbers, the terms of order $(N\kappa)^2$ become relevant. Given that the functions δ_1, δ_2, and δ_3 are positive, the quadratic terms in the photon number always reduce the measurement sensitivity. Moreover, they depend on γ through $A_\pm$. Accordingly, the loss of sensitivity appearing for higher received photon numbers depends on how the photons are distributed between the two sources.

Let us now consider the limiting case where the full HG basis is measured, i.e., $Q \to \infty$. In this limit, $F \to 1$, δ_2 and δ_3 vanish, and δ_1 remains finite and provides a correction to the measurement sensitivity, which takes the form

$$\begin{aligned} M_{\rm inf}[d,\theta,\hat{\mathbf{N}}] &= \lim_{Q\to\infty} M[d,\theta,\hat{\mathbf{N}}] \\ &= \frac{2N\kappa}{w^2}\left[1 - \frac{4(1-\gamma^2)N\kappa e^{-4x^2}x^2[(\gamma^2+1)N\kappa+1]}{(1-\gamma^4)N^2\kappa^2(1-e^{-4x^2})+2N\kappa+1}\right]. \end{aligned} \tag{65}$$

The equal-brightness case is obtained by setting $\gamma = 0$ in Eq. (65), which results in

$$M_{\rm inf}[d,\theta] = \frac{2N\kappa}{w^2} - \frac{8N^2\kappa^2x^2(N\kappa+1)e^{-4x^2}}{w^2[(N\kappa+1)^2 - N^2\kappa^2e^{-4x^2}]}, \tag{66}$$

which coincides with the quantum Fisher information for two equally bright thermal sources calculated in Refs. [25,26].

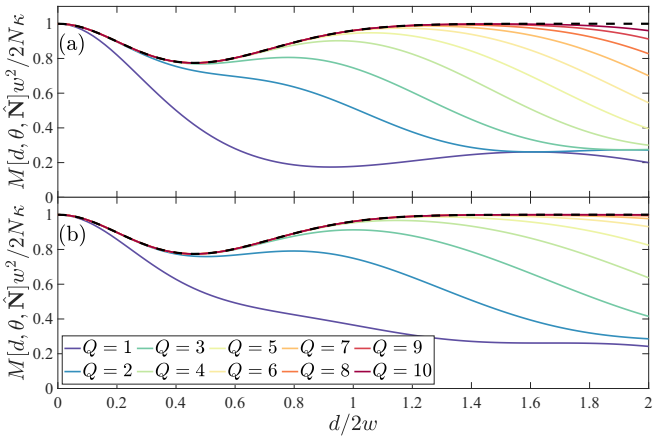

FIG. 7. Measurement sensitivity (62) for equally bright sources ($\gamma = 0$) and for different values of Q (solid lines) and alignment angle (a) $\theta = 0$ and (b) $\theta = \pi/4$. The black dashed line represents the quantum Fisher information (66). We assumed $N\kappa = 1.5$ for both panels.

Accordingly, our estimation strategy saturates the quantum Cramér-Rao bound in the asymptotic limit. The second term in Eqs. (65) and (66) reduces the estimation sensitivity for intermediate separations, i.e., a dip appears in Figs. 7 and 8. This dip gets deeper when increasing the mean photon number $N\kappa$ (see also [25,26]). Physically, this is due to the fact that for higher brightness of the sources there is a non-negligible probability of detecting multiple photons in the same modes which reduces the information that can be extracted from each photon.

The behavior of the measurement sensitivity (62) for $\gamma = 0$ and different values of Q is compared with the quantum Fisher information (66) in Fig. 7. While the asymptotic expression (66) is independent of the alignment angle θ, this is not true at finite Q. In particular, $M[d,\theta,\hat{\mathbf{N}}]$ is minimal when the sources are aligned with the axes ($\theta = 0, \pi/2$) and maximal when the sources are aligned along a bisector ($\theta = \pi/4, 3\pi/4$). This behavior can be observed in Fig. 7. In Fig. 7(a) ($\theta = 0$), for large separations d, the measurement sensitivity is clearly distinguishable from the quantum Fisher information for all

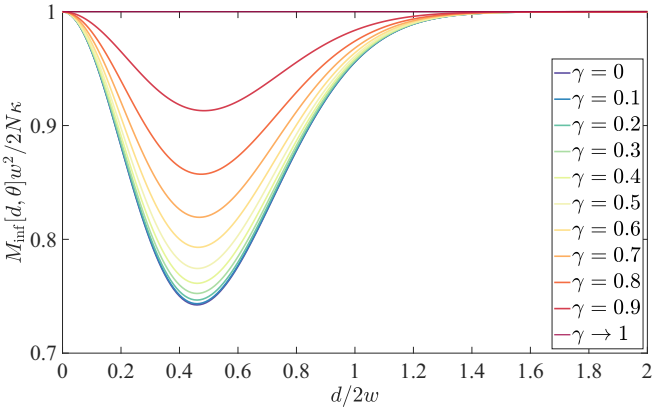

FIG. 8. Asymptotic measurement sensitivity (65) for fixed $N\kappa = 1.5$ and different brightness imbalances γ.

values of Q considered. On the other hand, in Fig. 7(b) ($\theta = \pi/4$), the measurement sensitivity cannot be discerned from the quantum Fisher information for $Q \geqslant 9$. Similar behavior was reported for the Fisher information, for the regime of low photon numbers in the image plane, in the Supplemental Material of [35]. The optimality of $\theta = \pi/4$ is due to the fact that we have assumed that we have a limited and equal number of modes available in each direction. Under this assumption, aligning the sources in one direction, we can use only Q of the Q^2 available modes. This limits the resolution especially for large separations. On the other hand, aligning the sources along the bisector is equivalent to rotating the basis in order to have all Q^2 available modes along the sources axis, which provides the optimal resolution. The asymptotic behavior ($Q \to \infty$) of the measurement sensitivity (65) for different brightnesses of the two sources ($\gamma \neq 0$) is illustrated in Fig. 8. We can see that the dip at intermediate separations gets shallower when increasing γ and it disappears in the limit $\gamma \to 1$. Accordingly, as observed for direct imaging in Sec. V, an intensity imbalance between the sources makes it easier to estimate their separation.

C. Measurements coefficients

Combining Eqs. (59) and (61) according to Eq. (16), we obtain the expression for the measurement coefficients (dashed lines in Fig. 11 in Sec. VII)

$$\frac{w}{\eta} m_{ij} = \frac{i+j-x^2}{x} - \frac{2\kappa N}{A_+ A_- - B^2} \times \{[(-1)^{i+j} A_+ - B]\mathcal{S}_+ - [(-1)^{i+j} B - A_-]\mathcal{S}_-\} \tag{67}$$

with the normalization constant

$$\eta[d, \theta, Q] = \left(\sum_{i,j=0}^{Q} m_{ij}^2 \right)^{-1/2}. \tag{68}$$

From Eq. (67) we note that the coefficients of the optimal observable only depend on the sum of the indices $i + j$. Accordingly, independently on the sources' orientations or their relative intensities, the method of moments prescribes measurement, with the same weight, the photon number in all HG modes of the same order. In other words, in the absence of noise, it is optimal to measure a circularly symmetric intensity distribution, which achieves the same sensitivity for each alignment angle θ.

