Quantum Theory of Superresolution for Two Incoherent Optical Point Sources

Mankei Tsang,1,2,* Ranjith Nair,1 and Xiao-Ming Lu1

1Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117583
2Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117551

(Received 8 November 2015; revised manuscript received 4 January 2016; published 29 August 2016)

Rayleigh’s criterion for resolving two incoherent point sources has been the most influential measure of optical imaging resolution for over a century. In the context of statistical image processing, violation of the criterion is especially detrimental to the estimation of the separation between the sources, and modern far-field superresolution techniques rely on suppressing the emission of close sources to enhance the localization precision. Using quantum optics, quantum metrology, and statistical analysis, here we show that, even if two close incoherent sources emit simultaneously, measurements with linear optics and photon counting can estimate their separation from the far field almost as precisely as conventional methods do for isolated sources, rendering Rayleigh’s criterion irrelevant to the problem. Our results demonstrate that superresolution can be achieved not only for fluorophores but also for stars.

I. INTRODUCTION

Rayleigh’s criterion for resolving two incoherent point sources, requiring them to be separated at least by a diffraction-limited spot size on the image plane [1,2], has been the most influential measure of optical imaging resolution for over a century. More recently, insights from quantum optics [3] and statistics [4] have led to revolutions in far-field superresolution techniques [5–7] beyond his criterion. The techniques proposed in Refs. [3,4] rely on locating a point source when no other nearby sources are radiating in the same optical mode. While such techniques have achieved spectacular success in microscopy, they require sophisticated control of the emission of special fluorophores and are irrelevant to astronomy and remote sensing.

For two sources with overlapping radiations on the image plane, studies have found that signal processing of the imaging data can still determine their locations, although the precision in the presence of photon shot noise quickly deteriorates when Rayleigh’s criterion is violated [8–10]. The precision degradation is mandated by the Cramér-Rao lower error bound [11], suggesting that the degradation is fundamental to direct imaging. Given such prior work, conventional wisdom thus suggests that the positions of two incoherent sources should become harder to estimate when their radiations overlap, a statistical phenomenon we call Rayleigh’s curse.

Since photon shot noise is now the dominant noise source in fluorescence microscopy [10,12] as well as stellar imaging [13–16], it is timely to inquire whether a quantum treatment can lead to new insights. Here, we attack the problem from the perspective of quantum metrology, a branch of quantum information theory relevant to sensing and imaging [17,18]. To be specific, we derive the fundamental quantum limit to the precision of locating two weak thermal optical point sources in the form of the quantum Cramér-Rao bound (QCRB) proposed by Helstrom [17]. Surprisingly, we find that the QCRB maintains a fairly constant value for any separation and shows no sign of Rayleigh’s curse. This behavior is in stark contrast to the QCRB for in-phase coherent sources, in which case Rayleigh’s curse is fundamental [19].

It is known mathematically that there exists a measurement scheme to attain the QCRB for one parameter asymptotically [20,21]. For a more concrete experimental implementation, here we propose the method of spatial-mode demultiplexing (SPADE). We show that SPADE can ideally estimate the separation between the two sources with quantum-optimal Fisher information, and we also propose linear optical system designs that can implement the measurement. Direct imaging is poor at localization of two close sources precisely because it estimates their separation poorly, and SPADE is able to overcome this problem and Rayleigh’s curse via further linear optical processing before photon counting.

The subject of quantum imaging has been extensively studied; see Appendix A for a review. Most prior proposals rely on nonclassical sources or multiphoton coincidence measurements, however, making them difficult and inefficient to use in practice. Incoherent sources, such as fluorophores and stars, are of course much more common,
and linear optical methods to enhance the localization precision for close incoherent sources will be of monumental interest to both localization microscopy [4,6,7] and astrometry [15,16]. The most relevant prior work remains the pioneering studies by Helstrom on thermal sources [17], yet he studied two sources only in the context of binary hypothesis testing and assumed a given separation in the two-source hypothesis [22]. As the separation is usually unknown and needs to be estimated in the first place [4,6–10,15,16], our parameter-estimation framework should be more useful.

II. QUANTUM OPTICS FOR WEAK THERMAL SOURCES

To illustrate the essential physics, we follow Lord Rayleigh’s lead [1] and assume quasimonochromatic scalar paraxial waves and one spatial dimension on the object and image planes. Within each short coherence time interval for a thermal source at an optical frequency, it is standard [13,14,23–27] to assume that the average photon number \( \epsilon \) arriving on the image plane is much smaller than 1, and useful information is obtained only after many photons have been measured over many such intervals. This means that the quantum density operator for the optical fields on the image plane in each coherence time interval can be well approximated as

\[
\rho = (1 - \epsilon) \rho_0 + \epsilon \rho_1 + O(\epsilon^2),
\]

where \( \rho_0 = \langle \text{vac} | \langle \text{vac} \rangle \) is the zero-photon state, \( \rho_1 \) is a one-photon state, and \( O(\epsilon^2) \) denotes terms on the order of \( \epsilon^2 \); see Appendix B for a detailed derivation. For the rest of the paper, we neglect the \( O(\epsilon^2) \) terms and use the \( \approx \) sign to denote the first-order approximations. Similar approximations were also used earlier to study stellar interferometry [26,27].

A connection with classical statistical optics can be made by observing that \( \rho_1 \) is related to the mutual coherence of the optical fields with respect to the Sudarshan-Glauber distribution. As shown in Appendix B, the one-photon state for two incoherent point sources and a diffraction-limited imaging system can be taken as

\[
\rho_1 \approx \frac{1}{2} (|\psi_1\rangle \langle \psi_1| + |\psi_2\rangle \langle \psi_2|),
\]

\[
|\psi_s\rangle = \int_{-\infty}^{\infty} dx |\psi_s(x)\rangle |x\rangle, \quad s = 1, 2,
\]

where \( x \) is the image-plane coordinate normalized with respect to the magnification factor of the imaging system [28], \( |\psi_1\rangle = a^\dagger(\mathbf{x})|\text{vac}\rangle \) is the photon image-plane position eigenket defined with respect to annihilation and creation operators that obey \( [a(x), a^\dagger(x')] = \delta(x - x') \) [29,30], and \( \psi_{s}(x) \) is the image-plane wave function from each source.

We can reproduce the standard Poisson model of direct image-plane photon counting [10,12–14,31–33] by considering the \( 1 - \epsilon \) probability of no photon count and the \( \epsilon \ll 1 \) probability of measuring a photon. If a photon is detected, the probability density of the photon position \( x \) is

\[
\Lambda(x) = \frac{1}{2} \left( |\langle x | \psi_1\rangle|^2 + |\langle x | \psi_2\rangle|^2 \right)
\]

\[
= \frac{1}{2} \left[ |\psi_1(x)|^2 + |\psi_2(x)|^2 \right].
\]

With \( \epsilon \ll 1 \), the photon count at each pixel with width \( dx \) can be approximated as Poisson with a mean given by \( \epsilon \Lambda(x) dx \). The total photon count over \( M \) coherence time intervals then remains approximately Poisson with a mean \( Me \Lambda(x) dx = M N \Lambda(x) dx \), where \( N \equiv Me \) is the average photon number collected over the \( M \) intervals and \( \Lambda(x) \) becomes the mean intensity profile. To illustrate, Fig. 1 depicts the wave functions and the mean intensity for a typical imaging system. Note the crucial point that \( \langle \psi_1 | \psi_2 \rangle = \int_{-\infty}^{\infty} dx \psi_1^* (x) \psi_2 (x) \neq 0 \), and the spatial modes excited by the two sources are, in general, not orthogonal, especially when Rayleigh’s criterion is violated. This overlap underlies all the physical and mathematical difficulties with the resolution problem, as it implies, on a fundamental level, that the two modes cannot be separated for independent measurements.

