PAPER • OPEN ACCESS

Subdiffraction incoherent optical imaging via spatial-mode demultiplexing

To cite this article: Mankei Tsang 2017 *New J. Phys.* **19** 023054

View the article online for updates and enhancements.

New Journal of Physics

The open access journal at the forefront of physics

PAPER

CrossMark

Subdiffraction incoherent optical imaging via spatial-mode demultiplexing

Mankei Tsang

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

E-mail: mankei@nus.edu.sg

Abstract

I propose a spatial-mode demultiplexing (SPADE) measurement scheme for the far-field imaging of spatially incoherent optical sources. For any object too small to be resolved by direct imaging under the diffraction limit, I show that SPADE can estimate its second or higher moments much more precisely than direct imaging can fundamentally do in the presence of photon shot noise. I also prove that SPADE can approach the optimal precision allowed by quantum mechanics in estimating the location and scale parameters of a subdiffraction object. Realizable with far-field linear optics and photon counting, SPADE is expected to find applications in both fluorescence microscopy and astronomy.

1. Introduction

Recent research, initiated by our group [1–7], has shown that far-field linear optical methods can significantly improve the resolution of two equally bright incoherent optical point sources with sub-Rayleigh separations [1–15], overcoming previously established statistical limits [16–19]. The rapid experimental demonstrations [12–15] have heightened the promise of our approach. An open problem, of fundamental interest in optics and monumental importance to astronomy and fluorescence microscopy, is whether these results can be generalized for an arbitrary distribution of incoherent sources. Here I take a major step towards solving the problem by proposing a generalized spatial-mode demultiplexing (SPADE) measurement scheme and proving its superiority over direct imaging via a statistical analysis.

The use of coherent optical processing to improve the lateral resolution of incoherent imaging has thus far received only modest attention, as prior proposals [13, 20–25] either did not demonstrate any substantial improvement or neglected the important effect of noise. Using quantum optics and parameter estimation theory, here I show that, for any object too small to be resolved by diffraction-limited direct imaging, SPADE can estimate its second or higher moments much more precisely than direct imaging can fundamentally do in the presence of photon shot noise. Moreover, I prove that SPADE can approach the optimal precision allowed by quantum mechanics in estimating the location and scale parameters of a subdiffraction object. Given the usefulness of moments in identifying the size and shape of an object [26], the proposed scheme, realizable with far-field linear optics and photon counting, should provide a major boost to incoherent imaging applications that are limited by diffraction and photon shot noise, including not only fluorescence microscopy [27–30] and space-based telescopes [31] but also modern ground-based telescopes [32–35].

This paper is organized as follows. Section 2 introduces the background theory of quantum optics and parameter estimation for incoherent imaging. Section 3 describes the SPADE scheme for general imaging. Section 4 presents the most important results of this paper, namely, a comparison between the statistical performances of direct imaging and SPADE in the subdiffraction regime, showing the possibility of giant precision enhancements for moment estimation, while appendix justifies an approximation made in section 4 in more detail. Section 5 presents a numerical example to illustrate the theory, comparing the errors in estimating

the first and second moments of subdiffraction objects using direct imaging and SPADE. Section 6 proves that SPADE is close to the quantum precision limits to location and scale estimation in the subdiffraction regime. Section 7 discusses other practical and open issues.

2. Background formalism

2.1. Quantum optics

I begin with the quantum formalism established in [1] to ensure correct physics. The quantum state of thermal light with M temporal modes and a bandwidth much smaller than the center frequency can be written as $\rho^{\otimes M}$, where

$$\rho = (1 - \epsilon)\rho_0 + \epsilon\rho_1 + O(\epsilon^2), \tag{2.1}$$

ϵ is the average photon number per mode assumed to be $\ll 1$ [36, 37], $\rho_0 = |\text{vac}\rangle \langle \text{vac}|$ is the vacuum state, ρ_1 is the one-photon state with a density matrix equal to the mutual coherence function, and $O(\epsilon^2)$ denotes second-order terms, which are neglected hereafter. It is standard to assume that the fields from incoherent objects, such as stellar or fluorescent emitters, are spatially uncorrelated at the source [37]. In a diffraction-limited imaging system, the fields then propagate as waves; the Van Cittert-Zernike theorem is the most venerable consequence [37]. At the image plane of a conventional lens-based two-dimensional imaging system in the paraxial regime [37, 38], this implies

$$\rho_1 = \int d^2\boldsymbol{R} F(\boldsymbol{R}) |\psi_{\boldsymbol{R}}\rangle \langle \psi_{\boldsymbol{R}}|, \qquad |\psi_{\boldsymbol{R}}\rangle = \int d^2\boldsymbol{r} \psi(\boldsymbol{r} - \boldsymbol{R}) |\boldsymbol{r}\rangle, \tag{2.2}$$

where $\boldsymbol{R} = (X, Y)$ is the object-plane position, the notation $(u_1, u_2, \ldots)$ denotes a column vector, $F(\boldsymbol{R})$ is the source intensity distribution with normalization $\int d^2\boldsymbol{R} F(\boldsymbol{R}) = 1$, $|\boldsymbol{r}\rangle = a^\dagger(\boldsymbol{r})|\text{vac}\rangle$ is a one-photon position eigenket on the image plane at position $\boldsymbol{r} = (x, y)$ with $[a(\boldsymbol{r}), a^\dagger(\boldsymbol{r}')] = \delta^2(\boldsymbol{r} - \boldsymbol{r}')$ [39], and $\psi(\boldsymbol{r})$ is the field point-spread function (PSF) of the imaging system. Without loss of generality, the image-plane position vector $\boldsymbol{r}$ has been scaled with respect to the magnification to follow the same scale as $\boldsymbol{R}$ [38]. For convenience, I also normalize the position vectors with respect to the width of the PSF to make them dimensionless.

Consider the processing and measurement of the image-plane field by linear optics and photon counting. The counting distribution for each ρ can be expressed as $\langle n_0, n_1, \ldots | \rho | n_0, n_1, \ldots \rangle$, where $|n_0, n_1, \ldots\rangle = (\prod_{j=0}^{\infty} b_j^{\dagger n_j} / \sqrt{n_j!})|\text{vac}\rangle$, $b_j \equiv \int d^2\boldsymbol{r} \phi_j^*(\boldsymbol{r}) a(\boldsymbol{r})$, $\phi_j(\boldsymbol{r})$ is the optical mode function that is projected to the jth output, and $[b_j, b_k^\dagger] = \int d^2\boldsymbol{r} \phi_j^*(\boldsymbol{r})\phi_k(\boldsymbol{r}) = \delta_{jk}$. With the negligence of multiphoton coincidences, the relevant projections are $\{|\text{vac}\rangle, |\phi_j\rangle\}$, with $|\phi_j\rangle \equiv |0, \ldots, n_j = 1, \ldots, 0\rangle = b_j^\dagger |\text{vac}\rangle = \int d^2\boldsymbol{r} \phi_j(\boldsymbol{r})|\boldsymbol{r}\rangle$. The zero-photon probability becomes $1 - \epsilon$ and the probability of one photon being detected in the jth mode becomes $\epsilon p(j)$, where

$$p(j) \equiv \langle \phi_j | \rho_1 | \phi_j \rangle = \int d^2\boldsymbol{R} F(\boldsymbol{R}) |\langle \phi_j | \psi_{\boldsymbol{R}} \rangle|^2 \tag{2.3}$$

is the one-photon distribution. A generalization of the measurement model using the concept of positive operator-valued measures is possible [1, 3] but not needed here.

For example, direct imaging can be idealized as a measurement of the position of each photon, leading to an expected image given by

$$f(\boldsymbol{r}) \equiv \langle \boldsymbol{r} | \rho_1 | \boldsymbol{r} \rangle = \int d^2\boldsymbol{R} F(\boldsymbol{R}) |\psi(\boldsymbol{r} - \boldsymbol{R})|^2, \tag{2.4}$$

which is a basic result in statistical optics [37, 38]. While equation (2.4) suggests that, similar to the coherent-imaging formalism, the PSF acts as a low-pass filter in the spatial frequency domain [38], the effect of more general optical processing according to equation (2.3) is more subtle and offers surprising advantages, as demonstrated by recent work [1–15] and elaborated in this paper.

Over M temporal modes, the probability distribution of photon numbers $m = (m_0, m_1, \ldots)$ detected in the respective optical modes becomes

$$P(m) = \sum_L \mathcal{M}(m|L)\mathcal{B}(L), \tag{2.5}$$

where $\mathcal{B}(L)$ is the binomial distribution for detecting L photons over M trials with single-trial success probability ϵ and $\mathcal{M}(m|L) = \delta_{L,\sum_j m_j} L! \prod_j [p(j)]^{m_j}/m_j!$ is the multinomial distribution of m given L total photons [40]. The average photon number in all modes becomes $N \equiv M\epsilon$. Taking the limit of $\epsilon \to 0$ while holding N constant, $\mathcal{B}(L)$ becomes Poisson with mean N, and $P(m) \to \exp(-N) \prod_j [Np(j)]^{m_j}/m_j!$, which is the widely used Poisson model of photon counting for incoherent sources at optical frequencies [3, 18, 27–31, 36].

