Subdiffraction incoherent optical imaging via spatial-mode demultiplexing: Semiclassical treatment

Mankei Tsang[*]

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

I present a semiclassical analysis of a spatial-mode demultiplexing (SPADE) measurement scheme for far-field incoherent optical imaging under the effects of diffraction and photon shot noise. Building on previous results that assume two point sources or the Gaussian point-spread function, I generalize SPADE for a larger class of point-spread functions and evaluate its errors in estimating the moments of an arbitrary subdiffraction object. Compared with the limits to direct imaging set by the Cramér-Rao bounds, the results show that SPADE can offer far superior accuracy in estimating second- and higher-order moments.

I. INTRODUCTION

Recent theoretical and experimental studies have shown that far-field optical methods can substantially improve subdiffraction incoherent imaging [1–20]. While most of the prior works focus on two point sources, Ref. [8] proposes a spatial-mode demultiplexing (SPADE) measurement technique that can enhance the estimation of moments for arbitrary subdiffraction objects. Although the predicted enhancements are promising for applications in both astronomy and fluorescence microscopy, such as size and shape estimation for stellar objects or fluorophore clusters, researchers in those fields may find it difficult to comprehend the quantum formalism used in Ref. [8]. One of the main goals of this work is therefore to introduce a more accessible semiclassical formalism that can reproduce the results there, assuming only a background knowledge of statistical optics on the level of Goodman [21,22] and parameter estimation on the level of Van Trees [23]. The formalism incorporates diffraction, photon shot noise, and—most importantly—coherent optical processing, which enables the enhancements proposed in Refs. [1–20]. This treatment thus sheds light on the physical origin of the enhancements, clarifying that no exotic quantum phenomenon is needed to explain or implement them.

As Ref. [8] assumes the Gaussian point-spread function (PSF) exclusively, another goal of this work is to generalize the results for a larger class of PSFs via the theory of orthogonal polynomials [15,24], affirming that enhancements remain possible in those cases. To set a benchmark for the proposed method, I derive limits to moment estimation via direct imaging in the form of Cramér-Rao bounds (CRBs) [23,25–29], which are original results in their own right and may be of independent interest to image-processing research [30–39]. On a more technical level, this work also investigates the estimation bias introduced by an approximation made in Ref. [8] and assures that it is harmless.

This paper is organized as follows. Section II introduces the background formalism of statistical optics, measurement noise, and CRBs. Section III presents the bounds for moment estimation via direct imaging of a subdiffraction object. Section IV introduces the theory of SPADE for a general class of PSFs and evaluates its biases and errors for moment estimation, showing that giant accuracy enhancements are possible for the second and higher-order moments. Section V revisits the case of Gaussian PSF studied in Ref. [8] and also proposes exactly unbiased estimators in the case of two dimensions. Section VI presents a Monte Carlo analysis to confirm the theory. Section VII concludes the paper, pointing out open questions and future directions. Appendixes A–H deal with mathematical issues that arise in the main text.

II. FORMALISM

A. Statistical optics

Consider an object emitting spatially incoherent light, a diffraction-limited imaging system, as depicted in Fig. 1, and the paraxial theory of quasimonochromatic scalar waves [21,22]. On the image plane, the mutual coherence function, also called the mutual intensity, can be expressed as [21,22]

$$\Gamma(x,x'|\theta) = \int dX \psi(x - X)\psi^*(x' - X)F(X|\theta), \quad (2.1)$$

where $x,x' \in \mathbb{R}^D$ are D-dimensional position vectors on the image plane, X is the object-plane position vector normalized with respect to the magnification factor, $F(X|\theta)$ is the object intensity function, $\theta = (\theta_0, \theta_1, \dots)$ is a vector of unknown parameters to be estimated, and $\psi(x)$ is the field PSF. To simplify the notations, I adopt the multi-index notation described in Appendix A and Ref. [24], such that D can be kept arbitrary, though $D = 1$ or 2 is typical in spectroscopy and imaging. Note that three-dimensional imaging requires a different formalism in the paraxial theory and is beyond the scope of this paper. The mean intensity on the image plane is

$$f(x|\theta) \equiv \Gamma(x,x|\theta) = \int dX |\psi(x - X)|^2 F(X|\theta), \quad (2.2)$$

which is a basic result in statistical optics [21,22].

---

[*]mankei@nus.edu.sg; http://mankei.tsang.googlepages.com/
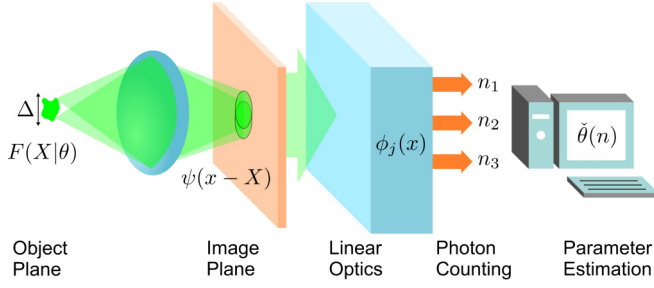
023830-1

FIG. 1. A far-field optical imaging system with additional optical processing after the image plane. See the main text for the definitions of the expressions.

For convenience, I normalize the position vectors with respect to the width of the PSF, such that the PSF width is equal to 1 in this unit. The PSF is assumed to obey the normalization

$$\int dx |\psi(x)|^2 = 1, \qquad (2.3)$$

such that

$$\theta_0 \equiv \int dX\, F(X|\theta) = \int dx\, f(x|\theta) \qquad (2.4)$$

is the mean optical power reaching the image plane.

Instead of intensity measurement on the image plane, consider the use of further linear optics to process the field followed by photon counting in each output channel, as depicted in Fig. 1. The mean power in each output channel can be expressed as

$$p_j(\theta) = \int dx \int dx'\, \phi_j^*(x)\phi_j(x')\Gamma(x,x'|\theta) \qquad (2.5)$$

$$= \int dX \left| \int dx\, \phi_j^*(x)\psi(x - X) \right|^2 F(X|\theta), \quad (2.6)$$

where $\phi_j^*(x)$ is a propagator that couples the image-plane field from position x to the jth output. If the optics after the image plane is passive, power conservation implies that

$$\sum_j p_j(\theta) \leqslant \theta_0. \qquad (2.7)$$

This can be satisfied if the set $\{\phi_j(x)\}$ is orthonormal, viz.,

$$\int dx\, \phi_j(x)\phi_k^*(x) = \delta_{jk}, \qquad (2.8)$$

by virtue of Bessel's inequality [40]. If $\{\phi_j(x)\}$ is also complete in the Hilbert space of image-plane fields, it becomes an orthonormal basis, and Parseval's identity leads to equality for Eq. (2.7) [40]. Physically, Eq. (2.5) implies that each output can be regarded as a projection of the image-plane field in a spatial mode. For example, direct imaging, which measures the spatial intensity on the image plane, can be modeled by taking $\phi_j(x) = \sqrt{dx^{(j)}}\delta(x^{(j)} - x)$, where $x^{(j)}$ is the position of each pixel with infinitesimal area $dx^{(j)}$, such that $p_j(\theta) = f(x^{(j)}|\theta)dx^{(j)}$. A generalization of the measurement model to deal with mode-dependent losses and nonorthogonal mode projections is possible via the concept of positive operator-valued measures [3] but not needed here.

In superresolution research, it is known that image processing can achieve arbitrary resolution if $f(x|\theta)$ is measured exactly and benign assumptions about the object are made [30,31,36]. The caveat is that the techniques are severely limited by noise, so the use of proper statistics is paramount in superresolution studies. For weak incoherent sources, such as astronomical optical sources and microscopic fluorophores, bunching or antibunching is negligible, and it is standard to assume a Poisson model for the photon counts $n = (n_1, n_2, \dots)$ at the output channels [3,22,25,26,28,32,41]. The Poisson distribution is

$$P(n|\theta) = \prod_j \exp[-\tau p_j(\theta)] \frac{[\tau p_j(\theta)]^{n_j}}{n_j!}, \qquad (2.9)$$

where

$$\tau \equiv \frac{\eta T}{\hbar\omega}, \qquad (2.10)$$

$\eta \in [0,1]$ is the detection efficiency, T is the integration time, and $\hbar\omega$ is the photon energy. The most important statistics here are the mean

$$\mathbb{E}(n_j) = \tau p_j(\theta), \qquad (2.11)$$

where $\mathbb{E}$ denotes the expectation with respect to P, and the covariance matrix

$$\mathbb{V}_{jk}(n) \equiv \mathbb{E}(n_j n_k) - \mathbb{E}(n_j)\mathbb{E}(n_k) = \mathbb{E}(n_j)\delta_{jk}, \qquad (2.12)$$

which is signal-dependent. If $\{\phi_j\}$ is an orthonormal basis, the mean photon number detected by the measurement is

$$N \equiv \sum_j \mathbb{E}(n_j) = \tau\theta_0. \qquad (2.13)$$

Conditioned on a total photon number $\sum_j n_j$, n obeys multinomial statistics, and the reconstruction of F via direct imaging becomes the density deconvolution problem in nonparametric statistics; see, for example, Ref. [39] and references therein.

The quantum formalism can arrive at the same Poisson model by assuming that the source is thermal, the mean photon number per spatiotemporal mode is much smaller than 1, and the photon count for each channel is integrated in time over many modes [1,8]. That said, an advantage of the semiclassical model besides simplicity is that it applies to any incoherent source that produces Poisson noise at the output, such as incoherent laser sources [22] and electron microscopy [41], without the need to satisfy all the assumptions of the quantum model.

B. Cramér-Rao bounds (CRBs)

To deal with the signal-dependent nature of Poisson noise, many existing approaches to computational superresolution [30,31,34–36] are inadequate. A more suitable tool to derive fundamental limits is the CRB, which is now standard in astronomy [25–27] and fluorescence microscopy [28,29]. For any estimator $\check{\theta}(n)$ that satisfies the unbiased condition

$$\mathbb{E}(\check{\theta}) = \theta, \qquad (2.14)$$

the mean-square error matrix is equal to its covariance, viz.,

$$\mathrm{MSE}_{\mu\nu}(\check{\theta}, \theta) \equiv \mathbb{E}(\check{\theta}_\mu - \theta_\mu)(\check{\theta}_\nu - \theta_\nu) = \mathbb{V}_{\mu\nu}(\check{\theta}), \qquad (2.15)$$

and the CRB is [23,25–29]

$$\mathrm{MSE}_{\mu\mu}(\check{\theta},\theta) \geqslant \mathrm{CRB}_{\mu\mu}(\theta), \tag{2.16}$$

where

$$\mathrm{CRB}(\theta) \equiv J^{-1}(\theta) \tag{2.17}$$

is the inverse of the Fisher information matrix defined as

$$J_{\mu\nu}(\theta) \equiv \sum_n P(n|\theta) \frac{\partial \ln P(n|\theta)}{\partial \theta_\mu} \frac{\partial \ln P(n|\theta)}{\partial \theta_\nu}. \tag{2.18}$$

In the limit of infinite trials, the maximum-likelihood estimator is asymptotically unbiased with a covariance reaching the CRB [23], so the bound is also useful as a measure of the achievable error in the asymptotic limit.

For the Poisson model, the Fisher information is

$$J_{\mu\nu}(\theta) = \tau \sum_j \frac{1}{p_j(\theta)} \frac{\partial p_j(\theta)}{\partial \theta_\mu} \frac{\partial p_j(\theta)}{\partial \theta_\nu}. \tag{2.19}$$

For example, the information for direct imaging with infinitesimal pixel size is

$$J_{\mu\nu}(\theta) = \tau \int dx \frac{1}{f(x|\theta)} \frac{\partial f(x|\theta)}{\partial \theta_\mu} \frac{\partial f(x|\theta)}{\partial \theta_\nu}. \tag{2.20}$$

The data-processing inequality [42] ensures that increasing the pixel size, or any processing of the image-plane intensity in general, cannot increase the amount of information. A simple extension of Eq. (2.19) for strong thermal sources with super-Poisson statistics can be found in Appendix C of Ref. [16].

An intuitive way of understanding Eq. (2.19) is to regard it as a signal-to-noise ratio: each derivative $\partial p_j / \partial \theta_\mu$ measures the sensitivity of an output to a parameter, while the denominator p_j is proportional to the Poisson variance and indicates the noise level. The form of Eq. (2.19) hence suggests that any parameter-insensitive background in p_j should be minimized. The nonlinear dependence of the Fisher information on p_j complicates the analysis, but also hints that coherent optical processing may lead to nontrivial effects.

The Bayesian CRB (BCRB) can be used to set more general limits for any biased or unbiased estimator [6,23,43–45]. Define the Bayesian mean-square error as

$$\mathrm{BMSE}(\check{\theta}) \equiv \int d\theta \, \Pi(\theta) \mathrm{MSE}(\check{\theta},\theta), \tag{2.21}$$

where $\Pi(\theta)$ is a prior probability density. For a prior that vanishes on the boundary of its domain, the BCRB is

$$\mathrm{BMSE}_{\mu\mu}(\check{\theta}) \geqslant \mathrm{BCRB}_{\mu\mu}, \tag{2.22}$$

$$\mathrm{BCRB} \equiv (\tilde{J} + K)^{-1}, \tag{2.23}$$

where

$$\tilde{J} \equiv \int d\theta \, \Pi(\theta) J(\theta) \tag{2.24}$$

is the Fisher information averaged over the prior and

$$K_{\mu\nu} \equiv \int d\theta \frac{1}{\Pi(\theta)} \frac{\partial \Pi(\theta)}{\partial \theta_\mu} \frac{\partial \Pi(\theta)}{\partial \theta_\nu} \tag{2.25}$$

is the prior information. Other Bayesian bounds for more general priors can be found in Ref. [45]. The BCRB also

applies to the worst-case error $\sup_\theta \mathrm{MSE}_{\mu\mu}(\check{\theta},\theta)$ for minimax estimation [6,44], since

$$\sup_\theta \mathrm{MSE}_{\mu\mu}(\check{\theta},\theta) \geqslant \mathrm{BMSE}_{\mu\mu}(\check{\theta}) \tag{2.26}$$

for any $\Pi(\theta)$, and the prior can be chosen to tighten the bound [6,44].