III. CLASSICAL AND QUANTUM CRAMÉR-RAO BOUNDS

To investigate the impact of measurement noise on parameter estimation, suppose that \( \rho \) depends on a set of unknown parameters denoted by \( \{\theta_{\mu}; \mu = 1, 2, \ldots\} \) [34],
and a quantum measurement is made on the image plane over the \( M \) intervals to estimate \( \theta \). Any quantum measurement can be mathematically described by a positive operator-valued measure (POVM) \( E(\mathcal{Y}) \) \cite{17}, such that the probability distribution of measurement outcome \( \mathcal{Y} \) is \( P(\mathcal{Y}) = \text{tr}(E(\mathcal{Y})\rho^{\otimes M}) \), with \( \text{tr} \) denoting the operator trace and \( \rho^{\otimes M} \) denoting a tensor product of \( M \) density operators. Let \( \hat{\theta}_\mu(\mathcal{Y}) \) be an estimator and

\[
\Sigma_{\mu\mu} = \int d\mathcal{Y} P(\mathcal{Y})[\hat{\theta}_\mu(\mathcal{Y}) - \theta_\mu][\hat{\theta}_\nu(\mathcal{Y}) - \theta_\nu],
\]

be the error covariance matrix. For any unbiased estimator, the Cramér-Rao bound is given by

\[
\Sigma_{\mu\mu} \geq(J^{-1})_{\mu\mu},
\]

where

\[
J_{\mu\nu} = \int d\mathcal{Y} \frac{1}{P(\mathcal{Y})} \left( \frac{\partial P(\mathcal{Y})}{\partial \theta_\mu} \right) \left( \frac{\partial P(\mathcal{Y})}{\partial \theta_\nu} \right)
\]

is the Fisher information matrix with respect to \( P(\mathcal{Y}) \) \cite{11}. I

For the Poisson model of direct imaging \cite{10,31–33},

\[
J^{(\text{direct})}_{\mu\nu} = N \int_{-\infty}^{\infty} dx \frac{1}{\Lambda(x)} \frac{\partial \Lambda(x)}{\partial \theta_\mu} \frac{\partial \Lambda(x)}{\partial \theta_\nu}.
\]

Alternatively, the same result can be derived without the Poisson approximation by considering the one-photon distribution given by Eq. (2.4) and no multiphoton coincidence. As the Cramér-Rao bound is asymptotically achievable \cite{11}, the Fisher information has become the standard precision measure in modern fluorescence microscopy \cite{10,31–33} as well as astronomy \cite{13,16,35,36}.

Direct imaging, though standard, is but one of the infinite measurement methods permitted by quantum mechanics. The ultimate performance of any quantum measurement and any unbiased estimator can be quantified using the quantum Cramér-Rao bound

\[
\Sigma_{\mu\mu} \geq(J^{-1})_{\mu\mu} \geq(\mathcal{K}^{-1})_{\mu\mu},
\]

where \( \mathcal{K} \) is the quantum Fisher information matrix in terms of \( \rho^{\otimes M} \) \cite{17}. To compute \( \mathcal{K} \) analytically, we assume a spatially invariant imaging system with \( \psi_s(x) = \psi(x - X_s) \), where \( \psi(x) \) is the point-spread function of the imaging system and \( X_s \) is the unknown position of each source \cite{28}. Both \( J^{(\text{direct})} \) and \( \mathcal{K} \) turn out to be diagonal if we redefine the parameters of interest as the centroid

\[
\theta_1 = \frac{X_1 + X_2}{2},
\]

and the separation

\[
\theta_2 = X_2 - X_1,
\]

as depicted in Fig. 1. We also assume, with little loss of generality, that the point-spread function has a constant \( x \)-independent phase, which can be easily implemented by a two-lens system \cite{28}. The phase is then irrelevant to \( \rho_1 \) in Eq. (2.2), and \( \psi(x) \) can be taken as real.

The computation of \( \mathcal{K} \) is described in Appendix C; the result is

\[
\mathcal{K}_{11} \approx 4N(\Delta k^2 - \gamma^2), \quad \mathcal{K}_{22} \approx N\Delta k^2,
\]

with \( \mathcal{K}_{12} = \mathcal{K}_{21} \approx 0 \), where

\[
\Delta k^2 \equiv \int_{-\infty}^{\infty} dx \left( \frac{\partial \psi(x)}{\partial x} \right)^2
\]

is the spatial-frequency variance of the real point-spread function set by the diffraction limit and

\[
\gamma \equiv \int_{-\infty}^{\infty} dx \frac{\partial \psi(x)}{\partial x} \psi(x - \theta_2)
\]

is a parameter that depends on \( \theta_2 \). The prefactor \( N \) indicates a shot-noise scaling with respect to the average photon number, as expected from classical sources \cite{18,19}. For \( \theta_2 \to 0 \), \( \gamma^2 \to 0 \), and we recover the standard shot-noise limit to the localization of isolated sources.

To compare the quantum Fisher information with the classical information for direct imaging, Fig. 2 plots the diagonal elements of \( J^{(\text{direct})} \) and \( \mathcal{K} \), assuming a Gaussian point-spread function \( \psi(x) = (2\pi\sigma^2)^{-1/4} \exp[-x^2/(4\sigma^2)] \), where \( \sigma = 1/(2\Delta k) = \lambda/(2\pi NA) \), \( \lambda \) is the free-space wavelength, and NA is the effective numerical aperture. The constant \( \mathcal{K}_{22} \), in particular, becomes

\[
\mathcal{K}_{22} \approx \frac{N}{4\sigma^2}.
\]

The Gaussian case is representative, and the same qualitative behaviors can be observed for other common point-spread functions. For the centroid, both the classical and quantum information are within a factor of 2 of the standard limit \( N/\sigma^2 \). \( J_{11}^{(\text{direct})} \leq \mathcal{K}_{11} \) as it should, but the small gap between the two means that there is little room for improvement.

The difference between the separation information quantities \( \mathcal{K}_{22} \) and \( J_{22}^{(\text{direct})} \) in Fig. 2 is much more dramatic. Both quantities approach the same limit \( N/(4\sigma^2) \) as \( \theta_2 \to 0 \), implying that direct imaging is quantum optimal for well-separated sources. For \( \theta_2/\sigma \to 0 \), however, the classical information \( J_{22}^{(\text{direct})} \) decreases to zero. This means that direct imaging is progressively worse at estimating the separation for closer sources, to the point that the information vanishes and the Cramér-Rao bound diverges at
θ_2 = 0. We call this divergent behavior due to overlapping wave functions “Rayleigh’s curse,” as it implies a severe penalty on the localization precision when the intensity profiles overlap significantly and Rayleigh’s criterion is violated for a given N.

Direct imaging suffers from Rayleigh’s curse for any point-spread function, as ∂Λ(x)/∂θ_2 vanishes at θ_2 = 0 while Λ(x) remains nonzero in regions of x where the derivative vanishes, causing J_22^{(direct)} to vanish via Eq. (3.4). This is the reason why the Cramér-Rao bounds derived in Refs. [8–10] for separation estimation all diverge when Rayleigh’s criterion is violated. Remarkably, the quantum information K_22 in Eq. (3.8) stays constant regardless of the separation. If the centroid θ_1 is known, there exists a POVM with error Σ_22 asymptotically attaining the single-parameter QCRB [20,21], viz.,

$$\Sigma_{22} \rightarrow \frac{1}{K_{22}} \approx \frac{1}{N\Delta k^2}, \quad N \rightarrow \infty. \quad (3.12)$$

This means that Rayleigh’s curse can be avoided for separation estimation, and considerable improvements can be obtained, if the optimal quantum measurement can be implemented. To expound the issue, Fig. 3 plots the quantum and classical Cramér-Rao bounds 1/K_{22} and 1/J_22^{(direct)} , demonstrating more dramatically the divergent error in the classical case and the substantial room for improvement offered by quantum mechanics.