The above observation does not apply when the sources are aligned along the x (y) axis. In these cases, as discussed in Sec. VI A, one needs only to measure modes with $i(j) = 0$. The correct coefficients can be still obtained from Eq. (67) by setting $i = 0$ ($j = 0$). Further comment on the behavior of the measurement coefficients (67) is postponed to Sec. VII, where we will compare them with their counterpart in the presence of noise.

VII. NOISY DEMULTIPLEXING

A. Measurement sensitivity

Let us start our discussion on the sensitivity of our estimation strategy in the presence of noise by commenting on its optimality. We have already observed in Sec. VI that for ideal demultiplexing our method approaches the quantum Cramér-Rao bound for arbitrary source separations and brightnesses, when a sufficiently large number of modes is measured. Moreover, when the number of received photons is low ($N\kappa \ll 1$), the covariance matrix (23) is dominated by its diagonal terms and we have

$$M[d, \theta, \hat{\mathbf{N}}] \approx \sum_{k=1} \frac{1}{N_k} \left(\frac{\partial N_k}{\partial d} \right)^2. \tag{69}$$

As discussed in Sec. IV B, crosstalk and misalignment only affect the overlap functions $f_{\pm,k}$, while dark counts modify the diagonal of the covariance matrix. As a consequence, when all noise sources are considered, Eq. (69) remains valid if we replace the mean photon number N_k with the mean photon number plus noise: $N_k' = N_k + N_k^{\mathrm{dc}}$. Equation (69) is equal to the Fisher information for demultiplexing calculated by assuming Poissonian sources in the low-brightness regime ($N\kappa \ll 1$) [18]. Accordingly, for $N\kappa \ll 1$ our estimation strategy saturates the Cramér-Rao bound even in the presence of noise. For thermal sources of arbitrary brightnesses and demultiplexing (noisy or not) in a finite number K of spatial modes, the Fisher information is unknown. In this regime, the sensitivity of our method provides a lower bound for the Fisher information. The latter can be saturated in practice by applying the simple estimation strategy presented in Sec. III to the optimal linear combination of mean photon numbers in the measurement modes discussed in Sec. VII B. The measurement sensitivity $M[d, \theta, \hat{\mathbf{N}}]$ is presented in Fig. 9 for different numbers of measured modes and different noise levels. We can see that all noise sources reduce the sensitivity with

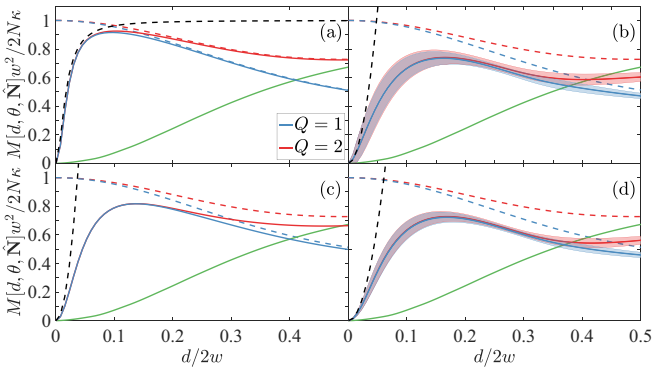

FIG. 9. Measurement sensitivity $M[d, \theta, \hat{\mathbf{N}}]$ for demultiplexing into HG modes $u_{nm}(\mathbf{r})$ with $n, m \leqslant Q$ [$K = (Q+1)^2$] with $Q = 1$ [red (upper) solid line] and $Q = 2$ [blue (middle) solid line] with different noise sources: (a) misalignment ($d_s/2w = 0.01$ and $\theta_s = \pi/4$), (b) crosstalk ($\langle |c_{ij}|^2 \rangle = 0.0017$), (c) dark counts ($\sigma_k = 0.001\,\forall k$), and (d) all three noise sources at the same time. Black dashed lines starting from the zero represent short-distance approximations; in particular, we used (a) Eq. (77), (b) Eq. (76) with $|t|^2 = \langle |c_{ij}|^2 \rangle$ and $\sigma = 0$, (c) Eq. (75) with $\sigma = 0.001$, and (d) Eq. (76) with $|t|^2 = \langle |c_{ij}|^2 \rangle$ and $\sigma = 0.001$. Dashed lines starting from one show the results for ideal measurements (62) for $Q = 1$ [red (upper) line] and $Q = 2$ [blue (lower) line]. The green (lower) solid line describes direct-imaging results (53). For all plots, we assumed $N\kappa = 1.5$, $\theta = \pi/4$, and $\gamma = 0$.

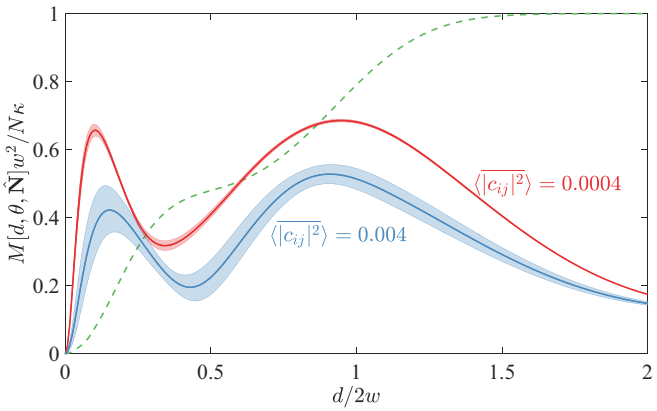

FIG. 10. Demultiplexing sensitivity $M[d,\theta,\hat{\mathbf{N}}]$ in the presence of dark counts $\sigma_k = 0.001\,\forall\, k$, misalignment $d_s/2w = 0.01$ and $\theta_s = \pi/4$, and two different crosstalk levels $\overline{\langle|c_{ij}|^2\rangle} = 0.0004$ [red (upper) solid line] and $\overline{\langle|c_{ij}|^2\rangle} = 0.004$ [blue (lower) solid line]. Solid lines and bands represent means and standard deviations computed from 500 realizations of the crosstalk matrices. The green dashed line represents the ideal direct-imaging sensitivity. All curves correspond to $\theta = \pi/4$, $N\kappa = 10$, and $\gamma = 0$.

respect to its ideal value (62) (red and blue dashed curves in Fig. 9). In particular, we have that for equally bright sources, $M[d,\theta,\hat{\mathbf{N}}]$ vanishes for $d \to 0$ and accordingly it gets harder to resolve small source separations.