The BCRB is close to the CRB if $J(\theta)$ is constant in the domain of the prior, such that $\tilde{J} = J$, and the prior information K is negligible relative to $\tilde{J}$, such that

$$\mathrm{BCRB} = (\tilde{J} + K)^{-1} \approx \tilde{J}^{-1} = J^{-1}. \tag{2.27}$$

A counterexample is the problem of two-point resolution [6], where J vanishes at a point in the parameter space and the BCRB becomes very sensitive to the choice of prior, as mentioned later in Sec. III B.

III. LIMITS TO DIRECT IMAGING

A. Error bounds

Define the object moments

$$\theta_\mu \equiv \int dX \, X^\mu F(X|\theta), \quad \mu \in \mathbb{N}_0^D, \tag{3.1}$$

as the parameters of interest. Note that the moments are unnormalized, unlike the definition in Ref. [8]. Under general conditions, the set of moments uniquely determine F [24], so there is little loss of generality with this parametrization. I will focus on moment estimation hereafter and not the pointwise reconstruction of F, however, for two reasons: the moments are more directly related to many useful parameters in practice, such as the brightness, location, size, and shape of an object [27,46], while the reconstruction of F without further prior information is ill-posed and a forlorn task in practice when noise is present [30,31,34,35,39], even with the techniques introduced in this work.

Expanding $|\psi(x - X)|^2$ in a Taylor series, the mean image given by Eq. (2.2) can be expressed in terms of θ as

$$f(x|\theta) = \sum_\mu \frac{\theta_\mu}{\mu!} (-\partial)^\mu |\psi(x)|^2. \tag{3.2}$$

The Fisher information given by Eq. (2.20) becomes

$$J_{\mu\nu}(\theta) = \tau \int dx \frac{[(-\partial)^\mu |\psi(x)|^2][(-\partial)^\nu |\psi(x)|^2]}{\mu!\nu! f(x|\theta)}. \tag{3.3}$$

Appendix B shows that this can be inverted analytically to give

$$\mathrm{CRB}_{\mu\nu}(\theta) = \frac{\theta_0^2}{N} \sum_{\xi,\zeta} (C^{-1})_{\mu\xi} M_{\xi\zeta}(\theta)(C^{-1})_{\nu\zeta}, \tag{3.4}$$

where N is the mean photon number given by Eq. (2.13),

$$M_{\mu\nu}(\theta) \equiv \frac{1}{\theta_0} \int dx \, f(x|\theta) x^{\mu+\nu} \tag{3.5}$$

is the normalized image moment matrix, the C matrix is defined as

$$C_{\mu\nu} \equiv \frac{1}{\nu!} \int dx |\psi(x)|^2 \partial^\nu x^\mu \qquad (3.6)$$

$$= \begin{cases} 0 & \text{if any } \nu_j > \mu_j, \\ \binom{\mu}{\nu} \Lambda_{\mu-\nu} & \text{otherwise,} \end{cases} \qquad (3.7)$$

and

$$\Lambda_\mu \equiv \int dx |\psi(x)|^2 x^\mu \qquad (3.8)$$

is a moment of the PSF. The lower-triangular property of C indicated by Eq. (3.7) means that C^{-1} is also lower-triangular and the low-order elements of the CRB can be computed from a finite number of low-order elements of M and C.

To proceed, I focus on the subdiffraction regime, which I define as the scenario in which the object support width Δ is much smaller than the PSF width. To be specific, the width is defined by

$$F(X|\theta) = 0 \text{ if } \max_j |X_j| > \frac{\Delta}{2}, \qquad (3.9)$$

and the subdiffraction regime is defined by the condition

$$\Delta \ll 1 \qquad (3.10)$$

in the dimensionless unit assumed here. This can be regarded as the extreme opposite to the sparse regime commonly assumed in compressed sensing [33–37] and can be ensured by prior information in practice. For example, a spot that resembles the PSF in a prior image indicates a subdiffraction object and can be studied further via the framework here; such spots are of course commonly found in both astronomical and microscopic imaging. In fluorescence microscopy, the subdiffraction support can even be enforced via stimulated-emission depletion (STED) [47], and the theory here can help STED microscopy gain more information about each spot beyond θ_0.

In the subdiffraction regime, the moments observe a magnitude hierarchy with respect to the order $|\mu|$, as

$$|\theta_\mu| \leqslant \int dX |X^\mu| F(X|\theta) \leqslant \theta_0 \left(\frac{\Delta}{2} \right)^{|\mu|}, \qquad (3.11)$$

and I can combine Eqs. (3.2), (3.5), and (3.8) to obtain

$$M_{\mu\nu}(\theta) = \frac{1}{\theta_0} \sum_{\xi=0}^{\mu+\nu} \theta_\xi \binom{\mu+\nu}{\xi} \Lambda_{\mu+\nu-\xi} \qquad (3.12)$$

$$= \Lambda_{\mu+\nu} + O(\Delta). \qquad (3.13)$$

In other words, the image is so blurred that it resembles the PSF to the zeroth order, and the image moments approach those of the PSF. The CRB hence becomes

$$\text{CRB}_{\mu\nu} = \frac{\theta_0^2}{N} \left[\sum_{\xi,\zeta} (C^{-1})_{\mu\xi} \Lambda_{\xi+\zeta} (C^{-1})_{\nu\zeta} + O(\Delta) \right]. \qquad (3.14)$$

This is the central result of Sec. III. An unbiased estimator that can approach this bound is described in Appendix C.

To set a more general limit for any biased or unbiased estimator, consider the BCRB described in Sec. II B. Since the
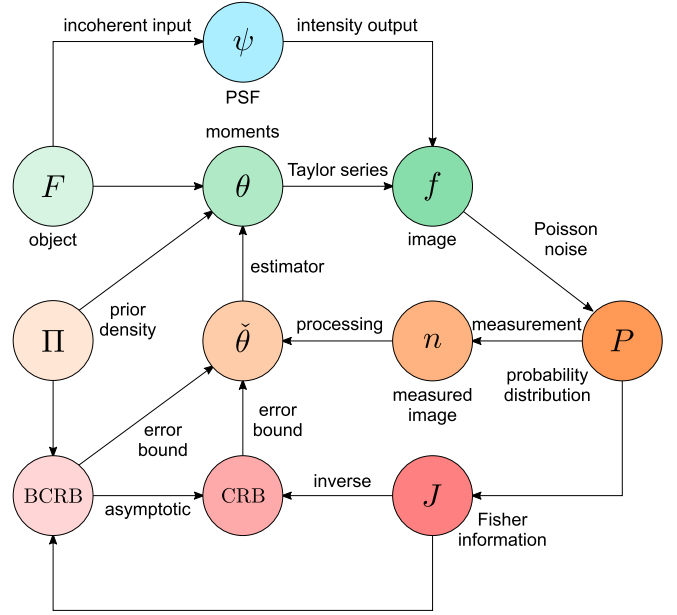

FIG. 2. A flowchart that summarizes the relationships among the various quantities defined for direct imaging in Sec. III.

Fisher information given by the inverse of Eq. (3.14) depends only on θ_0 and not the other parameters to the leading order, the average information $\tilde{J}$ defined by Eq. (2.24) is relatively insensitive to the choice of prior in the subdiffraction regime. For any reasonable prior that gives a finite prior information K, a long enough integration time can then make $\tilde{J}$ much larger than K in Eq. (2.23), leading to BCRB $\approx$ CRB, if θ_0 is replaced by a suitable prior value. The two bounds hence give similar results here in the asymptotic limit. Figure 2 summarizes the relationships among the various quantities defined for direct imaging in this section.

B. Special cases

The low-order elements of Eqs. (3.3) and (3.4) can be used to reproduce a few well-known results. For example, the CRB with respect to θ_0 can be derived from Eq. (3.4) and is given by

$$\text{CRB}_{00} = \frac{\theta_0^2}{N}, \qquad (3.15)$$

which is equal to the textbook result. Another example is point-source localization [26,28], for which known results can be retrieved from Eq. (3.3) by defining the location parameters as θ_μ/θ_0 for $|\mu| = |\nu| = 1$. To see this, assume $D = 1$ for simplicity, and the information with respect to $X = \theta_1/\theta_0$ in the $\Delta \to 0$, $f(x|\theta) \to \theta_0 |\psi(x)|^2$ limit becomes

$$J^{(X)} = \left(\frac{\partial \theta_1}{\partial X} \right)^2 J_{11} \to N \int dx \frac{[\partial |\psi(x)|^2]^2}{|\psi(x)|^2}, \qquad (3.16)$$

which is exact for one point source [26,28].

Considering $|\mu| = |\nu| = 2$, Eq. (3.3) can also reproduce the results in Refs. [41,48–50] regarding sub-Rayleigh two-point separation estimation. To see this, assume $D = 1$ again and that the centroid of the two point sources is at the origin. The second moment is then related to the separation d by $\theta_2 = \theta_0 d^2/4$. The

information with respect to d becomes

$$J^{(d)} = \left(\frac{\partial \theta_2}{\partial d}\right)^2 J_{22} \to \frac{Nd^2}{16} \int dx \frac{[\partial^2 |\psi(x)|^2]^2}{|\psi(x)|^2}. \quad (3.17)$$

This can be compared with a direct calculation of the information by considering the mean image

$$f(x|d) = \frac{\theta_0}{2}[|\psi(x - d/2)|^2 + |\psi(x + d/2)|^2], \quad (3.18)$$

and approximating it for sub-Rayleigh $d \ll 1$ as [16,41,49]

$$f(x|d) \approx \theta_0 \left[|\psi(x)|^2 + \frac{d^2}{8}\partial^2 |\psi(x)|^2\right]. \quad (3.19)$$

The information is then

$$J^{(d)} = \tau \int dx \frac{1}{f}\left(\frac{\partial f}{\partial d}\right)^2 \approx \frac{Nd^2}{16} \int dx \frac{[\partial^2 |\psi(x)|^2]^2}{|\psi(x)|^2}, \quad (3.20)$$

which coincides with Eq. (3.17). The vanishing $J^{(d)}$ and divergent $\text{CRB}^{(d)} = 1/J^{(d)}$ for $d \ll 1$ were first reported in Refs. [41,48–50] and called Rayleigh's curse in Ref. [1]. The BCRB becomes very sensitive to the choice of prior and produces a markedly different result from the CRB when applied to the worst-case error [6]. This issue depends on the parametrization [44] and does not arise for the moment parameters, however.

In the absence of a specific parametric model or equality parameter constraints [51], the full information matrix should be considered, and the CRB given by Eq. (3.4), which results from inverting the full information matrix, is a tighter limit [45] for general objects. Appendix D presents a limit of Eq. (3.4) when diffraction can be ignored, while Eq. (3.14) should be used in the subdiffraction regime.

This section has established fundamental limits to direct imaging in the subdiffraction and shot-noise-limited regime. The next sections show that coherent optical processing can beat them.

IV. SPATIAL-MODE DEMULTIPLEXING (SPADE)

A. Point-spread-function-adapted basis

References [1–20] have shown that SPADE, a technique of linear optics and photon counting with respect to a judiciously chosen basis of spatial modes, can substantially improve subdiffraction imaging. To generalize the use of the TEM basis in Ref. [8], I consider the point-spread-function-adapted (PAD) basis proposed by Rehacek *et al.* for the two-point problem [15] and apply it to more general objects. Denote the PAD basis by

$$\{\phi_q(x); q \in \mathbb{N}_0^D\}, \quad (4.1)$$

where the spatial modes are more conveniently defined in the spatial-frequency domain. Defining

$$\Phi_q(k) \equiv \frac{1}{(2\pi)^{d/2}} \int dk\, \phi_q(x) \exp(-ikx), \quad (4.2)$$

$$\Psi(k) \equiv \frac{1}{(2\pi)^{d/2}} \int dk\, \psi(x) \exp(-ikx), \quad (4.3)$$

$\Phi_q(k)$ can be expressed as

$$\Phi_q(k) = (-i)^{|q|} g_q(k) \Psi(k), \quad (4.4)$$

$$g_q(k) \equiv \sum_r G_{qr} k^r, \quad (4.5)$$

where $\{g_q(k); q \in \mathbb{N}_0^D\}$ is a set of real orthogonal polynomials with $|\Psi(k)|^2$ as the weight function [24], G is an invertible matrix that satisfies the lower-triangular property

$$G_{qr} = 0 \text{ if } r > q, \quad (4.6)$$

and the indices follow a total and degree-respecting order that obeys

$$r \geqslant q \Rightarrow |r| \geqslant |q|. \quad (4.7)$$

See Appendix B for more details about orthogonal polynomials. The polynomials are assumed to satisfy the orthonormal condition

$$\int dk\, \Phi_q^*(k)\Phi_r(k) = \int dk |\Psi(k)|^2 g_q(k) g_r(k) = \delta_{qr}, \quad (4.8)$$

which also ensures that $\{\phi_q\}$ is orthonormal. The completeness of $\{\phi_q\}$ can be proved along the lines of Ref. [24] but is not essential here. As $\phi_0(x) = \psi(x)$ and each higher-order mode in real space is a sum of $\psi(x)$ derivatives given by

$$\phi_q(x) = (-i)^{|q|} g_q(-i\partial)\psi(x), \quad (4.9)$$

the PAD basis can be regarded as a generalization of the binary SPADE concept in Ref. [1] and the derivative-mode concept in Ref. [14].