IV. SPATIAL-MODE DEMULTIPLEXING (SPADE)

Instead of measuring the position of each photon in the direct imaging method, we propose a discrimination in terms of the Hermite-Gaussian spatial modes [37] to estimate the separation. Consider the basis \{ |ϕ_q⟩; q = 0, 1, … \} with eigenkets given by

$$|ϕ_q⟩ = \int_{-∞}^{∞} dx φ_q(x) |x⟩, \quad q = 0, 1, … \quad (4.1)$$

$$φ_q(x) = \left(\frac{1}{2πσ^2}\right)^{1/4} \frac{1}{\sqrt{q!}} H_q\left(\frac{x}{\sqrt{2}σ}\right) \exp\left(-\frac{x^2}{4σ^2}\right), \quad (4.2)$$

where H_q is the Hermite polynomial [37]. The POVM for each coherence time interval can be expressed as projections

$$E_0 = |\text{vac}\rangle⟨\text{vac}|, \quad E_1(q) = |ϕ_q⟩⟨ϕ_q|. \quad (4.3)$$

Conditioned on a detection event, the probability of detecting the photon in the qth mode becomes

$$P_1(q) \approx \frac{1}{2} (|⟨ϕ_q|ψ_1⟩|^2 + |⟨ϕ_q|ψ_2⟩|^2). \quad (4.4)$$

Similar to direct imaging, ε ≪ 1 implies that, over M intervals, the total photon count m_q in each Hermite-Gaussian mode can be approximated as Poisson with a mean given by NP_1(q).
To proceed further, we assume that the centroid $\theta_1$ is known, and only $\theta_2$ is to be estimated. Since centroid estimation using direct imaging is relatively insensitive to the separation, the assumption of an accurately known centroid is not difficult to satisfy; see Appendix D for a detailed discussion. Under this assumption, we can assume $\theta_1 = 0$ without loss of generality, and the wave functions become $\psi_1(x) = \psi(x + \theta_2/2)$ and $\psi_2(x) = \psi(x - \theta_2/2)$. For simple analytic results, we further assume that the point-spread function is Gaussian. The overlap factors in Eq. (4.4) can then be evaluated by recognizing that $|\phi_q\rangle$ is mathematically equivalent to an energy eigenstate of a harmonic oscillator (in the configuration space of the photon), and $|\psi_1\rangle$ and $|\psi_2\rangle$ are equivalent to configuration-space coherent states with displacements $\pm \theta_2/(4\sigma)$. The result is

$$P_1(q) \approx |\langle \phi_q | \psi_1 \rangle|^2 = |\langle \phi_q | \psi_2 \rangle|^2 = \exp\left(-\frac{Q}{q^2}\right),$$

$$Q \equiv \frac{\theta_2^2}{16\sigma^2} .$$

This formula is valid even if the two sources have unequal intensities and $\rho_1$ is any mixture of $|\psi_1\rangle \langle \psi_1|$ and $|\psi_2\rangle \langle \psi_2|$. The classical Fisher information for the Hermite-Gaussian-basis measurement over $M$ intervals becomes

$$J_{22}^{\text{HG}} \approx N \sum_{q=0}^{\infty} P_1(q) \left[ \frac{\partial}{\partial \theta_2} \ln P_1(q) \right]^2 \approx \frac{N}{4\sigma^2} ,$$

which is equal to the quantum information given by Eq. (3.11) and also free of Rayleigh’s curse.

To measure in the Hermite-Gaussian basis, one needs to demultiplex the image-plane field in terms of the desired spatial modes before determining the outcome based on the mode in which the photon is detected. To do so with a high information-extraction efficiency, one should perform a one-to-one conversion of the Hermite-Gaussian modes into modes in a more accessible degree of freedom with minimal loss and measurements that capture as many photons as possible. For example, we can take advantage of the fact that the Hermite-Gaussian modes are waveguide modes of a quadratic-index waveguide [37]. Suppose that we couple the image-plane optical field into such a highly multimode waveguide centered at the centroid position, as shown in Fig. 4. Each mode with index $q$ acquires a different propagation constant $\beta_q$ along the longitudinal direction $z$. If a grating coupler [38] with spatial frequency $\kappa$ is then used to couple all the modes into free space, each mode will be coupled to a plane wave with a different spatial frequency $\beta_q - \kappa$ along the $z$ direction in free space, and a Fourier-transform lens can be used to focus the different plane waves onto different spots of a photon-counting array in the far field.

An alternative is to use evanescent coupling with different single-mode waveguides [39], as depicted in Fig. 5. If each single-mode waveguide is fabricated to have a propagation constant equal to a different value of $\beta_q$, the phase-matching condition will cause each mode in the multimode waveguide to be coupled to a specific fiber.

Given these physical setups, we can now explain the operation of SPADE in a more intuitive semiclassical optics language: It is based on the exquisite sensitivity of the mode-coupling efficiencies to the offset of the wave functions from the centroid. The incoherent sources are literally blinking on the fundamental coherence time scale, causing each image-plane photon to have a wave function given randomly by $\psi_1(x)$ or $\psi_2(x)$. Either wave function can excite the waveguide modes coherently with the same excitation probabilities, causing the final photon counts to be as sensitive to the offset for two sources as it is for one. Put another way, the incoherence between the two sources implies a random relative phase between the two fields and enables coupling into the first-order odd mode, which is the main spatial mode responsible for the high sensitivity to small offsets.
The use of photon counting is essential here to discriminate against the abundant but uninformative zero-photon events. If homodyne or heterodyne methods were used instead, they would suffer from excess vacuum fluctuations when no photon arrives. The poor performance of heterodyne methods for weak thermal sources is also known in the context of stellar interferometry [27,40].

The situation is different from measurements of coherent light, the density operator of which contains off-diagonal terms with respect to the photon-number basis, and the probabilistic photon picture is less adequate. Our preliminary calculations [41] confirm this expectation and suggest that heterodyne detection of the spatial modes still suffers from Rayleigh’s curse.

Suppose that a total of \( L \) photons are detected over the \( M \) trials. A record of the modes for the \( L \) photons \( \{ q_1, \ldots, q_L \} \) can be obtained, but in fact, a time-resolved record is not necessary, as \( \sum_i q_i \) is a sufficient statistic for estimating \( Q \). With \( \sum_i q_i \), meaning that the set of photon numbers \( \{ m_q = \sum_i \delta_{q,q}; q = 0, 1, \ldots \} \) detected in different modes is also sufficient. The maximum-likelihood estimator becomes

\[
\hat{Q}_{ML} = \frac{1}{L} \sum_q q m_q,
\]

\[
\hat{\theta}_{2ML} = 4\sigma \sqrt{\hat{Q}_{ML}}.
\]

which is straightforward to implement computationally. For \( L = 0 \), one can set \( \hat{\theta}_2 \) to a constant value; the \( L = 0 \) probability \( 1 - e^M \approx \exp(-N) \) is, in any case, negligible for large \( N \). Maximum-likelihood estimation can asymptotically saturate the Cramér-Rao bound \( \Sigma_{22} \geq 1/J_{22}^{(HG)} \) for large \( M \) [42]. With \( J_{22}^{(HG)} \approx K_{22} \), the QCRB is asymptotically attainable as well. Appendix E reports a Monte Carlo analysis of the maximum-likelihood estimator for SPADE, confirming that the Cramér-Rao bound remains close to the estimation error for finite photon numbers.

V. BINARY SPADE

Since direct imaging has trouble estimating the separation only when \( \theta_2/\sigma \) is small, and only low-order Hermite-Gaussian modes in SPADE are excited significantly in that case, we can focus on the discrimination of low-order modes to simplify the SPADE design. One such design is depicted in Fig. 6. where only the \( q = 0 \) component is coupled into the single-mode waveguide, while any photon in the higher-order modes remains in the multimode waveguide for subsequent detection. An alternative design is depicted in Fig. 7: The \( q = 0 \) mode is coupled to a single-mode waveguide, while higher-order modes are necessarily coupled to the leaky modes of the waveguide, which are also measured.

Conditioned on a detection event, the probability of detecting the photon in the \( q = 0 \) mode remains

\[
P_1(q \leq 0) \approx \exp(-Q),
\]

but now the higher-order modes cannot be discriminated, and the probability of detecting a photon in any higher-order mode becomes

\[
P_1(q > 0) = 1 - P_1(q = 0) \approx 1 - \exp(-Q).
\]

The Fisher information for this scheme is hence

\[
J_{22}^{(b)} \approx \frac{N}{4\sigma^2} \frac{Q \exp(-Q)}{1 - \exp(-Q)}.
\]

Figure 8 compares \( J_{22}^{(b)} \) with the optimal value \( J_{22}^{(HG)} \approx K_{22} \) as well as \( J_{22}^{(direct)} \) for direct imaging. It can be seen that binary SPADE gives significant information for small \( \theta_2/\sigma \), which happens to be the regime where direct imaging performs poorly. Binary SPADE actually works less well when the sources are far apart, and the two methods can complement each other to enhance the localization precision, as shown in Appendix D.

