



This is the accepted manuscript made available via CHORUS. The article has been published as:

Modern description of Rayleigh's criterion

Sisi Zhou and Liang Jiang

Phys. Rev. A **99**, 013808 — Published 3 January 2019

DOI: [10.1103/PhysRevA.99.013808](https://doi.org/10.1103/PhysRevA.99.013808)

A modern description of Rayleigh’s criterion

Sisi Zhou and Liang Jiang

*Departments of Applied Physics and Physics, Yale University, New Haven, Connecticut 06511, USA and
Yale Quantum Institute, Yale University, New Haven, Connecticut 06511, USA*

(Dated: October 29, 2018)

Rayleigh’s criterion states that it becomes essentially difficult to resolve two incoherent optical point sources separated by a distance below the width of point spread functions (PSF), namely in the subdiffraction limit. Recently, researchers have achieved superresolution for two incoherent point sources with equal strengths using a new type of measurement technique, surpassing Rayleigh’s criterion. However, situations where more than two point sources needed to be resolved have not been fully investigated. Here we prove that for any incoherent sources with arbitrary strengths, a one- or two-dimensional image can be precisely resolved up to its second moment in the subdiffraction limit, i.e. the Fisher information (FI) is non-zero. But the FI with respect to higher order moments always tends to zero polynomially as the size of the image decreases, for any type of non-adaptive measurement. We call this phenomenon a modern description of Rayleigh’s criterion. For PSFs under certain constraints, the optimal measurement basis estimating all moments in the subdiffraction limit for 1D weak-source imaging is constructed. Such basis also generates the optimal-scaling FI with respect to the size of the image for 2D or strong-source imaging, which achieves an overall quadratic improvement compared to direct imaging.

I. INTRODUCTION

Rayleigh’s criterion, as a long-standing textbook theorem, puts a fundamental limit on the power of optical resolution [1, 2]. It states that when two points are separated from each other by a distance smaller than the width of point-spread function (PSF) of the optical system [3], namely in the subdiffraction limit, it becomes essentially difficult to distinguish them. Recently however, researchers made a breakthrough towards surpassing Rayleigh’s criterion using a new type of measurement technique, by looking at the imaging problem from the perspective of quantum metrology [4–21].

In metrology, Fisher information (FI) characterizes the ultimate precision of parameter estimation through Cramér-Rao bound [22–24]. When estimating the separation between two equal strength incoherent sources, it can be shown that FI tends to zero as they become closer when we use direct imaging approach (i.e. counting photons at different positions on the imaging plane). However, the quantum Fisher information (QFI, equal to the maximum FI over all possible quantum measurements) remains a constant, implying the possibility of superresolution [4]. In fact, many types of measurement have been proposed to achieve this kind of superresolution [4–7, 25–29] and some of these approaches have already been demonstrated experimentally [30–33]. For example, when the PSF is Gaussian, it is possible to achieve the highest estimation precision by projecting the optical field onto Hermite-Gaussian modes [4, 27, 28].

While this new approach appears to be a promising candidate to substantially improve imaging resolution, many questions are yet to be answered: (1) What is the ultimate precision one can achieve, in a general imaging scenario, given experimentalists access to all types of measurement? (2) Which type of measurement achieves such precision? In this paper, we tackle these two ques-

tions by conducting a comprehensive Fisher information analysis in the general scenario where the incoherent source distribution on the object plane is arbitrary.

A direct way to parametrize an image is to use positions and intensities of each point as parameters to be estimated. However, it may not be the best choice because the position of one specific point does not tell much about the structure of the whole image. Instead, we can use moments to characterize an image which has wide applications in image analysis [34]. Since the difficulty involved in calculating QFI increases significantly as the number of sources increase, we only consider the limiting values of QFIs as the size of the image tends to zero (much smaller than the width of PSF) which we call “the subdiffraction limit”.

In this paper, we choose normalized moments (normalized so that it has dimension of length) as parameters to be estimated, where detailed calculations for Gaussian PSFs and the spatial-mode demultiplexing (SPADE) measurement scheme are contained in Refs. [8, 28]. We obtain the fundamental precision limit of estimating moments in the subdiffraction limit which formulated a modern description of Rayleigh’s criterion, as opposed to the traditional Rayleigh’s criterion restricted by direct imaging. We find that the FI with respect to (wrt) second moment remains a positive value in the subdiffraction limit, in accordance with previous work on estimating the separation between two coherent source. However, the FI wrt higher order moments always vanishes in the subdiffraction limit for non-adaptive measurements, answering question (1). This result shows the capability of going beyond direct imaging will not provide unlimited power and only push image resolution one step forward – from the first moment (the centroid of the image) to the second moment. To be specific, if we use s to represent the size of an image, the FI wrt to the K -th order moment vanishes as $O(s^{K-2})$ ($O(s^{K-1})$) when K is even

(odd), compared with $O(s^{2K-2})$ using direct imaging.

Based on the FI analysis, we also obtain optimal quantum measurements (in the subdiffraction limit) corresponding to the optimal FI. It is shown in this paper that when PSF is under certain constraints, the optimal measurement basis is strongly related to its derivatives. Roughly speaking, the probability from projecting the optical field onto the K -th order derivative of the PSF provides information of the $2K$ -th order moment of the image. And choosing derivatives as the measurement basis successfully classifies information of different moments into different measurement outcomes, which will provide optimal FIs wrt these moments in the subdiffraction limit. In this paper, we partially answer question (2) by first providing optimal quantum measurement scheme for second moment. For higher order moments, we prove the optimality of this scheme for 1D weak-source imaging. For 2D imaging or for strong-source imaging, such scheme only provides the optimal scaling of FI wrt s , but the coefficient may be further improved.

The paper is organized as follows. In Sec. II, we provide the formalism of the far-field imaging of incoherent optical sources. In Sec. III, we consider imaging for weak incoherent sources in one-dimensional imaging. We show that the Fisher information (FI) with respect to normalized moments decreases polynomially as the size of the image decrease, by order-of-magnitude analysis. To be specific, the FI wrt second moments remains a constant as the size of the image tends to zero, and the FI wrt to higher order moments drops to zero. In Sec. IV, we generalize the statement in Sec. III to sources with arbitrary strength, again by order-of-magnitude analysis. In Sec. V, we detail the FI analysis wrt to second moments by providing the exact value of FI and corresponding optimal measurements, as FI wrt second moment is not influenced by Rayleigh's criterion. In Sec. VI, we generalize all discussions about one-dimensional imaging to two-dimensional imaging, including calculating FI wrt to second moments in 2D. In Sec. VII, we detail the FI analysis wrt to all moments and show how the optimal scaling of FI can be achieved wrt all moments, which is improved quadratically when compared to direct imaging. Sec. VII also serves as a justification of the order-of-magnitude analysis in Sec. III and Sec. IV.

II. FORMALISM

Consider a one-dimensional object composed of J points on the object plane. The original field on the object plane can be expressed using P representation [35],

$$\rho_0 = \int D\alpha P_{\Gamma_0}(\alpha) |\alpha\rangle \langle \alpha|, \quad (1)$$

where $\alpha = (\alpha_1, \dots, \alpha_J)^T$ is the column vector of complex field amplitudes for J optical spatial modes and

$$|\alpha\rangle = \left(\prod_{j=1}^J e^{-\frac{|\alpha_j|^2}{2}} e^{\alpha_j a_j^\dagger} \right) |0\rangle, \quad (2)$$

where $|0\rangle$ is the vacuum state, a_j^\dagger and a_j are the canonical creation and annihilation operators at position x_j . Suppose the fields are uncorrelated at different points on the object plane, then $P_{\Gamma_0}(\alpha)$ is the independent Gaussian distribution of the J modes:

$$P_{\Gamma_0}(\alpha) = \prod_{j=1}^J \frac{1}{\pi(\Gamma_0)_j} \exp\left(-\sum_{j=1}^J \frac{|\alpha_j|^2}{(\Gamma_0)_j}\right), \quad (3)$$

where $(\Gamma_0)_j \geq 0$ is the average photon number emitted at the j th point and $\Gamma_0 = ((\Gamma_0)_1, \dots, (\Gamma_0)_J)^T$.

The imaging system maps the source operators a_j, a_j^\dagger into the image operators ψ_j, ψ_j^\dagger with an attenuation factor η :

$$a_j^\dagger \rightarrow \sqrt{\eta} \psi_j^\dagger + \sqrt{1-\eta} v_j^\dagger. \quad (4)$$

Here η is the transmission probability. $\psi_j^\dagger = \int dx \psi_{\text{PSF}}(x-x_j) a_x^\dagger$ is described by the point-spread function $\psi_{\text{PSF}}(x)$ (normalized) where a_x^\dagger is the canonical creation operator at position x and v_j^\dagger is the creation operator of the auxiliary environmental modes [7]. Moreover, we assume the PSF satisfies the following assumption

$$\int_{-\infty}^{\infty} \left(\frac{d^\ell}{dx^\ell} \psi_{\text{PSF}}^*(x) \right) \left(\frac{d^{\ell+1}}{dx^{\ell+1}} \psi_{\text{PSF}}(x) \right) dx = 0, \quad \forall \ell \geq 0, \quad (5)$$

which will later be used in determining the optimal measurement basis. This assumption is easily satisfied, for example, when PSFs are real (real PSFs can be implemented by a two-lens system [3]), e.g. $\psi_{\text{PSF}}(x) \propto e^{-x^2/4\sigma^2}$; or when they are even, e.g. $\psi_{\text{PSF}}(x) \propto e^{ikx^2/2z} \text{sinc}(x/\sigma)$.

After tracing out the environment, the field on the image plane is expressed as

$$\rho = \int D\alpha P_\Gamma(\alpha) |\psi_\alpha\rangle \langle \psi_\alpha|, \quad (6)$$

where

$$|\psi_\alpha\rangle = \frac{\prod_{j=1}^J e^{-\frac{|\alpha_j|^2}{2}} e^{\alpha_j \psi_j^\dagger} |0\rangle}{\left(\langle 0 | \prod_{j=1}^J e^{-|\alpha_j|^2} e^{\alpha_j^* \psi_j} e^{\alpha_j \psi_j^\dagger} |0\rangle \right)^{1/2}}, \quad (7)$$

and $\Gamma := \eta \Gamma_0$ is the average photon number received from each mode. We also define the average photon number on the image plane $\epsilon := \sum_{j=1}^J \Gamma_j$ (which is usually a small number) and the relative source strength $\gamma_j := \Gamma_j/\epsilon$ for later use. We can see that after integrating all phases in

α , only those photon number diagonal terms will survive and we may write

$$\rho = \sum_{m=0}^{\infty} \pi_m \rho_m, \quad (8)$$

where π_m is the probability of having m photons in the state and ρ_m is an m -photon multimode Fock state.

Our goal is to extract information of the image from ρ . We use a set of positive operators $\{E(n)\}$ satisfying $\sum_n E(n) = I$ to represent the positive-operator valued measure (POVM) performed on ρ [23, 36]. The resultant probability distributions are

$$P(n) = \text{tr}(\rho E(n)) \equiv \mathbb{E}[\langle \psi_\alpha | E(n) | \psi_\alpha \rangle], \quad (9)$$

where $\mathbb{E}[\cdot]$ represents expectation values under Gaussian distribution $P_\Gamma(\alpha)$.

The Cramér-Rao bound [22]

$$\Sigma \succeq \mathcal{F}^{-1} \quad (10)$$

provides the ultimate precision limit in terms of parameter estimation, where “ \succeq ” means the LHS minus the RHS is positive semi-definite, $\Sigma_{k\ell}$ is the error covariance matrix wrt parameters $\{M_k\}_{k \geq 1}$ and

$$\mathcal{F}_{k\ell} = \sum_n \frac{1}{P(n)} \frac{\partial P(n)}{\partial M_k} \frac{\partial P(n)}{\partial M_\ell} \quad (11)$$

is the corresponding Fisher information matrix (FIM). M_k are some functions of $\{x_j, \Gamma_j\}$, later chosen to be the normalized moments.

III. MODERN RAYLEIGH’S CRITERION FOR WEAK INCOHERENT SOURCES

The probability of measurement outcome n is

$$P(n) = \mathbb{E}[\langle \psi_\alpha | E(n) | \psi_\alpha \rangle] = \mathbb{E} \left[\frac{\langle 0 | e^{\alpha^\dagger \psi} E(n) e^{\psi^\dagger \alpha} | 0 \rangle}{\langle 0 | e^{\alpha^\dagger \psi} e^{\psi^\dagger \alpha} | 0 \rangle} \right], \quad (12)$$

where $\psi = (\psi_1, \dots, \psi_J)^T$ is the column vector of annihilation operators ψ_j . In the limit where the average photon number on the image plane ϵ is small (the value of ϵ is considered known because it is easy to measure), we can expand it as a series in ϵ :

$$P(n) = (1 - \epsilon) \langle 0 | E(n) | 0 \rangle + \epsilon p(n) + O(\epsilon^2), \quad (13)$$

where $p(n) := \epsilon \sum_{j=1}^J \gamma_j \langle 0 | \psi_j E(n) \psi_j^\dagger | 0 \rangle$. Since the first term contains no information of the object, the FIM will be dominated by the second term, which corresponds to the situation where only one photon is detected. To study the behavior of FIM in the subdiffraction limit, we expand ψ_j around its centroid \bar{X} . One should be careful with the convergence radius of the series expansion though, which has a lower bound independent of the

measurement $E(n)$ (see App. (B)). The second term in Eq. (13) becomes

$$\epsilon p(n) = \epsilon \sum_{k=0}^{\infty} \frac{p_k(n)}{k!} (M_k)^k, \quad (14)$$

where $p_k(n) = \sum_{j=1}^J \left(\frac{\partial}{\partial x_j} \right)^k p(n) \big|_{x_j = \bar{X}}$ is equal to the k -th order derivative of $\langle 0 | \psi_{\bar{X}} E(n) \psi_{\bar{X}}^\dagger | 0 \rangle$ wrt \bar{X} and M_k are normalized moments defined by

$$M_k = \left(\sum_{j=1}^J \gamma_j (x_j - \bar{X})^k \right)^{1/k} \quad (15)$$

for $k \geq 0$. Note that $(\cdot)^{1/k}$ is introduced here only to make sure M_k has dimension of length so that the estimation error should be comparable with the size of the image. Other methods to normalize moments, e.g. $M_k = (\sum_{j=1}^J \gamma_j (x_j - \bar{X})^k) / (\sum_{j=1}^J \gamma_j (x_j - \bar{X})^2)^{\frac{(k-1)}{2}}$ should also generate similar results. Here we wouldn’t worry about the phase of M_k because it is well defined locally. For example when $M_k = i |M_k|$, we can estimate $|M_k|$ instead so that all parameters are real.

Although $\{M_k\}_{k \geq 1}$ fully characterize the object configuration, they may not be independent given prior information of the object, but we can always choose a smaller set of independent moments as the parameters to be measured. For example, if the object contains only two points, there are only three degrees of freedom — the positions of two points and the ratio of their strengths, then we choose the first three moments as the parameters to be measured.

We use $s = \max_{i,j} |x_j - x_i|$ to characterize the size of the image and conduct FI analysis in the subdiffraction limit when $s \rightarrow 0$. Here we assume the centroid of the image $\bar{X} = \sum_{j=1}^J \gamma_j x_j$ is known accurately either based on existing telescopic data or pre-estimation. In this case, we have $M_1 = 0$.

In App. (H), we provide a measurement scheme for pre-estimation of \bar{X} . We only consider the situation where the centroid \bar{X} is already located around a reference point X_R such that $|\bar{X} - X_R| \lesssim s$ and analyze the resource required to precisely estimate \bar{X} and M_2 further. Half of the measurements are devoted to estimating \bar{X} and the other half to estimating M_2 using the optimal measurements wrt them respectively. We show that in 1D imaging, the scheme is optimal for weak sources and at least 93.0% efficient for strong sources. The methodology behind this scheme is presented in Sec. VII. Therefore we are not going to explain it in detail here.

Since any converging power series is dominated by its first non-zero term as $s \rightarrow 0$, we have

$$\frac{\partial P(n)}{\partial M_k} = O(s^{k-1}) \quad \text{and} \quad \frac{1}{P(n)} \frac{\partial P(n)}{\partial M_k} = O(s^{-1}). \quad (16)$$

Note that when the terms of lower order than k in $P(n)$ does not vanish, $\frac{1}{P(n)} \frac{\partial P(n)}{\partial M_k}$ should be bounded by a power

of s with higher order than $O(s^{-1})$. From Eq. (16), the FI for $k \geq 2$ would be

$$\mathcal{F}_{kk} = \sum_n \frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial M_k} \right)^2 = O(s^{k-2}), \quad (17)$$

which indicates the following theorem:

Theorem 1 (Modern Rayleigh's criterion for 1D imaging): *For imaging of incoherent point sources in the subdiffraction limit, the estimation variance of moment $M_{k>2}$ increases inverse-polynomially as s decreases; meanwhile, the estimation variance of the second moment M_2 is bounded by a constant independent of s .*

Note that we only need to bound the diagonal element of the FIM because the variance in estimation M_k satisfies

$$\Sigma_{kk} \geq (\mathcal{F}^{-1})_{kk} \geq \mathcal{F}_{kk}^{-1}. \quad (18)$$

where the equality holds true when \mathcal{F} is diagonal.