Despite this loss of sensitivity for small separations, even when all noise sources are considered at the same time, as in Fig. 9(d), demultiplexing outperforms ideal direct imaging (green lines). The regime where demultiplexing provides an advantage over direct imaging significantly depends on the brightnesses of the sources and the noise levels. In particular, for large mean photon numbers and low noise, there are multiple crossings between the demultiplexing and the ideal direct-imaging curves. To illustrate this behavior, in Fig. 10 we plot, together with the ideal direct-imaging curve (green), two curves corresponding to fixed misalignment and dark count levels but different crosstalk strengths for $N\kappa = 10$. We see that at low crosstalk levels [red (upper) solid curve in Fig. 10] the demultiplexing sensitivity crosses the ideal direct-imaging curve three times. Therefore, in the low-noise regime, there are regions of larger separations where demultiplexing outperforms direct imaging. On the other hand, higher-noise levels (blue curve in Fig. 10) reduce the demultiplexing sensitivity and cancel the larger-distance region where demultiplexing outperforms direct imaging. In Fig. 3 of [28], we investigated how the minimal separation at which the direct-imaging sensitivity crosses the demultiplexing sensitivity varies with the noise level for different source brightnesses.

To better understand how the different noise sources affect the sensitivity, we now perform a perturbative expansion of $M[d,\theta,\hat{\mathbf{N}}]$ for small separations ($x = d/2w \ll 1$). For simplicity, we focus on the case of equally bright thermal sources ($\gamma = 0$). Moreover, in practically relevant scenarios, we can assume that misalignment is of the order of the separation or smaller. Therefore, we assume $x = d/2w \ll 1$ and $x_s = d_s/2w \ll 1$ to be of order $a = \|(x, x_s)\| \ll 1$. From Eq. (28) we then obtain

$$\beta_{00}(\mathbf{r}_0 - \mathbf{r}_s) = 1 - \frac{x^2}{2} - 2x_s^2 - 2xx_s\cos(\theta - \theta_s) + O(a^3), \tag{70a}$$

$$\beta_{10}(\mathbf{r}_0 - \mathbf{r}_s) = 2x_s\cos\theta_s + x\cos\theta + O(a^3), \tag{70b}$$

$$\beta_{01}(\mathbf{r}_0 - \mathbf{r}_s) = 2x_s\sin\theta_s + x\sin\theta + O(a^3), \tag{70c}$$

$$\beta_{11}(\mathbf{r}_0 - \mathbf{r}_s) = (2x_s\sin\theta_s + x\sin\theta) \times (2x_s\cos\theta_s + x\cos\theta) + O(a^3), \tag{70d}$$

$$\beta_{20}(\mathbf{r}_0 - \mathbf{r}_s) = \frac{(2x_s\cos\theta_s + x\cos\theta)^2}{\sqrt{2}} + O(a^3), \tag{70e}$$

$$\beta_{02}(\mathbf{r}_0 - \mathbf{r}_s) = \frac{(2x_s\sin\theta_s + x\sin\theta)^2}{\sqrt{2}} + O(a^3), \tag{70f}$$

$$\beta_{nm}(\mathbf{r}_0 - \mathbf{r}_s) = O(a^3), \quad n + m \geqslant 3. \tag{70g}$$

Combining Eqs. (29) and (25) with the expansions (70), we get the expression for the derivative vector

$$\begin{aligned}\frac{w}{2N\kappa}D[d,\theta,\hat{\mathbf{N}}]_{nm} = x[&-2|c_{nm,00}|^2 + \cos^2\theta(2|c_{nm,10}|^2 + \sqrt{2}c_{nm,20}c^*_{nm,00} + \sqrt{2}c_{nm,00}c^*_{nm,20})\\ &+ \sin^2\theta(2|c_{nm,01}|^2 + \sqrt{2}c_{nm,02}c^*_{nm,00} + \sqrt{2}c_{nm,00}c^*_{nm,02})\\ &+ \sin(2\theta)(c_{nm,11}c^*_{nm,00} + c_{nm,10}c^*_{nm,01} + c_{nm,01}c^*_{nm,10} + c_{nm,00}c^*_{nm,11})] + O(a^2),\end{aligned} \tag{71}$$

where we note that up to first order in a, misalignment does not affect the derivative vector.

Let us now consider a generic weak crosstalk distribution, which corresponds to setting $\epsilon \ll 1$ in Eq. (30). In fact, for $\epsilon \ll 1$, we can write the crosstalk matrix as $C(\epsilon) \approx \mathbf{1} - i\epsilon G$, where the elements of the matrix G are of order unity. Accordingly, the off-diagonal elements of the crosstalk matrix, $c_{nm,kl}$ with $n \neq k$ and $m \neq l$, are of order ϵ, with $0 < \epsilon \ll 1$. Under this assumption, we can restrict ourselves to the smallest square covariance matrix containing all terms of order ϵ. Following this prescription, we obtain the 3×3 matrix

$$\Gamma' = \begin{pmatrix}\Gamma_{00,00} & \Gamma_{00,01} & \Gamma_{00,10}\\ \Gamma_{01,00} & \Gamma_{01,01} & \Gamma_{01,10}\\ \Gamma_{10,00} & \Gamma_{10,01} & \Gamma_{10,10}\end{pmatrix} + \begin{pmatrix}\Sigma & 0 & 0\\ 0 & \Sigma & 0\\ 0 & 0 & \Sigma\end{pmatrix}, \tag{72}$$

with $\Sigma = N^{\mathrm{dc}}(N^{\mathrm{dc}} + 1)$ the dark-count term, which we assume to be weak and the same for all modes. This truncated covariance matrix contains terms up to order ϵ^3. In particular, the leading order in ϵ is ϵ^0 for $\Gamma_{00,00}$; ϵ for $\Gamma_{01,00}$, $\Gamma_{10,00}$, $\Gamma_{00,01}$, $\Gamma_{00,10}$, $\Gamma_{10,10}$, and $\Gamma_{01,01}$; and ϵ^3 for $\Gamma_{10,01}$ and $\Gamma_{01,10}$.

The inverse of the 3×3 matrix in Eq. (72) can be obtained analytically and used to determine the measurement sensitivity, which results in

$$M[d,\theta,\hat{\mathbf{N}}] = \frac{2N\kappa}{w^2}\left(\frac{|c_{00,00}|^4}{2N\kappa(|c_{00,00}|^4+\sigma^2)+|c_{00,00}|^2+\sigma} + \frac{\sin^4\theta|c_{01,01}|^4+\cos^4\theta|c_{10,10}|^4}{2N\kappa\sigma^2+\sigma} + O(\epsilon)\right)x^2 + O(a^3), \tag{73}$$

where, under the above assumption that detection noise is the same in all modes, we have $\sigma = N^{\mathrm{dc}}/2N\kappa$. In the above expansion, we considered σ of order zero in ϵ. Accordingly, in Eq. (73), the behavior of the sensitivity for small x is dominated by dark counts (only diagonal terms of the crosstalk matrix appear).