For a total of \( L \) detected photons, \( m_0 \) (the number of photons detected in the \( q = 0 \) mode) and \( L \) are sufficient statistics for estimating \( Q \) and \( \theta_2 \), and \( m_0 \) follows the binomial distribution for \( L \) trials and success probability \( \exp(-Q) \) [42]. The maximum-likelihood estimator becomes

\[
\hat{Q}_{ML}^{(b)} = -\ln \frac{m_0}{L},
\]

\[
\hat{\theta}_{2ML}^{(b)} = 4\sigma \sqrt{\hat{Q}_{ML}^{(b)}}.
\]

For \( L = 0 \) or \( m_0 = 0 \), one can select finite values for \( \hat{\theta}_2 \) to regularize the estimator. Appendix E reports a Monte Carlo analysis of the resulting estimation error, confirming that it remains close to the Cramér-Rao bound for finite photon numbers.

Compared with the large amount of data generated by direct imaging and the complex algorithms needed to
FIG. 8. Fisher information for separation estimation versus normalized separation \( \theta_2/\sigma \) for a Gaussian point-spread function. \( J^{(HG)}_{22} \) is the information for the ideal Hermite-Gaussian-basis measurement, which is equal to the quantum value \( K_{22} \), \( J^{(direct)}_{22} \) is for direct imaging, and \( J^{(b)}_{22} \) is for binary SPADE. The vertical axis is normalized with respect to \( J^{(HG)}_{22} = K_{22} = N/(4\sigma^2) \).

process them, only two photon numbers are needed by binary SPADE to estimate the separation precisely. The highly compressed measurement output and computationally simple estimators, enabled by the coherent optical processing, come as bonuses with our schemes.

VI. OTHER POINT-SPREAD FUNCTIONS

Our analysis of SPADE so far relies on the assumption of a Gaussian point-spread function. For other point-spread functions, it is nontrivial to find a suitable basis of spatial modes, although we can still rely on the mathematical existence of a quantum-optimal measurement [20,21] to be sure that the QCRB can be saturated. For a more concrete method, the analysis of the binary SPADE schemes is fortunately still tractable, if we assume a single-mode waveguide with a mode profile that matches the point-spread function \( \psi(x) \) centered at the centroid position. Define \( |\psi⟩ = \int_{-\infty}^{\infty} dx |ψ(x)| x⟩ \) as the state of one photon in the waveguide mode. The efficiency of coupling a photon in state \( |\psi_1⟩ \) or \( |\psi_2⟩ \) into the waveguide mode becomes

\[
|\langle \psi|\psi_1⟩|^2 = |\langle \psi|\psi_2⟩|^2 = |\int_{-\infty}^{\infty} dx ψ^*(x)ψ(x + \theta_2/2)|^2 = |\int_{-\infty}^{\infty} dk |Ψ(k)|^2 \exp \left(\frac{i k \theta_2}{2}\right)|^2 = \Upsilon(\theta_2),
\]

(6.1)

where \( \Upsilon(\theta_2) \) is the mode overlap factor and \( Ψ(k) ≡ (2\pi)^{-1/2} \int_{-\infty}^{\infty} dx ψ(x) \exp(-ikx) \) is the optical transfer function of the imaging system before the image plane [28]. For the density operator in Eqs. (2.1), (2.2), and (2.3), or in fact any mixture of \( |\psi_1⟩⟨\psi_1| \) and \( |\psi_2⟩⟨\psi_2| \), the probability of finding a photon in the waveguide mode becomes \( P(\psi) ≈ \epsilon \Upsilon \), and the probability of finding a photon in any other mode is \( P(\bar{\psi}) ≈ \epsilon(1 - \Upsilon) \). The Fisher information over \( M \) intervals is then

\[
J^{(b)}_{22}(\theta_2 = 0) ≈ NΔk^2.
\]

(6.3)

Hence, \( J^{(b)}_{22} \) can reach the quantum information \( K_{22} = N\Delta k^2 \) at \( \theta_2 = 0 \), precisely where \( J^{(direct)}_{22} \) vanishes and Rayleigh’s curse is at its worst. For larger \( \theta_2 \), \( J^{(b)}_{22} \) is expected to decrease, as the scheme is unable to discriminate the higher-order modes that become more likely to be occupied. Figure 9 plots \( K_{22} \), the numerically computed \( J^{(direct)}_{22} \), and \( J^{(b)}_{22} \) for the sinc point-spread function [1] \( ψ(x) = (1/\sqrt{W}) \text{sinc}(x/W) \), where \( W = \lambda/(2NA) \), \( \text{sinc} u = \sin(\pi u)/(\pi u) \) for \( u \neq 0 \) and \( \text{sinc}(0) = 1 \). The information quantities demonstrate behaviors similar to the Gaussian case.
VII. TWO-DIMENSIONAL IMAGING

The essential physics remains unchanged when we consider two-dimensional imaging, and we discuss the generalization only briefly here; the details are given elsewhere [43]. The single-photon ket in Eq. (2.3) should now be expressed as \(|\psi_i\rangle = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \psi_i(x,y)|x,y\rangle\), where \(\langle x,y|x',y'\rangle = \delta(x-x')\delta(y-y')\) and \(\psi_i(x,y)\) is a two-dimensional wave function [29,30]. In terms of a point-spread function \(\varphi(x,y)\) and unknown positions \((X_1,Y_1)\) and \((X_2,Y_2)\), \(\psi_i(x,y) = \varphi(x-X_i,y-Y_i)\), and we can define the four centroid and separation parameters as \(\theta_1 = (X_1 + X_2)/2\), \(\theta_2 = X_2 - X_1\), \(\theta_3 = (Y_1 + Y_2)/2\), and \(\theta_4 = Y_2 - Y_1\). Here, \(\mathcal{J}^{\text{(direct)}}\) for the estimation of \(\theta_2\) and \(\theta_4\) decreases to zero when the sources are close, and Rayleigh’s curse still exists for direct imaging [9,10]. On the other hand, the quantum Fisher information matrix, to be reported in Ref. [43], again shows no sign of Rayleigh’s curse for two-dimensional separation estimation.

For SPADE, we can use the two-dimensional Hermite-Gaussian basis [37]. Assuming a Gaussian point-spread function and a known centroid, it is straightforward to show that a measurement of each photon in the Hermite-Gaussian basis with mode indices \(q\) and \(p\) obeys a two-variable Poisson distribution, and the classical Fisher information with respect to \(\theta_2\) and \(\theta_4\) remains a constant, free of Rayleigh’s curse, similar to the one-dimensional case. For other point-spread functions, such as the Airy disk [2,28], binary SPADE with a matching mode profile can estimate the separation without Rayleigh’s curse for small separations in the same way as the one-dimensional case, but information about the direction of the separation is lost. To obtain directional information, one needs to discriminate at least some of the higher-order modes in different directions.

A quadratic-index optical fiber can support two-dimensional Hermite-Gaussian modes, while a weakly guiding step-index fiber also has modes closely resembling the Hermite-Gaussian modes [38]. A complication arises for cylindrically symmetric fibers, as modes with the same total order \(q + p\) will have a degenerate propagation constant, causing multiple modes to satisfy the same phase-matching conditions in grating or evanescent coupling and preventing discrimination of modes with the same order. The net result is that directional information is compromised. One solution is to turn the point-spread function into an elliptic one with asymmetric widths and use an elliptic fiber to break the degeneracy.

VIII. CONCLUSION

We have presented two important results in this paper: the fundamental quantum limit to locating two incoherent optical point sources and the SPADE measurement schemes for quantum-optimal separation estimation. Our quantum bound sets the ultimate limit to localization precision in accordance with the fundamental laws of quantum mechanics, while SPADE can extract the full information offered by quantum mechanics concerning the separation parameter via linear photonics. The proposed SPADE schemes work well for close sources with significant overlap in their wave functions, avoiding Rayleigh’s curse and the divergent error that plagues direct imaging. The computational simplicity of the estimators is an additional advantage. Foreseeable applications include binary-star astrometry [15,16,44] and single-molecule imaging [7], either as a replacement of techniques based on fluorescence resonant energy transfer [12,45] or as an enhancement of localization microscopy [5–7,10,12] to provide complementary information about close pairs of fluorophores.