A simple schematic illustration of above theorem is shown in Fig. 1. Further justifications are contained in Sec. IV, Sec. V and Sec. VII. We discuss the validity of this order-of-magnitude analysis in App. (B). We emphasize here that the measurement is assumed to be non-adaptive in this paper and our analysis does not include the case where measurement can be adaptively modified (App. (B)) assuming prior knowledge on the moments to be estimated. And the adaptive measurement is excluded because it requires demanding experimental techniques. A more general analysis through direct calculation of quantum Fisher information, which can be applied to all type of measurement, can be found in Ref. [37].

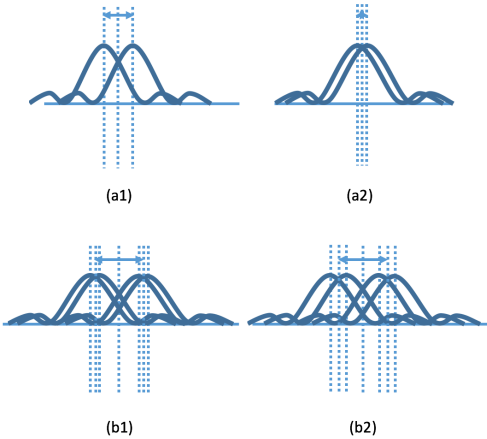


FIG. 1. (a) Images (a1) and (a2) have different M_2 . Consider two point sources with equal source strengths. The distance between them equal to $2M_2$ can be measured precisely, therefore it shall be easy to distinguish (a1) and (a2). (b) Images (b1) and (b2) have the same M_2 but different M_4 . Consider four point sources with equal source strengths. It is difficult to estimate the third and higher moments to distinguish the two images from each other.

IV. MODERN RAYLEIGH'S CRITERION FOR INCOHERENT SOURCES WITH ARBITRARY STRENGTHS

In this section, we generalize the discussion from the weak source limit to sources with arbitrary strengths.

In Eq. (12), we replace $\psi^\dagger \alpha$ with its expansion $\sum_{j=1}^J \alpha_j \int dx \psi_{\text{PSF}}(x - x_j) a_x^\dagger \equiv \sum_{k=0}^\infty \frac{A^{(k)}}{k!} \psi_{\bar{X}}^{(k)\dagger}$, where $A^{(k)} = \sum_{j=1}^J \alpha_j (x_j - \bar{X})^k$ and

$$\psi_{\bar{X}}^{(k)\dagger} = \frac{d^k}{d\bar{X}^k} \int dx \psi_{\text{PSF}}(x - \bar{X}) a_x^\dagger. \quad (19)$$

According to Wick's theorem (Isserlis' theorem) [38], any moment of Gaussian distributions can be expressed in terms of second order moments

$$\mathbb{E}[A^{(\ell_1)} A^{(\ell_2)*}] = \sum_{j=1}^J \Gamma_j (x_j - \bar{X})^{\ell_1 + \ell_2}; \quad (20)$$

$$\mathbb{E}[A^{(\ell_1)} A^{(\ell_2)}] = 0. \quad (21)$$

Here $\mathbb{E}[A^{(\ell_1)} A^{(\ell_2)}]$ vanishes when integrating wrt phases of α .

We observe that $P(n)$ can be decomposed into a power series of $O(s)$, like in Eq. (14),

$$P(n) = \sum_{k=0}^\infty Q_k(n; \{M_\ell, \ell \leq k\}), \quad (22)$$

where $Q_k(n) \equiv Q_k(n)$ is a function of the moments M_ℓ with $\ell \leq k$ so that $Q_k(n) = O(s^k)$. Explicit expressions of $Q_{0,1,2}(n)$ are provided in App. (C). For example,

$$Q_0(n) = \sum_{k=0}^\infty \frac{\epsilon^k}{(1+\epsilon)^{k+1}} \langle 0 | (\psi_{\bar{X}})^k E(n) (\psi_{\bar{X}}^\dagger)^k | 0 \rangle, \quad (23)$$

which is the probability of outcome n when all J points are located at the centroid \bar{X} with thermal average 'excitation' number ϵ . Hence, we have shown that order-of-magnitude analysis is still valid.

Specially, for $\epsilon \ll 1$, the expansion of $Q_k(n)$ depends solely on $p_k(n)$ and M_k :

$$Q_0(n) = \langle 0 | E(n) | 0 \rangle + O(\epsilon), \quad (24)$$

$$Q_k(n) = \epsilon \frac{p_k(n)}{k!} (M_k)^k + O(\epsilon^2), \quad \forall k \geq 1, \quad (25)$$

and Eq. (22) simplifies to Eq. (13) for weak incoherent sources. We also notice that $Q_2(n)/Q_0(n) = O(\epsilon s^2)$ (see App. (C)), which means the subdiffraction limit (requiring $Q_2(n) \ll Q_0(n)$) needs smaller s as ϵ increases.

V. FI WRT SECOND MOMENT AND CORRESPONDING OPTIMAL MEASUREMENT

In Sec. III, we have shown that there is a possibility to obtain a non-zero FI wrt M_2 . We are now going to find

the exact value of the optimal FI wrt second moment and corresponding measurement basis. First, let's consider the weak-source scenario,

$$\mathcal{F}_{22} = \sum_n \frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial M_2} \right)^2. \quad (26)$$

As $s \rightarrow 0$, $P(n)$ and $\frac{\partial P(n)}{\partial M_2}$ will be dominated by its first non-zero term, therefore according to Eq. (14),

$$\lim_{s \rightarrow 0} \mathcal{F}_{22} = \epsilon \sum_{n \in N_0^w} \frac{1}{\frac{p_2(n)}{2} (M_2)^2} (p_2(n) (M_2))^2 = 4\epsilon (\langle 0 | \psi_{\bar{X}}^{(1)} E(N_0^w) \psi_{\bar{X}}^{(1)\dagger} | 0 \rangle + \text{Re}[\langle 0 | \psi_{\bar{X}}^{(2)} E(N_0^w) \psi_{\bar{X}}^\dagger | 0 \rangle]), \quad (27)$$

where we define a set of 0-null measurement outcomes $N_0^w = \{n | \langle 0 | E(n) | 0 \rangle = \langle 0 | \psi_{\bar{X}} E(n) \psi_{\bar{X}}^\dagger | 0 \rangle = 0\}$ and $E(N_0^w) = \sum_{n \in N_0^w} E(n)$. We also note that $p_0(n) = 0$ implies $p_1(n) = 0$. Since $E(N_0^w)$ is Hermitian and non-negative, its eigenstates corresponding to non-vanishing eigenvalues must be orthogonal to $\psi_{\bar{X}}^\dagger | 0 \rangle$ and $\text{Re}[\langle 0 | \psi_{\bar{X}}^{(2)} E(N_0^w) \psi_{\bar{X}}^\dagger | 0 \rangle]$ must be zero. Therefore,

$$\max_{\{E(n)\}} \lim_{s \rightarrow 0} \mathcal{F}_{22} = 4\epsilon \int |\partial_x \psi_{\text{PSF}}(x)|^2 dx \equiv 4\epsilon \Delta k^2, \quad (28)$$

where the first equality is achieved when $\psi_{\bar{X}}^{(1)\dagger} | 0 \rangle$ is an eigenstate of $E(N_0^w)$ with an eigenvalue equal to one. For example,

$$E(N_0^w) = \frac{\psi_{\bar{X}}^{(1)\dagger} | 0 \rangle \langle 0 | \psi_{\bar{X}}^{(1)}}{\langle 0 | \psi_{\bar{X}}^{(1)} \psi_{\bar{X}}^{(1)\dagger} | 0 \rangle} \quad (29)$$

is optimal, in accordance with the optimality of the SPADE measurement scheme for Gaussian PSFs [4]. Furthermore, if $\psi_{\text{PSF}}(x)$ is an even function, its derivative will be odd and we can also choose $E(N_0^w)$ to be $\frac{I-\mathcal{P}}{2}$ where \mathcal{P} is the parity operator satisfying $\mathcal{P} \cdot f(x) = f(-x)$, which is the so-called SLIVER measurement scheme [25]. This type of measurement does not depend on the specific expressions of the point-spread functions.

We emphasize that above discussions are only applicable in the subdiffraction limit and the optimal measurement should be modified for finite s . When we consider the special case where there are only two equal strength point sources, however, Eq. (29) remains optimal even when s is large [4].

When we use direct imaging approach, i.e. $\{E(n)\} = \{a_x^\dagger a_x dx\}$, the 0-null measurement outcomes have zero measure and $\lim_{s \rightarrow 0} \mathcal{F}_{22} = 0$, because the probability density of the photon position x is

$$\langle 0 | \psi_{\bar{X}} a_x^\dagger a_x \psi_{\bar{X}}^\dagger | 0 \rangle = |\psi_{\text{PSF}}(x - \bar{X})|^2 \neq 0 \quad (30)$$

almost everywhere, which explains the traditional Rayleigh's criterion.

For an arbitrary source strength

$$\lim_{s \rightarrow 0} \mathcal{F}_{22} = \sum_{n \in N_0} \frac{1}{Q_2(n)} \left(\frac{\partial Q_2(n)}{\partial M_2} \right)^2, \quad (31)$$

where the 0-null measurement outcome $N_0 = \{n | Q_0(n) = \sum_{k=0}^{\infty} \frac{\epsilon^k}{k!(1+\epsilon)^{k+1}} \langle 0 | (\psi_{\bar{X}})^k E(n) (\psi_{\bar{X}}^\dagger)^k | 0 \rangle = 0\} = \{n | \langle 0 | (\psi_{\bar{X}})^k E(n) (\psi_{\bar{X}}^\dagger)^k | 0 \rangle = 0, \forall k\}$. We also note that $Q_0(n) = 0$ implies $Q_1(n) = 0$ (see App. (C)). A detailed calculation of Eq. (12) shows that when $n \in N_0$,

$$Q_2(n) = \sum_{k=0}^{\infty} \frac{\epsilon^{k+1}}{k!(1+\epsilon)^{k+1}} \langle 0 | (\psi_{\bar{X}})^k \psi_{\bar{X}}^{(1)} E(n) \psi_{\bar{X}}^{(1)\dagger} (\psi_{\bar{X}}^\dagger)^k | 0 \rangle M_2^2, \quad (32)$$

and hence

$$\max_{\{E(n)\}} \lim_{s \rightarrow 0} \mathcal{F}_{22} = 4\epsilon \int |\partial_x \psi_{\text{PSF}}(x)|^2 dx = 4\epsilon \Delta k^2. \quad (33)$$

It has the exact same expression as Eq. (28), meaning FI wrt the second moment grows linearly as the source strength grows, following the standard quantum limit [39]. Our results agree with previous work on estimating the separation between two incoherent sources for arbitrary source strengths [6, 7].

The measurement is optimal when $(\psi_{\bar{X}}^\dagger)^k \psi_{\bar{X}}^{(1)\dagger} | 0 \rangle$ are all eigenstates of $E(N_0)$ with eigenvalues equal to one. For example,

$$E(N_0) = \sum_{k=0}^{\infty} \frac{(\psi_{\bar{X}}^\dagger)^k \psi_{\bar{X}}^{(1)\dagger} | 0 \rangle \langle 0 | \psi_{\bar{X}}^{(1)} (\psi_{\bar{X}})^k}{k! \langle 0 | \psi_{\bar{X}}^{(1)} \psi_{\bar{X}}^{(1)\dagger} | 0 \rangle} \quad (34)$$

$$(\text{or when } \psi_{\text{PSF}}(x) \text{ is even, } E(N_0) = \frac{1 - \mathcal{P}}{2}) \quad (35)$$

is optimal, in accordance with the optimality of fin-SPADE and pix-SLIVER [6].

VI. GENERALIZATION TO 2D IMAGING

Results in previous sections can be directly generalized to two-dimensional imaging. Suppose there are J point sources at positions (x_j, y_j) . The normalized moments are redefined as following:

$$M_{k\ell} = \left(\sum_{j=1}^J \gamma_j (x_j - \bar{X})^k (y_j - \bar{Y})^\ell \right)^{\frac{1}{k+\ell}} \quad (36)$$

which fully characterizes the object configuration. Also, the size of the image $s := \max_{ij} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ and the centroid $(\bar{X}, \bar{Y}) := (\sum_{j=1}^J \gamma_j x_j, \sum_{j=1}^J \gamma_j y_j)$. We can expand the creation and annihilation operators

around the centroid ($\partial_{\bar{X}}^k$ denotes the k -th order derivative wrt \bar{X})

$$\begin{aligned}\psi_j^\dagger &= \int dx dy \psi_{\text{PSF}}(x - x_j, y - y_j) a_{xy}^\dagger \\ &\equiv \sum_{k, \ell=0}^{\infty} \frac{\psi_{\bar{X}\bar{Y}}^{(k\ell)\dagger}}{k!\ell!} (x_j - \bar{X})^k (y_j - \bar{Y})^\ell,\end{aligned}\quad (37)$$

and calculate the probability distribution $P(n)$ which is a series of $O(s^k)$

$$P(n) = \sum_{K=0}^{\infty} Q_K(n). \quad (38)$$

Similar to Eq. (16), we have the following order-of-magnitude analysis:

$$\frac{\partial P(n)}{\partial M_{k\ell}} = O(s^{k+\ell-1}) \text{ and } \frac{1}{P(n)} \frac{\partial P(n)}{\partial M_{k\ell}} = O(s^{-1}); \quad (39)$$

and similar to Eq. (17), the diagonal elements of the FI matrix is

$$\mathcal{F}_{k\ell k\ell} = \sum_n \frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial M_{k\ell}} \right)^2 = O(s^{k+\ell-2}), \quad (40)$$

thus extending the modern description of Rayleigh's criterion to 2D imaging:

Theorem 2 (Modern Rayleigh's criterion for 2D imaging): *For imaging of incoherent point sources in the subdiffraction limit, the estimation variance of any moment $M_{k\ell}$ with $k + \ell > 2$ increases inverse-polynomially as s decreases; however, the estimation variance of the second moment M_{20} , M_{11} and M_{02} are bounded by a constant independent of s .*

A simple schematic illustration above theorem is shown in Fig. 2. We are now going to find the exact values of FI wrt M_{20} , M_{11} and M_{02} and corresponding optimal measurements. For simplicity we consider the weak source scenario. For arbitrary source strength, the FIs are still the same and the optimal measurements $E(n)$ should be replaced with $\sum_{k=0}^{\infty} \frac{1}{k!} (\psi_{\bar{X}}^\dagger)^k E(n) (\psi_{\bar{X}})^k$ because

$$\begin{aligned}Q_2(n) &= \sum_{k=0}^{\infty} \frac{\epsilon^{k+1}}{k!(1+\epsilon)^{k+1}} \\ &\quad (\langle 0 | (\psi_{\bar{X}})^k \psi_{\bar{X}\bar{Y}}^{(10)} E(n) \psi_{\bar{X}\bar{Y}}^{(10)\dagger} (\psi_{\bar{X}}^\dagger)^k | 0 \rangle M_{20}^2 \\ &\quad + 2\text{Re}[\langle 0 | (\psi_{\bar{X}})^k \psi_{\bar{X}\bar{Y}}^{(10)} E(n) \psi_{\bar{X}\bar{Y}}^{(01)\dagger} (\psi_{\bar{X}}^\dagger)^k | 0 \rangle] M_{11}^2 \\ &\quad + \langle 0 | (\psi_{\bar{X}})^k \psi_{\bar{X}\bar{Y}}^{(01)} E(n) \psi_{\bar{X}\bar{Y}}^{(01)\dagger} (\psi_{\bar{X}}^\dagger)^k | 0 \rangle M_{02}^2),\end{aligned}\quad (41)$$

which is a generalization of Eq. (32) from 1D to 2D.