Another relevant scenario might be the one where crosstalk and dark counts give contributions of the same order to the photon counts N_{nm} in a given HG mode, i.e., when $2N\kappa|c_{ij}|^2 \sim N^{\mathrm{dc}}$, or, in other words, when $\sigma = N^{\mathrm{dc}}/2N\kappa$ is of order ϵ^2. In this case, the measurement sensitivity would be approximated by

$$M[d,\theta,\hat{\mathbf{N}}] = \frac{2N\kappa}{w^2}\left(\frac{\sin^4\theta|c_{01,01}|^4}{|c_{01,00}|^2+\sigma} + \frac{\cos^4\theta|c_{10,10}|^4}{|c_{10,00}|^2+\sigma} + O(\epsilon^{-1})\right)x^2 + O(a^3). \tag{74}$$

In this case, the probabilities of scattering from the first order modes u_{01} and u_{10} to the fundamental mode u_{00} enter in the behavior of the measurement sensitivity for small x. Setting $\sigma = 0$ (corresponding to no dark counts $N^{\mathrm{dc}} = 0$) in Eq. (74), we obtain the leading-order expansion of $M[d,\theta,\hat{\mathbf{N}}]$ when only crosstalk is present (see also the Supplemental Material of [35]).

Equations (73) and (74) can be simplified by considering a uniform crosstalk model [35], namely, by setting all diagonal entries of the crosstalk matrix to t and all the off-diagonal ones to r such that $|t|^2 + (D-1)|r|^2 = 1$. For weak crosstalk, we have in addition $|t|^2 \approx 1$ and $|r|^2 \ll 1$. Accordingly, Eqs. (73) and (74) become, respectively,

$$M_{\mathrm{dc}}[d,\theta,\hat{\mathbf{N}}] \approx \frac{2N\kappa}{w^2}\left(\frac{\cos(4\theta)+3}{8N\kappa\sigma^2+4\sigma} + \frac{1}{2N\kappa(\sigma^2+1)+\sigma+1}\right)x^2, \tag{75}$$

$$M_{\mathrm{ct}}[d,\theta,\hat{\mathbf{N}}] \approx \frac{2N\kappa}{w^2}\frac{\cos(4\theta)+3}{4(|r|^2+\sigma)}x^2. \tag{76}$$

We denote by M_{dc} (M_{ct}) the dark-count (crosstalk)-dominated sensitivity. Equation (75) remains valid also in the case when no crosstalk or misalignment is present and we expand to leading order in σ.

Interestingly enough, with the assumption that the misalignment is of the same order as the separation, in both Eqs. (73) and (74) the effect of misalignment is hidden by the other noise sources. However, in Fig. 9(a) we know that misalignment alone also causes the measurement sensitivity to go to zero for $x \to 0$. To illustrate how this happens, we repeat the procedure described above in the absence of crosstalk and dark counts. The measurement sensitivity, at leading order in a, results in

$$M_{\mathrm{mis}}[d,\theta,\hat{\mathbf{N}}] \approx \frac{2N\kappa}{w^2}\left(\frac{\sin^4\theta}{x^2\sin^2\theta+4x_s^2\sin^2\theta_s} + \frac{\cos^4\theta}{x^2\cos^2\theta+4x_s^2\cos^2\theta_s}\right)x^2. \tag{77}$$

The approximations of the measurement sensitivity in the different noise regimes described by Eqs. (73)–(77) are compared with the exact numerical results in Fig. 9. Interestingly, for all noise sources [Eqs. (73)–(77)] the measurement sensitivity goes to zero as x^2 for small separations. We observed, in Eq. (55), that also the direct-imaging sensitivity (for $\gamma = 0$) presents the same scaling: $w^2 M_{\mathrm{DI}}[d,\theta]/2N\kappa \sim 8x^2$. However, for typical noise levels (as reported in Fig. 9), demultiplexing has generally a more favorable coefficient.

B. Measurement coefficients

We now discuss the behavior of the optimal observable that allows us to reach the sensitivity bounds illustrated in the preceding section. For this goal, in Fig. 11 we study the dependence of the coefficients m_{ij} [see Eq. (16)] of the optimal linear combination of intensity measurements $\hat{X}_{\mathbf{m}} = \sum_{ij} m_{ij}N_{ij}$ in the HG modes $u_{ij}(\mathbf{r})$ as a function of the separation d between the two sources for $0 \leqslant i, j \leqslant Q$ with $Q = 1$ [Figs. 11(a)–11(d)] and $Q = 2$ [Fig. 11(e)].

Arguably, the most interesting feature of Fig. 11 is that when d is small compared to the diameter $2w$ of the PSF $u_0(\mathbf{r})$ ($d/2w \lesssim 0.2$), the coefficients m_{ij} depend very weakly on d. Accordingly, in the relevant regime of small separations, a fixed observable can be used to estimate a vast range of parameter values. An exception to this behavior is observed in Fig. 11(a) and in its close-up in Fig. 11(d), where we show the coefficients in the presence of misalignment only. In this case, we observe that for separations of the order of the misalignment, the measurement coefficients present some modulations. In particular, for $d \sim d_s$, the shifted centroid significantly impacts the image decomposition in the $u_{01}(\mathbf{r})$ and $u_{10}(\mathbf{r})$ modes. Accordingly, the coefficients for these modes are slightly depleted, while the coefficient of $u_{11}(\mathbf{r})$ mode is increased. However, when misalignment is combined with other imperfections [Fig. 11(e)] these modulations are washed out by the reduction in signal-to-noise ratio induced by crosstalk and dark counts in higher-order modes.

In Sec. VI we observed that, in the absence of noise, the measurement coefficients m_{ij} only depend on the sum $i+j$ [see Eq. (67)]. In Fig. 11 we show how different noise sources can break this degeneracy. This effect is especially clear for the coefficients m_{01} (blue) and m_{10} (green). In particular, we see that when the center of the measurement basis does not lie on the sources axis [$\theta_s \neq \theta$, dash-dotted lines in Figs. 11(a) and 11(d)] m_{01} and m_{10} deviate in opposite directions from the $\theta_d = \theta = \pi/4$ curve favoring the mode where the signal increases because of the centroid shift. Also crosstalk, which affects randomly the different modes, can remove the coeffi-