ACKNOWLEDGMENTS

We acknowledge useful discussions with Shan Zheng Ang and Shilin Ng. This work is supported by the Singapore National Research Foundation under NRF Grant No. NRF-NRF2011-07 and the Singapore Ministry of Education Academic Research Fund Tier 1 Project R-263-000-C06-112.

M. T. conceived the idea of applying quantum metrology to two-incoherent-source localization and developed the quantum optics formalism in Sec. II and Appendix B. R. N. derived the quantum Fisher information in Sec. III and Appendix C with X.-M. L. and M. T.’s inputs and checks. M. T. invented and analyzed the SPADE measurement schemes described in Secs. IV–VII, while R. N. first recognized the importance of the first-order Hermite-Gaussian mode to SPADE, as explained in Sec. IV. M. T. supervised the project and wrote the paper. All authors discussed extensively during the course of this work.

Note added.—Subsequent to the completion of this work, we developed a semiclassical but less general theory to explain our results here for pedagogy [46] and discovered an alternative scheme called super-localization via image-inversion interferometry (SLIVER) that can overcome Rayleigh’s curse without the need to tailor the device to the point-spread function [43,47]; we also derived the QCRB for thermal sources without the \(\epsilon \ll 1\) approximation using an alternative approach [48] and showed that variations of SPADE and SLIVER can attain the bound for arbitrary \(\epsilon\) [48], validating the results here. A generalization of our theory presented here to two dimensions, with similar conclusions, is described in detail elsewhere [43]. Following Ref. [49], Lupo and Pirandola have derived the ultimate quantum Fisher information for separation estimation with arbitrary quantum sources [50], including our independent result on thermal sources [48] as a special case. Experiments inspired by our theory have been reported in Refs. [51–54]; Refs. [52–54] also propose variations of SPADE that are easier to implement experimentally.
APPENDIX A: QUANTUM-IMAGING LITERATURE REVIEW

Helstrom pioneered the application of his quantum estimation and detection bounds to optical-imaging problems [17,22,55], focusing on coherent and thermal sources. In particular, the now well-known expression for the shot-noise-limited localization error for one classical source can be found in Ref. [55]; similar expressions in the context of direct imaging were later reported in Refs. [35,36,56]. For more recent studies of quantum metrology for coherent-state or nonclassical-state imaging, see, for example, Refs. [57–59]. For studies on the use of squeezed light for single-object localization, see, for example, Refs. [19,60–63]. None of these studies considered the problem of locating two close incoherent sources.

The standard quantum model of paraxial imaging and the use of nonclassical light for that purpose were proposed by Yuen and Shapiro [29]. This topic has been further investigated most notably by Kolobov and co-workers [64,65], who focused on coherent or squeezed light, homodyne detection, and field fluctuations. Such models are irrelevant to incoherent sources such as stars and fluorophores, for which the mean field is zero and photon counting is the more relevant method to minimize vacuum noise; to quote Helstrom [66],

With such incoherently illuminated or radiating objects, it is not the field of the light that is of interest, for that field is best described as a random process having zero mean value and a most erratic spatiotemporal variation. Rather it is the mean-square value of the field, averaged over many cycles of the dominant temporal frequency, that characterizes the object in the most informative way.

Subsequent work by Kolobov and co-workers [67–70] considered the squeezing and measurement of the eigenmodes of an imaging system for image-reconstruction superresolution. Again, these studies focused on coherent or squeezed light only. The “Rayleigh resolution limit” mentioned by many of these papers is a misnomer, as the resolution limit for coherent imaging should be attributed to Abbe, while Rayleigh’s criterion is defined for two incoherent sources [1,2] and is ill suited to coherent imaging [71]. Moreover, the imaging-system eigenmodes they studied have no relation to the spatial modes we propose for the two-source localization problem, and they did not use the more rigorous framework of statistical parameter estimation.

We can consider the schemes proposed in Refs. [72–79] as another class of superresolution imaging protocols, which require coherent or nonclassical sources and multiphoton coincidence measurements and do not consider statistical inference. It is well known in statistical optics that a multiphoton coincidence measurement, such as the obsolete Hanbury Brown-Twiss interferometry, fundamentally has a much poorer signal-to-noise ratio than amplitude interferometry because multiphoton coincidence events are rare for thermal optical sources [23,27]. The actual statistical resolution of this class of protocols is thus questionable, especially for weak optical sources, without further proofs in the context of inference accuracy. In recent years, there has also been significant interest in quantum lithography [80–83] and ghost imaging [81,84–86], although their applications are clearly different from our purpose and will not be elaborated here.

The relative neglect of incoherent sources in the quantum-imaging literature, despite their obvious importance, may be due to a lack of appreciation that quantum mechanics can be relevant to such highly classical light. Our work thus showcases quantum metrology as a powerful tool to discover the ultimate performance of sensing and imaging even for classical sources, providing not only rigorous quantum limits but also pleasant surprises for one of the most important applications in optics.

APPENDIX B: QUANTUM OPTICS: DERIVATION OF Eqs. (2.1)–(2.3)

Define \( \alpha = (\alpha_1, \ldots, \alpha_J) \) as a column vector of complex field amplitudes for \( J \) optical spatial modes on the image plane and \( |\alpha\rangle \) as a multimode coherent state with amplitude \( \alpha \). Any quantum state can be expressed as

\[
\rho = \int D\alpha \Phi(\alpha)|\alpha\rangle\langle\alpha|,
\]

where \( \Phi(\alpha) \) is the Sudarshan-Glauber representation and \( D\alpha \) is an appropriate measure [24]. For thermal sources, it is standard [24] to assume \( \Phi \) to be a zero-mean complex Gaussian given by

\[
\Phi(\alpha) = \frac{1}{\det(\pi\Gamma)} \exp(-\alpha^\dagger \Gamma^{-1} \alpha),
\]

where \( \alpha^\dagger = (\alpha_1^*, \ldots, \alpha_J^*) \) denotes the complex transpose of \( \alpha \),

\[
\Gamma = \mathbb{E}(\alpha\alpha^\dagger)
\]

is the image-plane mutual coherence matrix, and \( \mathbb{E}[f(\alpha)] = \int D\alpha \Phi(\alpha)f(\alpha) \) denotes the expectation of any function \( f \) with respect to the \( \Phi \) distribution. Writing the coherent state in terms of a superposition of Fock states and applying the Gaussian moment theorem [24] to Eq. (B1), we can express \( \rho \) as the incoherent mixture

\[
\rho = \sum_{n=0}^{\infty} \pi_n \rho_n, \quad n = \sum_j n_j,
\]

where \( \pi_n \) is the probability of having \( n \) total photons in the state and \( \rho_n \) is an \( n \)-photon multimode Fock state.
where tr denotes the operator trace, $a_j^\dagger$ is the annihilation operator for the $j$th mode, and $a_j$ is the creation operator. For example, a star with sunlike temperature 6000 K emits about $10^{-2}$ photons on average per mode at wavelength 500 nm, while the limited fraction of the coherence area captured by the telescope aperture further reduces the received photon number [23]. In microscopy, a typical fluorophore emits $< 10^7$ photons per second [31] with coherence time $< 50$ fs [12], leading to $\epsilon < 10^{-6}$ for two sources. The zero-photon probability given by

$$e \equiv \sum_j \text{tr} \rho a_j^\dagger a_j = \mathbb{E}(a_j^\dagger a_j) = \sum_j \Gamma_{jj} \ll 1, \quad (B5)$$

is then the highest, the one-photon probability given by

$$\sqrt{e} = \mathbb{E}(e^{-a_j^\dagger a_j} a_j) = e + O(e^2) \quad (B6)$$

is $e$ to the first order, and the multiphoton probability

$$\sum_{n=2}^\infty \sqrt{e}^n = 1 - e + O(e^2) \quad (B7)$$

is in the second order, leading to Eq. (2.1). As the vacuum state provides no information and multiphoton events are rare, we focus on the one-photon state $\rho_1$. This focus also makes our formalism applicable to inefficient single-photon emitters, which may have non-Poissonian multiphoton statistics but rare multiphoton events, and electron microscopy [9].