2.2. Parameter estimation

The central goal of imaging is to infer unknown properties of the source distribution $F(\boldsymbol{R})$ from the measurement outcome m. Here I frame it as a parameter estimation problem, defining $\theta = (\theta_1, \theta_2, \ldots)$ as a column vector of unknown parameters and assuming the source distribution $F(\boldsymbol{R}|\theta)$ to be a function of θ. Denote an estimator as $\check{\theta}(m)$ and its error covariance matrix as $\Sigma_{\mu\nu}(\theta) = \sum_m P(m|\theta)[\check{\theta}_\mu(m) - \theta_\mu][\check{\theta}_\nu(m) - \theta_\nu]$. For any unbiased estimator ($\sum_m \check{\theta}(m)P(m|\theta) = \theta$), the Cramér-Rao bound (CRB) is given by [40, 41]

$$\Sigma(\theta) \geqslant \text{CRB}(\theta) \equiv J^{-1}(\theta), \tag{2.6}$$

where $J(\theta)$ is the Fisher information matrix and the matrix inequality implies that $\Sigma - J^{-1}$ is positive-semidefinite, or equivalently $u^\top(\Sigma - J^{-1})u \geqslant 0$ for any real vector u. Assuming the model given by equation (2.5) and a known N, it can be shown [3] that

$$J_{\mu\nu}(\theta) = N \sum_j \frac{1}{p(j|\theta)} \frac{\partial p(j|\theta)}{\partial \theta_\mu} \frac{\partial p(j|\theta)}{\partial \theta_\nu}, \tag{2.7}$$

which is a well known expression [16–18, 28, 30, 36]. For example, the direct-imaging information, given equation (2.4) and the limit $p(j|\theta) \to \mathrm{d}^2 \boldsymbol{r} f(\boldsymbol{r}|\theta)$, is

$$J_{\mu\nu}^{(\text{direct})}(\theta) = N \int \mathrm{d}^2\boldsymbol{r} \frac{1}{f(\boldsymbol{r}|\theta)} \frac{\partial f(\boldsymbol{r}|\theta)}{\partial \theta_\mu} \frac{\partial f(\boldsymbol{r}|\theta)}{\partial \theta_\nu}. \tag{2.8}$$

For large N, the maximum-likelihood estimator is asymptotically normal with mean θ and covariance $\Sigma(\theta) = J^{-1}(\theta)$, even though it may be biased for finite N [40, 41]. Bayesian and minimax generalizations of the CRB for any biased or unbiased estimator are possible [5, 41] but not considered here as they offer qualitatively similar conclusions. The Fisher information is nowadays regarded as the standard precision measure in incoherent imaging research, especially in fluorescence microscopy [18, 28–30], where photon shot noise is the dominant noise source and a proper statistical analysis is essential.

Apart from the CRB, another useful property of the Fisher information is the data-processing inequality [42, 43], which mandates that, once the measurement is made, no further processing of the data can increase the information. For example, direct imaging with large pixels can be modeled as integrations of photon counts over groups of infinitesimally small pixels, so the information can never exceed equation (2.8). More generally, the data-processing inequality rules out the possibility of improving the information using any processing that applies to the direct-imaging intensity, such as the proposal by Walker *et al* for incoherent imaging in [20], even if the processing is done with optics. Hence, as argued by Tham *et al* [14], coherent processing that is sensitive to the phase of the field is the only way to improve upon equation (2.8). The information for any coherent processing and measurement is in turn limited by quantum upper bounds in terms of ρ_1 [1, 3, 6, 43–46].

3. Spatial-mode demultiplexing

SPADE is a technique previously proposed for the purpose of superresolving the separation between two incoherent point sources [1, 6, 7, 9, 13–15]. I now ask how SPADE can be generalized for the imaging of an arbitrary source distribution. Consider the transverse-electromagnetic (TEM) basis $\{|\boldsymbol{q}\rangle; \boldsymbol{q} = (q_x, q_y) \in \mathbb{N}^2\}$ [47], where

$$|\boldsymbol{q}\rangle = \int \mathrm{d}^2\boldsymbol{r}\, \phi_{\boldsymbol{q}}(\boldsymbol{r})|\boldsymbol{r}\rangle, \tag{3.1}$$

$$\phi_{\boldsymbol{q}}(\boldsymbol{r}) \equiv \frac{\mathrm{He}_{q_x}(x)\, \mathrm{He}_{q_y}(y)}{\sqrt{2\pi q_x! q_y!}} \exp\left(-\frac{x^2 + y^2}{4}\right), \tag{3.2}$$

and He_q is the Hermite polynomial [48, 49]. Assuming a Gaussian PSF given by $\psi(\boldsymbol{r}) = \phi_{00}(\boldsymbol{r})$, which is a common assumption in fluorescence microscopy [28, 30], $|\psi_{\boldsymbol{R}}\rangle$ is a coherent state [50], and the one-photon density matrix in the TEM basis becomes

$$g(\boldsymbol{q}, \boldsymbol{q}'|\theta) \equiv \langle \boldsymbol{q}|\rho_1(\theta)|\boldsymbol{q}'\rangle \tag{3.3}$$

$$= C(\boldsymbol{q}, \boldsymbol{q}') \int \mathrm{d}^2\boldsymbol{R}\, F(\boldsymbol{R}|\theta) e^{-(X^2 + Y^2)/4} X^{q_x + q_x'} Y^{q_y + q_y'}, \tag{3.4}$$

where

$$C(\boldsymbol{q}, \boldsymbol{q}') \equiv \frac{1}{2^{|\boldsymbol{q} + \boldsymbol{q}'|_1}\sqrt{\boldsymbol{q}! \boldsymbol{q}'!}}, \qquad |\boldsymbol{q}|_1 \equiv q_x + q_y, \qquad \boldsymbol{q}! \equiv q_x! q_y!. \tag{3.5}$$

Table 1. A list of measurement schemes, their projections, and the orders $\boldsymbol{\mu} = (\mu_X, \mu_Y)$ of the moment parameters $\Theta_{\boldsymbol{\mu}}$ to which they are sensitive.

Scheme	Projections	q_x	q_y	μ_X	μ_Y
TEM	$\lvert \boldsymbol{q} \rangle$	$\mathbb{N}$	$\mathbb{N}$	even	even
iTEM1	$[\lvert \boldsymbol{q} \rangle \pm \lvert \boldsymbol{q} + (1, 0)\rangle]/\sqrt{2}$	even	$\mathbb{N}$	1, 5, …	even
iTEM2	$[\lvert \boldsymbol{q} \rangle \pm \lvert \boldsymbol{q} + (0, 1)\rangle]/\sqrt{2}$	$\mathbb{N}$	even	even	1, 5, …
iTEM3	$[\lvert \boldsymbol{q} \rangle \pm \lvert \boldsymbol{q} + (1, -1)\rangle]/\sqrt{2}$	$\mathbb{N}$	odd	odd	1, 5, …
iTEM4	$[\lvert \boldsymbol{q} \rangle \pm \lvert \boldsymbol{q} + (1, 0)\rangle]/\sqrt{2}$	odd	$\mathbb{N}$	3, 7, …	even
iTEM5	$[\lvert \boldsymbol{q} \rangle \pm \lvert \boldsymbol{q} + (0, 1)\rangle]/\sqrt{2}$	$\mathbb{N}$	odd	even	3, 7, …
iTEM6	$[\lvert \boldsymbol{q} \rangle \pm \lvert \boldsymbol{q} + (1, -1)\rangle]/\sqrt{2}$	$\mathbb{N}$	even	odd	3, 7, …

To investigate the imaging capability of SPADE measurements, define

$$\Theta_{\boldsymbol{\mu}}(\theta) \equiv \int \mathrm{d}^2 \boldsymbol{R} F(\boldsymbol{R}|\theta) e^{-(X^2+Y^2)/4} X^{\mu_X} Y^{\mu_Y}, \tag{3.6}$$

with $\boldsymbol{\mu} = (\mu_X, \mu_Y)$, leading to a linear parameterization of g given by

$$g(\boldsymbol{q}, \boldsymbol{q}'|\theta) = C(\boldsymbol{q}, \boldsymbol{q}')\Theta_{\boldsymbol{q}+\boldsymbol{q}'}. \tag{3.7}$$

Notice that each $\Theta_{\boldsymbol{\mu}}$ is a moment of the source distribution filtered by a Gaussian. In particular, if the object is much smaller than the PSF width, the Gaussian can be neglected, and $\Theta_{\boldsymbol{\mu}}$ becomes a moment of the source distribution itself. This subdiffraction regime is of central interest to superresolution imaging and, as shown in section 4, also a regime in which direct imaging performs relatively poorly. Since a distribution is uniquely determined by its moments [51], $F(\boldsymbol{R}|\theta) \exp[-(X^2 + Y^2)/4]$ and therefore $F(\boldsymbol{R}|\theta)$ can be reconstructed given the moments, at least in principle. Note also that the object-moment order $\boldsymbol{\mu}$ is nontrivially related to the order of the matrix element via $\boldsymbol{\mu} = \boldsymbol{q} + \boldsymbol{q}'$, which is a peculiar feature of incoherent imaging.

A measurement in the TEM basis yields

$$p^{(\mathrm{TEM})}(\boldsymbol{q}|\theta) = C(\boldsymbol{q}, \boldsymbol{q})\Theta_{2\boldsymbol{q}}, \tag{3.8}$$

which is sensitive only to moments with even μ_X and μ_Y, as also recognized by Yang *et al* in [13]. This measurement is realized by demultiplexing the image-plane optical field in terms of the TEM basis via linear optics before photon counting for each mode and can be implemented by many methods, most commonly found in optical communications [1, 6, 15, 52–54]. To access the other moments, consider interferometry between two TEM modes that implements the projections

$$\lvert + \rangle \equiv \frac{1}{\sqrt{2}}(\lvert \boldsymbol{q} \rangle + \lvert \boldsymbol{q}' \rangle), \qquad \lvert - \rangle \equiv \frac{1}{\sqrt{2}}(\lvert \boldsymbol{q} \rangle - \lvert \boldsymbol{q}' \rangle). \tag{3.9}$$