In terms of the PAD basis, I can define a mutual coherence matrix as

$$\Gamma_{qq'}(\theta) \equiv \int dX\, h_q(X) h_{q'}^*(X) F(X|\theta), \quad (4.10)$$

$$h_q(X) \equiv \int dx\, \phi_q^*(x)\psi(x - X). \quad (4.11)$$

In particular, SPADE in terms of the PAD basis gives a set of output channels with powers

$$p_q(\theta) = \int dX |h_q(X)|^2 F(X|\theta) = \Gamma_{qq}(\theta), \quad (4.12)$$

and the Poisson photon counts $\{n_q; q \in \mathbb{N}_0^D\}$ have expected values

$$\mathbb{E}(n_q) = \tau_0 p_q(\theta), \quad \tau_0 \equiv \frac{\eta_0 T}{\hbar\omega}, \quad (4.13)$$

where η_0 is the efficiency of the PAD-basis measurement. An unbiased estimator of Γ_{qq} is

$$\check{\Gamma}_{qq} = \frac{n_q}{\tau_0}, \quad (4.14)$$

and its variance is

$$\mathbb{V}(\check{\Gamma}_{qq}) = \frac{\Gamma_{qq}}{\tau_0}. \quad (4.15)$$

In the context of the Gaussian PSF, Refs. [8,13] found that $p_q(\theta)$ is sensitive only to some of the object moments. To estimate the other moments, Ref. [8] further proposes measurements that access the off-diagonal elements of Γ. To measure an off-diagonal $\Gamma_{qq'}$, take two spatial modes with

indices q and q' from the PAD basis and interfere them, such that the outputs correspond to projections into the spatial modes

$$\varphi_{qq'}^+(x) = \frac{1}{\sqrt{2}}[\phi_q(x) + \phi_{q'}(x)], \qquad (4.16)$$

$$\varphi_{qq'}^-(x) = \frac{1}{\sqrt{2}}[\phi_q(x) - \phi_{q'}(x)], \qquad (4.17)$$

which I call interferometric-PAD (iPAD) modes. The powers at the two outputs are

$$p_{qq'}^+ = \frac{\Gamma_{qq} + \Gamma_{q'q'}}{2} + \text{Re}\,\Gamma_{qq'}, \qquad (4.18)$$

$$p_{qq'}^- = \frac{\Gamma_{qq} + \Gamma_{q'q'}}{2} - \text{Re}\,\Gamma_{qq'}. \qquad (4.19)$$

The photon counts, denoted by $n_{qq'}^+$ and $n_{qq'}^-$, have expected values

$$\mathbb{E}(n_{qq'}^+) = \tau_s p_{qq'}^+, \quad \mathbb{E}(n_{qq'}^-) = \tau_s p_{qq'}^-, \quad \tau_s \equiv \frac{\eta_s T}{\hbar\omega}, \qquad (4.20)$$

where η_s denotes the efficiency of the measurement that includes these two projections. Assume further that $|\Psi(k)|^2$ is centrosymmetric, as defined by

$$|\Psi(k)|^2 = |\Psi(-k)|^2, \qquad (4.21)$$

such that G, $h_q(X)$, and $\Gamma_{qq'}$ are all real, as shown in Appendix E and assumed hereafter. An unbiased estimator of $\Gamma_{qq'}$ is then

$$\check{\Gamma}_{qq'} = \frac{n_{qq'}^+ - n_{qq'}^-}{2\tau_s}, \qquad (4.22)$$

with

$$\mathbb{V}(\check{\Gamma}_{qq'}) = \frac{\Gamma_{qq} + \Gamma_{q'q'}}{4\tau_s}. \qquad (4.23)$$

The estimators $\check{\Gamma}_{qq'}$ given by Eqs. (4.14) and (4.22) will be used in Sec. IV B to construct moment estimators.

Since the iPAD modes are not orthogonal to the PAD modes, they cannot belong to the same orthonormal basis. This means that, if projections into both PAD and iPAD modes are desired, multiple measurements in different bases are needed and must be performed on different photons. This can be done either sequentially in time via configurable interferometers or on different beamsplitted parts of the light. If each measurement has an efficiency η_s, energy conservation mandates that

$$\sum_s \eta_s \leqslant 1. \qquad (4.24)$$

B. Moment estimation

To relate Γ to the object moments, use Eqs. (4.2)–(4.4) to rewrite the propagator $h_q(X)$ in Eq. (4.11) as

$$h_q(X) = i^{|q|} \int dk\,|\Psi(k)|^2 g_q(k)\exp(-ikX) \qquad (4.25)$$

$$= i^{|q|} \int dk\,|\Psi(k)|^2 g_q(k)\sum_r \frac{(-ik)^r X^r}{r!} \qquad (4.26)$$

$$= \sum_r H_{qr} X^r, \qquad (4.27)$$

where

$$H_{qr} \equiv \frac{i^{|q|}}{r!} \int dk\,|\Psi(k)|^2 g_q(k)(-ik)^r \qquad (4.28)$$

$$= \frac{i^{|q|}(-i)^{|r|}}{r!}(G^{-1})_{rq}, \qquad (4.29)$$

$$(H^{-1})_{qr} = q!\,i^{|q|}(-i)^{|r|} G_{rq}, \qquad (4.30)$$

as shown in Appendix E. Since G^{-1} and G are lower-triangular, H and H^{-1} are upper-triangular, satisfying

$$H_{qr} = 0, \ (H^{-1})_{qr} = 0 \text{ if } r < q. \qquad (4.31)$$

Substituting Eq. (4.27) into Eq. (4.10), $\Gamma_{qq'}$ can be related to the moments by

$$\Gamma_{qq'} = \sum_{r,r'} H_{qr} H_{q'r'} \theta_{r+r'}, \qquad (4.32)$$

which shows that each $\Gamma_{qq'}$ is sensitive to a combination of moments with orders at least as high as $|q + q'|$. Given the magnitudes of θ according to Eq. (3.11), the magnitude of $\Gamma_{qq'}$ can be expressed as

$$\Gamma_{qq'} = \theta_0 O(\Delta^{|q+q'|}), \qquad (4.33)$$

and the variances of the estimators given by Eqs. (4.15) and (4.23) become

$$\mathbb{V}(\check{\Gamma}_{qq'}) = \frac{\theta_0^2}{N_s} O(\Delta^{2\min(|q|,|q'|)}), \qquad (4.34)$$

$$N_s \equiv \tau_s \theta_0 = \frac{\eta_s T \theta_0}{\hbar\omega}. \qquad (4.35)$$

Equations (4.33) and (4.34) will be used to evaluate the errors of moment estimation.

Instead of computing the CRB and relying on asymptotic arguments, here I construct explicit moment estimators and evaluate their errors directly to demonstrate the achievable performance of SPADE. To begin, consider the inverse of Eq. (4.32) given by

$$\theta_{q+q'} = \sum_{r,r'} (H^{-1})_{qr}(H^{-1})_{q'r'}\Gamma_{rr'}, \qquad (4.36)$$

which implies that an unbiased estimator of $\theta_{q+q'}$ can be constructed from unbiased estimators of $\Gamma_{pp'}$ given by Eqs. (4.14) and (4.22), viz.,

$$\check{\theta}_{q+q'} = \sum_{r,r'} (H^{-1})_{qr}(H^{-1})_{q'r'}\check{\Gamma}_{rr'}. \qquad (4.37)$$

This estimator may not be realizable, however, as it may not be possible to group the needed projections into a reasonable number of bases. A fortuitous exception occurs for the Gaussian PSF, as elaborated later in Sec. V C.

To find a simpler estimator, I focus on the class of separable PSFs given by

$$|\Psi(k)|^2 = \prod_j |\Psi^{(j)}(k_j)|^2, \qquad (4.38)$$

where each $|\Psi^{(j)}(k_j)|^2$ is a one-dimensional function. Defining

$$g_{q_j}^{(j)}(k_j) = \sum_{r_j} G_{q_j r_j}^{(j)} k_j^{r_j} \qquad (4.39)$$

as the orthogonal polynomials with respect to each $|\Psi^{(j)}(k_j)|^2$, the natural orthogonal polynomials in the multivariate case are their products, viz.,

$$g_q(k) = \prod_j g_{q_j}^{(j)}(k_j). \tag{4.40}$$

As each $G_{q_j r_j}^{(j)}$ is lower-triangular, I obtain the condition

$$G_{qr} = \prod_j G_{q_j r_j}^{(j)} = 0 \text{ if any } r_j > q_j. \tag{4.41}$$

It follows from Eqs. (4.29) and (4.30) that H and H^{-1} are also separable and given by

$$H_{qr} = \prod_j \frac{i^{q_j}(-i)^{r_j}}{r_j!}[(G^{(j)})^{-1}]_{r_j q_j}, \tag{4.42}$$

$$\left(H^{-1}\right)_{qr} = \prod_j q_j! i^{q_j}(-i)^{r_j} G_{r_j q_j}^{(j)}. \tag{4.43}$$

Using the property

$$(H^{-1})_{qr} = 0 \text{ if any } q_j > r_j, \tag{4.44}$$

I can rewrite the sums in Eq. (4.36) as

$$\sum_r = \sum_{r_1=q_1}^{\infty} \cdots \sum_{r_D=q_D}^{\infty} \tag{4.45}$$

and obtain

$$\theta_{q+q'} = (H^{-1})_{qq}(H^{-1})_{q'q'}\Gamma_{qq'}$$
$$+ \sum_{|r+r'|>|q+q'|}(H^{-1})_{qr}(H^{-1})_{q'r'}\Gamma_{rr'}, \tag{4.46}$$

which consists of one $\theta_0 O(\Delta^{|q+q'|})$ term and higher-order terms, as ranked by Eq. (4.33). To evaluate the magnitude of the higher-order terms, note that, for a centrosymmetric $|\Psi(k)|^2$, $(H^{-1})_{qr} \propto G_{rq} = 0$ if $|r| = |q|+1, |q|+3, \ldots$ [24], so

$$\sum_{|r+r'|>|q+q'|}(H^{-1})_{qr}(H^{-1})_{q'r'}\Gamma_{rr'} = \theta_0 O(\Delta^{|q+q'|+2}), \tag{4.47}$$

which is smaller than the leading-order term by two orders of magnitude. A simplified estimator, involving only one $\check{\Gamma}_{qq'}$, can then be constructed as

$$\check{\theta}'_{q+q'} = (H^{-1})_{qq}(H^{-1})_{q'q'}\check{\Gamma}_{qq'} = \frac{\check{\Gamma}_{qq'}}{H_{qq}H_{q'q'}}, \tag{4.48}$$

where the last step uses the fact that $(H^{-1})_{qq} = 1/H_{qq}$ for a triangular matrix. The bias is then the negative of Eq. (4.47), viz.,

$$\mathbb{E}(\check{\theta}'_{q+q'}) - \theta_{q+q'} = \theta_0 O(\Delta^{|q+q'|+2}). \tag{4.49}$$

Figure 3 summarizes the relationships among the various quantities defined in this section, while Appendix G discusses a generalization of the estimator for nonseparable PSFs.

Given Eq. (4.34), the variance of the estimator is

$$\mathbb{V}(\check{\theta}'_{q+q'}) = \frac{\mathbb{V}(\check{\Gamma}_{qq'})}{H_{qq}^2 H_{q'q'}^2} = \frac{\theta_0^2}{N_s}O(\Delta^{2\min(|q|,|q'|)}). \tag{4.50}$$
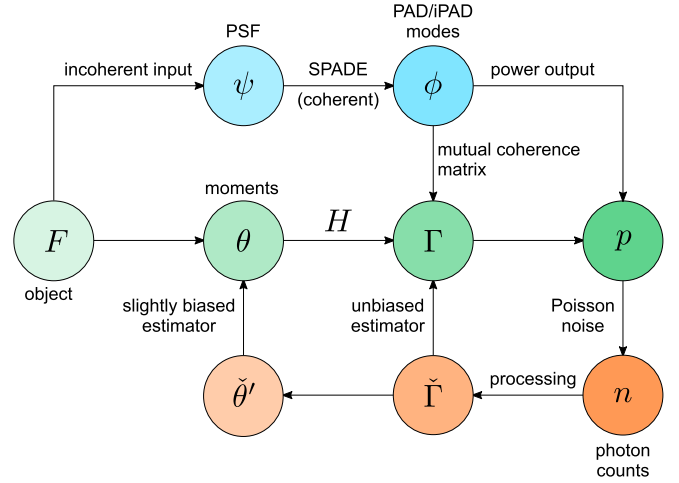

FIG. 3. A flowchart that summarizes the relationships among the various quantities defined for SPADE in Sec. IV.