The negligence of the $O(e^2)$ multiphoton probability leads to a Poisson photon-counting distribution [23], which ignores bunching or antibunching effects but remains an excellent empirical model for both astronomical optical sources [13,14,23–25] and fluorophores [10,12,31–33] by virtue of the $e \ll 1$ condition. To quote Mandel [25],

The light from these sources is always so weak that $\mathbb{E}(\Delta z_j^2)/T \ll 1$ [e in our terminology] and the degeneracy is unlikely to be detected in measurements on a single beam. The situation is, of course, improved when correlation measurements are undertaken on two or more coherent beams (Hanbury Brown and Twiss 1956), since these measurements single out the degenerate photons (Mandel 1958). Even so it is unlikely that any faint stars could be studied in this way.

Similarly, Goodman states that [23]

If the count degeneracy parameter [e in our terminology] is much less than 1, it is highly probable that there will be either zero or one count in each separate coherence interval of the incident classical wave. In such a case the classical intensity fluctuations have a negligible “bunching” effect on the photo-events, for (with high probability) the light is simply too weak to generate multiple events in a single coherence cell. If negligible bunching of the events takes place, the count statistics will be indistinguishable from those produced by stabilized single-mode laser radiation, for which no bunching occurs.

A more recent work by Zmuidzinas [13] also states that

It is well established that the photon counts registered by the detectors in an optical instrument follow statistically independent Poisson distributions, so that the fluctuations of the counts in different detectors are uncorrelated. To be more precise, this situation holds for the case of thermal emission (from the source, the atmosphere, the telescope, etc.) in which the mean photon occupation numbers of the modes incident on the detectors are low, $n \ll 1$ [e in our terminology]. In the high occupancy limit, $n \gg 1$, photon bunching becomes important in that it changes the counting statistics and can introduce correlations among the detectors. We will discuss only the first case, $n \ll 1$, which applies to most astronomical observations at optical and infrared wavelengths.

Define $|j\rangle = a_j^\dagger |\text{vac}\rangle$ as the ket with one photon only in the $j$th mode. Consider the one-photon matrix element

$$\langle j|\rho|k\rangle = \mathbb{E}(e^{-a_j^\dagger a_j} a_j^\dagger a_k) = \Gamma_{jk} + O(e^2). \quad (B9)$$

We can then assume, to the first order of $e$,

$$\sqrt{e} = \rho_1 \approx e, \quad \rho_1 \approx e \sum_{j,k} \Gamma_{jk} |j\rangle \langle k|. \quad (B10)$$

Similar approximations were also used in Refs. [26,27]. To derive $\Gamma$, let $\beta \equiv (\beta_1, \ldots, \beta_K)^\dagger$ be the field amplitudes for optical modes on the object plane, and consider the field propagation rule $a = S\beta$ for a linear optical system, where $S$ is the field scattering matrix. The image-plane mutual coherence $\Gamma$ is then related to the object-plane mutual coherence matrix $\Gamma^{(o)}$ by

$$\Gamma = S\Gamma^{(o)} S^\dagger. \quad (B11)$$

This propagation rule is a basic principle in both classical and quantum statistical optics [24].

In the paraxial regime, we can use localized wave-packet modes as a basis [29,30]. Let $u$ be the position index for a wave-packet mode on the one-dimensional object plane and consider two incoherent sources with equal intensities at positions $u = u_1$ and $u = u_2$. The fields are uncorrelated
at different points on the object plane, with nonzero intensities only at the sources. Then

\[ \Gamma_{uv}^{(0)} = e_0 \delta_{uv} (\delta_{uu_1} + \delta_{uu_2}), \quad (B12) \]

where \( e_0 \) is the average photon number from each source. On the image plane, the mutual coherence becomes

\[ \Gamma_{jk} = e_0 (S_{ju_1} S_{ku_1}^* + S_{ju_2} S_{ku_2}^*), \quad (B13) \]

and the average photon number can be expressed as

\[ e = 2\epsilon_0 \eta, \]

where \( \eta = \sum |S_{ju_1}|^2 \) is the quantum efficiency of the imaging system and we have made the reasonable assumption that the efficiency is the same for both sources. Equation (B10) can then be expressed as Eq. (2.2) if we define single-photon kets \( |\psi_j\rangle = \sum |j\rangle \psi(j, u_j)|j\rangle \) with normalized wave functions \( \psi(j, u_j) = S_{ju_1}/\sqrt{\eta} \). Assuming image-plane wave-packet positions \( x_j = x_0 + jdx \), position eigenkets \( |x_j\rangle = |j\rangle/\sqrt{dx} \), and wave functions \( \psi_j(x_j) = \psi(j, u_j)/\sqrt{dx} \), we arrive at Eq. (2.3) by taking the continuous-space limit with infinitesimal \( dx \) [29,30].

**APPENDIX C: QUANTUM METROLOGY: DERIVATION OF EQ. (3.8)**

The quantum Fisher information matrix with respect to \( \rho^{\otimes M} \) proposed by Helstrom [17] is defined as

\[ K_{\mu\nu}(\rho^{\otimes M}) = MK_{\mu\nu}(\rho) = M \text{Re} \text{tr} \mathcal{L}_\mu(\rho) \mathcal{L}_\nu(\rho), \quad (C1) \]

where \( \mathcal{L}_\mu(\rho) \) is a symmetric logarithmic derivative (SLD) of \( \rho \). Writing \( \rho \) in its eigenbasis as

\[ \rho = \sum_j D_j |e_j\rangle \langle e_j|, \quad (C2) \]

\( \mathcal{L}_\mu(\rho) \) can be expressed as [17,87,88]

\[ \mathcal{L}_\mu(\rho) = \sum_{j,k:D_k \neq 0} \frac{2}{D_j + D_k} \langle e_j | \frac{\partial \rho}{\partial \theta_\mu} | e_k \rangle \langle e_k| e_j \rangle. \quad (C3) \]

Given this definition and Eq. (B4), it can be shown that

\[ K(\rho) = \sum_n \pi_n K(\rho_n) \geq \pi_1 K(\rho_1), \quad (C4) \]

as each \( \rho_n \) is in an orthogonal subspace. Since the vacuum state \( \rho_0 = |\text{vac}\rangle \langle \text{vac}| \) contains no information and multiphoton events are rare, the total information will be dominated by that from the one-photon state \( \rho_1 \). We therefore focus on the one-photon component \( \pi_1 K(\rho_1) \) as a tight lower bound on the quantum information and assume, in the following,

\[ K(\rho) \approx \pi_1 K(\rho_1). \quad (C5) \]

With \( \pi_1 \approx \epsilon \) and the probability of multiphoton events being \( O(\epsilon^2) \) according to Eq. (B8), this approximation is accurate to the first order of \( \epsilon \).

To compute the quantum Fisher information matrix \( K(\rho_1) \) according to Eqs. (C1)–(C3), we first need to diagonalize the \( \rho_1 \) in Eqs. (2.2) and (2.3), noting that the eigenvectors should span the supports of \( \rho_1 \) and \( \partial \rho_1/\partial \theta_\mu \). The partial derivative of \( \rho_1 \) with respect to \( X_\mu \) can be expressed as

\[ \frac{\partial \rho_1}{\partial X_\mu} = \frac{\partial D_1}{\partial X_\mu} |e_1\rangle \langle e_1| + \frac{\partial D_2}{\partial X_\mu} |e_2\rangle \langle e_2| \]

\[ + \left[ D_1 \frac{\partial |e_1\rangle \langle e_1|}{\partial X_\mu} + D_2 \frac{\partial |e_2\rangle \langle e_2|}{\partial X_\mu} \right] + \text{H.c.}, \quad (C6) \]

where H.c. denotes the Hermitian conjugate. In addition to the support of \( \rho_1 \) spanned by \( |e_1\rangle \) and \( |e_2\rangle \), we also need to find more eigenvectors that span the support of \( \partial |e_1\rangle /\partial X_\mu \) and \( \partial |e_2\rangle /\partial X_\mu \).