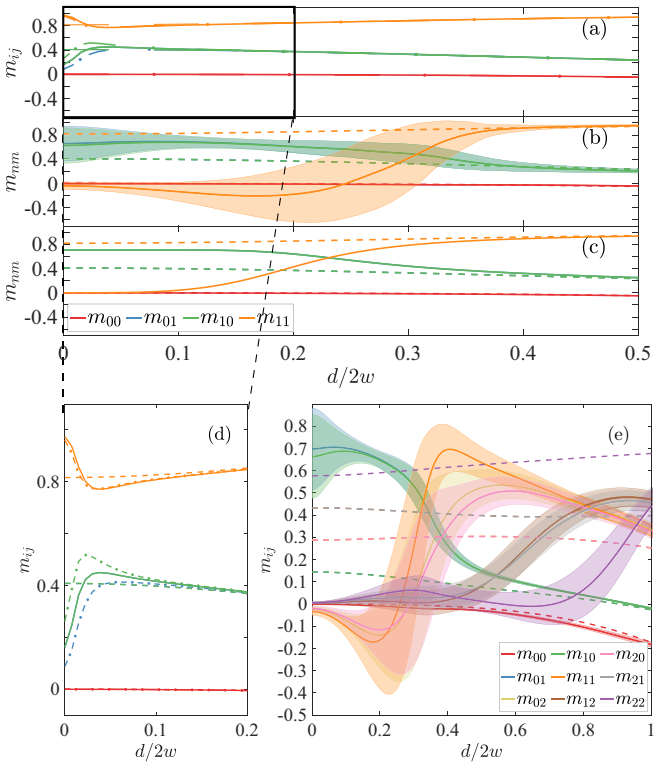

FIG. 11. Dependence on the source separation d of the measurement coefficients m_{ij} for intensity measurements in the HG modes basis $u_{ij}(\mathbf{r})$ with (a)–(d) $0 \leqslant i, j \leqslant 1$ ($K = 4$) and (e) $0 \leqslant i, l \leqslant 2$ ($K = 9$). Different noise sources are considered: (a) and (d) misalignment [$d_s/2w = 0.01$ and $\theta_s = \pi/4$ (solid) and $\theta_s = \pi/3$ (dotted)], (b) crosstalk ($\langle|c_{ij}|^2\rangle = 0.0017$), (c) dark counts ($\sigma = 0.001$), and (e) all three combined ($d_s/2w = 0.01$, $\theta_s = \pi/4$, $\langle|c_{ij}|^2\rangle = 0.0017$, and $\sigma = 0.001$). Dashed lines represents the coefficients in the noiseless case. Solid lines and bands in (b) and (d) represents the mean and one standard deviation computed over 500 crosstalk matrices. All plots correspond to $N\kappa = 1.5$ and $\theta = \pi/4$.

cients' degeneracy, as can be observed in the small difference between the green and blue lines in Fig. 11(b) or in the more evident separation of the yellow and pink curves (m_{02} and m_{20}) in Fig. 11(e). On the other hand, in Fig. 11(c), where the dark-count level is assumed to be the same in all modes, the coefficients m_{01} and m_{10} are perfectly degenerate. Of course, this would have not been the case if the electronic noise level had not been the same in the different modes. Finally, in Fig. 11(e) we notice that in the presence of weak noise, even though the degeneracy is removed, the coefficients m_{ij} still move in groups (green and blue for $i + j = 1$; orange, yellow, and pink for $i + j = 2$; brown and gray for $i + j = 3$; and purple for $i + j = 4$).

From Fig. 11(e) we can also see how the different coefficients change with the separation d between the two sources. First of all, for small separations, the mode u_{00} contains no information on d; accordingly $m_{00} = 0$ for a vast range of separations. Moreover, in the absence of noise [dashed lines Fig. 11(e)], all coefficients $m_{ij>0}$ [Eq. (67)] are different from zero for all values of d and higher-order modes have larger weights. In fact, even though, for small separations, the image of the two sources produces very low signals in the higher-order modes, these signals are noiseless. Accordingly, the optimal observable amplifies these small signals to extract the most information on the parameter d out of them. On the other hand, different noise sources introduce photon counts that contain no information on the separation. Accordingly, in the presence of noise [solid lines Fig. 11(e)], our optimal observable prescribes the measurement of only those modes where the signal-to-noise ratio is high enough. In particular, for small values of d, all coefficients are zero (or slightly negative) except m_{01} and m_{10}. Therefore, demultiplexing into the HG modes $u_{01}(\mathbf{r})$ and $u_{10}(\mathbf{r})$ (blue and green) is sufficient to achieve the optimal resolution. For larger separations, the other available modes start to become relevant, at first $u_{20}(\mathbf{r})$, $u_{11}(\mathbf{r})$, and $u_{02}(\mathbf{r})$ (orange, yellow, and pink), then $u_{12}(\mathbf{r})$ and $u_{21}(\mathbf{r})$ (brown and gray), and finally $u_{22}(\mathbf{r})$ (purple). Increasing the separation, the higher-order modes become dominant and all coefficients tend to their values in the absence of noise [the latter effect is visible for m_{01} and m_{10} in Fig. 11(e)].

VIII. MINIMAL RESOLVABLE DISTANCE

We conclude this work by discussing how the minimal distance that can be resolved with our moment-based estimation strategy scales with the number of detected photons. For the sake of simplicity, we focus here on thermal sources with equal brightness ($\gamma = 0$).

We consider the distance between the two sources to be resolvable as long as the estimation error Δd is smaller than the value d of the parameter itself, namely, when $d_{\min}/\Delta d \geqslant 1$ [35]. Following our moment-based estimation approach, we have $(\Delta d)^2 = 1/\mu M[d, \theta, \hat{\mathbf{N}}]$; accordingly, the minimal resolvable distance is defined by the condition

$$d_{\min}\sqrt{\mu M[d_{\min}, \theta, \hat{\mathbf{N}}]} = 1. \tag{78}$$

Given two thermal sources, each emitting on average N photons, an imaging system with a transmissivity κ, after μ measurements of our optimal observable, the mean number of detected photons is given by $N_{\text{det}} = \mu 2N\kappa$.[1] Accordingly, for a fixed imaging system, N_{det} is affected either by μ, which could be changed by increasing or decreasing the detection time, or by the brightness of the sources N. In Fig. 12 we study how for noisy demultiplexing the solution of Eq. (78) changes when we vary N_{det} either by increasing μ for $N\kappa = 1$ [blue (darker) curves] or by changing $N\kappa$ for $\mu = 1$ [red (lighter) curves]. While in the former case every noise source induces the same scaling $d_{\min} \sim N_{\text{det}}^{1/4}$, in the latter case the nontrivial dependence of $M[d_{\min}, \theta, \hat{\mathbf{N}}]$ on N leads to more complicated behaviors. This difference is due to the thermal statistics of our sources. On the other hand, in the case of

[1]To be more precise, $\mu 2N\kappa$ is the number of photons received in the image plane, which is always larger than or equal to the number of detected photons. The main cause of undetected photons is the finite number K of modes in the demultiplexing basis. However, at the level of the minimal resolvable distance, the population of the high-order modes is usually negligible. It is therefore justified to refer to $N_{\text{det}} = \mu(2N\kappa)$ as the number of detected photons.