To minimize the variance for a given moment θ_μ with $\mu = q + q'$, $\min(|q|,|q'|)$ should be made as high as possible. This can be accomplished by choosing

for each $j \in \{1, 2, \ldots, D\}$,

$$q_j = \begin{cases} \mu_j/2 & \text{if } \mu_j \text{ is even,} \\ \lfloor \mu_j/2 \rfloor & \text{if } \mu_j \text{ is the first odd number,} \\ \lfloor \mu_j/2 \rfloor & \text{if } \mu_j \text{ is odd and the last choice was } \lfloor \rfloor, \\ \lceil \mu_j/2 \rceil & \text{if } \mu_j \text{ is odd and the last choice was } \lceil \rceil. \end{cases} \tag{4.51}$$

The alternating floor ($\lfloor \rfloor$) and ceil ($\lceil \rceil$) operations keep $|q|$ high without exceeding $|q'|$. If $|\mu|$ is even, μ has an even number of odd elements, then $|q| = |q'| = |\mu|/2$. If $|\mu|$ is odd, μ has an odd number of odd elements, then $|q| = (|\mu|-1)/2$ and $|q'| = (|\mu|+1)/2$. Hence one can achieve

$$\min(|q|,|q'|) = \left\lfloor \frac{|\mu|}{2} \right\rfloor, \tag{4.52}$$

$$\mathbb{V}(\check{\theta}'_\mu) = \frac{\theta_0^2}{N_s}O(\Delta^{2\lfloor|\mu|/2\rfloor}), \tag{4.53}$$

and the mean-square error becomes

$$\text{MSE}(\check{\theta}'_\mu, \theta_\mu) = \mathbb{V}(\check{\theta}'_\mu) + [\mathbb{E}(\check{\theta}'_\mu) - \theta_\mu]^2 \tag{4.54}$$

$$= \frac{\theta_0^2}{N_s}O(\Delta^{2\lfloor|\mu|/2\rfloor}) + \theta_0^2 O(\Delta^{2|\mu|+4}). \tag{4.55}$$

Compared with the CRB for direct imaging given by Eq. (3.14), Eq. (4.55) can be much lower in the $\Delta \ll 1$ subdiffraction regime if $|\mu| \geqslant 2$, the bias is negligible, and η_s is on the same order of magnitude as the direct-imaging efficiency. This is the central result of Sec. IV. The conclusion holds also from the Bayesian or minimax perspective, since the BCRB for direct imaging is close to the CRB in the asymptotic limit, as argued in Sec. III A, while Eq. (4.55) also applies to the Bayesian or worst-case error for SPADE if θ_0 is replaced by a suitable prior value.

A heuristic explanation of the enhancements is as follows. Recall that Poisson noise is signal-dependent, and any

TABLE I. A list of measurement bases for moment estimation with a $D = 2$ separable PSF and their spatial modes. Measurement in each basis can provide a set of moment estimators $\{\check{\theta}'_\mu\}$ according to Eq. (4.48), where the set of $\mu = (\mu_1, \mu_2)$ indices is listed in the last two columns. The case of $D = 1$ can be retrieved by considering the PAD, iPAD1, and iPAD4 bases and q_1 and μ_1 only.

Basis	Spatial modes	q_1	q_2	$\mu_1 = q_1 + q'_1$	$\mu_2 = q_2 + q'_2$
PAD	$\phi_q(x)$	$\mathbb{N}_0$	$\mathbb{N}_0$	$2\mathbb{N}_0$	$2\mathbb{N}_0$
iPAD1	$\varphi^\pm_{qq'}(x); q' = q + (1,0)$	$2\mathbb{N}_0$	$\mathbb{N}_0$	$4\mathbb{N}_0 + 1$	$2\mathbb{N}_0$
iPAD2	$\varphi^\pm_{qq'}(x); q' = q + (0,1)$	$\mathbb{N}_0$	$2\mathbb{N}_0$	$2\mathbb{N}_0$	$4\mathbb{N}_0 + 1$
iPAD3	$\varphi^\pm_{qq'}(x); q' = q + (1,-1)$	$\mathbb{N}_0$	$2\mathbb{N}_0 + 1$	$2\mathbb{N}_0 + 1$	$4\mathbb{N}_0 + 1$
iPAD4	$\varphi^\pm_{qq'}(x); q' = q + (1,0)$	$2\mathbb{N}_0 + 1$	$\mathbb{N}_0$	$4\mathbb{N}_0 + 3$	$2\mathbb{N}_0$
iPAD5	$\varphi^\pm_{qq'}(x); q' = q + (0,1)$	$\mathbb{N}_0$	$2\mathbb{N}_0 + 1$	$2\mathbb{N}_0$	$4\mathbb{N}_0 + 3$
iPAD6	$\varphi^\pm_{qq'}(x); q' = q + (1,-1)$	$\mathbb{N}_0$	$2\mathbb{N}_0 + 2$	$2\mathbb{N}_0 + 1$	$4\mathbb{N}_0 + 3$

background in the signal increases the variance. In the subdiffraction regime, the direct image is so blurred that it resembles the PSF $|\psi(x)|^2$, and the fundamental mode $\phi_0(x) = \psi(x)$ acts as a background and the main contributor of noise. With SPADE, on the other hand, each moment estimator is designed to use spatial modes with the highest possible orders. The isolation from the lower-order modes, including the fundamental, substantially reduces the background and improves the signal-to-noise ratio.

C. Multimoment estimation

The remaining question is the number of bases needed to estimate all moments. For $D = 1$, three bases are enough: a measurement in the PAD basis provides

$$\{\check{\Gamma}_{qq}, q \in \mathbb{N}_0\} \text{ and } \{\check{\theta}'_\mu; \mu \in 2\mathbb{N}_0\}, \quad (4.56)$$

where $2\mathbb{N}_0 = \{0, 2, 4, \dots\}$, a measurement in the basis $\{\varphi^\pm_{q,q+1}(x); q \in 2\mathbb{N}_0\}$, provides

$$\{\check{\Gamma}_{q,q+1}, q \in 2\mathbb{N}_0\} \text{ and } \{\check{\theta}'_\mu; \mu \in 4\mathbb{N}_0 + 1\}, \quad (4.57)$$

where $4\mathbb{N}_0 + 1 = \{1, 5, 9, \dots\}$, and a measurement in the basis $\{\varphi^\pm_{q,q+1}(x); q \in 2\mathbb{N}_0 + 1\}$ provides

$$\{\check{\Gamma}_{q,q+1}, q \in 2\mathbb{N}_0 + 1\} \text{ and } \{\check{\theta}'_\mu; \mu \in 4\mathbb{N}_0 + 3\}, \quad (4.58)$$

where $2\mathbb{N}_0 + 1 = \{1, 3, 5, \dots\}$ and $4\mathbb{N}_0 + 3 = \{3, 7, 11, \dots\}$. If the light is split for measurements in all three basis, the condition of energy conservation given by Eq. (4.24) implies

$$\min(\eta_s) \leqslant \tfrac{1}{3}. \quad (4.59)$$

For $D = 2$, seven bases—defined by Table I and illustrated by Fig. 4—can do the job. I call these bases PAD and iPAD1–iPAD6, which generalize the TEM and iTEM1–iTEM6 bases proposed in Ref. [8] for the Gaussian PSF. Energy conservation now implies

$$\min(\eta_s) \leqslant \tfrac{1}{7} \quad (4.60)$$

if measurements in all seven bases are performed. The essential point is that the penalty in efficiency for multimoment estimation is only a constant factor, and significant enhancements over direct imaging remain possible.
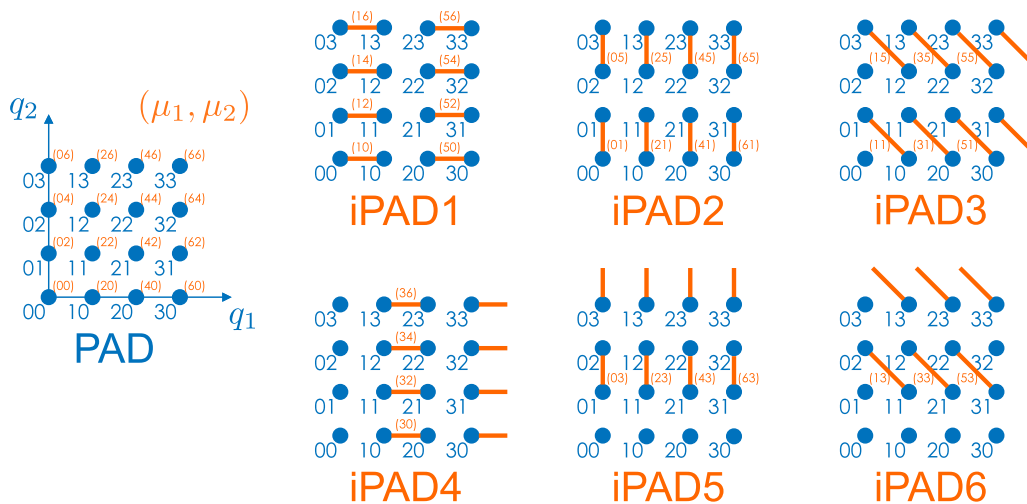

FIG. 4. An illustration of the PAD and iPAD1–iPAD6 bases in the mode-index space. Each dot in the (q_1, q_2) space represents a PAD mode, and the PAD modes form the PAD basis on the left. For each iPAD basis, a line connecting two dots represents an interference between the two PAD modes, producing two new modes that replace the original PAD modes in the basis. Each bracketed pair of numbers denote the order $(\mu_1, \mu_2) = (q_1 + q'_1, q_2 + q'_2)$ of the estimator $\check{\theta}'_\mu$ that a projection can provide via Eq. (4.48). In each iPAD basis, the unconnected dots represent the PAD modes that complete the basis and can also be measured to provide extra information.

D. Criterion for informative estimation

A word of caution is in order: even with SPADE, there are severe resolution limits. This is because the moments are inherently small parameters in the subdiffraction regime according to Eq. (3.11), and the error needs to be much smaller than the prior range of the parameter for the estimation to be informative. To evaluate the usefulness of an estimation relative to prior information, I adopt the Bayesian perspective [23,45,52] and consider the Bayesian error given by Eq. (2.21). In the absence of measurements, the error is determined by the prior and given by

$$\mathrm{BMSE}_{\mu\mu}^{(\Pi)} \equiv \mathbb{E}^{(\Pi)}[\theta_\mu - \mathbb{E}^{(\Pi)}(\theta_\mu)]^2 \leqslant \theta_0^2 \left(\frac{\Delta}{2}\right)^{2|\mu|}, \quad (4.61)$$

where $\mathbb{E}^{(\Pi)}$ denotes the expectation with respect to $\Pi(\theta)$, the upper bound comes from Eq. (3.11), and θ_0 is assumed to be given for simplicity. Using the bound as a conservative estimate of the prior error, a rule of thumb for informative estimation is

$$\frac{\mathrm{BMSE}_{\mu\mu}}{\theta_0^2 (\Delta/2)^{2|\mu|}} \ll 1. \quad (4.62)$$

The small prior error places a stringent requirement on the post-measurement error. For direct imaging, assuming the asymptotic limit where the BCRB is close to the CRB given by Eq. (3.14), the fractional BCRB is

$$\frac{\mathrm{BCRB}_{\mu\mu}}{\mathrm{BMSE}_{\mu\mu}^{(\Pi)}} \approx \frac{\mathrm{CRB}_{\mu\mu}}{\mathrm{BMSE}_{\mu\mu}^{(\Pi)}} = \frac{O(\Delta^{-2|\mu|})}{N}. \quad (4.63)$$

This value grows exponentially with the order $|\mu|$, meaning that the estimation of higher-order moments requires exponentially more photons to become informative.

For SPADE, an achievable Bayesian error can be obtained by averaging $\mathrm{MSE}(\check\theta'_\mu, \theta_\mu)$, and the magnitude is also given by Eq. (4.55). The fractional error becomes

$$\frac{\mathrm{BMSE}_{\mu\mu}}{\mathrm{BMSE}_{\mu\mu}^{(\Pi)}} = \frac{O(\Delta^{2\lfloor|\mu|/2\rfloor - 2|\mu|})}{N_s} + O(\Delta^4). \quad (4.64)$$

The $O(\Delta^4)$ relative bias is always much smaller than 1, but the fractional variance still grows with $|\mu|$ exponentially. Compared with direct imaging, the exponent is reduced for $|\mu| \geqslant 2$ and not as many photons are needed to achieve a small fractional error for a given moment, but higher-order moments remain more difficult to estimate.

This consideration suggests that SPADE is most useful for scenarios that depend on only a few low-order moments. For example, the two-point problem studied in Refs. [1–7,9–18,20] requires moments up to the second order only [8], the case of two unequal sources studied in Refs. [19,20] requires moments up to the third, and parametric object models with size and shape parameters [8,38] can also be related to low-order moments.

V. GAUSSIAN POINT-SPREAD FUNCTION

A. Direct imaging

For an illustrative example of the general theory, consider the Gaussian PSF,

$$\psi(x) = \frac{1}{(2\pi)^{d/4}} \exp\left(-\frac{||x||^2}{4}\right), \quad (5.1)$$

which is a common assumption in fluorescence microscopy [28,53]. The Hermite polynomials can be used to compute the CRB in the limit of $\Delta \to 0$, as shown in Appendix F. The result is

$$\mathrm{CRB}_{\mu\nu} \to \frac{\theta_0^2}{N} \mu! \delta_{\mu\nu}, \quad (5.2)$$

which coincides with the $D = 2$ theory in Ref. [8].