Assuming that \( \psi_\mu(x) = \psi(x - X_\mu) \) and the point-spread function \( \psi(x) \) has an \( x \)-independent phase, we can take \( \psi(x) \) to be real without loss of generality and choose the following orthonormal set of eigenvectors:

\[ |e_1\rangle = \frac{1}{\sqrt{2(1 - \delta)}} (|\psi_1\rangle - |\psi_2\rangle), \]

\[ |e_2\rangle = \frac{1}{\sqrt{2(1 + \delta)}} (|\psi_1\rangle + |\psi_2\rangle), \]

\[ |e_3\rangle = \frac{1}{c_3} \sqrt{\frac{k}{2}} (|\psi_{11}\rangle + |\psi_{22}\rangle) - \frac{\gamma}{\sqrt{1 - \delta}} |e_1\rangle, \]

\[ |e_4\rangle = \frac{1}{c_4} \sqrt{\frac{k}{2}} (|\psi_{11}\rangle - |\psi_{22}\rangle) + \frac{\gamma}{\sqrt{1 + \delta}} |e_2\rangle. \quad (C7) \]

where \( \Delta k^2 \) and \( \gamma \) are given by Eqs. (3.9) and (3.10), respectively,

\[ |\psi_{11}\rangle \approx \frac{1}{\Delta k} \int_{-\infty}^{\infty} dx \frac{\partial \psi(x - X_1)}{\partial X_1} |x\rangle, \]

\[ |\psi_{22}\rangle \approx \frac{1}{\Delta k} \int_{-\infty}^{\infty} dx \frac{\partial \psi(x - X_2)}{\partial X_2} |x\rangle, \]

\[ c_3 \equiv \left( \Delta k^2 + b^2 - \frac{\gamma^2}{1 - \delta} \right)^{1/2}, \]

\[ c_4 \equiv \left( \Delta k^2 + b^2 - \frac{\gamma^2}{1 + \delta} \right)^{1/2}, \]

\[ b^2 \equiv \int_{-\infty}^{\infty} dx \frac{\partial \psi(x - X_1)}{\partial X_1} \frac{\partial \psi(x - X_2)}{\partial X_2}, \]

\[ \delta \equiv \int_{-\infty}^{\infty} dx \psi(x - X_1) \psi(x - X_2), \quad (C8) \]

and the eigenvalues of \( \rho_1 \) are
\[ D_1 = \frac{1 - \delta}{2}, \quad D_2 = \frac{1 + \delta}{2}, \quad D_3 = D_4 = 0. \]  

(C9)

After more algebra, the SLD in Eq. (C3) with respect to the derivative in Eq. (C6) can be expressed as

\[ \mathcal{L}_\mu^{(X)} = \sum_{j,k} \mathcal{L}_{\mu,jk}^{(X)} \langle e_j | e_k \rangle \]  

(C10)

with a real and symmetric matrix \( \mathcal{L}_{\mu,jk}^{(X)} = \mathcal{L}_{\mu,kj}^{(X)} \), the nonzero and unique elements of which are found to be

\[
\mathcal{L}_{1,11}^{(X)} = -\mathcal{L}_{2,11}^{(X)} = \frac{\gamma}{1 - \delta}, \\
\mathcal{L}_{1,12}^{(X)} = \mathcal{L}_{2,12}^{(X)} = \frac{\gamma \delta}{\sqrt{1 - \delta^2}}, \\
\mathcal{L}_{1,13}^{(X)} = -\mathcal{L}_{2,13}^{(X)} = \frac{c_3}{\sqrt{1 - \delta}}, \\
\mathcal{L}_{1,14}^{(X)} = \mathcal{L}_{2,14}^{(X)} = \frac{c_4}{\sqrt{1 - \delta}}, \\
\mathcal{L}_{1,22}^{(X)} = -\mathcal{L}_{2,22}^{(X)} = -\frac{\gamma}{1 + \delta}, \\
\mathcal{L}_{1,23}^{(X)} = \mathcal{L}_{2,23}^{(X)} = \frac{c_3}{\sqrt{1 + \delta}}, \\
\mathcal{L}_{1,24}^{(X)} = -\mathcal{L}_{2,24}^{(X)} = \frac{c_4}{\sqrt{1 + \delta}}.
\]  

(C11)

In terms of the centroid and separation parameters given by \( \theta_1 = (X_1 + X_2)/2 \) and \( \theta_2 = X_2 - X_1 \), the SLDs become

\[ \mathcal{L}_1 = \mathcal{L}_1^{(X)} + \mathcal{L}_2^{(X)}, \quad \mathcal{L}_2 = \frac{\mathcal{L}_2^{(X)} - \mathcal{L}_1^{(X)}}{2}. \]  

(C12)

We can now substitute Eqs. (C9)–(C12) into Eq. (C1) to compute the quantum Fisher information matrix \( K(\rho_1) \). The final result, assuming \( M_{21} \equiv M_{e} = N_e \), is given by Eqs. (3.8) with zero off-diagonal terms.

APPENDIX D: UNKNOWN CENTROID AND MISALIGNMENT

Our analysis of SPADE in Secs. IV and V assumes that the centroid of the two sources is known exactly and the device is optimally aligned with the centroid. For astronomy, it is reasonable to assume that the centroid is known accurately, as extensive telescopic data on stellar objects should be readily available and conventional imaging is accurate in estimating the centroid. Even if the centroid is unknown, stellar objects usually shine long enough for one to collect ample prior information before aligning the SPADE device. For microscopy, however, biological samples may drift more quickly and fluorophores can bleach, giving little time and few photons for one to estimate both parameters. One option, to be explored in future work, is to scan the SPADE device across the image plane in a manner similar to the operation of a confocal microscope [12].

Another option, illustrated in Fig. 10, is to split the optical field by a beam splitter, measure one output port by a photon-counting array, and use the centroid estimate \( \tilde{\theta}_1 \) to align SPADE at the other port.

Let the center of a SPADE device be \( \theta_1 \) and consider \( \theta_1 \neq \tilde{\theta}_1 \) due to misalignment. For a Gaussian point-spread function and the Hermite-Gaussian-basis measurement, Eq. (4.4) should be generalized to

\[
P_1(q) \approx \frac{1}{2} \left[ \exp(-Q_1) \frac{Q_1^q}{q!} + \exp(-Q_2) \frac{Q_2^q}{q!} \right],
\]

\[
Q_1 = \frac{1}{4\sigma^2} \left( \tilde{\theta}_1 - \theta_1 + \frac{\theta_2}{2} \right)^2,
\]

\[
Q_2 = \frac{1}{4\sigma^2} \left( \tilde{\theta}_1 - \theta_1 - \frac{\theta_2}{2} \right)^2.
\]  

(D1)

Define the level of misalignment as

\[ \xi = \frac{|\tilde{\theta}_1 - \theta_1|}{\sigma}. \]  

(D2)

We treat \( \xi \) as a systematic error and \( \theta_2 \) as the parameter of interest for SPADE. Figure 11 plots the resulting Fisher information for several levels of misalignment.
It can be seen that the information degrades with misalignment, but appreciable enhancements over direct imaging are still present even if $\theta_2 \ll \sigma$ and the wave-function overlap is significant, as long as $\xi \ll 1$. Appendix E confirms this conclusion numerically for finite photon numbers.

To attain a tolerable level of misalignment, $\theta_1$ first needs to be estimated and $\hat{\theta}_1$ should be aligned with the estimate. With conventional imaging, the centroid estimation error is near optimal and on the order of $\sigma = \sqrt{N/p}$ in terms of the root-mean-square value, meaning that the number of extra photons $N_1$ needed to attain $\xi$ is roughly

$$N_1 \sim \frac{1}{\xi^2}. \quad (D3)$$

An even more realistic analysis would take $\hat{\theta}_1$ to be a stochastic waveform determined by the centroid measurements and the adaptive alignment control [89].