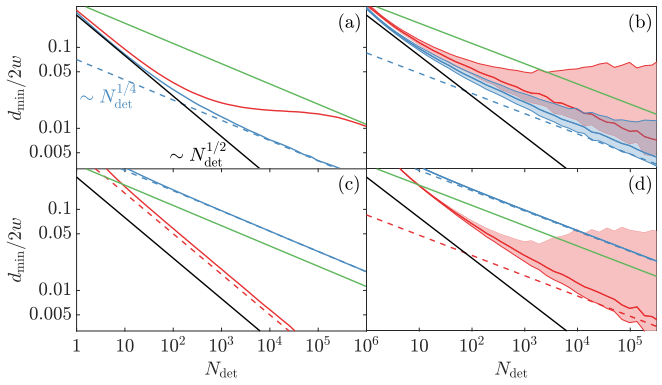

FIG. 12. Scaling of the minimal resolvable distance $d_{\min}$ with the number of detected photons $N_{\rm det} = \mu(2N\kappa)$ changed by varying either the number of measurements μ [blue (darker gray) line] or the brightness of the sources N [red (lighter gray) line]. We considered demultiplexing measurements into $K = (Q+1)^2 = 9$ HG modes $u_{ij}(\mathbf{r})$ with $0 \leqslant i, j \leqslant Q = 2$ and different noise sources: (a) misalignment ($d_s/2w = 0.01$ and $\theta_s = \pi/4$), (b) crosstalk ($\langle|c_{ij}|^2\rangle = 0.0017$), (c) dark counts ($N^{\rm dc} = 1$), and (d) all three combined. Black (lower) and green (upper) straight lines represent the $N_{\rm det}^{-1/2}$ scaling of ideal demultiplexing and the $N_{\rm det}^{-1/4}$ scaling of ideal direct imaging, respectively. Dashed lines represent analytical approximations for large $N_{\rm det}$ given by (a) Eq. (80), (b) Eq. (81), and (c) and (d) Eq. (82).

coherent sources there would not be any difference, since measuring μ copies of a coherent state $|\alpha\rangle$ gives the same results as measuring $|\mu\alpha\rangle$. To get an analytical understanding of these behaviors, we use the small-separation expansions of the measurement sensitivity $M[d, \theta, \hat{\mathbf{N}}]$. For ideal demultiplexing, our moment-based estimation strategy saturates the quantum Fisher information (62). In the small-d regime, the latter is dominated by the constant term $2N\kappa/w^2$, which defines the solution of Eq. (78),

$$d_{\min} = \frac{w}{\sqrt{N_{\rm det}}}. \tag{79}$$

Accordingly, ideal demultiplexing provides a "shot-noise" scaling of the minimal resolvable distance, which is represented by black lines in Fig. 12.

Let us now consider the impact of the different noise sources on the minimal resolvable distance achievable with demultiplexing measurements. For small values of $N_{\rm det}$, the minimal resolvable distance is determined by the expansion of $M[d, \theta, \hat{\mathbf{N}}]$ around a finite value of d. Independently of the noise level, the leading term of such an expansion is independent of d. As a consequence, we have the $d_{\min} \sim 1/\sqrt{N_{\rm det}}$ scaling we observed in the case of ideal demultiplexing. Increasing $N_{\rm det}$, the solution of Eq. (78) is determined by the quadratic behavior of $M[d, \theta, \hat{\mathbf{N}}]$ around $d \sim 0$ [see Eqs. (75)–(77)], which induces the changes of scaling observed in Fig. 12 (similar considerations can be found in Ref. [35]).

To evaluate the impact of misalignment on the large-$N_{\rm det}$ scaling minimal resolvable distance, we use Eq. (77), which we recall is valid for misalignment of the order of the separation $d_s \approx d \ll 1$, and no crosstalk and dark counts (see Table I). This expression leads to

$$d_{\min} = \frac{\sqrt{2d_s w}}{N_{\rm det}^{1/4}(\cos^4\theta \sec^2\theta_s + \sin^4\theta \csc^2\theta_s)^{1/4}}, \tag{80}$$

which is represented as a blue dashed line in Fig. 12(a). A similar $\sqrt{d_s}N_{\rm det}^{-1/4}$ scaling was reported in Ref. [33], where misalignment was studied in the low-brightness regime. We note that when N is increased for $\mu = 1$ the red (lighter) curve in Fig. 12(a) presents an almost flat region before falling on a $N_{\rm det}^{-1/4}$ line with a less favorable coefficient than the one predicted by Eq. (80). This different scaling is due to the second-order terms in $N\kappa$ which are present in the off-diagonal elements of the covariance matrix (23) whose inverse determines the measurement sensitivity $M[d, \theta, \hat{\mathbf{N}}]$. The analytical expression (80) cannot capture this behavior. In fact, it was derived from Eq. (77), which was obtained under the assumption that misalignment is of the same order of the separation $d \sim d_s \sim a$. The latter condition, which does not hold for the minimal resolvable distance when $N_{\rm det}$ is large, causes the higher terms in $N\kappa$ to disappear from $M[d, \theta, \hat{\mathbf{N}}]$ when expanding to leading order in a. On the other hand, for low photon numbers $N\kappa \lesssim 1$, Eq. (77) is accurate. Accordingly, Eq. (80) perfectly captures the scaling with the detected photon number when $N_{\rm det}$ is increased by changing the number of measurements μ with $N\kappa = 1$ (blue curve).

Let us now discuss the effect of crosstalk on the minimal resolvable distance. To do so, we consider the uniform crosstalk model and we insert Eq. (76) with $\sigma = 0$ (no dark counts) into Eq. (78), which results in

$$d_{\min} = \frac{w}{N_{\rm det}^{1/4}}\left(\frac{|r|^2}{3 + \cos(4\theta)}\right)^{1/4}. \tag{81}$$

Equation (81) is presented as blue dashed lines in Fig. 12(b) and coincides with the results obtained in Ref. [35] for Poissonian sources. We can see that the blue (darker) curve (fixed $N\kappa = 1$ and varying μ) approaches Eq. (81) for large N_d, while the red (lighter) one (fixed $\mu = 1$ and varying $N\kappa$) lies slightly above it. On the other hand, bright thermal states are more noisy, and this is reflected in the broad red error bands for large values of $N\kappa$.

Finally, let us focus on the role of dark counts. In this case, the large-$N_{\rm det}$ scaling of the minimal resolvable distance

TABLE I. Tabular representation of the role of the different noise sources in the perturbative expansions reported in Sec. VII A: $\surd$ indicates the dominant noise source, ($\surd$) indicates a noise source that is present but not dominant (in the case where the noise source affects the equation, its order is also included in the parentheses), and ✗ denotes a noise source that is not included. Equation numbers given in square brackets correspond to the uniform crosstalk model.