Suppose $[\psi_{\bar{X}\bar{Y}}^{(10)\dagger}, \psi_{\bar{X}\bar{Y}}] = [\psi_{\bar{X}\bar{Y}}^{(01)\dagger}, \psi_{\bar{X}\bar{Y}}] = 0$ and $\langle 0 | \psi_{\bar{X}\bar{Y}}^{(10)} \psi_{\bar{X}\bar{Y}}^{(01)\dagger} | 0 \rangle \in \mathbb{R}$. This assumption is satisfied, for

example, when the PSF is real. The second order term of $P(n)$ is

$$\begin{aligned}Q_2(n) &= \epsilon (\langle 0 | \psi_{\bar{X}\bar{Y}}^{(10)} E(n) \psi_{\bar{X}\bar{Y}}^{(10)\dagger} | 0 \rangle M_{20}^2 \\ &\quad + 2\text{Re}[\langle 0 | \psi_{\bar{X}\bar{Y}}^{(10)} E(n) \psi_{\bar{X}\bar{Y}}^{(01)\dagger} | 0 \rangle] M_{11}^2 \\ &\quad + \langle 0 | \psi_{\bar{X}\bar{Y}}^{(01)} E(n) \psi_{\bar{X}\bar{Y}}^{(01)\dagger} | 0 \rangle M_{02}^2) + O(\epsilon^2).\end{aligned}\quad (42)$$

We only consider 0-null measurement outcome $n \in N_0^w = \{n | \langle 0 | E(n) | 0 \rangle = \langle 0 | \psi_{\bar{X}\bar{Y}} E(n) \psi_{\bar{X}\bar{Y}}^\dagger | 0 \rangle = 0, \forall k\}$ because for $n \notin N_0^w$, the zeroth order term of $P(n)$ would be positive and does not contribute to the FI as $s \rightarrow 0$. Furthermore, we assume $E(n) = \Pi E(n) \Pi$ where Π is the projection onto the space $\text{span}\{\psi_{\bar{X}\bar{Y}}^{(10)\dagger} | 0 \rangle, \psi_{\bar{X}\bar{Y}}^{(01)\dagger} | 0 \rangle\}$ because any component of $E(n)$ perpendicular to it does not contribute to $Q_2(n)$ in the first order expansion of ϵ and consequently only affects the value of the FI in higher order terms of ϵ .

Then we can write every operator as a two-dimensional matrix in basis

$$\begin{aligned}\{|e_1\rangle &= \frac{1}{\sqrt{2(1+r)}} \left(\frac{\psi_{\bar{X}\bar{Y}}^{(10)\dagger}}{\Delta k_x} + \frac{\psi_{\bar{X}\bar{Y}}^{(01)\dagger}}{\Delta k_y} \right) | 0 \rangle, \\ |e_2\rangle &= \frac{1}{\sqrt{2(1-r)}} \left(\frac{\psi_{\bar{X}\bar{Y}}^{(10)\dagger}}{\Delta k_x} - \frac{\psi_{\bar{X}\bar{Y}}^{(01)\dagger}}{\Delta k_y} \right) | 0 \rangle\},\end{aligned}\quad (43)$$

where

$$\Delta k_x^2 := \langle 0 | \psi_{\bar{X}\bar{Y}}^{(10)} \psi_{\bar{X}\bar{Y}}^{(10)\dagger} | 0 \rangle = \int dx dy |\partial_x \psi_{\text{PSF}}(x, y)|^2, \quad (44)$$

$$\Delta k_y^2 := \langle 0 | \psi_{\bar{X}\bar{Y}}^{(01)} \psi_{\bar{X}\bar{Y}}^{(01)\dagger} | 0 \rangle = \int dx dy |\partial_y \psi_{\text{PSF}}(x, y)|^2, \quad (45)$$

$$\begin{aligned}r &:= \langle 0 | \psi_{\bar{X}\bar{Y}}^{(10)} \psi_{\bar{X}\bar{Y}}^{(01)\dagger} | 0 \rangle / (\Delta k_x \Delta k_y) \\ &= \frac{1}{\Delta k_x \Delta k_y} \int dx dy \partial_x \psi_{\text{PSF}}^*(x, y) \partial_y \psi_{\text{PSF}}(x, y).\end{aligned}\quad (46)$$

and $r \in (-1, 1)$. Therefore,

$$Q_2(n) \approx \epsilon \text{tr}(E(n) \rho_2), \quad (47)$$

where

$$\begin{aligned}\rho_2 &= \frac{1}{2} ((\Delta k_x^2 M_{20}^2 + \Delta k_y^2 M_{02}^2)(I + r\sigma_z) \\ &\quad + 2\Delta k_x \Delta k_y M_{11}^2 (rI + \sigma_z) \\ &\quad + \sqrt{1-r^2} (\Delta k_x^2 M_{20}^2 - \Delta k_y^2 M_{02}^2) \sigma_x).\end{aligned}\quad (48)$$

Note that ρ_2 depends not only on the PSF via $(\Delta k_x, \Delta k_y, r)$ but also on the second moments. The FIM can be then be calculated using $Q_2(n)$ for any specific POVM $\{E(n)\}$.

One way to parametrize the second moments is to define $M_{20} = X^2$, $M_{02} = Y^2$ and $M_{11} = \beta XY$, where X , Y is the standard deviation along x - and y - axis and β is the

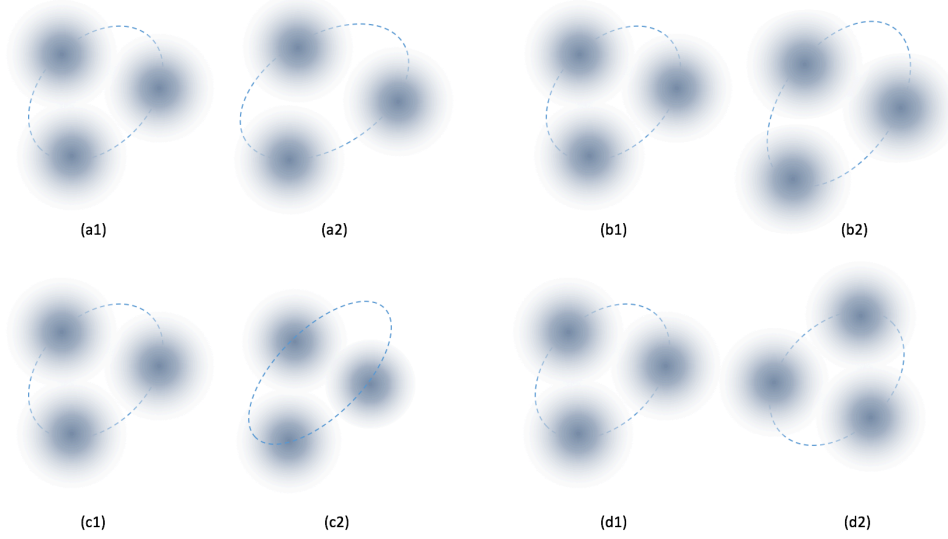


FIG. 2. Three point sources with equal source strengths. Given the values of (M_{20}, M_{11}, M_{02}) , three points are distributed on an ellipse $\frac{2}{3}(1 - \frac{M_{11}^2}{M_{20}^2 M_{02}^2}) = \frac{x^2}{M_{20}^2} + \frac{y^2}{M_{02}^2} - \frac{2M_{11}^2 xy}{M_{20}^2 M_{02}^2}$. (a) Images (a1) and (a2) are distinguishable due to different standard deviations along x -axis $X = M_{20}$ (b) Images (b1) and (b2) are distinguishable due to different standard deviations along y -axis $Y = M_{02}$. (c) Images (c1) and (c2) are distinguishable due to different x - y correlations $\beta = M_{11}^2/(M_{20}M_{02})$. (d) Images (d1) and (d2) have the same (M_{20}, M_{11}, M_{02}) . It is difficult to distinguish them from each other.

correlation between the distributions along x - and y - axis. If we approximate the image by a Gaussian distribution $P(x, y) = \frac{1}{2\pi XY\sqrt{1-\beta^2}} \exp(-\frac{1}{2(1-\beta^2)}(x \ y)\mathbf{C}^{-1}(x \ y)^T)$, where the covariance matrix is

$$\mathbf{C} = \begin{pmatrix} M_{20}^2 & M_{11}^2 \\ M_{11}^2 & M_{02}^2 \end{pmatrix} = \begin{pmatrix} X^2 & \beta XY \\ \beta XY & Y^2 \end{pmatrix}. \quad (49)$$

The contour lines of $P(x, y)$ will be ellipses described by $\frac{x^2}{X^2} + \frac{y^2}{Y^2} - \frac{2\beta xy}{XY} = \text{constant}$. Different distributions can be distinguished from each other if we can precisely estimate the values of (X, Y, β) . Another way to parametrize the second moments is to use the decomposition

$$\begin{aligned} \mathbf{C} &= \begin{pmatrix} M_{20}^2 & M_{11}^2 \\ M_{11}^2 & M_{02}^2 \end{pmatrix} \\ &= \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \Lambda_1^2 & 0 \\ 0 & \Lambda_2^2 \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \end{aligned} \quad (50)$$

The major and minor length of the ellipses correspond to the square root $\Lambda_{1,2}$ of the eigenvalues of \mathbf{C} and the orientation θ is associated with the direction of its eigenvectors. Estimation wrt $(\Lambda_1, \Lambda_2, \theta)$ is discussed in App. (D).

First let's consider the singular case where $\beta = 1$, $|M_{11}^2| = \sqrt{M_{20}^2 M_{02}^2}$ and ρ_2 is pure. It happens when all points sources are aligned on the same line, e.g. when there are only two point sources [5]. The optimal measurement can be determined by calculating quantum

Fisher information matrix (QFIM) wrt X and Y :

$$\mathcal{J}_{\mu\nu} = \epsilon \text{tr} \left(\frac{\mathcal{L}_\mu \mathcal{L}_\nu + \mathcal{L}_\nu \mathcal{L}_\mu}{2} \rho_2 \right), \quad \mu, \nu = X, Y \quad (51)$$

where the Hermitian operator \mathcal{L}_μ is the symmetric logarithmic derivative of ρ_2 wrt μ defined via $\partial_\mu \rho_2 = \frac{1}{2}(\mathcal{L}_\mu \rho_2 + \rho_2 \mathcal{L}_\mu)$ [24]. The QFIM derived from Eq. (48) is

$$\mathcal{J}[X, Y] = 4\epsilon \begin{pmatrix} \Delta k_x^2 & r \Delta k_x \Delta k_y \\ r \Delta k_x \Delta k_y & \Delta k_y^2 \end{pmatrix}. \quad (52)$$

The optimal measurement can be chosen to be any rank-one projection onto an orthonormal basis of the real space spanned by $\{|e_1\rangle, |e_2\rangle\}$, such as $\{E(n_1) = |e_1\rangle\langle e_1|, E(n_2) = |e_2\rangle\langle e_2|\}$ (the same as 2D-SPADE for Gaussian PSFs [5]) or $\{E(n_1) = |e_+\rangle\langle e_+|, E(n_2) = |e_-\rangle\langle e_-|\}$ where $|e_+\rangle = \frac{1}{\Delta k_x} \psi_{XY}^{(10)\dagger} |0\rangle$ and $|e_-\rangle = \frac{1}{\Delta k_y} \psi_{XY}^{(01)\dagger} |0\rangle$, because they will always satisfy the QFIM-achievable condition $E(n_i) \rho_2^{1/2} = c_{i,\mu} E(n_i) \mathcal{L}_\mu \rho_2^{1/2}$ for $i = 1, 2, \mu = X, Y$ and some real constant $c_{i,\mu}$.

For 2D PSF satisfying the following more strict assumption (generalized from Eq. (5)):

$$\begin{aligned} \int_{-\infty}^{\infty} \left(\frac{d^{\ell_1}}{dx^{\ell_1}} \frac{d^{\ell_2}}{dy^{\ell_2}} \psi_{\text{PSF}}^*(x, y) \right) \\ \left(\frac{d^{\ell_3}}{dx^{\ell_3}} \frac{d^{\ell_4}}{dy^{\ell_4}} \psi_{\text{PSF}}(x, y) \right) dx dy = 0, \\ \text{when } |\ell_1 - \ell_3| = 1 \text{ or } |\ell_2 - \ell_4| = 1, \end{aligned} \quad (53)$$

the FIM and corresponding measurement can be obtained in a simpler form (otherwise, the FIM can have off-diagonal terms). Eq. (53) is still quite general. When $\psi_{\text{PSF}}(x, y)$ is separable, i.e. $\psi_{\text{PSF}}(x, y) = \psi_{1,\text{PSF}}(x)\psi_{2,\text{PSF}}(y)$, Eq. (53) is automatically satisfied when $\psi_{1,\text{PSF}}(x)$ and $\psi_{2,\text{PSF}}(y)$ satisfy Eq. (5), e.g. $\psi_{\text{PSF}}(x, y) \propto e^{-(x^2+y^2)/4\sigma^2}$ or $\psi_{\text{PSF}}(x, y) \propto e^{ik(x^2+y^2)/2z} \text{sinc}(x/\sigma_1) \text{sinc}(y/\sigma_2)$. When $\psi_{\text{PSF}}(x, y)$ is a circularly-symmetric function, i.e. $\psi_{\text{PSF}}(x, y) = \psi_{\text{PSF}}(\sqrt{x^2+y^2})$, Eq. (53) is also true, e.g. $\psi_{\text{PSF}}(x, y) \propto e^{ik(x^2+y^2)/2z} \frac{J_1(\sqrt{x^2+y^2}/\sigma)}{\sqrt{x^2+y^2}}$ ($J_1(\cdot)$ is the first order Bessel function of the first kind). We assume from now on that Eq. (53) is satisfied for any 2D PSF. In this case, $r = 0$.

Note that if the projection $\{\Pi E(n) \Pi\}$ of measurements $\{E(n)\}$ onto the complex space spanned by $\{|e_1\rangle, |e_2\rangle\}$ is optimal, $\{E(n)\}$ is also optimal. In particular, when $\psi_{\text{PSF}}(x, y)$ is circularly symmetric, any measurement satisfying $\{\Pi E(n_1) \Pi = |e_+\rangle \langle e_+|, \Pi E(n_2) \Pi = |e_-\rangle \langle e_-|\}$ is optimal, including $\{E(n_1) = \frac{(I-\mathcal{P}_1)(I+\mathcal{P}_2)}{4}, E(n_2) = \frac{(I+\mathcal{P}_1)(I-\mathcal{P}_2)}{4}\}$ where the parity operators $\mathcal{P}_{1(2)}$ satisfies $\mathcal{P}_{1(2)}f(x, y) = f(-x, y)$ ($f(x, -y)$) (the same as 2D-SLIVER [5]). This type of measurement does not depend on the specific expressions of PSFs. In fact, any measurement $E(n) = \sum_{\mu\nu=+,-} m_{\mu,\nu} |e_\mu\rangle \langle e_\nu|$ can be transformed into a PSF-independent version by replacing $|e_+\rangle \langle e_+|$ with $\frac{(I-\mathcal{P}_1)(I+\mathcal{P}_2)}{4}$, $|e_+\rangle \langle e_-|$ with $\frac{(I-\mathcal{P}_1)(I+\mathcal{P}_2)}{4} \mathcal{S}_{12}$, $|e_-\rangle \langle e_+|$ with $\mathcal{S}_{12} \frac{(I-\mathcal{P}_1)(I+\mathcal{P}_2)}{4}$ and $|e_-\rangle \langle e_-|$ with $\frac{(I+\mathcal{P}_1)(I-\mathcal{P}_2)}{4}$ where $\mathcal{S}_{12}f(x, y) = f(y, x)$.

When M_{20} , M_{11} and M_{02} are independent parameters, $\beta < 1$, $|M_{11}^2| < M_{20}^2 M_{02}^2$ and ρ_2 is a mixed state. The QFIM wrt (X, Y, β) is

$$\mathcal{J}[X, Y, \beta] = 4\epsilon \begin{pmatrix} \Delta k_x^2 & 0 & 0 \\ 0 & \Delta k_y^2 & 0 \\ 0 & 0 & \frac{\Delta k_x^2 \Delta k_y^2 X^2 Y^2}{(\Delta k_x^2 X^2 + \Delta k_y^2 Y^2)(1-\beta^2)} \end{pmatrix}. \quad (54)$$

However, the QFIM is not simultaneously achievable for (X, Y, β) , meaning the quantum Cramér-Rao bound $\Sigma \succeq \mathcal{J}^{-1}$ is not attainable. The optimal measurement for (X, Y) is $\{\Pi E(n_1) \Pi = |e_+\rangle \langle e_+|, \Pi E(n_2) \Pi = |e_-\rangle \langle e_-|\}$ and the optimal measurement for β is $\{\Pi E(n_1) \Pi = |e'_1\rangle \langle e'_1|, \Pi E(n_2) \Pi = |e'_2\rangle \langle e'_2|\}$, where

$$|e'_1\rangle = \cos \theta' |e_1\rangle + \sin \theta' |e_2\rangle, \quad (55)$$

$$|e'_2\rangle = -\sin \theta' |e_1\rangle + \cos \theta' |e_2\rangle, \quad (56)$$

and $\theta' = \frac{1}{2} \tan^{-1} \left(\frac{\beta(X^2 - Y^2)}{2XY} \right)$. We note that when $\beta = 0$, the optimal measurement basis for (X, Y) and β are mutually unbiased. In fact, any three parameters characterizing ρ_2 can never be measured simultaneously using projection-valued measurement (PVM) because ρ_2 is only rank two. In practice, we can switch between different types of measurements during the measurement process. The resultant FIM will be the average of FIMs wrt each measurement.

VII. ESTIMATION OF ALL MOMENTS IN THE SUBDIFFRACTION LIMIT

Even though the information of normalized moments with an order higher than two is jeopardized in the subdiffraction limit, it is worth figuring out the maximum FI achievable and the optimal measurement corresponding to it as one may still need to measure the high-order normalized moments even when the FI is low and the estimation cost is expensive. In this section, we will assume all moments are independent variables and we only consider weak source scenario here. Generalization to sources with arbitrary strengths is contained in the App. (G). Ref. [28] contains a detailed discussion on the special case where the source is weak and the PSF is Gaussian, but the optimality was not proved there.