B. SPADE

The PSF in the spatial-frequency domain is

$$\Psi(k) = \left(\frac{2}{\pi}\right)^{d/4} \exp(-||k||^2). \quad (5.3)$$

A set of orthogonal polynomials with respect to $|\Psi(k)|^2$ is defined by

$$g_q(k) = \frac{1}{\sqrt{q!}} \mathrm{He}_q(2k), \quad (5.4)$$

and the PAD mode functions become

$$\Phi_q(k) = \left(\frac{2}{\pi}\right)^{d/4} \frac{(-i)^{|q|}}{\sqrt{q!}} \mathrm{He}_q(2k) \exp(-||k||^2), \quad (5.5)$$

$$\phi_q(x) = \frac{1}{(2\pi)^{d/4}\sqrt{q!}} \mathrm{He}_q(x) \exp\left(-\frac{||x||^2}{4}\right). \quad (5.6)$$

The PAD basis in this case is simply the TEM basis, as expected. The propagator given by Eq. (4.11) can be computed analytically with the help of the generating function for Hermite polynomials [54,55]; the result is

$$h_q(X) = H_{qq} \exp\left(-\frac{||X||^2}{8}\right) X^q, \quad (5.7)$$

$$H_{qq} = \frac{1}{2^{|q|}\sqrt{q!}}. \quad (5.8)$$

The mutual coherence matrix Γ defined by Eq. (4.10) becomes

$$\Gamma_{qq'} = H_{qq} H_{q'q'} \int dX \exp\left(-\frac{||X||^2}{4}\right) X^{q+q'} F(X|\theta). \quad (5.9)$$

Unbiased estimators of $\Gamma_{qq'}$ can be constructed from projections in the PAD and iPAD spatial modes according to Eqs. (4.14) and (4.22); the iPAD modes are called iTEM modes in Ref. [8]. The estimator variances are given by Eqs. (4.15) and (4.23), with magnitudes given by Eq. (4.34).

To estimate a given moment θ_μ, q and $q' = \mu - q$ can be chosen according to Eq. (4.51), the simplified estimator given by Eq. (4.48) can be used, and the error then agrees with Eq. (4.55). These results again agree with Ref. [8], except that Ref. [8] neglects the contribution of bias to the mean-square error and therefore does not include the second term in Eq. (4.55).

C. Exactly unbiased estimator

For $D = 2$, the PAD and iPAD1–iPAD6 bases described by Table I and Fig. 4 become the TEM and iTEM1–iTEM6 bases proposed in Ref. [8], and the estimator given by Eq. (4.48) is equivalent to the ones proposed in Ref. [8]. Interestingly, it is possible to go further than Ref. [8] and construct exactly unbiased moment estimators from these measurements. First note that Eq. (5.9) offers a shortcut to express each moment in terms of Γ as follows:

$$\theta_{q+q'} = \int dX \exp\left(\frac{||X||^2}{4}\right) \exp\left(-\frac{||X||^2}{4}\right) X^{q+q'} F(X|\theta) \tag{5.10}$$

$$= \int dX \sum_r \frac{X^{2r}}{r! 4^{|r|}} \exp\left(-\frac{||X||^2}{4}\right) X^{q+q'} F(X|\theta) \tag{5.11}$$

$$= \sum_r \frac{1}{r! 4^{|r|}} \int dX \exp\left(-\frac{||X||^2}{4}\right) X^{q+q'+2r} F(X|\theta) \tag{5.12}$$

$$= \sum_r \frac{\Gamma_{q+r,q'+r}}{r! 4^{|r|} H_{q+r,q+r} H_{q'+r,q'+r}}. \tag{5.13}$$

Combining Eqs. (4.48) and (5.13), it can then be shown that the estimator

$$\check{\theta}_\mu = \sum_r \frac{\check{\theta}'_{\mu+2r}}{r! 4^{|r|}} \tag{5.14}$$

is exactly unbiased. To construct

$$\{\check{\theta}_\mu; \mu \in (2\mathbb{N}_0) \times (2\mathbb{N}_0)\}, \tag{5.15}$$

one simply needs $\{\check{\theta}'_\mu; \mu \in (2\mathbb{N}_0) \times (2\mathbb{N}_0)\}$ from the PAD basis. To construct

$$\{\check{\theta}_\mu; \mu \in (2\mathbb{N}_0 + 1) \times (2\mathbb{N}_0)\}, \tag{5.16}$$

one needs $\{\check{\theta}'_\mu; \mu \in (2\mathbb{N}_0 + 1) \times (2\mathbb{N}_0)\}$, which can be obtained from the iPAD1 and iPAD4 bases. Similarly, to construct

$$\{\check{\theta}_\mu; \mu \in (2\mathbb{N}_0) \times (2\mathbb{N}_0 + 1)\}, \tag{5.17}$$

one needs $\{\check{\theta}'_\mu; \mu \in (2\mathbb{N}_0) \times (2\mathbb{N}_0 + 1)\}$, which can be obtained from the iPAD2 and iPAD5 bases. Finally, to construct

$$\{\check{\theta}_\mu; \mu \in (2\mathbb{N}_0 + 1) \times (2\mathbb{N}_0 + 1)\}, \tag{5.18}$$

one needs $\{\check{\theta}'_\mu; \mu \in (2\mathbb{N}_0 + 1) \times (2\mathbb{N}_0 + 1)\}$, which can be obtained from the iPAD3 and iPAD6 bases. The error matrix of the unbiased estimator becomes

$$\text{MSE}_{\mu\nu}(\check{\theta}, \theta) = \mathbb{V}_{\mu\nu}(\check{\theta}) = \frac{\theta_0^2}{\min(N_s)} O(\Delta^{2\lfloor |\mu|/2 \rfloor}) \delta_{\mu\nu}, \tag{5.19}$$

which remains on the same order of magnitude as the variance of the simplified estimator in Eq. (4.55), while the bias contribution is no longer present. The number of bases needed to achieve enhanced and exactly unbiased multimoment estimation for other PSFs and dimensions remains an open question.
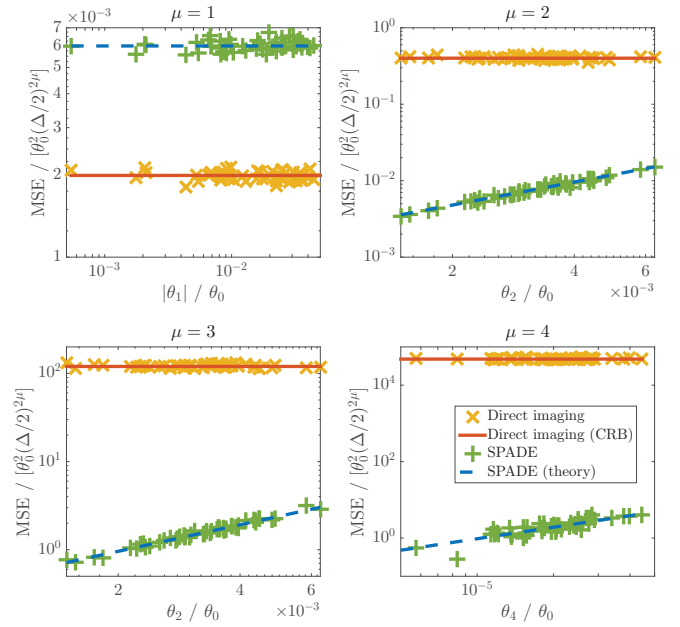

FIG. 5. Simulated and theoretical errors of moment estimation via direct imaging and SPADE for the Gaussian PSF. The discrete points are from the Monte Carlo simulations, and the lines are from the analytic theory. All axes are normalized, dimensionless, and in logarithmic scale. See the main text for details.

VI. NUMERICAL DEMONSTRATION

I now present Monte Carlo simulations to corroborate the theory. Assume $D = 1$. Each simulated object is an ensemble of $S = 5$ point sources with randomly generated positions $\{X_\sigma; \sigma = 1, \ldots, S\}$ within the interval

$$|X_\sigma| \leqslant \frac{\Delta}{2}, \quad \Delta = 0.2, \tag{6.1}$$

such that

$$F(X|\theta) = \frac{\theta_0}{S} \sum_{\sigma=1}^{S} \delta(X - X_\sigma). \tag{6.2}$$

50 objects are generated for each PSF under study. For direct imaging, I assume that the mean photon number is $N = 50\,000$, the pixel size is $dx = 0.1$, and 1000 samples of Poisson images are generated for each object. The estimator described in Appendix C is applied to each sample to estimate the moments θ_μ for $\mu = 1, 2, 3, 4$ (θ_0 can be estimated by summing all the photon counts and the results are trivial). The sample errors with respect to the true parameters are averaged to approximate the expected values. The averaged errors are then plotted for two different PSFs in Figs. 5 and 6 and compared with the CRB given by Eq. (3.14), omitting the $O(\Delta)$ correction.

To simulate SPADE according to Sec. IV, measurements in three different bases are simulated. The first basis is

$$\{\phi_0(x), \phi_1(x), \phi_2(x)\}, \tag{6.3}$$

with the simulated photon counts denoted by $\{n_0, n_1, n_2\}$, the second basis is

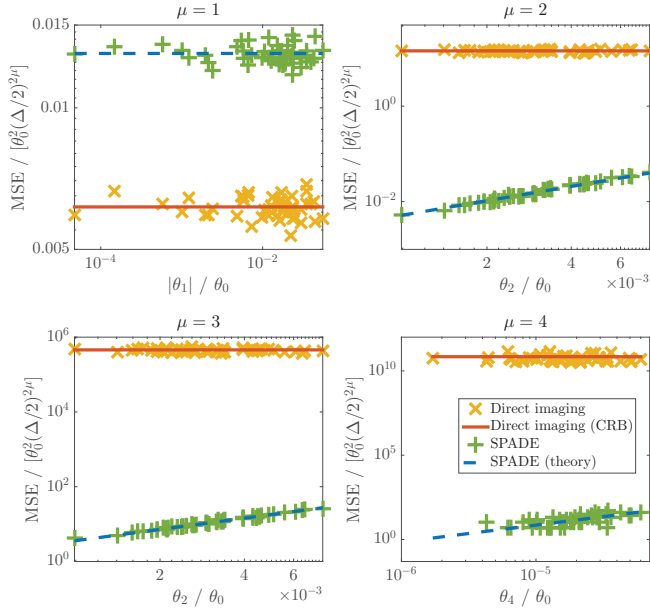$$\{\varphi_{01}^+(x), \varphi_{01}^-(x), \phi_2(x)\}, \tag{6.4}$$

FIG. 6. Simulated and theoretical errors of moment estimation via direct imaging and SPADE for the bump aperture given by Eq. (6.12). The format of the plots is identical to that of Fig. 5. See the main text for details.
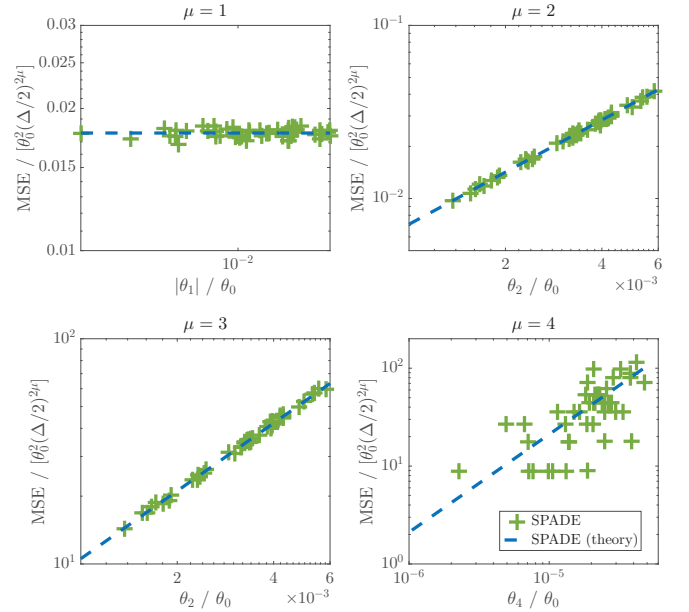

FIG. 7. Simulated and theoretical errors of moment estimation via SPADE for the rectangle aperture given by Eq. (6.13). The format of the plots is identical to that of Figs. 5 and 6. Note that the spread of errors for $\mu = 4$ looks more severe because the range of the vertical axis is smaller than those of the other $\mu = 4$ plots in Figs. 5 and 6. Also, relatively few photons from a subdiffraction object are coupled into the $\phi_2(x)$ mode, so the error itself has a high variance and more samples would be required for the average errors to get closer to the expected values.

with the photon counts denoted by $\{n_{01}^+, n_{01}^-, n_2'\}$, and the third basis is

$$\{\phi_0(x), \varphi_{12}^+(x), \varphi_{12}^-(x)\}, \qquad (6.5)$$

with the photon counts denoted by $\{n_0', n_{12}^+, n_{12}^-\}$. The light is split equally among the three measurements, such that $N_s = N/3$. All photons in higher-order modes are neglected.

To estimate the moments with SPADE, I use the simplified but biased estimator given by Eq. (4.48), with q given by Eq. (4.51). Using Eq. (4.22) for $\check{\Gamma}_{01}$, the estimator of θ_1 becomes

$$\check{\theta}_1' = \frac{\check{\Gamma}_{01}}{H_{00}H_{11}} = \frac{n_{01}^+ - n_{01}^-}{2H_{00}H_{11}\tau_s}. \qquad (6.6)$$

The estimator is applied to 1000 samples of the simulated photon counts for each object. The sample errors with respect to the true parameters are averaged and compared with the analytic expression

$$\mathrm{MSE}_{11} \approx \mathbb{V}(\check{\theta}_1') = \frac{\mathbb{V}(\check{\Gamma}_{01})}{H_{00}^2 H_{11}^2} \approx \frac{\Gamma_{00}}{4H_{00}^2 H_{11}^2 \tau_s} \approx \frac{\theta_0}{4H_{11}^2 \tau_s}, \qquad (6.7)$$

which neglects the bias and applies the approximations

$$\Gamma_{qq} + \Gamma_{q'q'} \approx \Gamma_{qq} \approx H_{qq}^2 \theta_{2q} \qquad (6.8)$$

to Eqs. (4.23) and (4.32). Similarly,

$$\check{\theta}_2' = \frac{n_1}{H_{11}^2 \tau_s}, \quad \mathrm{MSE}_{22} \approx \frac{\theta_2}{H_{11}^2 \tau_s}, \qquad (6.9)$$

$$\check{\theta}_3' = \frac{n_{12}^+ - n_{12}^-}{2H_{11}H_{22}\tau_s}, \quad \mathrm{MSE}_{33} \approx \frac{\theta_2}{4H_{22}^2 \tau_s}. \qquad (6.10)$$

To estimate θ_4, I use both of the photon counts that come from the two $\phi_2(x)$ projections to obtain

$$\check{\theta}_4' = \frac{n_2 + n_2'}{2H_{22}^2 \tau_s}, \quad \mathrm{MSE}_{44} \approx \frac{\theta_4}{2H_{22}^2 \tau_s}. \qquad (6.11)$$

There is no need to specify θ_0, τ, or τ_s individually if the errors are normalized with respect to θ_0^2. The simulated errors and the analytic expressions are plotted in Figs. 5–7 against the relevant parameters in log-log scale for the three PSFs. The three PSFs in the spatial-frequency domain under study and the associated PAD modes are plotted in Fig. 8.