This two-channel interferometric TEM (iTEM) measurement leads to

$$
\begin{aligned}
p^{(\boldsymbol{q},\boldsymbol{q}')}(+|\theta) &= \beta(\boldsymbol{q}, \boldsymbol{q}') + C(\boldsymbol{q}, \boldsymbol{q}')\Theta_{\boldsymbol{q}+\boldsymbol{q}'}, \\
p^{(\boldsymbol{q},\boldsymbol{q}')}(-|\theta) &= \beta(\boldsymbol{q}, \boldsymbol{q}') - C(\boldsymbol{q}, \boldsymbol{q}')\Theta_{\boldsymbol{q}+\boldsymbol{q}'},
\end{aligned}
\tag{3.10}
$$

$$\beta(\boldsymbol{q}, \boldsymbol{q}') \equiv \frac{1}{2}[C(\boldsymbol{q}, \boldsymbol{q})\Theta_{2\boldsymbol{q}} + C(\boldsymbol{q}', \boldsymbol{q}')\Theta_{2\boldsymbol{q}'}]. \tag{3.11}$$
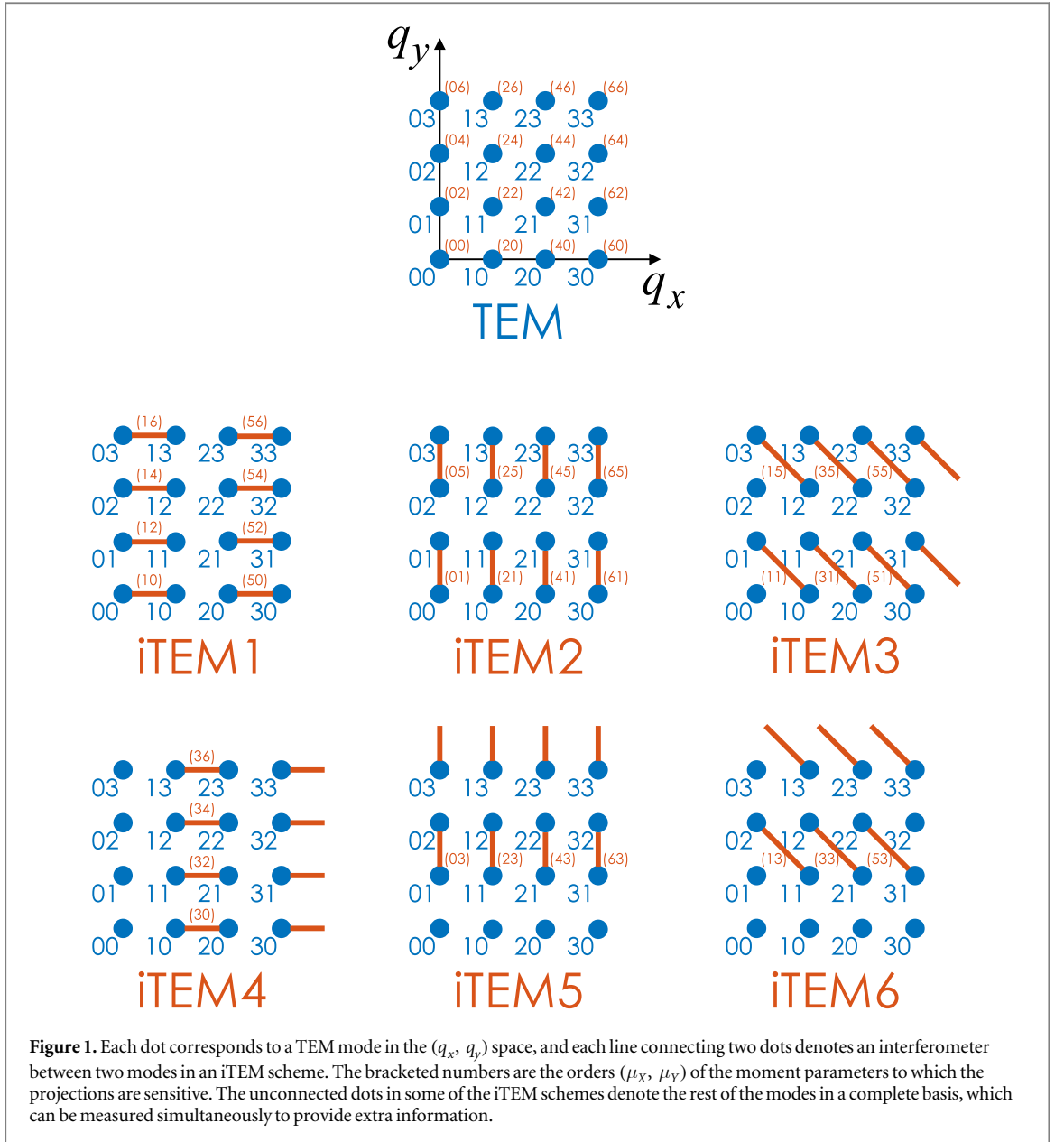
The dependence on $\Theta_{\boldsymbol{q}+\boldsymbol{q}'}$ is the main interest here, as it allows one to access any moment parameter.

For multiparameter estimation and general imaging, multiple TEM and iTEM measurements are needed. To be specific, table 1 lists a set of schemes that can be used together to estimate all the moment parameters, while figure 1 shows a graphical representation of the schemes in the (q_x, q_y) space. Neighboring modes are used in the proposed iTEM schemes because they maximize the Fisher information, as shown later in section 4. The bases in different schemes are incompatible with one another, so the photons have to be rationed among the 7 schemes, by applying the different schemes sequentially through reprogrammable interferometers or spatial-light modulators [15, 52–54] for example.

4. Statistical analysis

4.1. Direct imaging

Although the proposed SPADE method can in principle perform general imaging, its complexity would not be justifiable if it could not offer any significant advantage over direct imaging. To compare their statistical performances, consider first direct imaging with a Gaussian PSF. Expanding $\lvert \psi(\boldsymbol{r} - \boldsymbol{R})\rvert^2$ in a Taylor series, I obtain

Figure 1. Each dot corresponds to a TEM mode in the (q_x, q_y) space, and each line connecting two dots denotes an interferometer between two modes in an iTEM scheme. The bracketed numbers are the orders (μ_X, μ_Y) of the moment parameters to which the projections are sensitive. The unconnected dots in some of the iTEM schemes denote the rest of the modes in a complete basis, which can be measured simultaneously to provide extra information.

$$f(\boldsymbol{r}|\theta) = |\phi_{00}(\boldsymbol{r})|^2 \left[1 + \sum_\mu D_\mu(\boldsymbol{r})\theta_\mu\right], \tag{4.1}$$

$$D_\mu(\boldsymbol{r}) \equiv \frac{\mathrm{He}_{\mu_X}(x)\mathrm{He}_{\mu_Y}(y)}{\mu!}, \tag{4.2}$$

in terms of the moment parameters defined as

$$\theta_\mu \equiv \int \mathrm{d}^2\boldsymbol{R} F(\boldsymbol{R}|\theta) X^{\mu_X} Y^{\mu_Y}. \tag{4.3}$$

In terms of this parameterization, the Fisher information becomes

$$J_{\mu\nu}^{(\mathrm{direct})} = N \int \mathrm{d}^2\boldsymbol{r} \, |\phi_{00}(\boldsymbol{r})|^2 \frac{D_\mu(\boldsymbol{r})D_\nu(\boldsymbol{r})}{1 + \sum_\eta D_\eta(\boldsymbol{r})\theta_\eta}. \tag{4.4}$$

Assume now that the support of the source distribution is centered at the origin and has a maximum width Δ much smaller than the PSF width. Since the spatial dimensions have been normalized with respect to the PSF width, the PSF width is 1 in the dimensionless unit, and the assumption can be expressed as

$$\Delta \ll 1, \tag{4.5}$$

which defines the subdiffraction regime. The parameters are then bounded by

$$|\theta_\mu| \leqslant \left(\frac{\Delta}{2}\right)^{|\mu|_1}, \tag{4.6}$$

and the image is so blurred that it resembles the TEM_{00} mode rather than the object, viz.,
$f(\boldsymbol{r}|\theta) = |\phi_{00}(\boldsymbol{r})|^2[1 + O(\Delta)]$. Writing the denominator in equation (4.4) as $1 + O(\Delta)$ and applying the orthogonality of Hermite polynomials [48, 49], I obtain

$$J_{\mu\nu}^{(\mathrm{direct})} = \frac{N}{\mu!}[\delta_{\mu\nu} + O(\Delta)], \tag{4.7}$$

$$\mathrm{CRB}_{\mu\mu}^{(\mathrm{direct})} = \frac{\mu!}{N}[1 + O(\Delta)]. \tag{4.8}$$

This is a significant result in its own right, as it establishes a fundamental limit to superresolution algorithms for shot-noise-limited direct imaging [20, 55–57], generalizing the earlier results for two sources [16–18] and establishing that, at least for a Gaussian PSF, the moments are a natural, approximately orthogonal [58] set of parameters for subdiffraction objects.

4.2. SPADE

To investigate the performance of SPADE for moment estimation, note that, in the subdiffraction regime, equation (3.6) can be expressed as

$$\Theta_\mu = \theta_\mu + O(\Delta^{|\mu|_1+2}), \tag{4.9}$$

where $O(\Delta^{|\mu|_1+2})$ is a linear combination of moments that are at least two orders above μ and therefore much smaller than θ_μ. Approximating Θ_μ with θ_μ greatly simplifies the analysis below; appendix contains a more detailed justification of this approximation. For the TEM scheme, taking $\Theta_{2q} = \theta_{2q}$ in equation (3.8) makes the information matrix diagonal, with the nonzero elements given by

$$J_{\mu\mu}^{(\mathrm{TEM})} = \frac{N^{(\mathrm{TEM})}C(\boldsymbol{q}, \boldsymbol{q})}{\theta_{2q}}, \quad \mu = 2\boldsymbol{q}, \tag{4.10}$$

where $N^{(\mathrm{TEM})}$ is the average photon number available to the TEM scheme. The relevant CRB components are hence

$$\mathrm{CRB}_{\mu\mu}^{(\mathrm{TEM})} = \frac{\theta_{2q}}{N^{(\mathrm{TEM})}C(\boldsymbol{q}, \boldsymbol{q})}, \quad \mu = 2\boldsymbol{q}. \tag{4.11}$$

Defining the photon count of the $\boldsymbol{q}$th channel as $m_{\boldsymbol{q}}$ with expected value $N^{(\mathrm{TEM})}p^{(\mathrm{TEM})}(\boldsymbol{q}|\theta)$, it is straightforward to show that the estimator

$$\check{\theta}_{2q} = \frac{m_{\boldsymbol{q}}}{N^{(\mathrm{TEM})}C(\boldsymbol{q}, \boldsymbol{q})} \tag{4.12}$$

is unbiased and achieves the error given by equation (4.11) under the assumption $\Theta_{2q} = \theta_{2q}$.