Eq. (5) and Eq. (53) are satisfied for PSFs in this section and the main result in this section can be summarized in this theorem:

Theorem 3 (Optimal scaling wrt all moments): For any moment M_K ($M_{L(K-L)}$) in 1D (2D) imaging with arbitrary source strengths, the estimation variance is at least $O(s^{2-K})$ when K is even and $O(s^{1-K})$ when K is odd.

For directly imaging, the denominator in Eq. (17) and Eq. (40) are always $O(1)$ and the Fisher information wrt M_K or $M_{L(K-L)}$ will be $O(s^{2K-2})$ which is $O(s^K)$ ($O(s^{K-1})$) times smaller than the maximum FI $O(s^{K-2})$ ($O(s^{K-1})$) for even (odd) moments we obtain here.

For simplicity, let's first look at the one-dimensional case with weak sources ($\epsilon \ll 1$), (The analysis for arbitrary source strengths is detailed in App. (G).) According to Eq. (16) and Eq. (17), the lowest power of s $\mathcal{F}_{k\ell}$ can attain is $\max\{k, \ell\} - 2$ if and only if there is an $E(n)$ such that $P(n)$ is zero until the $\min\{k, \ell\}$ -th order of s . However, this condition is not necessarily satisfiable for each moments.

In order for $p_0(n) = \langle 0 | \psi_{\bar{X}} E(n) \psi_{\bar{X}}^\dagger | 0 \rangle = 0$, $E(n)$ has to be orthogonal to $\psi_{\bar{X}}^\dagger | 0 \rangle$ ($\psi_{\bar{X}}^\dagger | 0 \rangle$ is not in the support of $E(n)$). Similarly, according to Eq. (32), in order for $Q_2(n)$ (up to the first order of ϵ) to be zero, $E(n)$ has to be orthogonal to $\psi_{\bar{X}}^{(1)\dagger} | 0 \rangle$. We define ℓ -null measurement outcomes

$$N_\ell^w = \{n | \langle 0 | E(n) | 0 \rangle = \langle 0 | \psi_{\bar{X}}^{(k)} E(n) \psi_{\bar{X}}^{(k)\dagger} | 0 \rangle = 0, \forall k \leq \ell\}, \quad (57)$$

and we have $N_\ell^w \subseteq N_{\ell-1}^w$ for all ℓ , that is, ℓ -null measurement is $(\ell-1)$ -null. Then for all $\ell \geq 0$, $Q_{2\ell} = O(\epsilon^2)$ requires $n \in N_\ell^w$. Suppose $n \in N_{\ell-1}^w$, then

$$Q_k(n) = O(\epsilon^2), \quad \forall k \leq 2\ell - 1 \quad (58)$$

and

$$Q_{2\ell}(n) = \frac{\epsilon}{\ell!^2} \langle 0 | \psi_{\bar{X}}^{(\ell)} E(n) \psi_{\bar{X}}^{(\ell)\dagger} | 0 \rangle (M_{2\ell})^{2\ell} + O(\epsilon^2). \quad (59)$$

We assume derivatives of the PSF $\{\partial_{\bar{X}}^k \psi_{\text{PSF}}(x - \bar{X}), k \geq 0\}$ form a linear independent subset in $L^2(\mathbb{C})$. An orthonormal set $\{b^{(k)}(x), k \geq 0\}$ can be generated via Gram-Schmidt process such that $b^{(\ell)}(x)$ is orthogonal to every $\partial_{\bar{X}}^k \psi_{\text{PSF}}(x - \bar{X})$ with $k \leq \ell - 1$ and

$$q_\ell := \frac{1}{\ell!} \int b^{(\ell)*}(x) \partial_{\bar{X}}^\ell \psi_{\text{PSF}}(x - \bar{X}) dx \in \mathbb{R}. \quad (60)$$

For example, when the PSF is Gaussian, $\{b^{(k)}(x), k \geq 0\}$ are the Hermite-Gaussian modes; when the PSF is a sinc function, $\{b^{(k)}(x), k \geq 0\}$ are the spherical Bessel functions of the first kind. We also notice that, according to Eq. (5), $\text{span}\{b^{(k)}(x), k \text{ is even}\} = \text{span}\{\partial_{\bar{X}}^k \psi_{\text{PSF}}(x - \bar{X}), k \text{ is even}\}$, $\text{span}\{b^{(k)}(x), k \text{ is odd}\} = \text{span}\{\partial_{\bar{X}}^k \psi_{\text{PSF}}(x - \bar{X}), k \text{ is odd}\}$ and they are orthogonal subspaces.

Then $\mathcal{F}_{2\ell 2\ell}$ is maximized when $b_{\bar{X}}^{(\ell)\dagger}|0\rangle$ is an eigenstate of $E(n)$ with an eigenvalue equal to one. The resultant FI is

$$\max_{\{E(n)\}} \mathcal{F}_{2\ell 2\ell} \approx \epsilon q_\ell^2 (2\ell)^2 (M_{2\ell})^{2\ell-2} = O(s^{2\ell-2}). \quad (61)$$

For example when $\ell = 1$, $b^{(1)}(x) = \frac{1}{\Delta k} \partial_{\bar{X}} \psi_{\text{PSF}}(x - \bar{X})$ and Eq. (61) gives Eq. (28).

We can show that it is possible for the FI to attain the lowest power of s (the highest precision) for even moments. To be specific, if $k = 2\ell$ is even, by projecting quantum states on the image plane onto basis $\{b_{\bar{X}}^{(\ell)\dagger}|0\rangle, \ell \geq 0\}$ ($b_{\bar{X}}^{(\ell)\dagger} = \int dx b^{(\ell)}(x) a_x^\dagger$), \mathcal{F}_{kk} is maximized and proportional to the $(2\ell - 2)$ -th power of s , as indicated in Eq. (17). Moreover, according to the Cramér-Rao bound (Eq. (10)),

$$\Sigma_{2\ell 2\ell} \geq (\mathcal{F}^{-1})_{2\ell 2\ell} \geq (\mathcal{F}_{2\ell 2\ell})^{-1}. \quad (62)$$

The estimation precision of $M_{2\ell}$ is lower bounded by the value of $(\mathcal{F}_{2\ell 2\ell})^{-1}$. Meanwhile, the choice of measurement basis $\{b_{\bar{X}}^{(\ell)\dagger}|0\rangle, \ell \geq 0\}$ not only minimizes the value of $(\mathcal{F}_{2\ell 2\ell})^{-1}$ but also makes \mathcal{F} diagonal, which means that the second equality in Eq. (62) holds true. Therefore, we conclude that $\{b_{\bar{X}}^{(\ell)\dagger}|0\rangle, \ell \geq 0\}$ is an optimal basis for estimation of even moments for weak incoherent sources. Note that $\{b_{\bar{X}}^{(\ell)\dagger}|0\rangle, \ell \geq 0\}$ may not be a complete basis, but any POVM is optimal as long as it contains projections onto them and other terms $E(n)$ contained in $\{E(n)\}$ is irrelevant because they do not affect the FIM in the lowest order approximation. We do not write out the irrelevant part of POVM in our discussion.

For odd moments, however, the above arguments do not apply. If we require $n \in N_\ell^w$ to satisfy

$$Q_{2\ell}(n) = O(\epsilon^2), \quad (63)$$

then $E(n)$ is not supported by $\psi_{\bar{X}}^{(k)\dagger}|0\rangle$ for all $k \leq \ell$.

Consequently, we have

$$Q_{2\ell+1}(n) = \frac{2\epsilon (M_{2\ell+1})^{2\ell+1}}{\ell!(\ell+1)!} \text{Re}[\langle 0|\psi_{\bar{X}}^{(\ell)} E(n) \psi_{\bar{X}}^{(\ell+1)\dagger}|0\rangle] + O(\epsilon^2) = O(\epsilon^2), \quad (64)$$

which implies negligible contribution from weak sources.

Therefore, in order to take odd moments into account, we need to relax Eq. (63) by choosing $n \in N_{\ell-1}^w \setminus N_{\ell+1}^w$ to keep the $O(\epsilon)$ term in $Q_{2\ell+1}(n)$. The coefficient of $(M_{2\ell+1})^{2\ell+1}$ can be non-zero when $E(n)$ is supported by both $\psi_{\bar{X}}^{(\ell+1)\dagger}|0\rangle$ and $\psi_{\bar{X}}^{(\ell)\dagger}|0\rangle$. Meanwhile,

$$Q_{2\ell}(n) = \frac{\epsilon}{\ell!^2} \langle 0|\psi_{\bar{X}}^{(\ell)} E(n) \psi_{\bar{X}}^{(\ell)\dagger}|0\rangle (M_{2\ell})^{2\ell} + O(\epsilon^2) \quad (65)$$

would be non-zero at $O(\epsilon)$ too. In the subdiffraction limit ($s \rightarrow 0$), the denominator in Eq. (17) is dominated by $Q_{2\ell}$ when $n \in N_{\ell-1}^w \setminus N_{\ell+1}^w$. As shown in App. (E), we can maximize $\mathcal{F}_{2\ell+1 2\ell+1}$ and in the meantime make the estimation of odd moments independent from the estimation of even moments (by letting $\mathcal{F}_{2\ell+1 2\ell} = \mathcal{F}_{2\ell 2\ell+1} = O(s^{2\ell})$). Then analogous to Eq. (62), $\mathcal{F}_{2\ell+1 2\ell+1}$ fully characterizes the estimation precision of $M_{2\ell+1}$. It is maximized when $E(n)$ are projections onto $\{\frac{b_{\bar{X}}^{(\ell)\dagger} \pm b_{\bar{X}}^{(\ell+1)\dagger}}{\sqrt{2}}|0\rangle\}$. Up to the lowest order of s and ϵ ,

$$\max_{\{E(n)\}} \mathcal{F}_{2\ell+1 2\ell+1} \approx 4\epsilon q_{\ell+1}^2 (2\ell+1)^2 \frac{(M_{2\ell+1})^{4\ell}}{(M_{2\ell})^{2\ell}} = O(s^{2\ell}). \quad (66)$$

In the meantime, we can also calculate $\mathcal{F}_{2\ell 2\ell}$ which is exactly its optimal value as in Eq. (61). Therefore, $\{\frac{b_{\bar{X}}^{(\ell)\dagger} \pm b_{\bar{X}}^{(\ell+1)\dagger}}{\sqrt{2}}|0\rangle\}$ achieves the optimal precision for both $M_{2\ell}$ and $M_{2\ell+1}$ simultaneously.

To conclude, we can use the following two subsets of measurement basis: $B_1^w = \{\frac{b_{\bar{X}}^{(\ell)\dagger} \pm b_{\bar{X}}^{(\ell+1)\dagger}}{\sqrt{2}}|0\rangle, \ell \text{ is even}\}$ and $B_2^w = \{\frac{b_{\bar{X}}^{(\ell)\dagger} \pm b_{\bar{X}}^{(\ell+1)\dagger}}{\sqrt{2}}|0\rangle, \ell \text{ is odd}\}$ (divided into two subsets so that they don't overlap) to estimate $\{M_k|k = 4k' \text{ or } 4k'+1, k \geq 1\}$ and $\{M_k|k = 4k'+2 \text{ or } 4k'+3, k' \geq 0\}$, respectively. Each moment can be measured with the optimal precision and independently from other moments (the FIM is diagonal). However, each one of $B_{1,2}^w$ can only extract half of the whole moment information: B_1^w estimates moments with orders equal to multiples of 4 plus 0 or 1; B_2^w estimates moments with orders equal to multiples of 4 plus 2 or 3. If one only needs to estimate even moments, $B_0^w = \{b_{\bar{X}}^{(\ell)\dagger}|0\rangle, \ell \geq 0\}$ is optimal.

Now let's consider the two-dimensional case. Similar to the one-dimensional case, we define

$$N_K^w = \{\langle 0|E(n)|0\rangle = \langle 0|\psi_{\bar{X}}^{(k\ell)} E(n) \psi_{\bar{X}}^{(k\ell)\dagger}|0\rangle = 0, \forall k, \ell, \text{ s.t. } 0 \leq k + \ell \leq K\}. \quad (67)$$

Type	Measurement basis	L	Moments estimated
B_0^w	$b_{\bar{X}\bar{Y}}^{(L\ K-L)^\dagger} 0\rangle$	N	$M_{2L,2K-2L}$
B_1^w	$\frac{1}{\sqrt{2}}(b_{\bar{X}\bar{Y}}^{(L\ K-L)^\dagger} \pm b_{\bar{X}\bar{Y}}^{(L+1\ K-L-1)^\dagger}) 0\rangle$	even	$M_{2L+1,2K-2L-1},$
B_2^w	$\frac{1}{\sqrt{2}}(b_{\bar{X}\bar{Y}}^{(L\ K-L)^\dagger} \pm b_{\bar{X}\bar{Y}}^{(L+1\ K-L-1)^\dagger}) 0\rangle$	odd	$(q_{L,K-L}^2(M_{2L,2K-2L})^{2K} + q_{L+1,K-L-1}^2(M_{2L+2,2K-2L-2})^{2K})^{\frac{1}{2K}}$
B_3^w	$\frac{1}{\sqrt{2}}(b_{\bar{X}\bar{Y}}^{(L\ K-L)^\dagger} \pm b_{\bar{X}\bar{Y}}^{(L+1\ K-L)^\dagger}) 0\rangle$	even	$M_{2L+1,2K-2L},$
B_4^w	$\frac{1}{\sqrt{2}}(b_{\bar{X}\bar{Y}}^{(L\ K-L)^\dagger} \pm b_{\bar{X}\bar{Y}}^{(L+1\ K-L)^\dagger}) 0\rangle$	odd	$M_{2L,2K-2L}$
B_5^w	$\frac{1}{\sqrt{2}}(b_{\bar{X}\bar{Y}}^{(K-L\ L)^\dagger} \pm b_{\bar{X}\bar{Y}}^{(K-L\ L+1)^\dagger}) 0\rangle$	even	$M_{2K-2L,2L+1},$
B_6^w	$\frac{1}{\sqrt{2}}(b_{\bar{X}\bar{Y}}^{(K-L\ L)^\dagger} \pm b_{\bar{X}\bar{Y}}^{(K-L\ L+1)^\dagger}) 0\rangle$	odd	$M_{2K-2L,2L}$

TABLE I. Measurement basis and corresponding moments in 2D imaging, details in App. (F).

Suppose $n \in N_{K-1}^w$, the $O(s^{2K})$ term in $P(n)$ would be

$$Q_{2K}(n) = \sum_{\ell, \ell'=0}^K \frac{\epsilon}{\ell! \ell'! (K-\ell)! (K-\ell')!} \langle 0 | \psi_{\bar{X}\bar{Y}}^{(\ell\ K-\ell)} E(n) \psi_{\bar{X}\bar{Y}}^{(\ell'\ K-\ell')\dagger} | 0 \rangle (M_{(\ell+\ell')(2K-\ell-\ell')})^{2K} + O(\epsilon^2). \quad (68)$$

Q_{2K} is derived from Taylor expansion of Eq. (12). We notice that Q_{2K} can be written as $\mathbb{E}[\langle \Psi_K | E(n) | \Psi_K \rangle]$ for some unnormalized state $|\Psi_K\rangle$. Hence Q_{2K} is always non-negative and is equal to zero (up to the first order of ϵ) if and only if $n \in N_K^w$. Based on the method of induction, we conclude that $Q_{2K} = O(\epsilon^2)$ if and only if $n \in N_K^w$. Therefore, by choosing proper measurement basis for $n \in N_{K-1}^w \setminus N_K^w$, one can estimate $M_{L\ 2K-L}$ with an FI up to $O(s^{2K-2})$ for all $0 \leq L \leq 2K$. In general, the optimal measurement basis depends on the value of each moments.

For $M_{L\ 2K+1-L}$, consider the $O(s^{2K+1})$ term in $P(n)$:

$$Q_{2K+1}(n) = \sum_{\ell, \ell'=0}^K \frac{2\epsilon}{\ell! \ell'! (K-\ell)! (K+1-\ell')!} \text{Re}[\langle 0 | \psi_{\bar{X}\bar{Y}}^{(\ell\ K-\ell)} E(n) \psi_{\bar{X}\bar{Y}}^{(\ell'\ K+1-\ell')\dagger} | 0 \rangle] (M_{(\ell+\ell')(2K+1-\ell-\ell')})^{2K+1} + O(\epsilon^2). \quad (69)$$

Clearly, if $n \in N_K^w$, $Q_{2K+1}(n) = 0$. Therefore we should focus on measurement $E(n)$ such that $n \in N_{K-1}^w \setminus N_{K+1}^w$. Similar to 1D imaging, the optimal scaling we can obtain for $M_{L\ 2K+1-L}$ is $O(s^{2K})$.