Figure 5 plots the results for the Gaussian PSF described in Sec. V. The simulated errors all match the theory, despite the approximations in the analytic expressions. In particular, the agreement confirms that the contribution of bias to the errors of SPADE is negligible. For $\mu = 1$, SPADE uses one-third of the photons only, and its errors are three times those of direct imaging. For higher moments, however, SPADE outperforms direct imaging by orders of magnitude.

It is important to note that the plotted mean-square errors are normalized with respect to $\theta_0^2 (\Delta/2)^{2\mu}$, which is the square of the prior limit given by Eq. (3.11), and only the normalized errors for $\mu = 1, 2$ go significantly below 1. According to the discussion in Sec. IV D, this implies that only the estimation for $\mu \leqslant 2$ is informative, while the estimation for $\mu \geqslant 3$ would require a lot more photons to become informative. The high variances of the estimators for $\mu \geqslant 3$ also suggest that, for the given photon number, replacing them with Bayesian estimators
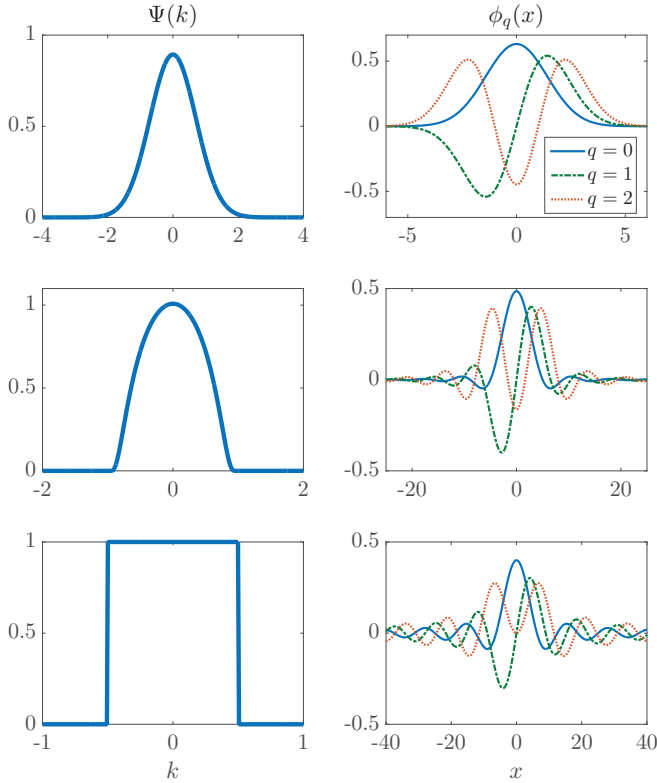
FIG. 8. Left column: the aperture functions for the three PSFs under study: the Gaussian (first row), the bump given by Eq. (6.12) (second row), and the rectangle given by Eq. (6.13) (third row). Right column: the PAD modes $\phi_0(x)$, $\phi_1(x)$, and $\phi_2(x)$ for each PSF. All the axes follow the dimensionless units defined in the main text.

[23,45,52] can reduce their errors to the vicinity of the prior levels given by Eq. (4.61), although the bias will go up a lot.

The second PSF under study is the "bump" aperture function [40]

$$\Psi(k) = \begin{cases} \Psi(0) \exp\left(-\frac{k^2}{1-k^2}\right), & |k| < 1, \\ 0, & |k| \geqslant 1, \end{cases} \quad (6.12)$$

where $\Psi(0) \approx 1.0084$ is a normalization constant. The compact support models a hard bandwidth limit, while the infinite differentiability of $\Psi(k)$ ensures that all the moments of $|\psi(x)|^2$ are finite and the direct-imaging theory in Sec. III is valid, as discussed in Appendix H. The simulated errors, plotted in Fig. 6, behave similarly to those in the Gaussian case, except that the direct-imaging errors are substantially higher for higher moments. The enhancements by SPADE appear even bigger, though not big enough to bring the errors for $\mu \geqslant 3$ down to the informative regime for the given photon number.

The final PSF is the textbook rectangle aperture function

$$\Psi(k) = \begin{cases} 1, & |k| < 1/2, \\ 0, & |k| \geqslant 1/2. \end{cases} \quad (6.13)$$

The second and higher moments of $|\psi(x)|^2$ are infinite, meaning that the direct-imaging theory in Sec. III is inapplicable, as discussed in Appendix H. Fortunately, the orthogonal polynomials with respect to $|\Psi(k)|^2$ and therefore the PAD basis remain well-defined [15]. Figure 7 plots the results for

SPADE, which are similar to those for the bump aperture in Fig. 6. Although these results have no direct-imaging limits to compare with, the earlier results on the two-point problem for this PSF [1,14,15] suggest that significant improvements remain likely.

VII. CONCLUSION

The semiclassical treatment complements the quantum approach in Ref. [8] by offering a shortcut to the Poisson photon-counting model for incoherent sources, passive linear optics, and photon counting. Besides pedagogy, this work generalizes the results in Refs. [1–20] for more general objects and PSFs in the context of moment estimation, demonstrating that the giant enhancements by SPADE are not limited to the case of two point sources or Gaussian PSF considered in prior works.

Many open problems remain, such as extensions for more general PSFs, more complex objects, and three-dimensional imaging; the effect of excess statistical and systematic errors, such as dark counts, aberrations, turbulence, and nonparaxial effects [56]; the application of more advanced Bayesian or minimax statistics [23,33–39,45]; and the quantum optimality of the measurements [1,4–10,15,18–20]. Experimental implementation is another important future direction. For proof-of-concept demonstrations, it should be possible to use the same setups described in Refs. [11–14] to estimate at least the second moments of more general objects. For practical applications in astronomy and fluorescence microscopy, efficient demultiplexing for broadband sources is needed. The technical challenge is by no means trivial, but the experimental progress on spatial-mode demultiplexers has been encouraging [11–14,57–63], and the promise of giant imaging enhancements using simply far-field linear optics should motivate further efforts.

APPENDIX A: MULTI-INDEX NOTATION

A D-dimensional vector of continuous variables is written as

$$x = (x_1, x_2, \ldots, x_D) \in \mathbb{R}^D. \quad (A1)$$

For such a vector, the following notations are assumed:

$$dx \equiv \prod_{j=1}^{D} dx_j, \quad \int dx \equiv \int_{\mathbb{R}^D} dx,$$

$$\delta(x - x') \equiv \prod_{j=1}^{D} \delta(x_j - x'_j), \quad \partial_x \equiv \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_D}\right),$$

$$kx \equiv \sum_{j=1}^{D} k_j x_j, \quad ||x||^2 \equiv xx. \quad (A2)$$

If the subscript is omitted in ∂, derivatives with respect to x are assumed.

A vector of integer indices, on the other hand, is defined as

$$\mu = (\mu_1, \mu_2, \ldots, \mu_D) \in \mathbb{N}_0^D. \tag{A3}$$

For such a vector, the following notations are assumed:

$$0 \equiv (0, \ldots, 0), \quad |\mu| \equiv \sum_{j=1}^{D} |\mu_j|,$$

$$\sum_{\mu} \equiv \sum_{\mu \in \mathbb{N}_0^D}, \quad \sum_{\mu=\nu}^{\xi} \equiv \sum_{\mu=\nu_1}^{\xi_1} \cdots \sum_{\mu=\nu_D}^{\xi_D},$$

$$\mu! \equiv \prod_{j=1}^{D} \mu_j!, \quad \binom{\mu}{\nu} \equiv \frac{\mu!}{(\mu-\nu)!\nu!}. \tag{A4}$$

Note that the one-norm is assumed for index vectors. Other useful notations include

$$x^\mu \equiv \prod_{j=1}^{D} x_j^{\mu_j}, \quad \partial_x^\mu \equiv \prod_{j=1}^{D} \frac{\partial^{\mu_j}}{\partial x_j^{\mu_j}}. \tag{A5}$$

APPENDIX B: CRB FOR DIRECT IMAGING

It is useful to define a Hilbert space

$$\mathcal{H} \equiv \overline{\text{span}} \left\{ b_\mu(x); \mu \in \mathbb{N}_0^D \right\} \tag{B1}$$

with respect to

$$b_\mu(x) \equiv \frac{(-\partial)^\mu |\psi(x)|^2}{\mu! \tilde{f}(x|\theta)}, \quad \tilde{f}(x|\theta) \equiv \frac{f(x|\theta)}{\theta_0}, \tag{B2}$$

and the weighted inner product

$$\langle u, v \rangle \equiv \int dx \, \tilde{f}(x|\theta) u(x) v(x), \tag{B3}$$

where $\overline{\text{span}}$ is the closed linear span inside the $L^2(\tilde{f})$ space [24,40] and $\tilde{f}(x|\theta)$ is the normalized image. In other words, any function in $\mathcal{H}$ can be expressed as a linear combination of $\{b_\mu(x)\}$. Equation (3.3) becomes

$$J_{\mu\nu} = \frac{\tau}{\theta_0} \langle b_\mu, b_\nu \rangle. \tag{B4}$$

This can be inverted with the help of orthogonal polynomials. Define

$$a \equiv \left\{ a_\mu(x); \mu \in \mathbb{N}_0^D \right\}, \tag{B5}$$

where $a_\mu(x)$ is a real polynomial with degree $|\mu|$ and the orthonormal condition is

$$\langle a_\mu, a_\nu \rangle = \delta_{\mu\nu}. \tag{B6}$$

For orthogonal polynomials to exist, the moment matrix M given by Eq. (3.5) should be positive-definite [24], or equivalently

$$\int dx \, \tilde{f}(x|\theta) \mathcal{P}^2(x) > 0 \tag{B7}$$

for any polynomial $\mathcal{P}$. The strict positiveness can be satisfied as long as the support of $\tilde{f}(x|\theta)$ is an infinite set, as $\mathcal{P}^2(x)$ has a finite number of zeros only.

The orthogonal polynomials can be computed by applying the Gram-Schmidt procedure to the set of monomials

$\{x^\mu; \mu \in \mathbb{N}_0^D\}$ if the set is totally ordered [24]. For $D = 1$, the natural order $\{1, x, x^2, \ldots\}$ leads to a unique set of orthogonal polynomials for a given weight function. For $D \geqslant 2$, however, the situation is more complicated. A useful requirement is that the order should respect the degree in the sense of

$$\nu \geqslant \mu \Rightarrow |\nu| \geqslant |\mu|. \tag{B8}$$

An example is the graded lexicographical order, defined by

$$\nu > \mu \Leftrightarrow |\nu| > |\mu|, \text{ or if } |\nu| = |\mu|,$$
$$\text{the first nonzero } \nu_j - \mu_j > 0. \tag{B9}$$

For $D = 2$, for example, the order is

$$(0,0) < (0,1) < (1,0) < (0,2) < (1,1) < (2,0)$$
$$< \cdots (0,|\mu|) < (1,|\mu|-1) < \cdots < (|\mu|,0) < \cdots, \tag{B10}$$

but one should see in this example that indices with the same total degree $|\mu|$ may be ordered in other ways and there is no single compelling choice; a different choice will lead to a different set of orthogonal polynomials. In the following, I assume simply that a degree-respecting order has been chosen; the analysis is valid regardless of the choice.

Express each polynomial as

$$a_\mu(x) = \sum_\nu A_{\mu\nu} x^\nu, \tag{B11}$$

where A is a matrix that satisfies the lower-triangular property

$$A_{\mu\nu} = 0 \text{ if } \nu > \mu. \tag{B12}$$

Combining Eqs. (3.5), (B6), and (B11), I obtain

$$\sum_{\xi,\zeta} A_{\mu\xi} M_{\xi\zeta} A_{\nu\zeta} = \delta_{\mu\nu}. \tag{B13}$$

Given a total order of the indices, the matrices can be rasterized into two-dimensional matrices. Equation (B14) can then be written more compactly as

$$AMA^\top = I, \tag{B14}$$

where $\top$ denotes the matrix transpose and I is the identity matrix. As M is positive-definite, A can be obtained from the Cholesky decomposition

$$M = LL^\top, \tag{B15}$$

where L is a real lower-triangular matrix with positive diagonal elements [64]. Since the diagonal elements of a triangular matrix are also its eigenvalues, L is invertible, L^{-1} is also lower-triangular, and setting

$$A = L^{-1} \tag{B16}$$

leads to

$$M = (A^{-1})(A^{-1})^\top, \tag{B17}$$

which satisfies Eq. (B14).