A precision enhancement factor can be defined as the ratio of equations (4.8) to (4.11), viz.,

$$\frac{\mathrm{CRB}_{\mu\mu}^{(\mathrm{direct})}}{\mathrm{CRB}_{\mu\mu}^{(\mathrm{TEM})}} \approx \frac{N^{(\mathrm{TEM})}}{N}\frac{\mu!}{2^{|\mu|_1}(\mu/2)!\theta_\mu}. \tag{4.13}$$

Apart from a factor $N^{(\mathrm{TEM})}/N$ determined by the different photon numbers detectable in each method, the important point is that the factor scales inversely with $\theta_\mu = O(\Delta^{|\mu|_1})$, so the enhancement is enormous in the $\Delta \ll 1$ subdiffraction regime. The prefactor in equation (4.13) also increases with increasing μ.

To investigate the errors in estimating the other moments via the iTEM schemes, assume $\Theta_{q+q'} = \theta_{q+q'}$ in equations (3.10). The dependence of equations (3.10) on $\theta_{q+q'}$ is the main interest, while I treat $\beta(\boldsymbol{q}, \boldsymbol{q}')$ as an unknown nuisance parameter [59]; the TEM scheme can offer additional information about $\beta(\boldsymbol{q}, \boldsymbol{q}')$ via θ_{2q} and $\theta_{2q'}$ but it is insignificant and neglected here to simplify the analysis. The information matrix with respect to $\{\theta_{q+q'}, \beta(\boldsymbol{q}, \boldsymbol{q}')\}$ is block-diagonal and consists of a series of two-by-two matrices, each of which can be determined from equations (3.10) for two parameters $(\theta_{q+q'}, \beta(\boldsymbol{q}, \boldsymbol{q}'))$ and is given by

$$J^{(q,q')} = \frac{2N^{(\mathrm{iTEM})}}{\beta^2(\boldsymbol{q}, \boldsymbol{q}') - C^2(\boldsymbol{q}, \boldsymbol{q}')\theta_{q+q'}^2}\begin{pmatrix}\beta(\boldsymbol{q}, \boldsymbol{q}')C^2(\boldsymbol{q}, \boldsymbol{q}') & -C^2(\boldsymbol{q}, \boldsymbol{q}')\theta_{q+q'} \\ -C^2(\boldsymbol{q}, \boldsymbol{q}')\theta_{q+q'} & \beta(\boldsymbol{q}, \boldsymbol{q}')\end{pmatrix}, \tag{4.14}$$

where $N^{(\mathrm{iTEM})}$ is the average photon number available to the iTEM scheme. The CRB component with respect to $\theta_{q+q'}$ is hence obtained by taking the inverse of equation (4.14) and extracting the relevant term; the result is

$$\text{CRB}_{\mu\mu}^{(\text{iTEM})} = \frac{\beta(\boldsymbol{q}, \boldsymbol{q}')}{2N^{(\text{iTEM})}C^2(\boldsymbol{q}, \boldsymbol{q}')}, \qquad \boldsymbol{\mu} = \boldsymbol{q} + \boldsymbol{q}'. \tag{4.15}$$

Defining the two photon counts of the $(\boldsymbol{q}, \boldsymbol{q}')$ iTEM channels as $m_+^{(\boldsymbol{q},\boldsymbol{q}')}$ and $m_-^{(\boldsymbol{q},\boldsymbol{q}')}$ with expected values $N^{(\text{iTEM})}p^{(\boldsymbol{q},\boldsymbol{q}')}(+|\theta)$ and $N^{(\text{iTEM})}p^{(\boldsymbol{q},\boldsymbol{q}')}(-|\theta)$, respectively, it can be shown that the estimator

$$\check{\theta}_{\boldsymbol{q}+\boldsymbol{q}'} = \frac{m_+^{(\boldsymbol{q},\boldsymbol{q}')} - m_-^{(\boldsymbol{q},\boldsymbol{q}')}}{2N^{(\text{iTEM})}C(\boldsymbol{q}, \boldsymbol{q}')} \tag{4.16}$$

is unbiased and achieves the error given by equation (4.15) under the assumption $\Theta_{\boldsymbol{q}+\boldsymbol{q}'} = \theta_{\boldsymbol{q}+\boldsymbol{q}'}$. The iTEM schemes can also offer information about $\theta_{2\boldsymbol{q}}$ and $\theta_{2\boldsymbol{q}'}$ via the background parameter $\beta(\boldsymbol{q}, \boldsymbol{q}')$, but the additional information is inconsequential and neglected here.

An enhancement factor can again be expressed as

$$\frac{\text{CRB}_{\mu\mu}^{(\text{direct})}}{\text{CRB}_{\mu\mu}^{(\text{iTEM})}} \approx \frac{N^{(\text{iTEM})}}{N} \frac{\mu!}{2^{2|\mu|_1 - 1}\boldsymbol{q}!(\boldsymbol{\mu} - \boldsymbol{q})!\beta(\boldsymbol{q}, \boldsymbol{\mu} - \boldsymbol{q})}. \tag{4.17}$$

With the background parameter $\beta(\boldsymbol{q}, \boldsymbol{\mu} - \boldsymbol{q})$ on the order of $\Delta^{\min[|2\boldsymbol{q}|_1, |2(\mu-q)|_1]}$, both $1/\beta$ and the $\mu!/[\boldsymbol{q}!(\boldsymbol{\mu} - \boldsymbol{q})!]$ coefficient can be maximized by choosing $\boldsymbol{q}$ to be as close to $\boldsymbol{\mu}/2$ as possible. This justifies the pairing of neighboring modes in the iTEM schemes listed in figure 1 and table 1. With iTEM1, iTEM2, iTEM4, and iTEM5, $|\boldsymbol{\mu}|_1$ is odd, and

$$\beta = O(\Delta^{|\boldsymbol{\mu}|_1 - 1}). \tag{4.18}$$

With iTEM3 and iTEM6, $|\boldsymbol{\mu}|_1$ is even, and

$$\beta = O(\Delta^{|\boldsymbol{\mu}|_1}). \tag{4.19}$$

The enhancements, being inversely proportional to β, can again be substantial for higher moments. The only exception is the estimation of the first moments θ_{10} and θ_{01}, for which the right-hand side of equation (4.17) becomes $N^{(\text{iTEM})}/N$ and the iTEM schemes offer no advantage.

These results can be compared with [1, 6] for the special case of two equally bright point sources. If the origin of the image plane is aligned with their centroid and their separation along the X direction is d, $\theta_{20} = d^2/4$, and a reparameterization leads to a transformed Fisher information $\mathcal{J}^{(\text{direct})}(d) \approx Nd^2/8$ and $\mathcal{J}^{(\text{TEM})}(d) \approx N/4$ for the estimation of d, in accordance with the results in [1, 6] to the leading order of d. The experiments reported in [13–15] serve as demonstrations of the proposed scheme in this special case.

4.3. Elementary explanation

The enhancements offered by SPADE can be understood by considering the signal-to-noise ratio (SNR) of a measurement with Poisson statistics. Suppose for simplicity that the mean count of an output can be written as $Np(j|\theta) = A\theta + B$, which consists of a signal component $A\theta$ and a background B. The variance is $A\theta + B$, so the SNR can be expressed as $(A\theta)^2/(A\theta + B)$. To maximize it, the background B should be minimized to reduce the variance. For direct imaging, the background according to equation (4.1) is dominated by the TEM$_{00}$ mode, whereas each output of SPADE is able to filter out that mode as well as other irrelevant low-order modes to minimize the background without compromising the signal. To wit, equation (3.8) for TEM measurements has no background, while equations (3.10) for iTEM also have low backgrounds in the subdiffraction regime. The Fisher information given by equation (2.7) is simply a more rigorous statistical tool that formalizes the SNR concept and provides error bounds; reducing the background likewise improves the information by reducing the denominator in equation (2.7).

In this respect, the proposed scheme seems to work in a similar way to nulling interferometry for exoplanet detection [60]. The nulling was proposed there for the special purpose of blocking the glare of starlight, however, and there had not been any prior statistical study of nulling for subdiffraction objects to my knowledge. The surprise here is that coherent processing in the far field can vastly improve general incoherent imaging even in the subdiffraction regime and in the presence of photon shot noise, without the need to manipulate the sources as in prior superresolution microscopic methods [61–66] or detect evanescent waves via lossy or unrealistic materials [67, 68].