Again we assume derivatives of the PSF $\{\partial_{\bar{X}}^k \partial_{\bar{Y}}^\ell \psi_{\text{PSF}}(x - \bar{X}, y - \bar{Y}), k, \ell \geq 0\}$ form a linear independent subset in $L^2(\mathbb{C})$. An orthonormal set $\{b^{(k\ell)}(x), k \geq 0\}$ can be generated such that $b^{(k\ell)}(x)$ is orthogonal to every $\partial_{\bar{X}}^{k'} \partial_{\bar{Y}}^{\ell'} \psi_{\text{PSF}}(x - \bar{X}, y - \bar{Y})$ with $k + \ell \leq k' + \ell'$, $(k, \ell) \neq (k', \ell')$,

$$q_{k\ell} := \frac{1}{k! \ell!} \int b^{(k\ell)*}(x) \partial_{\bar{X}}^k \partial_{\bar{Y}}^\ell \psi_{\text{PSF}}(x - \bar{X}, y - \bar{Y}) dx dy \quad (70)$$

and $q_{k\ell} \in \mathbb{R}$. Suppose $\psi_{\text{PSF}}(x, y)$ is separable and $\psi_{\text{PSF}}(x, y) = \psi_{1, \text{PSF}}(x) \psi_{2, \text{PSF}}(y)$. One can generate two orthonormal sets $\{b_{1(2)}^{(k)}(x), \forall k \geq 0\}$ via Gram-Schmidt process from the derivatives of $\psi_{1(2), \text{PSF}}(x)$ as in 1D imaging. Then we have $b^{(k\ell)}(x) = b_1^{(k)}(x) b_2^{(\ell)}(x)$.

Similar to 1D imaging, one can project ρ onto these basis to extract information of moments (see Table I) and achieve the optimal scaling of s (but not necessarily the optimal coefficients). As before, one type of measurement can only estimate part of all the moments (1/4 to be specific) and by combining different types of measurements one can grasp information of all moments. In practice, combining $\{B_i^w\}_{i=1}^6$ will be enough to extract all the information of moments from ρ . For further justifications and calculations of FIs see App. (F).

In the case of sources with arbitrary strengths, we show in App. (G) that the same scaling wrt s is still achievable by replacing every $E(n)$ with $\sum_{k=0} \frac{1}{k!} (\psi_{\bar{X}}^\dagger)^k E(n) (\psi_{\bar{X}})^k$ (or $\sum_{k=0} \frac{1}{k!} (\psi_{\bar{X}\bar{Y}}^\dagger)^k E(n) (\psi_{\bar{X}\bar{Y}})^k$ for 2D imaging) which also give the same FIs as in the weak source scenario. However, the coefficient may be further improved using other basis, due to the fact that information of high order moments can be obtained by detecting several low order derivative operators simultaneously, which is neglectable when the source is weak. In contrast to estimation of the second moment, when estimating higher order moments, the optimal precision increases superlinearly (instead of linearly) as the source strength grows in the subdiffraction limit.

VIII. CONCLUSION AND OUTLOOK

We have performed a comprehensive Fisher information analysis on general imaging scenarios in the subdiffraction limit, where the improvement of image resolution is considered difficult due to the positive width of point spread functions. We conclude that, for any incoherence sources, a 1D or 2D image can be precisely estimated up to its second moment and the higher order moments are difficult to estimate in the sense that the

error increase inverse-polynomially as the size of image decreases. The imaging situation considered in the paper is quite general where both the number of point sources and source strengths can be arbitrary. The problem of pre-estimation of centroid is also worked out.

For real point spread functions, we put forward a measurement scheme which provides the optimal Fisher information in the subdiffraction limit. The measurement basis is constructed based on the derivatives of the point spread function, which are closely related to moments of an image. The optimal measurement scheme for second moment is discussed in detail. For higher order moments, compared with direct imaging approach, our measurement scheme guarantees at least a quadratic improvement of Fisher information in terms of the scaling wrt the size of the image. The coefficient of Fisher information is also optimal for weak sources, but can be further improved for strong sources. It is not clear, though, which measurement basis is optimal in terms of the exact value of Fisher information for strong sources.

The generality of our results has a cost though — the Fisher information is only calculated in the limiting case where the size of the image tends to zero. Direct calculations for a positive size can be difficult and it remains unsolved how to identify the optimal measurement scheme

when the size is not too small (in the subdiffraction limit) and also not too large (the point spread function can be viewed as a delta function). Our results, however, is an important theoretical result towards the ultimate resolution limit for incoherent optical imaging.

Note added.—Recently, Ref. [37] appeared, which directly calculates an upper bound of the quantum Fisher information wrt moments for subdiffraction incoherent optical imaging. This approach can be applied to all types of measurements, without the non-adaptivity restriction in our analysis. Our results on arbitrary source strength, generalization to two-dimensional imaging and optimal scaling achieving measurement, however, are not covered in Ref. [37].

ACKNOWLEDGMENTS

We acknowledge support from the ARL-CDQI (W911NF-15-2-0067), ARO (W911NF-14-1-0011, W911NF-14-1-0563), ARO MURI (W911NF-16-1-0349), AFOSR MURI (FA9550-14-1-0052, FA9550-15-1-0015), NSF (EFMA-1640959), the Alfred P. Sloan Foundation (BR2013-049), and the Packard Foundation (2013-39273).

-
- [1] L. Rayleigh, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **8**, 261 (1879).
 - [2] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light* (Elsevier, 2013).
 - [3] J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company Publishers, 2005).
 - [4] M. Tsang, R. Nair, and X.-M. Lu, Physical Review X **6**, 031033 (2016).
 - [5] S. Z. Ang, R. Nair, and M. Tsang, Physical Review A **95**, 063847 (2017).
 - [6] R. Nair and M. Tsang, Physical review letters **117**, 190801 (2016).
 - [7] C. Lupo and S. Pirandola, Physical review letters **117**, 190802 (2016).
 - [8] A. Chrostowski, R. Demkowicz-Dobrzanski, M. Jarzyna, and K. Banaszek, arXiv preprint arXiv:1709.08392 (2017).
 - [9] J. Rehacek, Z. Hradil, B. Stoklasa, M. Paur, J. Grover, A. Krzic, and L. Sanchez-Soto, arXiv preprint arXiv:1709.07705 (2017).
 - [10] X.-M. Lu, H. Krovi, R. Nair, S. Guha, and J. H. Shapiro, arXiv preprint arXiv:1802.02300 (2018).
 - [11] Z. Yu and S. Prasad, arXiv preprint arXiv:1805.09227 (2018).
 - [12] S. Prasad and Z. Yu, arXiv preprint arXiv:1807.09853 (2018).
 - [13] M. Tsang, Journal of Modern Optics **65**, 1385 (2018).
 - [14] J. Rehacek, Z. Hradil, B. Stoklasa, M. Paúr, J. Grover, A. Krzic, and L. Sánchez-Soto, Physical Review A **96**, 062107 (2017).
 - [15] J. Rehacek, Z. Hradil, D. Koutný, J. Grover, A. Krzic, and L. Sánchez-Soto, Physical Review A **98**, 012103 (2018).
 - [16] M. P. Backlund, Y. Shechtman, and R. L. Walsworth, arXiv preprint arXiv:1803.01776 (2018).
 - [17] J. M. Donohue, V. Ansari, J. Rehacek, Z. Hradil, B. Stoklasa, M. Paúr, L. L. Sánchez-Soto, and C. Silberhorn, arXiv preprint arXiv:1805.02491 (2018).
 - [18] C. Napoli, T. Tufarelli, S. Piano, R. K. Leach, and G. Adesso, arXiv preprint arXiv:1805.04116 (2018).
 - [19] S. Ram, E. S. Ward, and R. J. Ober, Proceedings of the National Academy of Sciences **103**, 4457 (2006).
 - [20] J. Chao, E. S. Ward, and R. J. Ober, JOSA A **33**, B36 (2016).
 - [21] A. von Diezmann, Y. Shechtman, and W. Moerner, Chemical reviews **117**, 7244 (2017).
 - [22] H. Kobayashi, B. L. Mark, and W. Turin, *Probability, random processes, and statistical analysis: applications to communications, signal processing, queueing theory and mathematical finance* (Cambridge University Press, 2011).
 - [23] C. W. Helstrom, *Quantum detection and estimation theory* (Academic press, 1976).
 - [24] S. L. Braunstein and C. M. Caves, Physical Review Letters **72**, 3439 (1994).
 - [25] R. Nair and M. Tsang, Optics express **24**, 3684 (2016).
 - [26] F. Yang, R. Nair, M. Tsang, C. Simon, and A. I. Lvovsky, arXiv preprint arXiv:1706.08633 (2017).
 - [27] J. Rehacek, M. Paur, B. Stoklasa, L. Motka, Z. Hradil, and L. Sanchez-Soto, arXiv preprint arXiv:1607.05837 (2016).

- [28] M. Tsang, New Journal of Physics **19**, 023054 (2017).
- [29] M. Parniak, S. Borówka, K. Boroszko, W. Wasilewski, K. Banaszek, and R. Demkowicz-Dobrzański, arXiv preprint arXiv:1803.07096 (2018).
- [30] Z. S. Tang, K. Durak, and A. Ling, Optics express **24**, 22004 (2016).
- [31] F. Yang, A. Tashchilina, E. S. Moiseev, C. Simon, and A. I. Lvovsky, Optica **3**, 1148 (2016).
- [32] W.-K. Tham, H. Ferretti, and A. M. Steinberg, Physical Review Letters **118**, 070801 (2017).
- [33] M. Paúr, B. Stoklasa, Z. Hradil, L. L. Sánchez-Soto, and J. Rehacek, Optica **3**, 1144 (2016).
- [34] J. Flusser, T. Suk, and B. Zitová, *2D and 3D Image Analysis by Moments* (John Wiley & Sons, 2016).
- [35] D. F. Walls and G. J. Milburn, *Quantum optics* (Springer Science & Business Media, 2007).
- [36] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge university press, 2010).
- [37] M. Tsang, arXiv preprint arXiv:1806.02781 (2018).
- [38] L. Isserlis, Biometrika **12**, 134 (1918).
- [39] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).
- [40] *Radius of convergence*, URL <http://mathworld.wolfram.com/RadiusofConvergence.html>.

Appendix A: Summary

We summarize the contents of each appendix here:

- **App. (B)** discusses the condition under which the series expansion of probabilities and FIs. For a well-behaved point spread function, the series expansion of probability converges uniformly and therefore the FIs can also be expanded wrt different orders of the size of the image. We also point out that our analysis can only be applied to non-adaptive measurements in order for the series expansion to be valid.
- **App. (C)** provides the first three terms in the series expansion of measurement probability for arbitrary incoherent sources, which is not explicitly given in **Sec. IV**.
- **App. (D)** provides an alternative way to parametrize second moments in 2D imaging, as opposed to the one in **Sec. VI**.
- **App. (E)**, **App. (F)** and **App. (G)** complement discussions in **Sec. VII** in terms of optimizing FI wrt odd moments for weak incoherent sources in 1D imaging, 2D imaging, generalization to arbitrary strengths.
- **App. (H)** discusses the pre-estimation of the centroid. We provide a measurement scheme which is optimal for weak sources and at least 93.0% efficient for strong sources.

The main results in this paper are also summarized in **Table II** and **Table III** for reference.

Appendix B: Validity of series expansions of probability and FIM

In this section, we justify the series expansion of the probability $P(n)$ around its centroid. For simplicity, we only consider weak sources in 1D imaging. For single-photon measurement,

$$P(n) = \epsilon \sum_{j=1}^J \gamma_j \langle 0 | \psi_j E(n) \psi_j^\dagger | 0 \rangle. \quad (\text{B1})$$

We want to know when the following series will converge uniformly to $P(n)$:

$$\sum_{k=0}^{\infty} P_k(n) (M_k)^k \stackrel{?}{=} P(n), \quad (\text{B2})$$

where

$$P_k(n) = \frac{\epsilon}{k!} \frac{\partial^k}{\partial x_j^k} \langle 0 | \psi_j E(n) \psi_j^\dagger | 0 \rangle \Big|_{x_j = \bar{x}}. \quad (\text{B3})$$

Let the radius of convergence $R = (\limsup_{k \rightarrow \infty} |P_k(n)|^{1/k})^{-1}$, then **Eq. (B2)** converges uniformly as long as $s < R$ [40].

	Weak source ($\epsilon \ll 1$)	Strong source (arbitrary ϵ)
Moments	$M_k = (\sum_j \gamma_j (x_j - \bar{X})^k)^{1/k}$ Eq. (15)	
Probability for outcome n	$P(n) = \mathbb{E}[\langle \psi_\alpha E(n) \psi_\alpha \rangle]$ Eq. (12)	
	$P(n) = (1 - \epsilon) \langle 0 E(n) 0 \rangle + \epsilon p(n) + O(\epsilon^2)$ Eq. (13)	$P(n) = \sum_{k=0}^{\infty} Q_k(n)$ Eq. (22)
	$p(n) = \sum_{k=0}^{\infty} \frac{p_k(n)}{k!} (M_k)^k$ Eq. (14)	App. (C)
FI	$\mathcal{F}_{k\ell} = \sum_n \frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial M_k} \right)^2 = O(s^{k-2})$ Eq. (17)	
Maximum FI	$\max_{\{E(n)\}} \mathcal{F}_{kk} = \begin{cases} O(s^{k-2}) & k \text{ is even,} \\ O(s^{k-1}) & k \text{ is odd.} \end{cases}$	
	$\max_{\{E(n)\}} \lim_{s \rightarrow 0} \mathcal{F}_{2\ell 2\ell} = \epsilon q_\ell^2 (2\ell)^2 (M_{2\ell})^{2\ell-2}$ Eq. (61)	$\max_{\{E(n)\}} \lim_{s \rightarrow 0} \mathcal{F}_{22} = 4\epsilon \Delta k^2$
	$\max_{\{E(n)\}} \lim_{s \rightarrow 0} \mathcal{F}_{2\ell+1 2\ell+1} = 4\epsilon q_{\ell+1}^2 (2\ell+1)^2 \frac{(M_{2\ell+1})^{4\ell}}{(M_{2\ell})^{2\ell}}$ Eq. (66)	Eq. (33)
Optimal Measurement	$B_0^w, B_1^w \text{ and } B_2^w$ Sec. VII	For $M_2, E(N_0) = \sum_{k=0}^{\infty} \frac{(\psi_X^\dagger)^k \psi_X^{(1)\dagger} 0\rangle \langle 0 \psi_X^{(1)} (\psi_{\bar{X}})^k}{k! \langle 0 \psi_X^{(1)} \psi_X^{(1)\dagger} 0 \rangle}$ Eq. (34) & App. (G)

TABLE II. A summary of the main results (1D)

	Weak source ($\epsilon \ll 1$)	Strong source (arbitrary ϵ)
Moments	$M_{k\ell} = \left(\sum_j \gamma_j (x_j - \bar{X})^k (y_j - \bar{Y})^\ell\right)^{1/(k+\ell)}$ <p style="text-align: center;">Eq. (36)</p>	
Probability for outcome n	$P(n) = \mathbb{E}[\langle \psi_\alpha E(n) \psi_\alpha \rangle]$ <p style="text-align: center;">Eq. (12)</p>	
	$P(n) = (1 - \epsilon) \langle 0 E(n) 0 \rangle + \epsilon p(n) + O(\epsilon^2)$ <p style="text-align: center;">Eq. (13)</p>	$P(n) = \sum_{K=0}^\infty Q_K(n)$
	$p(n) = \sum_{k\ell=0}^\infty \frac{p_{k\ell}(n)}{k!\ell!} (M_{k\ell})^{k+\ell}$ $p_{k\ell}(n) = \partial_{\bar{X}}^k \partial_{\bar{Y}}^\ell \langle 0 \psi_{\bar{X}\bar{Y}} E(n) \psi_{\bar{X}\bar{Y}}^\dagger 0 \rangle$	
FI	$\mathcal{F}_{k\ell k'\ell'} = \sum_n \frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial M_{k\ell}}\right)^2 = O(s^{k+\ell-2}) \quad , \text{Eq. (40)}$	
Maximum FI	$\max_{\{E(n)\}} \mathcal{F}_{L(K-L)L(K-L)} = \begin{cases} O(s^{K-2}) & k \text{ is even,} \\ O(s^{K-1}) & k \text{ is odd.} \end{cases}$ <p style="text-align: center;">Sec. VII & App. (F)</p>	
Optimal Measurement	$B_{0,1,2,3,4,5,6}^w$ <p style="text-align: center;">Table I</p>	For M_{20}, M_{11} and M_{02} , see Sec. VI. <p style="text-align: center;">App. (G)</p>

TABLE III. A summary of the main results (2D)

Next we show that the radius of convergence $R \geq R_0$ where R_0 independent of $E(n)$.

$$R_0 = \left(\sup_\ell \left(\frac{\|\psi_{\text{PSF}}^{(\ell)}\|}{\ell!} \right)^{1/\ell} \right)^{-1}, \quad (\text{B4})$$

where $\psi_{\text{PSF}}^{(\ell)}$ represents the ℓ -th order derivative of f and $\|\psi_{\text{PSF}}^{(\ell)}\| = \sqrt{\int_{-\infty}^{\infty} |\psi_{\text{PSF}}^{(\ell)}(x)|^2 dx}$. Then

$$R^{-1} = \limsup_{k \rightarrow \infty} |P_k(n)|^{1/k} \leq \limsup_{k \rightarrow \infty} \left| \sum_{\ell=0}^k \frac{1}{\ell!(k-\ell)!} \|\psi_{\text{PSF}}^{(k-\ell)}\| \|\psi_{\text{PSF}}^{(\ell)}\| \right|^{1/k} \leq R_0^{-1}. \quad (\text{B5})$$

Therefore when $s < R_0 \leq R$, Eq. (B2) uniformly converges. For example, for a Gaussian PSF

$$\psi_{\text{PSF}}(x) = \frac{1}{(2\pi\sigma^2)^{1/4}} \exp\left(-\frac{x^2}{4\sigma^2}\right). \quad (\text{B6})$$

From

$$\int_{-\infty}^{+\infty} e^{x^2} \left(\left(\frac{d}{dx} \right)^\ell e^{-x^2} \right)^2 dx = \sqrt{\pi} \ell! 2^\ell, \quad (\text{B7})$$

we see that $R_0 \geq \sigma$ from Eq. (B4). Therefore in the subdiffraction limit ($s \ll \sigma$), the series expansion is always valid. However, things may break down when $s > R_0$ which may happen if $\psi_{\text{PSF}}(x)$ has complex sub-wavelength structure.