For binary SPADE, Eqs. (5.1) and (5.2) should be generalized to

$$P_1(q = 0) \approx \frac{1}{2} [\exp(-Q_1) + \exp(-Q_2)], \quad (D4)$$

$$P_1(q > 0) \approx 1 - \frac{1}{2} \exp(-Q_1) - \frac{1}{2} \exp(-Q_2). \quad (D5)$$

Figure 12 plots the Fisher information for misaligned binary SPADE, showing a similar degradation behavior to that in Fig. 11 for nonzero $\xi$. Significant improvements over direct imaging are still possible for small separations and $\xi \ll 1$. For two-parameter estimation, consider the hybrid scheme in Fig. 10, assuming 50-50 beam splitting and binary SPADE, for example. For simplicity, assume that the binary-SPADE output is used only for separation estimation, such that the total information matrix with respect to $\theta_1$ and $\theta_2$ remains diagonal. Compared with direct imaging, the centroid information for the hybrid scheme is halved, viz.,

$$J_{11}^{(\text{hybrid})} = \frac{J_{11}^{(\text{direct})}}{2}, \quad (D6)$$

but the separation information gained by binary SPADE can be appreciable, with

$$J_{22}^{(\text{hybrid})} = \frac{J_{22}^{(\text{direct})}}{2} + \frac{J_{22}^{(b)}}{2}. \quad (D7)$$

The net performance of the hybrid scheme can be quantified in terms of the Cramér-Rao bounds for locating $X_1$ and $X_2$. For a diagonal information matrix $\mathcal{J}$ with respect to $\theta_1$ and $\theta_2$, the bound on the mean-square error $\Sigma^{(x)}$ of estimating either $X_1 = \theta_1 - \theta_2/2$ or $X_2 = \theta_1 + \theta_2/2$ is simply

$$\Sigma^{(x)}_{ss} \geq \frac{1}{\mathcal{J}_{11}} + \frac{1}{4 \mathcal{J}_{22}}, \quad s = 1, 2. \quad (D8)$$

which demonstrates the detrimental effect of small $\mathcal{J}_{22}$ for localization. Figure 13 compares the localization bounds for the hybrid scheme and direct imaging in log-log scale. For small separations, it can be seen that the increased separation information in the hybrid scheme more than...
compensates for the reduced centroid information and allows localization errors substantially lower than those for direct imaging. With a higher \( N_1 = N/2 \), more accurate centroid information from the imaging port can be used to reduce the misalignment at the SPADE port, and performance converging to the ideal \( \xi = 0 \) case in Fig. 13 can be expected for high \( N \).

**APPENDIX E: MONTE CARLO ANALYSIS**

To confirm that the classical Cramér-Rao bounds satisfactorily represent the actual performance of SPADE for finite photon numbers, here we simulate the device output data numerically, apply maximum-likelihood estimation, and investigate the resulting error. To refine our error analysis, we condition our results on the total number of detected photons \( L \), which is obtained after an experiment, rather than the average photon number \( N \) [8]. It is not difficult to show that, conditioned on \( L \), the classical and quantum Fisher information retain their expressions except that \( N \) is replaced by \( L \). The error bounds become

\[
\frac{1}{\mathcal{J}_2^{(\text{Hy})}} \approx \frac{1}{\mathcal{K}_2^{(\text{Hy})}} \approx \frac{4\sigma^2}{L^2}. \tag{E1}
\]

It can also be shown that the sufficient statistic \( \sum qm_m \) in the maximum-likelihood estimator for SPADE in Eq. (4.7) is Poisson with mean \( LQ \), so it is simple to generate samples of the maximum-likelihood estimates \( \hat{Q}_{\text{ML}} \) and \( \hat{\theta}_{\text{ML}} \) according to Eq. (4.7).

Figure 14 plots the simulated mean-square errors, normalized with respect to Eq. (E1), for several values of \( L \). It is intriguing to see that, as \( \theta_2/\sigma \to 0 \), the errors go below the Cramér-Rao bounds. This is a well-known statistical phenomenon called superefficiency [90,91], as the maximum-likelihood estimator here is actually biased for finite samples, and the simple Cramér-Rao bounds considered here need not apply. In asymptotic frequentist statistics, superefficiency is not regarded as an important idea [90] because a superefficient estimator can beat the Cramér-Rao bound only on a set of points with zero measure in the asymptotic limit [90,91], suggesting that any region of superefficiency should shrink for larger samples, as also shown in Fig. 14, and its usefulness is increasingly limited. A Bayesian version of the Cramér-Rao bound [11] can also be used to bound the global or minimax error of any biased or unbiased estimator; the Fisher information still plays a decisive role in the Bayesian bound and its significance as a precision measure remains strong in Bayesian and minimax statistics [92].

For our present purpose, the main point of Fig. 14 is that the errors remain less than twice the Cramér-Rao bound at worst and even offer the pleasant surprise of superefficiency for small separations. The overall closeness of the errors to the Cramér-Rao bounds supports our use of the Fisher information to represent the performance of SPADE.

For binary SPADE, the Fisher information conditioned on \( L \) has the same form as Eq. (5.3), and the Cramér-Rao bound can be expressed as

\[
\frac{1}{\mathcal{J}_2^{(b)}} \approx \frac{4\sigma^2}{L} \frac{1 - \exp(-Q)}{Q \exp(-Q)}. \tag{E2}
\]

The sufficient statistic \( m_0 \) in \( \hat{\theta}_{\text{ML}} \) given by Eq. (5.4) is binomial and also simple to generate. In case \( m_0 = 0 \), we...
SPADE with otherwise the same parameters as those for the errors begin to follow the rising Cramér-Rao bound trends as their counterparts in Fig. 14, and for larger Figure 15 plots the simulated mean-square errors for binary

\[ \frac{\sigma^2}{\bar{\theta}_2} = \frac{1}{\mathcal{J}_{22}^{(b)}} \]

set \( \bar{\theta}_2 = 2\sigma \), the maximum of our considered range of \( \theta_2 \). Figure 15 plots the simulated mean-square errors for binary SPADE with maximum-likelihood estimation, conditioned on \( L \) detected photons. Note that the vertical axis is normalized with respect to \( 4\sigma^2/L \), so the plotted values are the actual errors magnified by \( L/(4\sigma^2) \). Each error is computed by averaging \( 10^5 \) simulations, and the lines connecting the data points are guides for the eye.

FIG. 15. Simulated mean-square errors for binary SPADE with maximum-likelihood estimation, conditioned on \( L \) detected photons. Note that the vertical axis is normalized with respect to \( 4\sigma^2/L \). Each error is computed by averaging \( 10^5 \) simulations, and the lines connecting the data points are guides for the eye.

according to Eq. (E2). This supports our use of the Fisher information to represent the performance of binary SPADE.

To investigate the effect of misalignment described in Appendix D, Fig. 16 plots the simulated errors for binary SPADE with a misalignment level defined in Eq. (D2) given by \( \xi = 0.1 \). The overhead photon number required to achieve \( \xi = 0.1 \) is \( N_1 \sim 100 \) according to Eq. (D3) and negligible if \( L \gg N_1 \). Since \( \xi \) is unknown in reality, the maximum-likelihood estimator used in the simulations assumes zero misalignment for simplicity. Despite the model mismatch, the errors remain close to the Cramér-Rao bound, especially for larger \( L \), and substantially below the bound for direct imaging.

For a given \( N \), \( L \) has a mean \( M e = N \) and standard deviation \( \sqrt{Me(1-e)} \approx \sqrt{N} \). This means that the distribution of \( L \) becomes increasingly sharp around the mean at \( L = N \) for large \( N \), and we can expect the performance for a given \( L = N \) to be an increasingly accurate approximation of the average performance in the given-\( N \), random-\( L \) scenario.

FIG. 16. Simulated mean-square errors of binary SPADE with misalignment level \( \xi = 0.1 \), conditioned on \( L \) detected photons. The maximum-likelihood estimator that assumes no misalignment is used. Note that the vertical axis is normalized with respect to \( 4\sigma^2/L \). Each error is computed by averaging \( 10^5 \) simulations, and the lines connecting the data points are guides for the eye. The errors are substantially below the Cramér-Rao bound for direct imaging (dash-dotted curve).