5. Numerical demonstration

Here I present a numerical study to illustrate the proposal and confirm the theory. Assume an object that consists of 5 equally bright point sources with random positions within the square $-0.3 \leqslant X \leqslant 0.3$ and $-0.3 \leqslant Y \leqslant 0.3$. The average photon number is assumed to be $N = 5 \times 10\,000$ in total. Figure 2 shows an example of the generated source positions and a direct image with pixel size $\delta x \delta y = 0.1 \times 0.1$ and Poisson
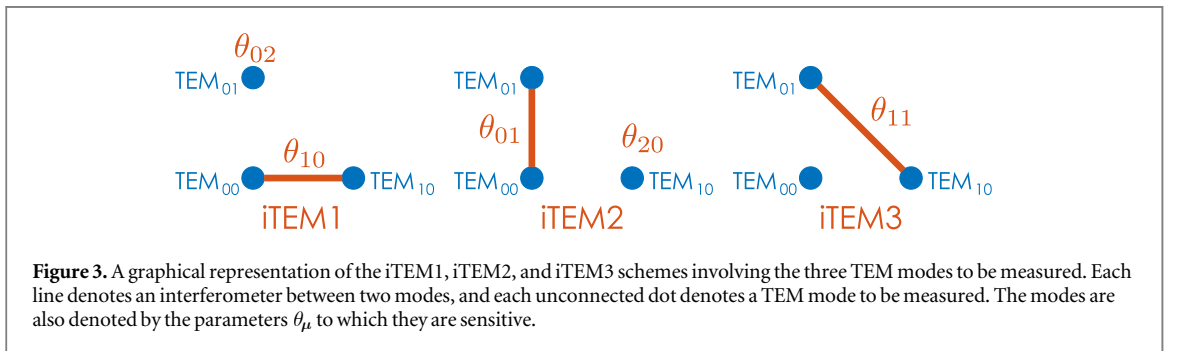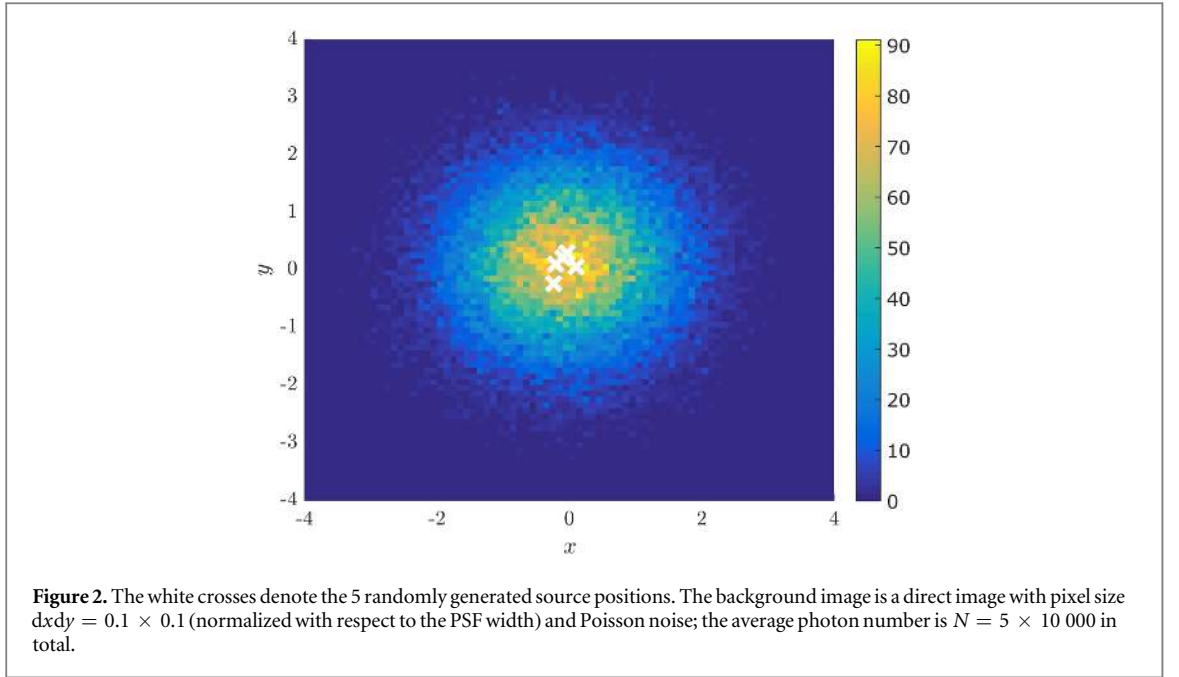
Figure 2. The white crosses denote the 5 randomly generated source positions. The background image is a direct image with pixel size $dx dy = 0.1 \times 0.1$ (normalized with respect to the PSF width) and Poisson noise; the average photon number is $N = 5 \times 10\,000$ in total.

Figure 3. A graphical representation of the iTEM1, iTEM2, and iTEM3 schemes involving the three TEM modes to be measured. Each line denotes an interferometer between two modes, and each unconnected dot denotes a TEM mode to be measured. The modes are also denoted by the parameters θ_μ to which they are sensitive.

Table 2. The projections for the SPADE measurement scheme depicted in figure 3. $|0, 0\rangle$ corresponds to the TEM_{00} mode, $|1, 0\rangle$ corresponds to the TEM_{10} mode, and $|0, 1\rangle$ corresponds to the TEM_{01} mode.

iTEM1	iTEM2	iTEM3
$(|0, 0\rangle + |1, 0\rangle)/\sqrt{2}$	$(|0, 0\rangle + |0, 1\rangle)/\sqrt{2}$	$(|1, 0\rangle + |0, 1\rangle)/\sqrt{2}$
$(|0, 0\rangle - |1, 0\rangle)/\sqrt{2}$	$(|0, 0\rangle - |0, 1\rangle)/\sqrt{2}$	$(|1, 0\rangle - |0, 1\rangle)/\sqrt{2}$
$|0, 1\rangle$	$|1, 0\rangle$	$|0, 0\rangle$

noise. I focus on the estimation of the first and second moments of the source distribution given by $\{\theta_\mu; \boldsymbol{\mu} = (1, 0), (0, 1), (2, 0), (0, 2), (1, 1)\}$. For direct imaging, I use the estimator $\check{\theta}_\mu = \mu! \sum_j D_\mu(\boldsymbol{r}_j) m(\boldsymbol{r}_j)/N$, where $m(\boldsymbol{r}_j)$ is the photon count at a pixel positioned at $\boldsymbol{r}_j$. It can be shown that, in the small-pixel limit, this estimator is unbiased and approaches the CRB given by equation (4.8) for $\Delta \ll 1$.

For SPADE, I consider only the TEM_{00}, TEM_{10}, and TEM_{01} modes, and the photons in all the other modes are discarded. As illustrated in figure 3, the iTEM1, iTEM2, and iTEM3 schemes suffice to estimate the parameters of interest. Table 2 lists the projections, and figure 4 plots the spatial wave functions for the projections. The light is assumed to be split equally among the three schemes, leading to 9 outputs; figure 5 shows a sample of the photon counts drawn from Poisson statistics. For the estimators, I use equations (4.12) and (4.16). Compared with the large number of pixels in direct imaging, the compressive nature of SPADE for moment estimation is an additional advantage.

Figure 6 plots the numerically computed mean-square errors (MSEs) for 100 randomly generated objects versus true parameters in log–log scale. Each error value for a given object is computed by averaging the squared difference between the estimator and the true parameter over 500 samples of Poissonian outputs. For
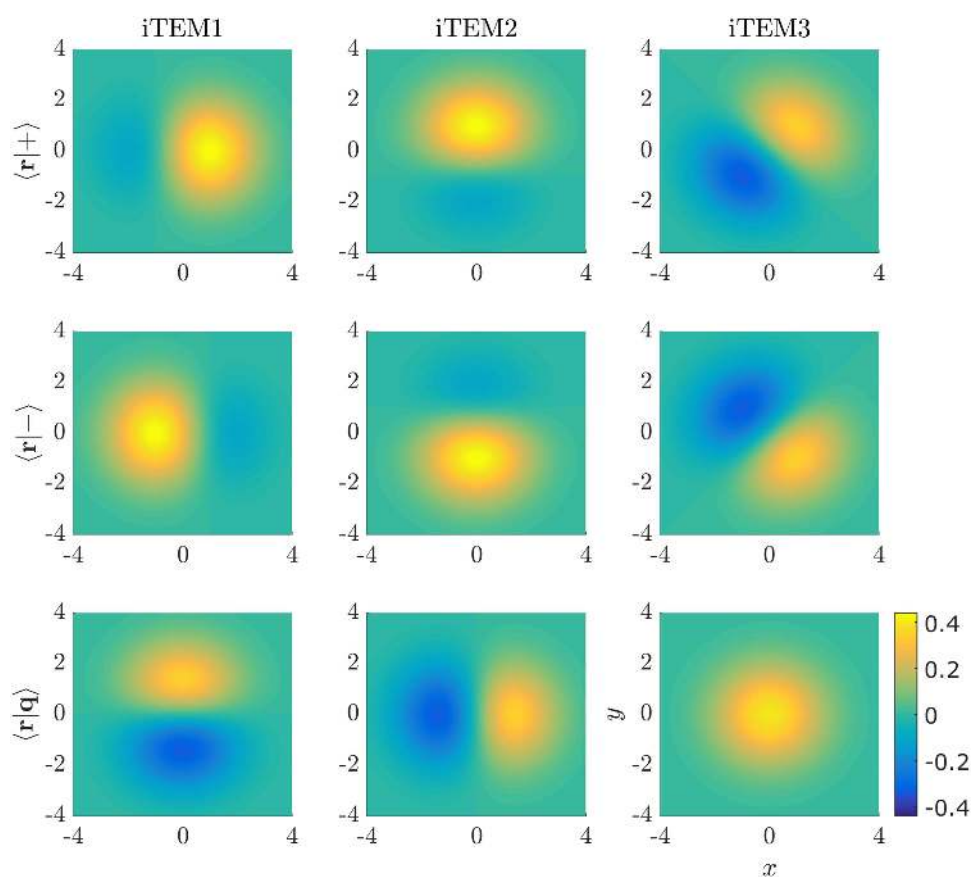
Figure 4. The spatial wave functions $\langle r|\phi_j \rangle$ for the projections listed in table 2. x and y are image-plane coordinates normalized with respect to the PSF width and the color code corresponds to amplitudes of normalized wave functions.
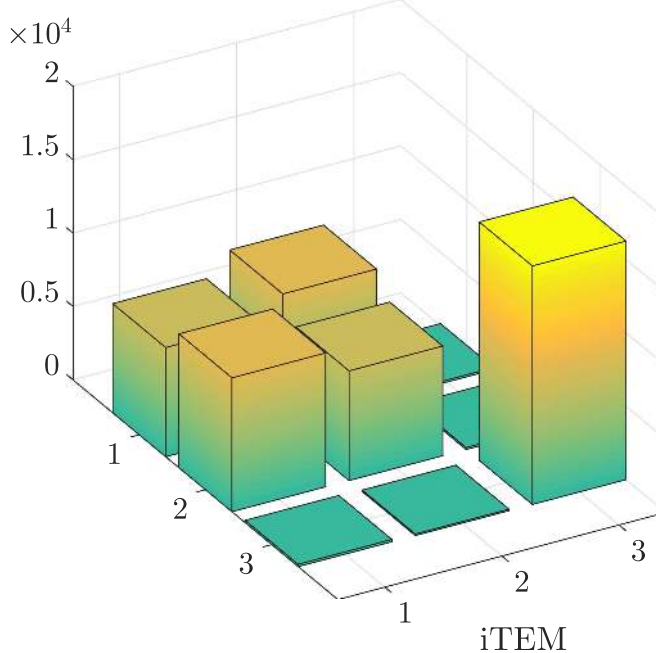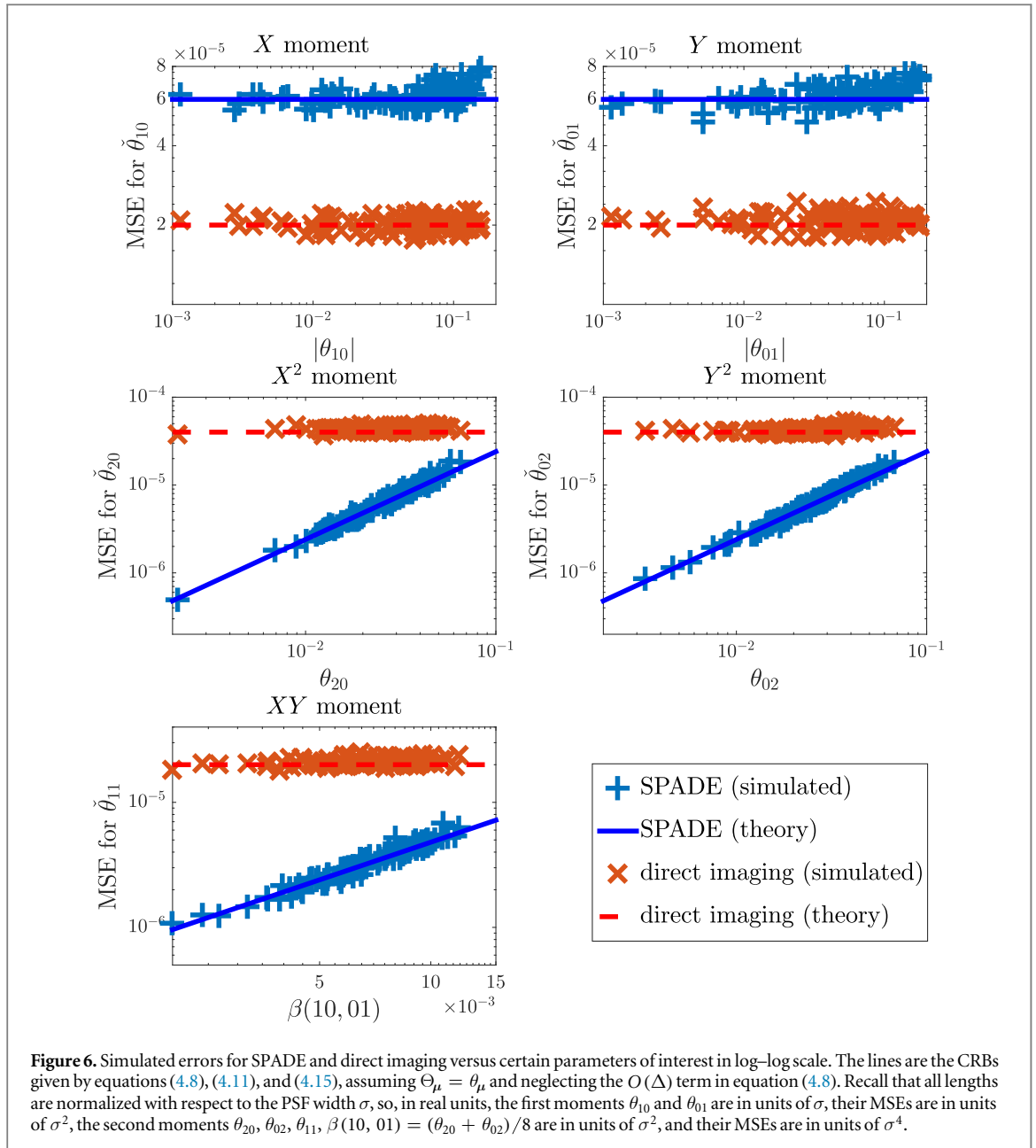