When $s < R_0$, the diagonal element of the Fisher information matrix is

$$\begin{aligned} \mathcal{F}_{kk} &= \sum_n \frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial M_k} \right)^2 \\ &= \sum_n \frac{|P_k(n) k M_k^{k-1}|^2}{|P_k(n) M_k^k|} \frac{b_k^2}{a_k}, \end{aligned} \quad (\text{B8})$$

where we assume

$$\left| \frac{P(n)}{P_k(n) M_k^k} \right| = a_k, \quad \text{and} \quad \left| \frac{\partial P(n)/\partial M_k}{P_k(n) k M_k^{k-1}} \right| = b_k. \quad (\text{B9})$$

Suppose $\frac{b_k^2}{a_k} \leq c_k$, we have

$$\begin{aligned} \mathcal{F}_{kk} &< \sum_n \frac{(P_k(n) k |M_k|^{k-1})^2}{|P_k(n)| |M_k|^k} c_k = c_k k^2 |M_k|^{k-2} \sum_n |P_k(n)| \\ &= c_k k^2 |M_k|^{k-2} \left(\frac{\epsilon}{k!} \frac{\partial^k}{\partial x_j^k} \langle 0 | \psi_j (E(N_+) - E(N_-)) \psi_j^\dagger | 0 \rangle \Big|_{x_j = \bar{x}} \right) \\ &\leq 2c_k k^2 \left(\sum_{\ell=0}^k \frac{\epsilon}{\ell!(k-\ell)!} \|\psi_{\text{PSF}}^{(k-\ell)}\| \|\psi_{\text{PSF}}^{(\ell)}\| \right) |M_k|^{k-2} = O(s^{k-2}), \end{aligned} \quad (\text{B10})$$

where $N_+ = \{n : P_k(n) \geq 0\}$, $N_- = \{n : P_k(n) < 0\}$ and $E(N_\pm) = \sum_{n \in N_\pm} E(n)$.

The order-of-magnitude analysis above is valid only when

$$c_k = \left| \frac{P_k(n) (M_k)^k}{\sum_{k'=0}^{\infty} P_{k'}(n) (M_{k'})^{k'}} \right| \left(\sum_{k'=k}^{\infty} P_{k'}(n) \frac{\partial (M_{k'})^{k'}}{\partial M_k} \right)^2 \Big/ (P_k(n) k (M_k)^{k-1})^2 \quad (\text{B11})$$

is reasonably small when s is small. We argue that this is usually true for non-adaptive measurements:

- Consider first the case when $|P_k(n) (M_k)^k| \gg \left| \sum_{k' > k} P_{k'}(n) (M_{k'})^{k'} \right|$, then clearly

$$c_k \approx \frac{|P_k(n) (M_k)^k|}{\sum_{k'=0}^{\infty} P_{k'}(n) (M_{k'})^{k'}} \cdot 1 \lesssim 1. \quad (\text{B12})$$

- When $|P_k(n)(M_k)^k| \ll \left| \sum_{k' > k} P_{k'}(n)(M_{k'})^{k'} \right| = O(s^{k+1})$,

$$c_k \approx \frac{O(s^{k+1})}{|P_k(n)k(M_k)^{k-1}|} \quad (\text{B13})$$

may be large. However, the contribution to \mathcal{F}_{kk}

$$\frac{|P_k(n)kM_k^{k-1}|^2}{|P_k(n)M_k^k|} c_k = O(s^k) \quad (\text{B14})$$

is negligible.

- When $|P_k(n)(M_k)^k| \approx \left| \sum_{k' > k} P_{k'}(n)(M_{k'})^{k'} \right|$ and (when $P_{k'}(n) = 0$ for all $k' \leq k$) the first and second terms in

$$P(n) = P_k(n)(M_k)^k + \sum_{k' > k} P_{k'}(n)(M_{k'})^{k'} \quad (\text{B15})$$

cancel each other out, up to the lowest order of s . Above analysis could become invalid. However, it requires a special design of measurement based on prior knowledge of the moments. We exclude this type of adaptive measurement in our discussion.

Appendix C: First three terms in the series expansion of measurement probability for arbitrary incoherent sources

We aim to expand $P(n)$ in series of $O(s^k)$ where s is the size of the image. To do this we replace $\psi^\dagger \alpha$ with $\sum_{k=0}^{\infty} \frac{A^{(k)}}{k!} \psi_X^{(k)\dagger}$ in Eq. (12), where $A^{(k)} = \sum_{j=1}^J \alpha_j (x_j - \bar{X})^k$ and $\psi_X^{(k)\dagger} = \frac{d^k}{d\bar{X}^k} \int dx \psi_{\text{PSF}}(x - \bar{X}) a_x^\dagger$.

First of all, we calculate the value of denominator which gives

$$\langle 0 | e^{\alpha^\dagger \psi} e^{\psi^\dagger \alpha} | 0 \rangle = e^{\int dx |\sum_j \alpha_j \psi_{\text{PSF}}(x - x_j)|^2}. \quad (\text{C1})$$

Therefore,

$$P(n) = \mathbb{E}[e^{-\int dx |\sum_j \alpha_j \psi_{\text{PSF}}(x - x_j)|^2} \sum_{k=0}^{\infty} \frac{1}{k!^2} \langle 0 | (\alpha^\dagger \psi)^k E(n) (\psi^\dagger \alpha)^k | 0 \rangle]. \quad (\text{C2})$$

The zeroth order term is

$$\begin{aligned} Q_0(n) &= \sum_{k=0}^{\infty} \frac{1}{k!^2} \mathbb{E}[e^{-|A^{(0)}|^2} |A^{(0)}|^{2k}] \langle 0 | \psi_X^k E(n) (\psi_X^\dagger)^k | 0 \rangle \\ &= \sum_{k=0}^{\infty} \frac{\epsilon^k}{k!(1+\epsilon)^{k+1}} \langle 0 | \psi_X^k E(n) (\psi_X^\dagger)^k | 0 \rangle, \end{aligned} \quad (\text{C3})$$

where we use $\mathbb{E}[e^{-|A^{(0)}|^2} |A^{(0)}|^{2k}] = \frac{k! \epsilon^k}{(1+\epsilon)^{k+1}}$.

The first order term is

$$\begin{aligned} Q_1(n) &= \sum_{k=1}^{\infty} \frac{1}{k!^2} (2k) \mathbb{E}[(e^{-|A^{(0)}|^2}) (A^{(0)*})^{k-1} A^{(1)*} (A^{(0)})^k] \text{Re}[\langle 0 | (\psi_X)^{k-1} \psi_X^{(1)} E(n) (\psi_X^\dagger)^k | 0 \rangle] \\ &= \sum_{k=0}^{\infty} \frac{2\epsilon^{k+1}}{k!(1+\epsilon)^{k+2}} \text{Re}[\langle 0 | (\psi_X)^k \psi_X^{(1)} E(n) (\psi_X^\dagger)^{k+1} | 0 \rangle] M_1, \end{aligned} \quad (\text{C4})$$

where we use $\mathbb{E}[(e^{-|A^{(0)}|^2}) (A^{(0)*})^{k-1} A^{(1)*} (A^{(0)})^k] = \frac{k! \epsilon^k M_1}{(1+\epsilon)^{k+1}}$.

The second order term is

$$\begin{aligned}
Q_2(n) &= \sum_{k=0}^{\infty} \frac{1}{k!^2} \mathbb{E}[(e^{-|A^{(0)}|^2}) (\text{Re}[A^{(0)*} A^{(2)}] + |A^{(1)}|^2) |A^{(0)}|^{2k}] \langle 0 | \psi_{\bar{X}}^{(1)} \psi_{\bar{X}}^{(1)\dagger} | 0 \rangle \langle 0 | (\psi_{\bar{X}})^k E(n) (\psi_{\bar{X}}^\dagger)^k | 0 \rangle \\
&+ \sum_{k=1}^{\infty} \frac{1}{k!^2} (k \mathbb{E}[(e^{-|A^{(0)}|^2}) (A^{(0)*})^{k-1} A^{(2)*} (A^{(0)})^k] \text{Re}[\langle 0 | (\psi_{\bar{X}})^{k-1} \psi_{\bar{X}}^{(2)} E(n) (\psi_{\bar{X}}^\dagger)^k | 0 \rangle] \\
&\quad + k(k-1) \mathbb{E}[(e^{-|A^{(0)}|^2}) (A^{(0)*})^{k-2} (A^{(1)*})^2 (A^{(0)})^k] \text{Re}[\langle 0 | (\psi_{\bar{X}})^k E(n) (\psi_{\bar{X}}^{(1)\dagger})^2 (\psi_{\bar{X}}^\dagger)^{k-2} | 0 \rangle] \\
&\quad + k^2 \mathbb{E}[(e^{-|A^{(0)}|^2}) |A^{(0)}|^{2k-2} |A^{(1)}|^2] \langle 0 | (\psi_{\bar{X}})^{k-1} \psi_{\bar{X}}^{(1)} E(n) \psi_{\bar{X}}^{(1)\dagger} (\psi_{\bar{X}}^\dagger)^{k-1} | 0 \rangle) \\
&= \sum_{k=0}^{\infty} \frac{\epsilon^{k+1}}{k! (1+\epsilon)^{k+1}} ((\langle 0 | (\psi_{\bar{X}})^k \psi_{\bar{X}}^{(1)} E(n) \psi_{\bar{X}}^{(1)\dagger} (\psi_{\bar{X}}^\dagger)^k | 0 \rangle + \text{Re}[\langle 0 | (\psi_{\bar{X}})^k \psi_{\bar{X}}^{(2)} E(n) (\psi_{\bar{X}}^\dagger)^{k+1} | 0 \rangle]) \\
&\quad - (k+2) \langle 0 | \psi_{\bar{X}}^{(1)} \psi_{\bar{X}}^{(1)\dagger} | 0 \rangle \langle 0 | (\psi_{\bar{X}})^k E(n) (\psi_{\bar{X}}^\dagger)^k | 0 \rangle) M_2^2) \\
&+ \sum_{k=0}^{\infty} \frac{\epsilon^{k+2}}{k! (1+\epsilon)^{k+3}} \langle 0 | (\psi_{\bar{X}})^{k+2} E(n) (\psi_{\bar{X}}^{(1)\dagger})^2 (\psi_{\bar{X}}^\dagger)^k | 0 \rangle M_1^2 + \sum_{k=0}^{\infty} \frac{\epsilon^{k+1} (k-\epsilon)}{k! (1+\epsilon)^{k+2}} \langle 0 | (\psi_{\bar{X}})^k \psi_{\bar{X}}^{(1)} E(n) \psi_{\bar{X}}^{(1)\dagger} (\psi_{\bar{X}}^\dagger)^k | 0 \rangle M_1^2,
\end{aligned} \tag{C5}$$

where we use $\mathbb{E}[(e^{-|A^{(0)}|^2}) |A^{(0)}|^{2(k-1)} |A^{(1)}|^2] = \frac{(k-1)! \epsilon^k (M_2^2 - M_1^2)}{(1+\epsilon)^k} + \frac{k! \epsilon^k M_1^2}{(1+\epsilon)^{k+1}}$, $\mathbb{E}[(e^{-|A^{(0)}|^2}) |A^{(0)}|^{2(k-1)} A^{(2)*} A^{(0)}] = \frac{k! \epsilon^k M_2^2}{(1+\epsilon)^{k+1}}$ and $\mathbb{E}[(e^{-|A^{(0)}|^2}) (A^{(0)*})^{k-2} (A^{(1)*})^2 (A^{(0)})^k] = \frac{k! \epsilon^k M_2^2}{(1+\epsilon)^{k+1}}$. Suppose the centroid is accurately known, we have $M_1 = 0$ and $Q_1(n) = 0$. If we define $N_0 = \{n | Q_0(n) = \sum_{k=0}^{\infty} \frac{\epsilon^k}{k! (1+\epsilon)^{k+1}} \langle 0 | (\psi_{\bar{X}})^k E(n) (\psi_{\bar{X}}^\dagger)^k | 0 \rangle = 0\} = \{n | \langle 0 | (\psi_{\bar{X}})^k E(n) (\psi_{\bar{X}}^\dagger)^k | 0 \rangle = 0, \forall k\}$. For $n \in N_0$, $Q_0(n) = Q_1(n) = 0$ and only the first term in $Q_2(n)$ survives, which gives Eq. (32). The second term $\text{Re}[\langle 0 | (\psi_{\bar{X}})^k \psi_{\bar{X}}^{(2)} E(n) (\psi_{\bar{X}}^\dagger)^{k+1} | 0 \rangle]$ in Eq. (C5) vanishes for $n \in N_0$ because $E(n)$ is Hermitian and non-negative and its eigenstates corresponding to non-vanishing eigenvalues must be orthogonal to $(\psi_{\bar{X}}^\dagger)^k | 0 \rangle$ for all k .

Appendix D: An alternative way to parametrize second moments in 2D imaging

Here we calculation the optimal FIM wrt $(\Lambda_1, \Lambda_2, \theta)$ as defined in Sec. VI. We only consider the situation where $\Delta k_x = \Delta k_y = \Delta k$ and $r = 0$ as the form of FIM becomes quite complicated otherwise and provides no physical intuition. The QFIM wrt $(\Lambda_1, \Lambda_2, \theta)$ calculated from Eq. (48) is

$$\mathcal{J}[\Lambda_1, \Lambda_2, \theta] = \begin{pmatrix} 4\epsilon \Delta k^2 & 0 & 0 \\ 0 & 4\epsilon \Delta k^2 & 0 \\ 0 & 0 & \frac{4\epsilon \Delta k^2 (\Lambda_1^2 - \Lambda_2^2)^2}{\Lambda_1^2 + \Lambda_2^2} \end{pmatrix}. \tag{D1}$$

It is clear from Eq. (D1) that when $\Lambda_1 = \Lambda_2$, the QFI is zero, which means when the image is circular-uniformly distributed (up to its second moment), we are not able to estimate θ in the subdiffraction limit.

The corresponding optimal measurements found from the QFIM calculation are

$$\begin{aligned}
E(n_1) &= (\cos(\theta + \pi/4) |e_1\rangle - \sin(\theta + \pi/4) |e_2\rangle) (\cos(\theta + \pi/4) \langle e_1| - \sin(\theta + \pi/4) \langle e_2|), \\
E(n_2) &= (\sin(\theta + \pi/4) |e_1\rangle + \cos(\theta + \pi/4) |e_2\rangle) (\sin(\theta + \pi/4) \langle e_1| + \cos(\theta + \pi/4) \langle e_2|).
\end{aligned} \tag{D2}$$

for estimation of (Λ_1, Λ_2) and

$$\begin{aligned}
E(n_1) &= (\cos \theta |e_1\rangle - \sin \theta |e_2\rangle) (\cos \theta \langle e_1| - \sin \theta \langle e_2|), \\
E(n_2) &= (\sin \theta |e_1\rangle + \cos \theta |e_2\rangle) (\sin \theta \langle e_1| + \cos \theta \langle e_2|).
\end{aligned} \tag{D3}$$

for estimation of θ . We note here that Eq. (D2) and Eq. (D3) are mutually unbiased.