Equation	Misalignment	Crosstalk	Dark counts
Eq. (73) [Eq. (75)]	($\surd$)	($\surd, \epsilon \ll \sigma$)	$\surd$
Eq. (74) [Eq. (76)]	($\surd$)	$\surd$	($\surd, \sigma \lesssim \epsilon^2$)
Eq. (77)	$\surd$	✗	✗

can be obtained by inserting Eq. (75) into Eq. (78). Here we consider the dark counts level N^{dc} to be constant as we increase N_{det}, which results in

$$d_{\min} = \frac{\sqrt{2}w}{(N^2\kappa^2\mu)^{1/4}}\left(\frac{\cos(4\theta)+3}{4N^{\mathrm{dc}}(N^{\mathrm{dc}}+1)} + \frac{1}{N^{\mathrm{dc}}(N^{\mathrm{dc}}+1)+2N\kappa(2N\kappa+1)}\right)^{-1/4}. \quad (82)$$

For large values of $N\kappa$, Eq. (82) simplifies to

$$d_{\min} = \frac{\sqrt{2}w}{\sqrt{N\kappa}\mu^{1/4}}\left(\frac{\cos(4\theta)+3}{N^{\mathrm{dc}}(N^{\mathrm{dc}}+1)}\right)^{-1/4}. \quad (83)$$

From this expression we can observe that when the number of detected photons N_{det} is increased by increasing μ, the minimal resolvable distance scales as $d_{\min} \sim N_{\mathrm{det}}^{-1/4}$. On the other hand, when N_{det} is increased with the source brightness, we obtain the same scaling that we had in the ideal case $N_{\mathrm{det}}^{-1/2}$. This has a very clear physical interpretation: If we detect N_{det} photons by accumulating several measurements ($\mu \gg 1$) at low brightnesses ($N\kappa \lesssim 1$), we also accumulate dark counts. On the other hand, in a single shot measurement ($\mu \gg 1$) at high brightness ($N\kappa \gg 1$) with the same detector dark counts becoming negligible, we obtain the scaling of an ideal measurement. These different behaviors can be observed by comparing the scaling of the red ($\mu = 1$, variable N) and blue lines ($2N\kappa = 1$, variable μ) in Fig. 12(c).

Finally, let us mention that even though Eq. (83) predicts the correct scaling with N and μ, it tends to underestimate $d_{\min}$. This is due to the quadratic approximation (75) which, as visible in Fig. 9(c), grows faster than the exact measurement sensitivity, and consequently $d\sqrt{\mu M_{\mathrm{dc}}[d,\theta,\hat{\mathbf{N}}]}$ crosses 1 [see Eq. (78)] earlier than the exact curve $d\sqrt{\mu M[d,\theta,\hat{\mathbf{N}}]}$. This underestimation is particularly relevant for low dark counts levels $N^{\mathrm{dc}} \ll 1$ and in particular it leads to an unphysical better-than-ideal scaling for $N^{\mathrm{dc}} < (\sqrt{4+2/\mu}-2)/4$.

In Fig. 12(d) we present the behavior of the minimal resolvable distance when all noise sources are present at the same time. As discussed above, when the number of detected photons N_{det} is increased by accumulating many measurements μ (blue curves), we are also increasing the dark counts. In this case, dark counts are the dominant noise source [compare with the blue curve in Fig. 12(c)] and the error bands due to crosstalk become negligible. On the other hand, when N_{det} is increased with the brightness of the sources N [red curve in Fig. 12(d)], dark counts quickly become negligible and the dominant noise source is crosstalk [compare with the red curve in Fig. 12(c)].

In conclusion, let us compare the noisy demultiplexing results with those of direct imaging. The large-N_d scaling of the minimal resolvable distance for noiseless direct imaging can be obtained from Eq. (55), which for $\gamma = 0$ results in $M[d,\theta,N_p] \approx 8x^2$ and leads to (see also [35])

$$d_{\min} = \frac{w}{(N_d)^{1/4}}\left(\frac{1}{2}\right)^{1/4}. \quad (84)$$

Equation (84) is plotted as green lines in Fig. 12 and for a vast range of noise conditions it stays above the noisy demultiplexing curves. An exception is represented by the blue curves in Figs. 12(c) and 12(d). However, in Figs. 12(c) and 12(d) we chose $N^{\mathrm{dc}} = 1$, for which the approximation (83) gives sensible results. For large source brightnesses, this value corresponds to a reasonable ratio between dark counts and photons counts in each demultiplexed mode. On the other hand, it is overpessimistic for $2N\kappa = 1$, as it can be quickly understood by considering, for example, that $N_{01}/2N\kappa \approx 10^{-5}$ at $x = d/2w \approx 10^{-2}$. Therefore, we can conclude that, in most practical situations, spatial-mode demultiplexing allows us to resolve significantly smaller separations than direct imaging.

IX. CONCLUSION

We have discussed in detail a method to extract the separation of two thermal sources, with arbitrary and possibly different brightnesses, from a single optimized measurement observable. For this imaging application, we considered different measurements such as direct imaging with pixel detectors and realistic spatial-mode demultiplexing. Our results show how, even in the presence of different relevant noise sources, demultiplexing allows for better resolutions than direct imaging. In several realistic scenarios, we constructed the optimal observable for demultiplexing measurements, which depends very weakly on the separation d, providing stability of the estimation procedure over a vast domain of separations. Moreover, in the limiting case of noiseless demultiplexing, for arbitrary received photon numbers and separations, the optimized observable approaches the quantum Cramér-Rao bound if sufficiently many modes are measured. Finally, for low photon numbers in the image plane, the sensitivity $M[d,\theta,\hat{\mathbf{N}}]$ of our method saturates the Fisher information even for noisy demultiplexing.

From our results it is evident that the coefficients of the optimal linear combination of photon-number measurements are severely affected by noise. This dependence must be taken into account if one wants to achieve optimality. This could be done by calculating the coefficients from experimental measurements of the covariance matrix and the derivative vector, obtained with the help of two test sources. Alternatively, one could characterize the different noise sources and then compute the coefficients theoretically. In the latter case, precise measures of the crosstalk matrix of a demultiplexer and the dark count level of detectors are not problematic. On the other hand, misalignment errors are mainly due to an imperfect knowledge of the source centroid. These errors may be limited by scanning the multiplexer position or by using adaptive methods such as the one proposed in Ref. [32].

Finally, let us point out that in experimental realizations (see, for example, [24]) all available observables can be measured at the same time, which allows us to look at arbitrary linear combinations in postprocessing. Therefore, the estimation technique presented here can be understood as the optimal postprocessing technique that makes the best use of the available data. As a consequence, our approach fits particularly well for the estimation of a dynamically changing parameter. In fact, provided the photon detectors are faster than the typical timescale of the parameter changes, the optimal measurement coefficients at each time can be selected in postprocessing.

ACKNOWLEDGMENTS

G.S. acknowledges financial support from ONERA, the French aerospace laboratory. M.G. acknowledges funding from the LabEx ENS-ICFP Grants No. ANR-10-LABX-0010 and No. ANR-10-IDEX-0001-02 PSL*. This work was partially funded by French ANR under COSMIC Project No. ANR-19-ASTR-0020-01. This work received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 899587. This work was supported by the European Union's Horizon 2020 research and innovation program under the QuantERA program through the project ApresSF.