To invert Eq. (B4), I also need to prove that a is an orthonormal basis in $\mathcal{H}$. The orthonormality given by Eq. (B6) is satisfied by definition, while the completeness follows from

the fact that the only function $u(x) = \sum_\nu \lambda_\nu b_\nu(x)$ in $\mathcal{H}$ that is orthogonal to a in the sense of

$$\langle a_\mu, u \rangle = \sum_\nu \langle a_\mu, b_\nu \rangle \lambda_\nu = 0, \quad \mu \in \mathbb{N}_0^D, \tag{B18}$$

is the zero function, provided that

$$B_{\mu\nu} \equiv \langle a_\mu, b_\nu \rangle = \frac{1}{\nu!} \int dx\, a_\mu(x)(-\partial)^\nu |\psi(x)|^2 \tag{B19}$$

is an invertible matrix. To prove so, apply integration by parts to Eq. (B19) to obtain

$$B_{\mu\nu} = \frac{1}{\nu!} \int dx |\psi(x)|^2 \partial^\nu a_\mu(x) = \sum_\xi A_{\mu\xi} C_{\xi\nu}, \tag{B20}$$

$$B = AC, \tag{B21}$$

where C is defined by Eq. (3.6). Since A is invertible, it suffices to prove that C is also invertible. Consider the term $\partial^\nu x^\mu$ in $C_{\mu\nu}$. $\nu > \mu$ in a degree-respecting order implies $|\nu| > |\mu|$, or $|\nu| = |\mu|$ and $\nu \neq \mu$. In either case, there exists at least one $\nu_j > \mu_j$ that makes $\partial^\nu x^\mu$ vanish, resulting in

$$C_{\mu\nu} = 0 \text{ if } \nu > \mu, \tag{B22}$$

meaning that C is lower-triangular. The eigenvalues of C are then the diagonal elements and given by

$$C_{\mu\mu} = \frac{1}{\mu!} \int dx |\psi(x)|^2 \partial^\mu x^\mu = \int dx |\psi(x)|^2 = 1. \tag{B23}$$

Hence C is invertible. Since both A and C are lower-triangular and invertible, $B = AC$ is also lower-triangular and invertible, and

$$B^{-1} = C^{-1} A^{-1} \tag{B24}$$

is lower-triangular as well.

I can now use the a basis to express Eq. (B4) as

$$J_{\mu\nu} = \frac{\tau}{\theta_0} \sum_\xi \langle b_\mu, a_\xi \rangle \langle a_\xi, b_\nu \rangle = \frac{\tau}{\theta_0} \sum_\xi B_{\xi\nu} B_{\xi\nu}. \tag{B25}$$

In matrix form,

$$J = \frac{\tau}{\theta_0} B^\top B, \tag{B26}$$

and the CRB becomes

$$\text{CRB} = J^{-1} = \frac{\theta_0}{\tau} B^{-1} (B^{-1})^\top \tag{B27}$$

$$= \frac{\theta_0}{\tau} C^{-1} M (C^{-1})^\top, \tag{B28}$$

where I have applied Eqs. (B24) and (B17).

APPENDIX C: A MOMENT ESTIMATOR FOR DIRECT IMAGING

Let a' be a set of orthogonal polynomials with respect to $\tilde{f}(x|\theta')$ at a trial parameter value θ'. Each polynomial can be expressed as

$$a'_\mu(x) = \sum_\nu A'_{\mu\nu} x^\mu. \tag{C1}$$

The orthonormal condition is

$$\int dx\, \tilde{f}(x|\theta') a'_\mu(x) a'_\nu(x) = \delta_{\mu\nu}. \tag{C2}$$

The mean image given by Eq. (3.2) can be expanded in the a' basis as

$$f(x|\theta) = \tilde{f}(x|\theta') \sum_\mu a'_\mu(x) \beta_\mu, \tag{C3}$$

which is an example of orthogonal-series expansion in non-parametric density estimation [65,66], and

$$\beta_\mu \equiv \int dx\, f(x|\theta) a'_\mu(x) = \sum_\nu B'_{\mu\nu} \theta_\nu \tag{C4}$$

are the generalized Fourier coefficients, where

$$B'_{\mu\nu} = \int dx\, a'_\mu(x)(-\partial)^\nu |\psi(x)|^2 = \sum_\xi A'_{\mu\xi} C_{\xi\nu}. \tag{C5}$$

Let $\{n(\mathcal{S}); \mathcal{S} \subseteq \mathbb{R}^D\}$ be the Poisson process [67,68] obtained by direct imaging with infinitesimal pixel size. The expected value of n over an area $\mathcal{S}$ is

$$\mathbb{E}[n(\mathcal{S})] = \tau \int_\mathcal{S} dx\, f(x|\theta), \tag{C6}$$

and $\{n(\mathcal{S}_1), n(\mathcal{S}_2), \dots\}$ are independent Poisson variables if $\{\mathcal{S}_1, \mathcal{S}_2, \dots\}$ are disjoint subsets. Equations (C2), (C3), and (C6) imply that

$$\check{\beta}_\mu = \frac{1}{\tau} \int n(dx) a'_\mu(x) \tag{C7}$$

is an unbiased estimator of β and

$$\check{\theta}_\mu = \sum_\nu (B'^{-1})_{\mu\nu} \check{\beta}_\nu \tag{C8}$$

is an unbiased estimator of θ. The covariances of the estimators are

$$\mathbb{V}_{\mu\nu}(\check{\beta}) = \frac{1}{\tau} \int dx\, f(x|\theta) a'_\mu(x) a'_\nu(x), \tag{C9}$$

$$\mathbb{V}(\check{\theta}) = (B'^{-1}) \mathbb{V}(\check{\beta}) (B'^{-1})^\top. \tag{C10}$$

Suppose for the moment that the trial θ' happens to coincide with the true θ. Then $\tilde{f}(x|\theta') = \tilde{f}(x|\theta)$, $a' = a$, and $B' = B$, resulting in

$$\mathbb{V}_{\mu\nu}(\check{\beta}) = \frac{1}{\tau} \int dx\, f(x|\theta) a_\mu(x) a_\nu(x) = \frac{\theta_0}{\tau} \delta_{\mu\nu}, \tag{C11}$$

$$\mathbb{V}(\check{\theta}) = \frac{\theta_0}{\tau} (B^{-1})(B^{-1})^\top, \tag{C12}$$

which is equal to the CRB given by Eq. (B27). In reality, θ is unknown of course, but $\tilde{f}(x|\theta)$ can be approximated in the subdiffraction regime as

$$\tilde{f}(x|\theta) = |\psi(x)|^2 + O(\Delta), \tag{C13}$$

and I can set

$$\tilde{f}(x|\theta') = |\psi(x)|^2. \tag{C14}$$

This gives $A'M^{(0)}A'^\top = I$, where $M_{\mu\nu}^{(0)} \equiv \Lambda_{\mu+\nu}$ is the PSF moment matrix. The covariances become

$$\mathbb{V}_{\mu\nu}(\check{\beta}) = \frac{\theta_0}{\tau} \int dx [|\psi(x)|^2 + O(\Delta)] a'_\mu(x) a'_\nu(x) \quad \text{(C15)}$$

$$= \frac{\theta_0}{\tau} [\delta_{\mu\nu} + O(\Delta)], \quad \text{(C16)}$$

$$\mathbb{V}(\check{\theta}) = \frac{\theta_0}{\tau} [(B'^{-1})(B'^{-1})^\top + O(\Delta)] \quad \text{(C17)}$$

$$= \frac{\theta_0}{\tau} [(C^{-1})M^{(0)}(C^{-1})^\top + O(\Delta)], \quad \text{(C18)}$$

which coincides with the CRB given by Eq. (3.14) in the leading order. An iterative approach that replaces $\tilde{f}(x|\theta')$ with an estimated $\tilde{f}(x|\theta)$ in each step should further improve the estimation, though it is unnecessary for Sec. VI.

APPENDIX D: CRB FOR DIRECT IMAGING IN THE DIFFRACTION-UNLIMITED REGIME

Suppose that the PSF $|\psi(x)|^2 = \delta(x)$ is infinitely sharp and $f(x|\theta) = F(x|\theta)$. The image moments given by Eq. (3.5) become identical to those of the object, viz.,

$$M_{\mu\nu} = \frac{\theta_{\mu+\nu}}{\theta_0}, \quad \text{(D1)}$$

the C matrix given by Eq. (3.6) becomes

$$C_{\mu\nu} = \frac{1}{\nu!} \int dx \delta(x) \partial^\nu x^\mu = \delta_{\mu\nu}, \quad \text{(D2)}$$

and the CRB given by Eq. (3.4) becomes

$$\text{CRB}_{\mu\nu} = \frac{\theta_{\mu+\nu}}{\tau}. \quad \text{(D3)}$$

This represents an ideal scenario in which the imaging is limited only by shot noise and not by diffraction. Equation (D3) also serves as a general lower bound on the CRB given by Eq. (2.19) for any linear-optical processing, as Eq. (2.6) is a Markov chain on $F(X|\theta)$ and the data-processing inequality [42] can be invoked.

To verify Eq. (D3), suppose that F consists of isolated point sources, viz.,

$$F(X|\theta) = \sum_\sigma \vartheta_\sigma \delta(X - X_\sigma), \quad \text{(D4)}$$

and since $|\psi(x)|^2 = \delta(x)$, their positions can be perfectly resolved. The unknowns are then ϑ, and the CRB with respect to ϑ is

$$J_{\sigma\gamma}^{(\vartheta)} = \frac{\tau}{\vartheta_\sigma} \delta_{\sigma\gamma}, \quad \text{CRB}_{\sigma\gamma}^{(\vartheta)} = \frac{\vartheta_\sigma}{\tau} \delta_{\sigma\gamma}. \quad \text{(D5)}$$

Expressing the moments as

$$\theta_\mu = \sum_\sigma \vartheta_\sigma X_\sigma^\mu, \quad \text{(D6)}$$

I can compute the CRB with respect to the moments via the transformation

$$\text{CRB}_{\mu\nu} = \sum_{\sigma,\gamma} \frac{\partial \theta_\mu}{\partial \vartheta_\sigma} \text{CRB}_{\sigma\gamma}^{(\vartheta)} \frac{\partial \theta_\nu}{\partial \vartheta_\gamma} = \frac{\theta_{\mu+\nu}}{\tau}, \quad \text{(D7)}$$

which coincides with Eq. (D3).

APPENDIX E: PROPERTIES OF MATRICES IN SEC. IV

Equation (4.5) can be inverted to give

$$k^r = \sum_s (G^{-1})_{rs} g_s(k). \quad \text{(E1)}$$

Substituting this in Eq. (4.28) and using the orthonormality given by Eq. (4.8), I obtain

$$H_{qr} = \frac{i^{|q|}(-i)^{|r|}}{r!} \sum_s (G^{-1})_{rs} \int dk |\Psi(k)|^2 g_q(k) g_s(k) \quad \text{(E2)}$$

$$= \frac{i^{|q|}(-i)^{|r|}}{r!} (G^{-1})_{rq}. \quad \text{(E3)}$$

The inverse is given by Eq. (4.30), which can be confirmed by directly computing HH^{-1} or $H^{-1}H$. Since G^{-1} and G are lower-triangular, H and H^{-1} are upper-triangular.

If $|\Psi(k)|^2$ is centrosymmetric according to Eq. (4.21), Ref. [24] shows that $g_q(k)$ consists of only even-order monomials $\{k^r; |r| \text{ even}\}$ if $|q|$ is even and only odd-order monomials $\{k^r; |r| \text{ odd}\}$ if $|q|$ is odd. Thus

$$G_{qr} = 0 \text{ if } |q| - |r| \text{ is odd}, \quad \text{(E4)}$$

$$g_q(k) = (-1)^{|q|} g_q(-k). \quad \text{(E5)}$$

Substituting k with $-k$ in the integral in Eq. (4.25) yields

$$h_q(X) = i^{|q|} \int dk |\Psi(-k)|^2 g_q(-k) \exp(ikX) \quad \text{(E6)}$$

$$= (-i)^{|q|} \int dk |\Psi(k)|^2 g_q(k) \exp(ikX) \quad \text{(E7)}$$

$$= h_q^*(X), \quad \text{(E8)}$$

and $h_q(X)$ is real. It follows that H and H^{-1} are real as well.

APPENDIX F: CRB FOR DIRECT IMAGING WITH THE GAUSSIAN PSF

In the limit of $\Delta \to 0$,

$$\tilde{f}(x|\theta) = |\psi(x)|^2 = \frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{||x||^2}{2}\right). \quad \text{(F1)}$$

A set of orthogonal polynomials are

$$a_\mu(x) = \frac{1}{\sqrt{\mu!}} \text{He}_\mu(x), \quad \text{(F2)}$$

where

$$\text{He}_\mu(x) \equiv \prod_{j=1}^D \text{He}_{\mu_j}(x_j), \quad \text{(F3)}$$

and the definition of the single-variable Hermite polynomials can be found, for example, in Refs. [54,55]. The B matrix defined by Eq. (B19) can then be computed by substituting the identity

$$(-\partial)^\nu |\psi(x)|^2 = |\psi(x)|^2 \text{He}_\nu(x) \quad \text{(F4)}$$

for Hermite polynomials [54,55] and using the orthonormality of a. The result is

$$B_{\mu\nu} = \frac{1}{\sqrt{\mu!}} \delta_{\mu\nu}, \quad \text{(F5)}$$

which can be substituted into Eq. (B27) to give Eq. (5.2).