Appendix E: Optimization of FI wrt odd moments for weak incoherent sources in 1D imaging

Up to the lowest order of s and ϵ ,

$$\begin{aligned}\mathcal{F}_{2\ell+1\ 2\ell+1} &\approx \sum_{n \in N_{\ell-1}^w \setminus N_{\ell+1}^w} \frac{1}{Q_{2\ell}(n)} \left(\frac{\partial Q_{2\ell+1}(n)}{\partial M_{2\ell+1}} \right)^2 \\ &= \frac{4(2\ell+1)^2 \epsilon (M_{2\ell+1})^{4\ell}}{(\ell+1)!^2 (M_{2\ell})^{4\ell}} \sum_{n \in N_{\ell-1}^w \setminus N_{\ell+1}^w} \frac{(\text{Re}[\langle 0 | \psi_{\bar{X}}^{(\ell)} E(n) \psi_{\bar{X}}^{(\ell+1)\dagger} | 0 \rangle])^2}{\langle 0 | \psi_{\bar{X}}^{(\ell)} E(n) \psi_{\bar{X}}^{(\ell)\dagger} | 0 \rangle}\end{aligned}\quad (\text{E1})$$

First we note that, in order to maximize $\mathcal{F}_{2\ell+1\ 2\ell+1}$, we can assume $E(n)$ is a rank-one projector for each n , because for any $E(n) = \sum_i p_i |\Phi_i\rangle \langle \Phi_i|$,

$$\frac{(\sum_i p_i \text{Re}[\langle 0 | \psi_{\bar{X}}^{(\ell)} |\Phi_i\rangle \langle \Phi_i| \psi_{\bar{X}}^{(\ell+1)\dagger} | 0 \rangle])^2}{\sum_i p_i \langle 0 | \psi_{\bar{X}}^{(\ell)} |\Phi_i\rangle \langle \Phi_i| \psi_{\bar{X}}^{(\ell)\dagger} | 0 \rangle} \leq \sum_i p_i \frac{(\text{Re}[\langle 0 | \psi_{\bar{X}}^{(\ell)} |\Phi_i\rangle \langle \Phi_i| \psi_{\bar{X}}^{(\ell+1)\dagger} | 0 \rangle])^2}{\langle 0 | \psi_{\bar{X}}^{(\ell)} |\Phi_i\rangle \langle \Phi_i| \psi_{\bar{X}}^{(\ell)\dagger} | 0 \rangle} \quad (\text{E2})$$

according to Cauchy-Schwarz inequality. Therefore deviding any POVM into corresponding projective measurements will only increase FI. Furthermore, if $E(n) = |\Phi_n\rangle \langle \Phi_n|$,

$$\frac{(\text{Re}[\langle 0 | \psi_{\bar{X}}^{(\ell)} |\Phi_n\rangle \langle \Phi_n| \psi_{\bar{X}}^{(\ell+1)\dagger} | 0 \rangle])^2}{\langle 0 | \psi_{\bar{X}}^{(\ell)} |\Phi_n\rangle \langle \Phi_n| \psi_{\bar{X}}^{(\ell)\dagger} | 0 \rangle} \leq \frac{|\langle 0 | \psi_{\bar{X}}^{(\ell)} |\Phi_n\rangle \langle \Phi_n| \psi_{\bar{X}}^{(\ell+1)\dagger} | 0 \rangle|^2}{\langle 0 | \psi_{\bar{X}}^{(\ell)} |\Phi_n\rangle \langle \Phi_n| \psi_{\bar{X}}^{(\ell)\dagger} | 0 \rangle} = \langle 0 | \psi_{\bar{X}}^{(\ell+1)} |\Phi_n\rangle \langle \Phi_n| \psi_{\bar{X}}^{(\ell+1)\dagger} | 0 \rangle. \quad (\text{E3})$$

We can, for example, choose the measurement basis to be $|\Phi_{\pm}\rangle = \frac{b_{\bar{X}}^{(\ell)\dagger} \pm b_{\bar{X}}^{(\ell+1)\dagger}}{\sqrt{2}} |0\rangle$ (other real superposition of $b_{\bar{X}}^{(\ell)\dagger} |0\rangle$ and $b_{\bar{X}}^{(\ell+1)\dagger} |0\rangle$ also works) which achieves the optimal FI

$$\max_{\{E(n)\}} \mathcal{F}_{2\ell+1\ 2\ell+1} = 4(2\ell+1)^2 \epsilon q_{\ell+1}^2 \frac{(M_{2\ell+1})^{4\ell}}{(M_{2\ell})^{4\ell}}. \quad (\text{E4})$$

Here we use the property that $b^{(\ell)}(x)$ is orthogonal to $\partial_{\bar{X}}^{\ell+1} \psi_{\text{PSF}}(x - \bar{X})$ (based on [Eq. \(5\)](#)). Moreover, according to [Eq. \(10\)](#),

$$\Sigma_{2\ell+1\ 2\ell+1} \geq (\mathcal{F}^{-1})_{2\ell+1\ 2\ell+1} \geq (\mathcal{F}_{2\ell+1\ 2\ell+1})^{-1}. \quad (\text{E5})$$

The measurement basis $|\Phi_{\pm}\rangle$ also leads to $\mathcal{F}_{2\ell+1\ 2\ell} = \mathcal{F}_{2\ell\ 2\ell+1} = O(s^{2\ell})$ which means \mathcal{F} is effectively diagonal and the second equality in the above inequality holds, because up to the lowest order of s we have

$$Q_{2\ell}(n_+) = Q_{2\ell}(n_-), \quad (\text{E6})$$

$$Q_{2\ell+1}(n_+) = -Q_{2\ell+1}(n_-), \quad (\text{E7})$$

$$\mathcal{F}_{2\ell+1\ 2\ell} \approx \sum_{\substack{E(n)=|\Phi_+\rangle \langle \Phi_+| \\ E(n)=|\Phi_-\rangle \langle \Phi_-|}} \frac{1}{Q_{2\ell}(n)} \left(\frac{\partial Q_{2\ell+1}(n)}{\partial M_{2\ell+1}} \right) \left(\frac{\partial Q_{2\ell}(n)}{\partial M_{2\ell}} \right) = 0. \quad (\text{E8})$$

Appendix F: Measurement basis and corresponding FIs for weak incoherent sources in 2D imaging

According to [Eq. \(68\)](#), by choosing measurement basis

$$B_0^w = \{b_{\bar{X}\bar{Y}}^{(L\ K-L)\dagger} |0\rangle, \forall K \geq 0, 0 \leq L \leq K\} \quad (\text{F1})$$

where $b_{\bar{X}\bar{Y}}^{(k\ell)\dagger} = \int dx dy b_1^{(k)}(x - \bar{X}) b_2^{(\ell)}(y - \bar{Y}) a_{xy}^\dagger$, one can achieve the optimal scaling of s (but not necessarily the optimal coefficients) for FIs wrt $M_{2L\ 2K-2L}$ for all K and $L \leq K$:

$$\mathcal{F}_{2L\ 2K-2L, 2L\ 2K-2L} |_{B_0^w} \approx \epsilon q_{L, K-L}^2 (2K)^2 (M_{2L\ 2K-2L})^{2K-2} = O(s^{2K-2}). \quad (\text{F2})$$

By choosing measurement basis

$$B_1^w = \left\{ \frac{1}{\sqrt{2}} (b_{\bar{X}\bar{Y}}^{(L K-L)\dagger} \pm b_{\bar{X}\bar{Y}}^{(L+1 K-L-1)\dagger}) |0\rangle, \forall K \geq 0, 0 \leq L \leq K-1 \text{ is even} \right\} \quad (\text{F3})$$

(or $B_2^w = \left\{ \frac{1}{\sqrt{2}} (b_{\bar{X}\bar{Y}}^{(L K-L)\dagger} \pm b_{\bar{X}\bar{Y}}^{(L+1 K-L-1)\dagger}) |0\rangle, \forall K \geq 0, 0 \leq L \leq K-1 \text{ is odd} \right\}$), one can achieve the optimal scaling of s for FIs wrt $M_{2L+1 2K-(2L+1)}$ for all K is even (or odd) and $L < K$:

$$\begin{aligned} \mathcal{F}_{2L+1 2K-(2L+1), 2L+1 2K-(2L+1)}|_{B_{1,2}^w} &\approx \\ \frac{4\epsilon(q_{L,K-L}^2(M_{2L,2K-2L})^{2K} + q_{L+1,K-L-1}^2(M_{2L+2,2K-2L-2})^{2K})q_{L,K-L}^2q_{L+1,K-L-1}^2(2K)^2(M_{2L+1 2K-2L-1})^{4K-2}}{(q_{L,K-L}^2(M_{2L,2K-2L})^{2K} + q_{L+1,K-L-1}^2(M_{2L+2,2K-2L-2})^{2K})^2 - 4q_{L,K-L}^2q_{L+1,K-L-1}^2(M_{2L+1,2K-2L-1})^{4K}} \\ &= O(s^{2K-2}). \quad (\text{F4}) \end{aligned}$$

Meanwhile, $(q_{L,K-L}^2(M_{2L,2K-2L})^{2K} + q_{L+1,K-L-1}^2(M_{2L+2,2K-2L-2})^{2K})^{\frac{1}{2K}}$ as a parameter can be estimated simultaneously with precision $O(s^{-2K})$ and independently of $M_{2L+1 2K-(2L+1)}$. Here we have used the property that $b^{(k\ell)}(x)$ is orthogonal to $\partial_{\bar{X}}^{k'} \partial_{\bar{Y}}^{\ell'} \psi_{\text{PSF}}(x - \bar{X}, y - \bar{Y})$ as long as k and k' (or ℓ and ℓ') do not have the same parity (i.e. are not both even or odd). To conclude, $B_{1,2}^w$ cover the estimation of moments whose orders on x - and y -axis are both even or both odd. The optimal FI scaling is $O(s^{2K-2})$ in this case, where $2K$ is the sum of orders on x - and y -axis.

For moments who have different parities on x - and y -axis, we can use basis

$$B_3^w = \left\{ \frac{1}{\sqrt{2}} (b_{\bar{X}\bar{Y}}^{(L K-L)\dagger} \pm b_{\bar{X}\bar{Y}}^{(L+1 K-L)\dagger}) |0\rangle, \forall K \geq 0, 0 \leq L \leq K \text{ is even} \right\}; \quad (\text{F5})$$

$$B_4^w = \left\{ \frac{1}{\sqrt{2}} (b_{\bar{X}\bar{Y}}^{(L K-L)\dagger} \pm b_{\bar{X}\bar{Y}}^{(L+1 K-L)\dagger}) |0\rangle, \forall K \geq 0, 0 \leq L \leq K \text{ is odd} \right\}; \quad (\text{F6})$$

$$B_5^w = \left\{ \frac{1}{\sqrt{2}} (b_{\bar{X}\bar{Y}}^{(K-L L)\dagger} \pm b_{\bar{X}\bar{Y}}^{(K-L L+1)\dagger}) |0\rangle, \forall K \geq 0, 0 \leq L \leq K \text{ is even} \right\}; \quad (\text{F7})$$

$$B_6^w = \left\{ \frac{1}{\sqrt{2}} (b_{\bar{X}\bar{Y}}^{(K-L L)\dagger} \pm b_{\bar{X}\bar{Y}}^{(K-L L+1)\dagger}) |0\rangle, \forall K \geq 0, 0 \leq L \leq K \text{ is odd} \right\}. \quad (\text{F8})$$

Based on Eq. (69), we can calculate the following FIs (up to the lowest order of s):

$$\mathcal{F}_{2L+1 2K-2L, 2L+1 2K-2L}|_{B_{3,4}^w} \approx 4\epsilon q_{L+1,K-L}^2 (2K+1)^2 \frac{(M_{2L+1 2K-2L})^{4K}}{(M_{2L 2K-2L})^{2K}} = O(s^{2K}); \quad (\text{F9})$$

$$\mathcal{F}_{2K-2L 2L+1, 2K-2L 2L+1}|_{B_{5,6}^w} \approx 4\epsilon q_{K-L,L+1}^2 (2K+1)^2 \frac{(M_{2K-2L 2L+1})^{4K}}{(M_{2K-2L 2L})^{2K}} = O(s^{2K}). \quad (\text{F10})$$

and $M_{2L 2K-2L}$ ($M_{2K-2L 2L}$) can be estimated simultaneously and independently with $M_{2L+1 2K-2L}$ ($M_{2K-2L 2L+1}$):

$$\mathcal{F}_{2L 2K-2L, 2L 2K-2L}|_{B_{3,4}^w} \approx \epsilon q_{L,K-L}^2 (2K)^2 (M_{2L 2K-2L})^{2K-2} = O(s^{2K-2}); \quad (\text{F11})$$

$$\mathcal{F}_{2K-2L 2L, 2K-2L 2L}|_{B_{5,6}^w} \approx \epsilon q_{K-L,L}^2 (2K)^2 (M_{2K-2L 2L})^{2K-2} = O(s^{2K-2}). \quad (\text{F12})$$

which are exactly their optimal values as in Eq. (F2).

Appendix G: Estimation of higher order moments with arbitrary source strengths

Here we only consider 1D imaging, the discussion can be easily generalized to 2D imaging. As already shown in Sec. V. Only 0-null measurement $n \in N_0 = \{n | Q_0(n) = \langle 0 | (\psi_{\bar{X}})^k E(n) (\psi_{\bar{X}}^\dagger)^k | 0 \rangle = 0, \forall k\}$ contributes to the FI wrt M_2 . Using the method of induction, we define ℓ -null measurement

$$N_\ell = \{n | \langle \Phi | E(n) | \Phi \rangle = 0, \forall |\Phi\rangle \in B^{(k)}, k \leq \ell\}, \quad (\text{G1})$$

where $B^{(\ell)} = \{(\prod_k \psi_{\bar{X}}^{(\ell_k)\dagger}) | 0 \rangle, \forall \{\ell_k \geq 0, k \in \mathbb{N}\}, \text{s.t. } \sum_k \ell_k = \ell\}$. When $n \in N_{\ell-1}$, $M_{2\ell}$ first appears in $Q_{2\ell}(n, \{M_k, k \leq 2\ell\})$. Up to the lowest order of s ,

$$\mathcal{F}_{2\ell 2\ell} = \sum_n \frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial M_{2\ell}} \right)^2 \approx \sum_{n \in N_{\ell-1}} \frac{1}{Q_{2\ell}(n, \{M_k, k \leq 2\ell\})} \left(\frac{\partial Q_{2\ell}(n, \{M_k, k \leq 2\ell\})}{\partial M_{2\ell}} \right)^2, \quad (\text{G2})$$

where $Q_{2\ell}(n, \{M_k, k \leq 2\ell\})$ is the $O(s^{2\ell})$ order term of

$$P(n) = \mathbb{E}\left[\frac{\langle 0|e^{\alpha^\dagger \psi} E(n) e^{\psi^\dagger \alpha}|0\rangle}{\langle 0|e^{\alpha^\dagger \psi} e^{\psi^\dagger \alpha}|0\rangle}\right] = \mathbb{E}[e^{-\int dx |\sum_j \alpha_j \psi_{\text{PSF}}(x-x_j)|^2} \sum_{k=0}^{\infty} \frac{1}{k!^2} \langle 0|(\alpha^\dagger \psi)^k E(n) (\psi^\dagger \alpha)^k |0\rangle]. \quad (\text{G3})$$

When $n \in N_{\ell-1}$, $Q_{2\ell}(n)$ has the following form:

$$Q_{2\ell}(n) = \frac{1}{\ell!^2} \sum_{k=0}^{\infty} \frac{\epsilon^{k+1}}{k!(1+\epsilon)^{k+1}} \langle 0|(\psi_{\bar{X}})^k \psi_{\bar{X}}^{(\ell)} E(n) \psi_{\bar{X}}^{(\ell)\dagger} (\psi_{\bar{X}}^\dagger)^k |0\rangle (M_{2\ell})^{2\ell} + Q_{2\ell}^R(n), \quad (\text{G4})$$

where the remainder term $Q_{2\ell}^R(n)$ contains only moments with orders lower than 2ℓ . We note that $Q_{2\ell}(n, \{M_k, k \leq 2\ell\})$ contains only terms like (summing over $k \geq \max\{K_0^+, K_0^-\}$)

$$\frac{1}{(\ell_1^+! \dots \ell_{m_+}^+!)(\ell_1^-! \dots \ell_{m_-}^-!)} \frac{1}{(K_1^+! \dots K_{m_+}^+!(k-K_0^+)!)(K_1^-! \dots K_{m_-}^-!(k-K_0^-)!)}$$

$$\mathbb{E}[e^{-|A^{(0)}|^2} A^{(0)*k-K_0^+} A^{(\ell_1^+)*K_1^+} \dots A^{(\ell_{m_+}^+)*K_{m_+}^+} A^{(0)k-K_0^-} A^{(\ell_1^-)K_1^-} \dots A^{(\ell_{m_-}^-)K_{m_-}^-}]$$

$$\langle 0|(\psi_{\bar{X}})^{k-K_0^+} (\psi_{\bar{X}}^{\ell_1^+})^{K_1^+} \dots (\psi_{\bar{X}}^{\ell_{m_+}^+})^{K_{m_+}^+} E(n) (\psi_{\bar{X}}^{\ell_1^-})^{K_1^-} \dots (\psi_{\bar{X}}^{\ell_{m_-}^-})^{K_{m_-}^-} (\psi_{\bar{X}}^\dagger)^{k-K_0^-} |0\rangle, \quad (\text{G5})$$

where $K_{m'}^\pm, \ell_{m'}^\pm \in \mathbb{N}^+$, $m' = 1, \dots, m_\pm$, $K_0^\pm = \sum_{m'=1}^{m_\pm} K_{m'}^\pm$ and $\ell = \sum_{m'=1}^{m_\pm} \ell_{m'}^\pm$. From Wick's theorem and Eq. (20), it is clear that the only term dependent on $M_{2\ell}$ corresponds to $K_0^+ = K_0^- = 1$, $m_\pm = 1$ and $\ell_{m_\pm} = \ell$. When $m_\pm = 1$ and $\ell_{m_\pm} = \ell$, we have

$$\mathbb{E}[e^{-|A^{(0)}|^2} |A^{(0)}|^{2(k-1)} |A^{(\ell)}|^2] = \frac{(k-1)! \epsilon^k (M_{2\ell})^{2\ell}}{(1+\epsilon)^k} + \frac{(k-1)!(k-1-\epsilon) \epsilon^k M_\ell^{2\ell}}{(1+\epsilon)^{k+1}}, \quad (\text{G6})$$

proving Eq. (G4). Therefore, by choosing the modified measurement

$$B_0 = \left\{ \sum_{k=0}^{\infty} \frac{1}{k!} (\psi_{\bar{X}}^\dagger)^k b_{\bar{X}}^{(\ell)\dagger} |0\rangle \langle 0| b_{\bar{X}}^{(\ell)} (\psi_{\bar{X}})^k, \forall \ell \geq 0 \right\}, \quad (\text{G7})$$

$Q_{2\ell}^R(n) = 0$ and the same FI (Eq. (61)) wrt $M_{2\ell}$ is recovered using the modified measurement B_0 . We can also modify other basis analogously by allowing multi-photon detection of $\psi_{\bar{X}}^\dagger$ and it will provide the same expression of FIs as in the weak source scenario. Note that here each component of B_0 is not a POVM but a PVM because we don't need to distinguish the number of $\psi_{\bar{X}}^\dagger$ photon we detect. However if we choose to distinguish them, that is, using

$$B'_0 = \left\{ \frac{1}{k!} (\psi_{\bar{X}}^\dagger)^k b_{\bar{X}}^{(\ell)\dagger} |0\rangle \langle 0| b_{\bar{X}}^{(\ell)} (\psi_{\bar{X}})^k, \forall k, \ell \geq 0 \right\}, \quad (\text{G8})$$

the FI will be no smaller (easily proven using Cauchy-Schwarz inequality) and the FIM is still effectively diagonal.