[1] E. Abbe, Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung, Arch. Mikrosk. Anat. **9**, 413 (1873).

[2] Lord Rayleigh F.R.S, Investigations in optics, with special reference to the spectroscope, Philos. Mag. **8**, 261 (1879).

[3] J. W. Goodman, *Statistical Optics* (Wiley, New York, 2015).

[4] S. W. Hell and J. Wichmann, Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy, Opt. Lett. **19**, 780 (1994).

[5] T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell, Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission, Proc. Natl. Acad. Sci. USA **97**, 8206 (2000).

[6] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, Imaging intracellular fluorescent proteins at nanometer resolution, Science **313**, 1642 (2006).

[7] M. T. L. Hsu, V. Delaubert, P. K. Lam, and W. P. Bowen, Optimal optical measurement of small displacements, J. Opt. B **6**, 495 (2004).

[8] V. Delaubert, N. Treps, C. C. Harb, P. K. Lam, and H.-A. Bachor, Quantum measurements of spatial conjugate variables: Displacement and tilt of a Gaussian beam, Opt. Lett. **31**, 1537 (2006).

[9] O. Pinel, J. Fade, D. Braun, P. Jian, N. Treps, and C. Fabre, Ultimate sensitivity of precision measurements with intense Gaussian quantum light: A multimodal approach, Phys. Rev. A **85**, 010101(R) (2012).

[10] C. Helstrom, Resolution of point sources of light as analyzed by quantum detection theory, IEEE Trans. Inf. Theory **19**, 389 (1973).

[11] M. Tsang, Resolving starlight: A quantum perspective, Contemp. Phys. **60**, 279 (2019).

[12] C. W. Helstrom and C. W. Helstrom, Quantum Detection and Estimation Theory (Academic, New York, 1976), Vol. 3.

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

[14] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Springer Science + Business Media, New York, 2011), Vol. 1.

[15] M. G. A. Paris, Quantum estimation for quantum technology, Int. J. Quantum Inf. **07**, 125 (2009).

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

[17] L. Pezzè and A. Smerzi, in *Quantum Theory of Phase Estimation, Proceedings of the International School of Physics "Enrico Fermi," Course CLXXXVIII, Varenna, 2014*, edited by G. M. Tino and M. A. Kasevich (IOS, Amsterdam, 2014), pp. 691–741.

[18] M. Tsang, R. Nair, and X.-M. Lu, Quantum Theory of Superresolution for Two Incoherent Optical Point Sources, Phys. Rev. X **6**, 031033 (2016).

[19] M. Paúr, B. Stoklasa, Z. Hradil, L. L. Sánchez-Soto, and J. Rehacek, Achieving the ultimate optical resolution, Optica **3**, 1144 (2016).

[20] Z. S. Tang, K. Durak, and A. Ling, Fault-tolerant and finite-error localization for point emitters within the diffraction limit, Opt. Express **24**, 22004 (2016).

[21] F. Yang, A. Tashchilina, E. S. Moiseev, C. Simon, and A. I. Lvovsky, Far-field linear optical superresolution via heterodyne detection in a higher-order local oscillator mode, Optica **3**, 1148 (2016).

[22] W.-K. Tham, H. Ferretti, and A. M. Steinberg, Beating Rayleigh's Curse by Imaging Using Phase Information, Phys. Rev. Lett. **118**, 070801 (2017).

[23] J.-F. Morizur, L. Nicholls, P. Jian, S. Armstrong, N. Treps, B. Hage, M. Hsu, W. Bowen, J. Janousek, and H.-A. Bachor, Programmable unitary spatial mode manipulation, J. Opt. Soc. Am. A **27**, 2524 (2010).

[24] P. Boucher, C. Fabre, G. Labroille, and N. Treps, Spatial optical mode demultiplexing as a practical tool for optimal transverse distance estimation, Optica **7**, 1621 (2020).

[25] R. Nair and M. Tsang, Far-Field Superresolution of Thermal Electromagnetic Sources at the Quantum Limit, Phys. Rev. Lett. **117**, 190801 (2016).

[26] C. Lupo and S. Pirandola, Ultimate Precision Bound of Quantum and Subwavelength Imaging, Phys. Rev. Lett. **117**, 190802 (2016).

[27] J. H. Shapiro, The quantum theory of optical communications, IEEE J. Sel. Top. Quantum Electron. **15**, 1547 (2009).

[28] G. Sorelli, M. Gessner, M. Walschaers, and N. Treps, Optimal Observables and Estimators for Practical Superresolution Imaging, Phys. Rev. Lett. **127**, 123604 (2021).

[29] M. Gessner, A. Smerzi, and L. Pezzè, Metrological Nonlinear Squeezing Parameter, Phys. Rev. Lett. **122**, 090503 (2019).

[30] A. Kholevo, A generalization of the Rao-Cramer inequality, Theory Probab. Appl. **18**, 359 (1974).

[31] F. Fröwis, R. Schmied, and N. Gisin, Tighter quantum uncertainty relations following from a general probabilistic bound, Phys. Rev. A **92**, 012102 (2015).

[32] M. R. Grace, Z. Dutton, A. Ashok, and S. Guha, Approaching quantum-limited imaging resolution without prior knowledge of the object location, J. Opt. Soc. Am. A **37**, 1288 (2020).

[33] J. O. de Almeida, J. Kołodyński, C. Hirche, M. Lewenstein, and M. Skotiniotis, Discrimination and estimation of incoherent sources under misalignment, Phys. Rev. A **103**, 022406 (2021).

[34] J. Řehaček, Z. Hradil, B. Stoklasa, M. Paúr, J. Grover, A. Krzic, and L. L. Sánchez-Soto, Multiparameter quantum metrology of incoherent point sources: Towards realistic superresolution, Phys. Rev. A **96**, 062107 (2017).

[35] M. Gessner, C. Fabre, and N. Treps, Superresolution Limits from Measurement Crosstalk, Phys. Rev. Lett. **125**, 100501 (2020).

[36] R. A. Bertlmann and P. Krammer, Bloch vectors for qudits, J. Phys. A: Math. Theor. **41**, 235303 (2008).

[37] Y. L. Len, C. Datta, M. Parniak, and K. Banaszek, Resolution limits of spatial mode demultiplexing with noisy detection, Int. J. Quantum Inf. **18**, 1941015 (2020).

[38] C. Lupo, Subwavelength quantum imaging with noisy detectors, Phys. Rev. A **101**, 022323 (2020).

[39] C. Oh, S. Zhou, Y. Wong, and L. Jiang, Quantum Limits of Superresolution in a Noisy Environment, Phys. Rev. Lett. **126**, 120502 (2021).

[40] W. W. Hager, Updating the inverse of a matrix, SIAM Rev. **31**, 221 (1989).