Figure 5. A sample of the simulated photon counts from SPADE. The order of the matrix elements follows table 2 and figure 4. Note how the counts for the antisymmetric modes are much lower as a result of filtering out the lower-order modes. As argued in section 4.3, such dark ports enable a higher SNR by reducing the background without compromising the signal.

Figure 6. Simulated errors for SPADE and direct imaging versus certain parameters of interest in log–log scale. The lines are the CRBs given by equations (4.8), (4.11), and (4.15), assuming $\Theta_\mu = \theta_\mu$ and neglecting the $O(\Delta)$ term in equation (4.8). Recall that all lengths are normalized with respect to the PSF width σ, so, in real units, the first moments θ_{10} and θ_{01} are in units of σ, their MSEs are in units of σ^2, the second moments $\theta_{20}, \theta_{02}, \theta_{11}, \beta(10, 01) = (\theta_{20} + \theta_{02})/8$ are in units of σ^2, and their MSEs are in units of σ^4.

comparison, figure 6 also plots the CRBs given by equations (4.8), (4.11), and (4.15), assuming $\Theta_\mu = \theta_\mu$ and neglecting the $O(\Delta)$ term in equation (4.8). A few observations can be made:

1. As shown by the plots in the first row of figure 6, SPADE is 3 times worse than direct imaging at estimating the first moments. This is because SPADE uses only 1/3 of the available photons to estimate each first moment.

2. The theory suggests that the advantage of SPADE starts with the second moments, and indeed the other plots show that SPADE is substantially more precise at estimating them, even though SPADE uses only a fraction of the available photons to estimate each moment. This enhancement is a generalization of the recent results on two sources [1–6, 8, 9, 12–15].

3. The errors are all remarkably tight to the CRBs, despite the simplicity of the estimators and the approximations in the bounds. In particular, the excellent performance of the SPADE estimator in the subdiffraction regime justifies its assumption of $\Theta_\mu = \theta_\mu$.

6. Quantum limits

In the diffraction-*un*limited regime, it is not difficult to prove that direct imaging achieves the highest Fisher information allowed by quantum mechanics. To be precise, that regime can be defined as one in which the PSF is

so sharp relative to the source distribution that $\{|\psi_{\boldsymbol{R}}\rangle; \boldsymbol{R} \in \mathrm{supp}(F)\}$ can be approximated as the orthogonal position basis $\{|\boldsymbol{r}\rangle\}$. ρ_1 becomes diagonal in that basis, and the quantum Fisher information [1, 3, 43–46] is equal to the direct-imaging information given by equation (2.8). The physics in the opposite subdiffraction regime is entirely different, however, as diffraction causes $\{|\psi_{\boldsymbol{R}}\rangle\}$ to have significant overlaps with one another, and more judicious measurements can better deal with the resulting indistinguishability, as demonstrated in sections 4 and 5.

I now prove that SPADE is in fact near-quantum-optimal in estimating location and scale parameters of a source distribution in the subdiffraction regime. Suppose that the distribution has the form

$$F(\boldsymbol{R}|\theta) = F(\boldsymbol{R}(\boldsymbol{\xi}|\theta)), \tag{6.1}$$

such that θ parameterizes a coordinate transformation $\boldsymbol{R} = \boldsymbol{R}(\boldsymbol{\xi}|\theta)$, and the transformation leads to a reference measure $F_0(\boldsymbol{\xi})$ that is independent of θ. Taking $\boldsymbol{R}$ as a column vector, I can rewrite equation (2.2) as

$$\rho_1(\theta) = \mathbb{E}_0(|\psi_{\boldsymbol{R}}\rangle\langle\psi_{\boldsymbol{R}}|), \tag{6.2}$$

$$|\psi_{\boldsymbol{R}}\rangle = e^{-i\boldsymbol{k}^\top \boldsymbol{R}(\boldsymbol{\xi}|\theta)}|\psi_0\rangle, \tag{6.3}$$

where $\mathbb{E}_0(\cdot) \equiv \int d\boldsymbol{\xi} F_0(\boldsymbol{\xi})(\cdot)$ and $\boldsymbol{k}$ is the momentum operator in a column vector. I can now use the quantum upper bound on the Fisher information [1, 43, 46] and the convexity of the quantum Fisher information [69, 70] to prove that the Fisher information for any measurement is bounded as

$$J(\theta) \leqslant K(\rho_1(\theta)) \leqslant \tilde{K}(\theta) \equiv N\mathbb{E}_0[K(|\psi_{\boldsymbol{R}}\rangle\langle\psi_{\boldsymbol{R}}|)], \tag{6.4}$$

where K is the quantum Fisher information proposed by Helstrom [44]. For the pure state, K can be computed analytically to give

$$K(|\psi_{\boldsymbol{R}}\rangle\langle\psi_{\boldsymbol{R}}|) = 4\frac{\partial \boldsymbol{R}^\top(\boldsymbol{\xi}|\theta)}{\partial\theta_\mu}\langle\psi_0|\Delta\boldsymbol{k}\Delta\boldsymbol{k}^\top|\psi_0\rangle\frac{\partial \boldsymbol{R}(\boldsymbol{\xi}|\theta)}{\partial\theta_\nu}, \tag{6.5}$$

$$\Delta\boldsymbol{k} \equiv \boldsymbol{k} - \langle\psi_0|\boldsymbol{k}|\psi_0\rangle, \tag{6.6}$$

leading to

$$\tilde{K}_{\mu\nu} = N\mathbb{E}_0\left[\frac{\partial \boldsymbol{R}^\top(\boldsymbol{\xi}|\theta)}{\partial\theta_\mu}\frac{\partial \boldsymbol{R}(\boldsymbol{\xi}|\theta)}{\partial\theta_\nu}\right] \tag{6.7}$$

for the gaussian PSF. For example, a location parameter can be expressed as $\boldsymbol{R} = \boldsymbol{\xi} - (\theta, 0)$. Equation (6.7) then gives

$$\tilde{K} = N. \tag{6.8}$$

This can be attained by either direct imaging or iTEM1 in the subdiffraction regime.