APPENDIX G: AN ESTIMATOR FOR SPADE WITH NONSEPARABLE PSFs

The simple estimator given by Eq. (4.48) relies on the strong upper-triangular property of H given by Eq. (4.44) for separable PSFs. Without it, the weaker property given by Eq. (4.31) for a degree-respecting order still implies that the $\sum_r$ sum in Eq. (4.36) can be separated into an $|r| = |q|$ group and and an $|r| > |q|$ group, viz.,

$$\sum_r = \sum_{|r|=|q|} + \sum_{|r|>|q|}, \tag{G1}$$

and Eq. (4.36) becomes

$$\theta_{q+q'} = \sum_{|r|=|q|,|r'|=|q'|} (H^{-1})_{qr}(H^{-1})_{q'r'}\Gamma_{rr'}$$
$$+ \sum_{|r+r'|>|q+q'|} (H^{-1})_{qr}(H^{-1})_{q'r'}\Gamma_{rr'}. \tag{G2}$$

If I assume the estimator

$$\check{\theta}'_{q+q'} = \sum_{|r|=|q|,|r'|=|q'|} (H^{-1})_{qr}(H^{-1})_{q'r'}\check{\Gamma}_{rr'}, \tag{G3}$$

the bias is also given by Eq. (4.49), while the variance is

$$\mathbb{V}(\check{\theta}'_{q+q'}) = \sum_{|r|=|q|,|r'|=|q'|} [(H^{-1})_{qr}]^2[(H^{-1})_{q'r'}]^2\mathbb{V}(\check{\Gamma}_{rr'}) \tag{G4}$$

$$= \frac{\theta_0^2}{\min(N_s)}O(\Delta^{|q+q'|}), \tag{G5}$$

which can still be minimized by choosing q and q' according to Eq. (4.51).

A problem with Eq. (G3) is that, for a given $|q|$ and $|q'|$, the number of (r,r') indices with $|r| = |q|$ and $|r'| = |q'|$ is

$$\binom{|q| + D - 1}{|q|} \times \binom{|q'| + D - 1}{|q'|}, \tag{G6}$$

so the estimator may require a large number of $\check{\Gamma}_{rr'}$'s and a large number of bases to implement for a high-order moment, leading to a reduction in $\min(N_s)$. This difficulty is compounded by the fact that, for $D \geqslant 2$, there exist infinitely many sets of orthogonal polynomials for a given weight function, as pointed out in Appendix B, leading to infinite possible choices of the c polynomials and the PAD basis. For separable PSFs, the choice of the separable PAD basis in Sec. IV B fortunately leads to only one term in Eq. (G3), but it remains an open question whether Eq. (G3) can be further simplified via a more specific choice of the PAD basis for nonseparable PSFs.

APPENDIX H: CONDITIONS FOR FINITE IMAGE MOMENTS

Given Eqs. (3.11) and (3.12), M is finite if all the PSF moments $\{\Lambda_\mu; \mu \in \mathbb{N}_0^D\}$ are finite. Consider

$$\Lambda_\mu = \int dk\, \Psi^*(k)(i\,\partial_k)^\mu\Psi(k) \tag{H1}$$

in terms of the Fourier transform given by Eq. (4.3). A sufficient condition for Λ to be finite is that $\Psi(k)$ is infinitely differentiable and has compact support; an example is the bump function given by Eq. (6.12).

If any Λ_μ is infinite, the C matrix given by Eq. (3.7) and the CRB given by Eq. (3.14) also have infinite elements, and the direct-imaging theory in Sec. III and Appendix B breaks down. This happens for the rectangle aperture function given by Eq. (6.13). A solution, not explored in this work, may be to smooth $\Psi(k)$ by convolving it with a bump function with support width w, such that the smoothed $\Psi(k)$ becomes infinitely differentiable but remains compactly supported. When $w \ll 1$, the result should offer a good approximation of that for the original $\Psi(k)$.

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

[2] R. Nair and M. Tsang, Interferometric superlocalization of two incoherent optical point sources, Opt. Express **24**, 3684 (2016).

[3] M. Tsang, R. Nair, and X.-M. Lu, Quantum information for semiclassical optics, in Proceedings of SPIE, Quantum and Nonlinear Optics IV Vol. 10029 (SPIE, Bellingham, WA, 2016), p. 1002903.

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

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

[6] M. Tsang, Conservative classical and quantum resolution limits for incoherent imaging, J. Mod. Opt. **65**, 104 (2018).

[7] S. Z. Ang, R. Nair, and M. Tsang, Quantum limit for two-dimensional resolution of two incoherent optical point sources, Phys. Rev. A **95**, 063847 (2017).

[8] M. Tsang, Subdiffraction incoherent optical imaging via spatial-mode demultiplexing, New J. Phys. **19**, 023054 (2017).

[9] H. Krovi, S. Guha, and J. H. Shapiro, Attaining the quantum limit of passive imaging, arXiv:1609.00684 (physics, physics:quant-ph).

[10] X.-M. Lu, R. Nair, and M. Tsang, Quantum-optimal detection of one-versus-two incoherent sources with arbitrary separation, arXiv:1609.03025 (quant-ph).

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

[12] 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).

[13] 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).

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

[15] J. Rehacek, M. Paúr, B. Stoklasa, Z. Hradil, and L. L. Sánchez-Soto, Optimal measurements for resolution beyond the Rayleigh limit, Opt. Lett. **42**, 231 (2017).

[16] F. Yang, R. Nair, M. Tsang, C. Simon, and A. I. Lvovsky, Fisher information for far-field linear optical superresolution via homodyne or heterodyne detection in a higher-order local oscillator mode, Phys. Rev. A **96**, 063829 (2017).

[17] R. Kerviche, S. Guha, and A. Ashok, Fundamental limit of resolving two point sources limited by an arbitrary point spread function, in *2017 IEEE International Symposium on Information Theory (ISIT)* (IEEE, Piscataway, NJ, 2017), pp. 441.

[18] A. Chrostowski, R. Demkowicz-Dobrzański, M. Jarzyna, and K. Banaszek, On super-resolution imaging as a multiparameter estimation problem, Int. J. Quantum. Inform. **15**, 1740005 (2017).

[19] J. Řeháček, 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).

[20] J. Rehacek, Z. Hradil, D. Koutny, J. Grover, A. Krzic, and L. L. Sanchez-Soto, Optimal measurements for multiparameter quantum metrology with mixed states, arXiv:1712.08524 (quant-ph).

[21] J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 2004).

[22] J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985).

[23] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York, 2001).

[24] C. F. Dunkl and Y. Xu, *Orthogonal Polynomials of Several Variables* (Cambridge University Press, Cambridge, 2001).

[25] J. Zmuidzinas, Cramér-Rao sensitivity limits for astronomical instruments: Implications for interferometer design, J. Opt. Soc. Am. A **20**, 218 (2003).

[26] *Observing Photons in Space: A Guide to Experimental Space Astronomy*, edited by M. C. E. Huber, A. Pauluhn, J. L. Culhane, J. G. Timothy, K. Wilhelm, and A. Zehnder (Springer, New York, 2013).

[27] E. D. Feigelson and G. J. Babu, *Modern Statistical Methods for Astronomy* (Cambridge University Press, Cambridge, 2012).

[28] J. Chao, E. Sally Ward, and R. J. Ober, Fisher information theory for parameter estimation in single molecule microscopy: Tutorial, J. Opt. Soc. Am. A **33**, B36 (2016).

[29] A. von Diezmann, Y. Shechtman, and W. E. Moerner, Three-dimensional localization of single molecules for super-resolution imaging and single-particle tracking, Chem. Rev. **117**, 7244 (2017).

[30] D. J. Brady, *Optical Imaging and Spectroscopy* (Wiley, Hoboken, NJ, 2009).

[31] G. de Villiers and E. R. Pike, *The Limits of Resolution* (CRC, Boca Raton, FL, 2016).

[32] *Handbook of Biological Confocal Microscopy*, edited by J. B. Pawley (Springer, New York, 2006).

[33] M. Raginsky, R. M. Willett, Z. T. Harmany, and R. F. Marcia, Compressed sensing performance bounds under Poisson noise, IEEE Trans. Signal Proc. **58**, 3990 (2010).

[34] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, Maximum entropy and the nearly black object, J. R. Stat. Soc. Ser. B **54**, 41 (1992).

[35] E. J. Candès and C. Fernandez-Granda, Towards a mathematical theory of super-resolution, Commun. Pure Appl. Math. **67**, 906 (2014).

[36] G. Schiebinger, E. Robeva, and B. Recht, Superresolution without separation, Inf. Infer.: J. IMA iax006 (2017).

[37] L. Zhu, W. Zhang, D. Elnatan, and B. Huang, Faster STORM using compressed sensing, Nat. Methods **9**, 721 (2012).

[38] M. Bierbaum, B. D. Leahy, A. A. Alemi, I. Cohen, and J. P. Sethna, Light Microscopy at Maximal Precision, Phys. Rev. X **7**, 041007 (2017).

[39] A. Meister, *Deconvolution Problems in Nonparametric Statistics* (Springer, Berlin, 2009).

[40] L. Debnath and F. A. Shah, *Wavelet Transforms and Their Applications* (Springer, New York, 2015).

[41] S. Van Aert, A. J. den Dekker, D. Van Dyck, and A. van den Bos, High-resolution electron microscopy and electron tomography: Resolution versus precision, J. Struct. Biol. **138**, 21 (2002).

[42] R. Zamir, A proof of the Fisher information inequality via a data processing argument, IEEE Trans. Inf. Theor. **44**, 1246 (1998).

[43] M. P. Schützenberger, A generalization of the Fréchet-Cramér inequality to the case of Bayes estimation, Bull. Am. Math. Soc. **63**, 142 (1957).

[44] R. D. Gill and B. Y. Levit, Applications of the Van Trees inequality: A Bayesian Cramér-Rao bound, Bernoulli **1**, 59 (1995).

[45] *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*, edited by H. L. Van Trees and K. L. Bell (Wiley-IEEE, Piscataway, NJ, 2007).

[46] P. R. Nicovich, D. M. Owen, and K. Gaus, Turning single-molecule localization microscopy into a quantitative bioanalytical tool, Nat. Protocols **12**, 453 (2017).

[47] S. W. Hell, Far-field optical nanoscopy, Science **316**, 1153 (2007).

[48] M.-J. Tsai and K.-P. Dunn, Performance limitations on parameter estimation of closely spaced optical targets using shot-noise detector model, Tech. Rep. ADA073462 (Lincoln Laboratory, MIT, 1979).

[49] E. Bettens, D. Van Dyck, A. J. den Dekker, J. Sijbers, and A. van den Bos, Model-based two-object resolution from observations having counting statistics, Ultramicroscopy **77**, 37 (1999).

[50] S. Ram, E. S. Ward, and R. J. Ober, Beyond Rayleigh's criterion: A resolution measure with application to single-molecule microscopy, Proc. Natl. Acad. Sci. (USA) **103**, 4457 (2006).

[51] J. D. Gorman and A. O. Hero, Lower bounds for parametric estimation with constraints, IEEE Trans. Inf. Theor. **36**, 1285 (1990).

[52] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis* (Springer-Verlag, New York, 1985).

[53] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, Gaussian approximations of fluorescence microscope point-spread function models, Appl. Opt. **46**, 1819 (2007).

[54] DLMF, NIST Digital Library of Mathematical Functions, http://dlmf.nist.gov/, Release 1.0.11 of 2016-06-08 (2016), online companion to [55] .

[55] *NIST Handbook of Mathematical Functions*, edited by F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark (Cambridge University Press, New York, 2010), print companion to [54].

[56] K. I. Mortensen, L. S. Churchman, J. A. Spudich, and H. Flyvbjerg, Optimized localization analysis for single-molecule tracking and super-resolution microscopy, Nat. Methods **7**, 377 (2010).

[57] 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).

[58] G. Li, N. Bai, N. Zhao, and C. Xia, Space-division multiplexing: The next frontier in optical communication, Adv. Opt. Photon. **6**, 413 (2014).

[59] L.-W. Luo, N. Ophir, C. P. Chen, L. H. Gabrielli, C. B. Poitras, K. Bergmen, and M. Lipson, WDM-compatible mode-division multiplexing on a silicon chip, Nat. Commun. **5**, 3069 (2014).

[60] A. Mohanty, M. Zhang, A. Dutt, S. Ramelow, P. Nussenzveig, and M. Lipson, Quantum interference between transverse spatial waveguide modes, Nat. Commun. **8**, 14010 (2017).

[61] L. Martin, D. Mardani, H. E. Kondakci, W. D. Larson, S. Shabahang, A. K. Jahromi, T. Malhotra, A. N. Vamivakas, G. K. Atia, and A. F. Abouraddy, Basis-neutral Hilbert-space analyzers, Sci. Rep. **7**, 44995 (2017).

[62] M. Mirhosseini, M. Malik, Z. Shi, and R. W. Boyd, Efficient separation of the orbital angular momentum eigenstates of light, Nat. Commun. **4**, 2781 (2013).

[63] Y. Zhou, M. Mirhosseini, D. Fu, J. Zhao, S. M. Hashemi Rafsanjani, A. E. Willner, and R. W. Boyd, Sorting Photons by Radial Quantum Number, Phys. Rev. Lett. **119**, 263602 (2017).

[64] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1985).

[65] S. T. Buckland, Fitting density functions with polynomials, J. R. Stat. Soc. Ser. C **41**, 63 (1992).

[66] S. Efromovich, Orthogonal series density estimation, Wiley Interdisc. Rev.: Computat. Stat. **2**, 467 (2010).

[67] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space* (Springer-Verlag, New York, 1991).

[68] E. Çınlar, *Probability and Stochastics* (Springer, New York, 2011).