The advantage of SPADE starts with the second moments, which are particularly relevant to scale estimation. Let $\boldsymbol{R} = \theta\boldsymbol{\xi}$, which results in

$$\tilde{K} = N\mathbb{E}_0(\boldsymbol{\xi}^\top\boldsymbol{\xi}). \tag{6.9}$$

For the TEM measurement, on the other hand,

$$p^{(\mathrm{TEM})}(\boldsymbol{q}|\theta) = \mathbb{E}_0\left(e^{-Q_X}\frac{Q_X^{q_x}}{q_x!}e^{-Q_Y}\frac{Q_Y^{q_y}}{q_y!}\right), \tag{6.10}$$

$$Q_X \equiv \frac{\theta^2\xi_X^2}{4}, \quad Q_Y \equiv \frac{\theta^2\xi_Y^2}{4}. \tag{6.11}$$

Defining

$$\bar{\boldsymbol{q}} \equiv \sum_{\boldsymbol{q}}\boldsymbol{q}\,p^{(\mathrm{TEM})}(\boldsymbol{q}|\theta) = \frac{\theta^2}{4}\mathbb{E}_0\begin{pmatrix}\xi_X^2 \\ \xi_Y^2\end{pmatrix}, \tag{6.12}$$

$$\boldsymbol{V} \equiv \sum_{\boldsymbol{q}}(\boldsymbol{q} - \bar{\boldsymbol{q}})(\boldsymbol{q} - \bar{\boldsymbol{q}})^\top p^{(\mathrm{TEM})}(\boldsymbol{q}|\theta) \tag{6.13}$$

$$= \frac{\theta^2}{4}\mathbb{E}_0\begin{pmatrix}\xi_X^2 & 0 \\ 0 & \xi_Y^2\end{pmatrix} + O(\theta^4), \tag{6.14}$$

and using the lower bound by Stein *et al* [71], I obtain

$$J^{(\mathrm{TEM})} \geqslant N\frac{\partial\bar{\boldsymbol{q}}^\top}{\partial\theta}\boldsymbol{V}^{-1}\frac{\partial\bar{\boldsymbol{q}}}{\partial\theta} \to N\mathbb{E}_0(\boldsymbol{\xi}^\top\boldsymbol{\xi}), \tag{6.15}$$

which approaches the quantum limit given by equation (6.9) for $\theta \to 0$. The argument can be made more precise if the form of $F_0(\boldsymbol{\xi})$ is known, as the extended convexity of the quantum Fisher information can be used

to obtain a tighter upper bound [70, 72], while the $O(\theta^4)$ term in equation (6.14) can be computed to obtain an explicit lower bound for any θ.

7. Discussion

Though promising, the giant precision enhancements offered by SPADE do not imply unlimited imaging resolution for finite photon numbers. The higher moments are still more difficult to estimate even with SPADE in terms of the fractional error, which is $\sim \mathrm{CRB}_{\mu\mu}/\theta_\mu^2 = 1/O(N\Delta^{|\mu|_1})$ for even $|\mu|_1$ and $1/O(N\Delta^{|\mu|_1+1})$ for odd $|\mu|_1$, meaning that more photons are needed to attain a desirable fractional error for higher-order moments. Intuitively, this is because of the inherent inefficiency of subdiffraction objects to couple to higher-order modes, and the need to accumulate enough photons in those modes to achieve an acceptable SNR. A related issue is the reconstruction of the full source distribution, which requires all moments in principle. A finite number of moments cannot determine the distribution uniquely by themselves [51], although a wide range of regularization methods, such as maximum entropy and basis pursuit, are available for more specific scenarios [51, 55, 57, 73, 74].

Despite these limitations, the fact remains that direct imaging is an even poorer choice of measurement for subdiffraction objects and SPADE can extract much more information, simply from the far field. For example, the size and shape of a star, a planetary system, a galaxy, or a fluorophore cluster that is poorly resolved under direct imaging can be identified much more accurately through the estimation of the second or higher moments by SPADE. Alternatively, SPADE can be used to reach a desirable precision with far fewer photons or a much smaller aperture, enhancing the speed or reducing the size of the imaging system for the more special purposes. In view of the statistical analysis in [2, 4, 6], the image-inversion interferometers proposed in [2, 4, 6, 23–25] are expected to be similarly useful for estimating the second moments. For larger objects, scanning in the manner of confocal microscopy [27] or adaptive alignment [1] should be helpful.

Many open problems remain; chief among them are the incorporation of prior information, generalizations for non-Gaussian PSFs, the derivation of more general quantum limits, the possibility of even better measurements, and experimental implementations. The quantum optimality of SPADE for general imaging is in particular an interesting question. These daunting problems may be attacked by more advanced methods in quantum metrology [43–46, 75–78], quantum state tomography [79–81], compressed sensing [55–57, 81], and photonics design [52–54].

Appendix. Nuisance parameters

Instead of assuming $\Theta_\mu = \theta_\mu$ as in section 4, I consider here the exact relation given by equation (3.6), which can be expressed as

$$\Theta_\mu = \theta_\mu - \frac{\theta_{\mu+(2,0)}}{4} - \frac{\theta_{\mu+(0,2)}}{4} - \ldots \tag{A1}$$

For the TEM scheme, this implies that each qth channel contains information about not only θ_{2q} but also the higher-order moments. If I assume that each qth channel is used to estimate only θ_{2q}, however, then the higher-order moments act only as nuisance parameters [59] to the estimation of θ_{2q}. This is a conservative assumption, as the data-processing inequality [42, 43] implies that neglecting outputs can only reduce the information, but the assumption also means that I do not need to consider any channel with order lower than q to compute the CRB with respect to θ_{2q}, simplifying the analysis below.

Given the above assumption, I can compute the information matrix with respect to the parameters $(\theta_{2q}, \theta_{2q+(2,0)}, \theta_{2q+(0,2)}, \ldots)$ by considering only the qth and higher-order channels; the result is

$$J^{(\mathrm{TEM})} = \begin{pmatrix} J^{(\mathrm{TEM})}_{2q,2q} & J^{(\mathrm{TEM})}_{2q,2q+(2,0)} & \cdots \\ \hline J^{(\mathrm{TEM})}_{2q+(2,0),2q} & J^{(\mathrm{TEM})}_{2q+(2,0),2q+(2,0)} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \equiv N^{(\mathrm{TEM})} \left(\begin{array}{c|c} \alpha & \eta^\top \\ \hline \eta & \gamma \end{array} \right), \tag{A2}$$

where $\alpha = C(q, q)/\Theta_{2q}$ and $\eta = O(\Delta^{-2|q|_1})$ are determined only by the qth channel, while $\gamma = O(\Delta^{-(2|q|_1+2)})$ is mainly determined by the higher-order channels. The CRB with respect to θ_{2q} becomes [59]

$$\mathrm{CRB}^{(\mathrm{TEM})}_{2\boldsymbol{q},2\boldsymbol{q}} = \frac{1}{N^{(\mathrm{TEM})}} (\alpha - \eta^{\top}\gamma^{-1}\eta)^{-1}. \tag{A3}$$

The key here is that $\eta\gamma^{-1}\eta^{\top} = O(\Delta^{-2|\boldsymbol{q}|_1+2})$ is smaller than $\alpha = O(\Delta^{-2|\boldsymbol{q}|_1})$ by two orders of Δ, so

$$\mathrm{CRB}^{(\mathrm{TEM})}_{2\boldsymbol{q},2\boldsymbol{q}} = \frac{1}{N^{(\mathrm{TEM})}\alpha}[1 + O(\Delta^2)] \tag{A4}$$

$$= \frac{\Theta_{2\boldsymbol{q}}}{N^{(\mathrm{TEM})}C(\boldsymbol{q},\boldsymbol{q})}[1 + O(\Delta^2)] \tag{A5}$$

$$= \frac{\theta_{2\boldsymbol{q}}}{N^{(\mathrm{TEM})}C(\boldsymbol{q},\boldsymbol{q})}[1 + O(\Delta^2)], \tag{A6}$$

which is consistent with equation (4.11).

An intuitive way of understanding this result is to rewrite equation (A1) as

$$\theta_{\boldsymbol{\mu}} = \Theta_{\boldsymbol{\mu}} + \frac{\theta_{\boldsymbol{\mu}+(2,0)}}{4} + \frac{\theta_{\boldsymbol{\mu}+(0,2)}}{4} + \ldots, \tag{A7}$$

which implies that the total error in $\theta_{\boldsymbol{\mu}}$ consists of the error in $\Theta_{\boldsymbol{\mu}}$ as well as the errors in the higher-order moments. The higher-order moments can be estimated much more accurately via the higher-order channels, so the effect of their uncertainties on the estimation of $\theta_{\boldsymbol{\mu}}$ is negligible. A similar exercise can be done for the iTEM schemes, with similar results.

In practice, such a careful treatment of the nuisance parameters is unlikely to be necessary in the subdiffraction regime, as the numerical analysis in section 5 shows that excellent results can be obtained simply by taking $\Theta_{\boldsymbol{\mu}} = \theta_{\boldsymbol{\mu}}$ without any correction.

References

[1] Tsang M, Nair R and Lu X-M 2016 Quantum theory of superresolution for two incoherent optical point sources *Phys. Rev.* X **6** 031033
[2] Nair R and Tsang M 2016 Interferometric superlocalization of two incoherent optical point sources *Opt. Express* **24** 3684–701
[3] Tsang M, Nair R and Lu X-M 2016 Quantum information for semiclassical optics *Proc. SPIE* **10029** 1002903
[4] Nair R and Tsang M 2016 Far-field superresolution of thermal electromagnetic sources at the quantum limit *Phys. Rev. Lett.* **117** 190801
[5] Tsang M 2016 Conservative classical and quantum resolution limits for incoherent imaging arXiv:1605.03799
[6] Ang S Z, Nair R and Tsang M 2016 Quantum limit for two-dimensional resolution of two incoherent optical point sources arXiv:1606.00603
[7] Lu X-M, Nair R and Tsang M 2016 Quantum-optimal detection of one-versus-two incoherent sources with arbitrary separation arXiv:1609.03025
[8] Lupo C and Pirandola S 2016 Ultimate precision bound of quantum and subwavelength imaging *Phys. Rev. Lett.* **117** 190802
[9] Rehacek J, Paúr M, Stoklasa B, Hradil Z and Sánchez-Soto L L 2017 Optimal measurements for resolution beyond the Rayleigh limit *Opt. Lett.* **42** 231–4
[10] Krovi H, Guha S and Shapiro J H 2016 Attaining the quantum limit of passive imaging arXiv:1609.00684
[11] Kerviche R, Guha S and Ashok A 2017 Fundamental limit of resolving two point sources limited by an arbitrary point spread function arXiv:1701.04913
[12] Tang Z S, Durak K and Ling A 2016 Fault-tolerant and finite-error localization for point emitters within the diffraction limit *Opt. Express* **24** 22004
[13] Yang F, Tashchilina A, Moiseev E S, Simon C and Lvovsky A I 2016 Far-field linear optical superresolution via heterodyne detection in a higher-order local oscillator mode *Optica* **3** 1148
[14] Tham W K, Ferretti H and Steinberg A M 2016 Beating Rayleigh's curse by imaging using phase information *Phys. Rev. Lett.* **118** 070801
[15] Paúr M, Stoklasa B, Hradil Z, Sánchez-Soto L L and Rehacek J 2016 Achieving the ultimate optical resolution *Optica* **3** 1144
[16] Bettens E, van Dyck D, den Dekker A J, Sijbers J and van den Bos A 1999 Model-based two-object resolution from observations having counting statistics *Ultramicroscopy* **77** 37–48
[17] van Aert S, den Dekker A J, van Dyck D and van den Bos A 2002 High-resolution electron microscopy and electron tomography: resolution versus precision *J. Struct. Biol.* **138** 21–33
[18] Ram S, Ward E S and Ober R J 2006 Beyond Rayleigh's criterion: a resolution measure with application to single-molecule microscopy *Proc. Natl Acad. Sci. USA* **103** 4457–62
[19] Acuna C O and Horowitz J 1997 A statistical approach to the resolution of point sources *J. Appl. Stat.* **24** 421–36
[20] Walker J G, Brakenhoff G J, Bertero M, Pike E R, Davies R E and Young M R 1993 Superresolving scanning optical microscopy using holographic optical processing *J. Opt. Soc. Am.* A **10** 59–64
[21] Tamburini F, Anzolin G, Umbriaco G, Bianchini A and Barbieri C 2006 Overcoming the Rayleigh criterion limit with optical vortices *Phys. Rev. Lett.* **97** 163903
[22] Sandeau N and Giovannini H 2006 Increasing the lateral resolution of 4pi fluorescence microscopes *J. Opt. Soc. Am.* A **23** 1089
[23] Wicker K and Heintzmann R 2007 Interferometric resolution improvement for confocal microscopes *Opt. Express* **15** 12206
[24] Wicker K, Sindbert S and Heintzmann R 2009 Characterisation of a resolution enhancing image inversion interferometer *Opt. Express* **17** 15491
[25] Weigel D, Foerster R, Babovsky H, Kiessling A and Kowarschik R 2011 Enhanced resolution of microscopic objects by image inversion interferometry *Opt. Express* **19** 26451
[26] Prokop R J and Reeves A P 1992 A survey of moment-based techniques for unoccluded object representation and recognition *CVGIP, Graph. Models Image Process.* **54** 438–60
[27] Pawley J B (ed) 2006 *Handbook of Biological Confocal Microscopy* (New York: Springer)

[28] Deschout H, Zanacchi F C, Mlodzianoski M, Diaspro A, Bewersdorf J, Hess S T and Braeckmans K 2014 Precisely and accurately localizing single emitters in fluorescence microscopy *Nat. Methods* **11** 253–66

[29] Mortensen K I, Churchman L S, Spudich J A and Flyvbjerg H 2010 Optimized localization analysis for single-molecule tracking and super-resolution microscopy *Nat. Methods* **7** 377–81

[30] Chao J, Ward E S and Ober R J 2016 Fisher information theory for parameter estimation in single molecule microscopy: tutorial *J. Opt. Soc. Am.* A **33** B36

[31] Huber M C E *et al* (ed) 2013 *Observing Photons in Space: A Guide to Experimental Space Astronomy* (New York: Springer)

[32] Christou J C *et al* 2016 Adaptive optics capabilities at the Large Binocular Telescope Observatory *Proc. SPIE* **9909** 99092E

[33] Sanders G H 2013 The Thirty Meter Telescope (TMT): an international observatory *J. Astrophys. Astron.* **34** 81–6

[34] Bernstein R A *et al* 2014 Overview and status of the Giant Magellan Telescope project *Proc. SPIE* **9145** 91451C

[35] Davies R *et al* 2016 MICADO: first light imager for the E-ELT *Proc. SPIE* **9908** 99081Z

[36] Zmuidzinas J 2003 Cramér-Rao sensitivity limits for astronomical instruments: implications for interferometer design *J. Opt. Soc. Am.* A **20** 218–33

[37] Goodman J W 1985 *Statistical Optics* (New York: Wiley)

[38] Goodman J W 2004 *Introduction to Fourier Optics* (New York: McGraw-Hill)

[39] Shapiro J H 2009 The quantum theory of optical communications *IEEE J. Sel. Top. Quantum Electron.* **15** 1547–69

[40] Wasserman L A 2004 *All of Statistics* (New York: Springer)

[41] van Trees H L 2001 *Detection, Estimation, and Modulation Theory, Part I* (New York: Wiley)

[42] Zamir R 1998 A proof of the Fisher information inequality via a data processing argument *IEEE Trans. Inf. Theory* **44** 1246–50

[43] Hayashi M 2006 *Quantum Information* (Berlin: Springer)

[44] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)

[45] Holevo A S 2011 *Probabilistic and Statistical Aspects of Quantum Theory* (Pisa, Italy: Edizioni della Normale)

[46] Hayashi M (ed) 2005 *Asymptotic Theory of Quantum Statistical Inference: Selected Papers* (Singapore: World Scientific)

[47] Yariv A 1989 *Quantum Electronics* (New York: Wiley)

[48] DLMF, NIST Digital Library of Mathematical Functions, http://dlmf.nist.gov/, Release 1.0.11 of 2016-06-08, online companion to [49]

[49] Olver F W J, Lozier D W, Boisvert R F and Clark C W (ed) 2010 *NIST Handbook of Mathematical Functions* (New York, NY: Cambridge University Press) print companion to [48]

[50] Mandel L and Wolf E 1995 *Optical Coherence and Quantum Optics* (Cambridge: Cambridge University Press)

[51] Bertero M 1989 Linear inverse and iII-posed problems *Advances in Electronics and Electron Physics* vol 75 ed P W Hawkes (New York: Academic) pp 1–120

[52] Armstrong S, Morizur J-F, Janousek J, Hage B, Treps N, Lam P K and Bachor H-A 2012 Programmable multimode quantum networks *Nat. Commun.* **3** 1026

[53] Miller D A B 2013 Self-configuring universal linear optical component [Invited] *Photon. Res.* **1** 1–15

[54] Li G, Bai N, Zhao N and Xia C 2014 Space-division multiplexing: the next frontier in optical communication *Adv. Opt. Photon.* **6** 413–87

[55] de Castro Y and Gamboa F 2012 Exact reconstruction using Beurling minimal extrapolation *J. Math. Anal. Appl.* **395** 336–54

[56] Candès E J and Fernandez-Granda C 2014 Towards a mathematical theory of super-resolution *Commun. Pure Appl. Math.* **67** 906–56

[57] Schiebinger G, Robeva E and Recht B 2015 Superresolution without Separation arXiv:1506.03144

[58] Cox D R and Reid N 1987 Parameter orthogonality and approximate conditional inference *J. R. Stat. Soc.* B **49** 1–39 http://www.jstor.org/stable/2345476

[59] van Trees H L and Bell K L (ed) 2007 *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking* (Piscataway, NJ: Wiley)

[60] Labeyrie A, Lipson S G and Nisenson P 2006 *An Introduction to Optical Stellar Interferometry* (Cambridge: Cambridge University Press)

[61] Moerner W E 2007 New directions in single-molecule imaging and analysis *Proc. Natl Acad. Sci.* **104** 12596–602

[62] Betzig E, Patterson G H, Sougrat R, Lindwasser O W, Olenych S, Bonifacino J S, Davidson M W, Lippincott-Schwartz J and Hess H F 2006 Imaging intracellular fluorescent proteins at nanometer resolution *Science* **313** 1642–5

[63] Hell S W 2007 Far-field optical nanoscopy *Science* **316** 1153–8

[64] Kolobov M I (ed) 2007 *Quantum Imaging* (New York: Springer)

[65] Taylor M A, Janousek J, Daria V, Knittel J, Hage B, Bachor H-A and Bowen W P 2013 Biological measurement beyond the quantum limit *Nat. Photon.* **7** 229–33

[66] Tsang M 2015 Quantum limits to optical point-source localization *Optica* **2** 646–53

[67] Pendry J B 2000 Negative refraction makes a perfect lens *Phys. Rev. Lett.* **85** 3966–9

[68] Tsang M and Psaltis D 2008 Magnifying perfect lens and superlens design by coordinate transformation *Phys. Rev.* B **77** 035122

[69] Fujiwara A 2001 Quantum channel identification problem *Phys. Rev.* A **63** 042304

[70] Ng S, Ang S Z, Wheatley T A, Yonezawa H, Furusawa A, Huntington E H and Tsang M 2016 Spectrum analysis with quantum dynamical systems *Phys. Rev.* A **93** 042121

[71] Stein M, Mezghani A and Nossek J A 2014 A lower bound for the fisher information measure *IEEE Signal Process. Lett.* **21** 796–9

[72] Alipour S and Rezakhani A T 2015 Extended convexity of quantum fisher information in quantum metrology *Phys. Rev.* A **91** 042104

[73] Gull S F and Skilling J 1984 Maximum entropy method in image processing *IEE Proc.* F **131** 646–59

[74] Milanfar P, Karl W C and Willsky A S 1996 A moment-based variational approach to tomographic reconstruction *IEEE Trans. Image Process.* **5** 459–70

[75] Kahn J and Guţă M 2009 Local asymptotic normality for finite dimensional quantum systems *Commun. Math. Phys.* **289** 597–652

[76] Berry D W, Tsang M, Hall M J W and Wiseman H M 2015 Quantum Bell-Ziv-Zakai bounds and Heisenberg limits for waveform estimation *Phys. Rev.* X **5** 031018

[77] Lu X-M and Tsang M 2016 Quantum Weiss-Weinstein bounds for quantum metrology *Quantum Sci. Technol.* **1** 015002

[78] Gagatsos C N, Branford D and Datta A 2016 Gaussian systems for quantum-enhanced multiple phase estimation *Phys. Rev.* A **94** 042342

[79] Ježek M and Hradil Z 2004 Reconstruction of spatial, phase, and coherence properties of light *J. Opt. Soc. Am.* A **21** 1407

[80] Lvovsky A I and Raymer M G 2009 Continuous-variable optical quantum-state tomography *Rev. Mod. Phys.* **81** 299–332

[81] Gross D, Liu Y-K, Flammia S T, Becker S and Eisert J 2010 Quantum state tomography via compressed sensing *Phys. Rev. Lett.* **105** 150401