However, even B'_0 is not optimal when estimating $M_{2\ell}$. Physically, the reason is that the information of high order moments can be obtained by detecting several low order derivative operators simultaneously, which is neglectable when the source is weak. For $\ell = 1$, the only lower order moment is $M_1 = 0$, therefore strong source strength does not make a difference when calculating the FI, as shown in Sec. V.

We provide a simple example showing B'_0 is not an optimal measurement basis by replacing it with a better basis. Consider $\ell = 4$ (and we want to estimate the value of $M_{2\ell} = M_8$). Suppose $s > 0$. For simplicity, we only consider the replacement in 2-photon subspace, i.e. we don't change any $k+1$ -photon basis in B'_0 with $k \neq 1$ and their contributions to \mathcal{F}_{88} will remain the same. For 2-photon subspace, we consider the possibility of choosing another basis in $B_{4,2^2} = \text{span}\{\psi_{\bar{X}}^\dagger b^{(4)\dagger} |0\rangle, \frac{1}{\sqrt{2}} b^{(2)\dagger} b^{(2)\dagger} |0\rangle\} \equiv \text{span}\{\mathbf{b}_4, \mathbf{b}_{2^2}\}$. After some calculations, we have $Q_8(n)$ in this 2-dimensional subspace

$$Q_8(n) = \text{tr} \left(E(n) \begin{pmatrix} A_{44} & A_{42^2} \\ A_{2^2 4} & A_{2^2 2^2} \end{pmatrix} \right), \quad (\text{G9})$$

where

$$A_{44} = q_4^2 \mathbb{E}[e^{-|A^{(0)}|^2} |A^{(4)}|^2 |A^{(0)}|^2] = q_4^2 \frac{\epsilon^2}{(1+\epsilon)^2} \left((M_8^8 - M_4^8) + \frac{2}{1+\epsilon} M_4^8 \right) > 0, \quad (\text{G10})$$

$$A_{42^2} = A_{2^2 4} = q_4 \left(\frac{q_2^2}{4} \right) \mathbb{E}[e^{-|A^{(0)}|^2} A^{(4)} A^{(0)} (A^{(2)*})^2] = q_4 \left(\frac{q_2^2}{4} \right) \frac{2\epsilon^2}{(1+\epsilon)^2} \left((M_6^6 M_2^2 - M_4^4 M_2^4) + \frac{1}{1+\epsilon} M_4^4 M_2^4 \right) > 0, \quad (\text{G11})$$

$$A_{2^2 2^2} = \left(\frac{q_2^2}{4} \right)^2 \mathbb{E}[e^{-|A^{(0)}|^2} |A^{(2)}|^4] = \left(\frac{q_2^2}{4} \right)^2 \left(\frac{2\epsilon^2}{1+\epsilon} M_4^8 - \frac{4\epsilon^3}{(1+\epsilon)^2} M_4^4 M_2^4 + \frac{2\epsilon^4}{(1+\epsilon)^3} M_2^8 \right) > 0. \quad (\text{G12})$$

We can easily find a non-trivial image such that $A_{44} A_{2^2 2^2} - A_{42^2}^2 > 0$, then we maximize \mathcal{F}_{88} in this 2-dimensional subspace by doing QFI calculation, which gives

$$\begin{aligned} \max_{E(n) \text{ in } B_{4,2^2}} \frac{1}{Q_8(n)} \left(\frac{\partial Q_8(n)}{\partial M_8} \right)^2 &= \left(q_4^2 8 M_8^7 \frac{\epsilon^2}{(1+\epsilon)^2} \right)^2 \frac{A_{44} A_{2^2 2^2} - A_{42^2}^2 + A_{2^2 2^2}^2}{(A_{44} + A_{2^2 2^2})(A_{44} A_{2^2 2^2} - A_{42^2}^2)} \\ &> \frac{1}{Q_8(n)} \left(\frac{\partial Q_8(n)}{\partial M_8} \right)^2 \Big|_{E(n)=\mathbf{b}_4 \mathbf{b}_4^\dagger} = \left(q_4^2 8 M_8^7 \frac{\epsilon^2}{(1+\epsilon)^2} \right)^2 \frac{1}{A_{44}}. \end{aligned} \quad (\text{G13})$$

Now we've proven $\mathbf{b}_4 \mathbf{b}_4^\dagger$ does not generate the maximum FI wrt M_8 and B'_0 is not optimal. Meanwhile, we also note that the FIM is effectively diagonal in the subdiffraction limit, thus \mathcal{F}_{88} fully characterizes the measurement precision of M_8 . In general, any non-zero off-diagonal term (A_{42^2} in this case) in the same photon number subspace would lead to the same result. It means the precision of high order moments estimation could be enhanced by utilizing the detection of several low order derivative operators simultaneously.

For odd moments $M_{2\ell+1}$ ($\ell \geq 1$), suppose $n \in N_{\ell-1} \setminus N_{\ell+1}$, we have

$$Q_{2\ell+1}(n) = \frac{1}{\ell!(\ell+1)!} \sum_{k=0}^{\infty} \frac{\epsilon^{k+1}}{k!(1+\epsilon)^{k+1}} \text{Re}[\langle 0 | (\psi_{\bar{X}})^k \psi_{\bar{X}}^{(\ell+1)} E(n) \psi_{\bar{X}}^{(\ell)\dagger} (\psi_{\bar{X}}^\dagger)^k | 0 \rangle] (M_{2\ell+1})^{2\ell+1} + Q_{2\ell+1}^R(n). \quad (\text{G14})$$

The modified measurement

$$B_{1(2)} = \left\{ \sum_{k=0}^{\infty} \frac{1}{k!} (\psi_{\bar{X}}^\dagger)^k \frac{b_{\bar{X}}^{(\ell)\dagger} \pm b_{\bar{X}}^{(\ell+1)\dagger}}{\sqrt{2}} | 0 \rangle \langle 0 | \frac{b_{\bar{X}}^{(\ell)} \pm b_{\bar{X}}^{(\ell+1)}}{\sqrt{2}} (\psi_{\bar{X}})^k, \forall \ell \text{ is odd (or even)} \right\}, \quad (\text{G15})$$

also leads to the same FI Eq. (66), as for the even moments.

Appendix H: Pre-estimation of the centroid

The procedure to estimate the centroid can be divided into two step: (1) Find a reference point X_R such that $|\bar{X} - X_R| \lesssim s$; (2) Precisely locate \bar{X} within the subdiffraction limit. Assuming the resource required for step (1) is neglectable (it's a coarse estimation), we only consider the resource required for step (2). Normally, to fully resolve an image, we need to achieve a degree of precision where $\delta M_k \ll s$ ($k \geq 2$) and here we analyze the resource required to achieve $\delta \bar{X} \ll s$ so that it won't induce a significant error in the estimation of higher other moments.

We first consider 1D weak source scenario. After step (1), we are already in the subdiffraction regime and we can expand $P(n)$ around X_R up to $O(s^2)$, which gives

$$\begin{aligned} P(n) \approx Q_0(n) + Q_1(n) + Q_2(n) &= \epsilon \langle 0 | \psi_{X_R} E(n) \psi_{X_R}^\dagger | 0 \rangle + 2\epsilon \text{Re}[\langle 0 | \psi_{X_R}^{(1)} E(n) \psi_{X_R}^\dagger | 0 \rangle] \tilde{M}_1 \\ &+ \epsilon (\langle 0 | \psi_{X_R}^{(1)} E(n) \psi_{X_R}^{(1)\dagger} | 0 \rangle + 2\text{Re}[\langle 0 | \psi_{X_R}^{(2)} E(n) \psi_{X_R}^\dagger | 0 \rangle]) \tilde{M}_2^2 + O(\epsilon^2). \end{aligned} \quad (\text{H1})$$

Here \tilde{M}_1 and \tilde{M}_2 is redefined using X_R as the centroid. According to App. (E), the optimal measurement in terms of estimating $\tilde{M}_1 = \bar{X} - X_R$ can be an arbitrary projection onto two orthonormal basis in the real span of $\{\psi_{X_R}^\dagger | 0 \rangle, \psi_{X_R}^{(1)\dagger} | 0 \rangle\}$ as long as $Q_0(n) \gg Q_1(n) \gg Q_2(n)$ is satisfied. For example,

$$E(n_\pm) = \left(\frac{\psi_{X_R}^\dagger \pm \frac{1}{\Delta k} \psi_{X_R}^{(1)\dagger}}{\sqrt{2}} \right) | 0 \rangle \langle 0 | \left(\frac{\psi_{X_R} \pm \frac{1}{\Delta k} \psi_{X_R}^{(1)}}{\sqrt{2}} \right) \quad (\text{H2})$$

is optimal. The corresponding FI is

$$\mathcal{F}_{11} = 4\epsilon\Delta k^2. \quad (\text{H3})$$

which is the same as Eq. (28). Therefore, if we want to estimate both the second moment M_2 and the centroid \bar{X} , a straightforward method is to first use half of the whole resource to locate \bar{X} such that $\delta\bar{X} \ll s$ and then use the rest half to estimate M_2 as described in Sec. V. The effective FIM would be half of the optimal ones Eq. (H3) and Eq. (28),

$$\mathcal{F}(\tilde{M}_1, M_2) = \begin{pmatrix} 2\epsilon\Delta k^2 & 0 \\ 0 & 2\epsilon\Delta k^2 \end{pmatrix}, \quad (\text{H4})$$

which is only half of the QFIM [4]

$$\mathcal{J}(\tilde{M}_1, M_2) = \begin{pmatrix} 4\epsilon\Delta k^2 & 0 \\ 0 & 4\epsilon\Delta k^2 \end{pmatrix}. \quad (\text{H5})$$

When we want to estimate even higher order moments, the resource required to locate \bar{X} is neglectable.

Now we show Eq. (H4) is the optimal precision we can get (in the subdiffraction limit) and the QFIM Eq. (H5) is not achievable. For any POVM $\{E(n)\}$, the only case when $P(n)$ does not lead to a zero-FI wrt M_2 is when there is an $E(n)$ such that

$$P(n) \approx A_0(n) + A_1(n)\tilde{M}_1 + A_2(n)(M_2^2 + \tilde{M}_1^2) = O(s^2) \quad (\text{H6})$$

where

$$A_0(n) = \epsilon \langle 0 | \psi_{X_R} E(n) \psi_{X_R}^\dagger | 0 \rangle = O(s^2), \quad (\text{H7})$$

$$A_1(n) = 2\epsilon \text{Re}[\langle 0 | \psi_{X_R}^{(1)} E(n) \psi_{X_R}^\dagger | 0 \rangle] = O(s), \quad (\text{H8})$$

$$A_2(n) = \epsilon(\langle 0 | \psi_{X_R}^{(1)} E(n) \psi_{X_R}^{(1)\dagger} | 0 \rangle + 2\text{Re}[\langle 0 | \psi_{X_R}^{(2)} E(n) \psi_{X_R}^\dagger | 0 \rangle]) = O(1), \quad (\text{H9})$$

and we use the relation $\tilde{M}_2^2 = M_2^2 + M_1^2$. Note that $A_2(n) \approx \epsilon \langle 0 | \psi_{X_R}^{(1)} E(n) \psi_{X_R}^{(1)\dagger} | 0 \rangle$, because $\text{Re}[\langle 0 | \psi_{X_R}^{(2)} E(n) \psi_{X_R}^\dagger | 0 \rangle] = O(s)$ can be neglected. Since

$$\frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial \tilde{M}_1} \right)^2 = \frac{(A_1(n) + 2A_2(n)\tilde{M}_1)^2}{A_0(n) + A_1(n)\tilde{M}_1 + A_2(n)(M_2^2 + \tilde{M}_1^2)}, \quad (\text{H10})$$

$$\frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial M_2} \right)^2 = \frac{4A_2(n)^2 M_2^2}{A_0(n) + A_1(n)\tilde{M}_1 + A_2(n)(M_2^2 + \tilde{M}_1^2)} \quad (\text{H11})$$

and $A_1(n)^2 \leq 4A_2(n)A_0(n)$, we have

$$\frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial M_2} \right)^2 + \frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial \tilde{M}_1} \right)^2 \lesssim 4A_2(n) \approx 4\epsilon \langle 0 | \psi_{X_R}^{(1)} E(n) \psi_{X_R}^{(1)\dagger} | 0 \rangle. \quad (\text{H12})$$

When $P(n)$ is dominated by $Q_0(n)$, we also have

$$\frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial M_2} \right)^2 + \frac{1}{P(n)} \left(\frac{\partial P(n)}{\partial \tilde{M}_1} \right)^2 \approx \frac{A_1(n)^2}{A_0(n)} \leq 4\epsilon \langle 0 | \psi_{X_R}^{(1)} E(n) \psi_{X_R}^{(1)\dagger} | 0 \rangle. \quad (\text{H13})$$

Therefore, any achievable FIM must satisfies

$$\mathcal{F}_{11} + \mathcal{F}_{22} \leq 4\epsilon \sum_n \langle 0 | \psi_{X_R}^{(1)} E(n) \psi_{X_R}^{(1)\dagger} | 0 \rangle = 4\epsilon\Delta k^2, \quad (\text{H14})$$

and

$$(\delta\tilde{M}_1)^2 + (\delta M_2)^2 = \text{tr}(\Sigma) \geq \text{tr}(\mathcal{F}^{-1}) \geq \sum_{i=1}^2 \mathcal{F}_{ii}^{-1} \geq \frac{4}{\text{tr}(\mathcal{F})} \geq \frac{1}{\epsilon\Delta k^2}. \quad (\text{H15})$$

Clearly the last three equalities are simultaneously satisfied when FIM is Eq. (H4), implying the optimality of our measurement scheme.

The situation becomes a bit more complicated for arbitrary source strengths. First, we expand $P(n)$ around X_R up to $O(s)$

$$P(n) \approx Q_0(n) + Q_1(n) = \sum_{k=0}^{\infty} \frac{\epsilon^k}{k!(1+\epsilon)^{k+1}} \langle 0 | \psi_{X_R}^k E(n) (\psi_{X_R}^\dagger)^k | 0 \rangle + \sum_{k=0}^{\infty} \frac{2\epsilon^{k+1}}{k!(1+\epsilon)^{k+2}} \text{Re}[\langle 0 | \psi_{X_R}^k \psi_{X_R}^{(1)} E(n) (\psi_{X_R}^\dagger)^{k+1} | 0 \rangle] \tilde{M}_1. \quad (\text{H16})$$

Since the quantum state is diagonal in photon number basis, the optimal measurement estimating \tilde{M}_1 must also be diagonal in photon number basis [24], that is, $\{E(n)\}$ should contains $\{E(n_{k,i}), k \geq 1\}$ where $E(n_k) = \Pi_{\mathbf{k}} E(n_{k,i}) \Pi_{\mathbf{k}}$ and $\Pi_{\mathbf{k}}$ is projection onto \mathbf{k} -photon subspace. In this case, we shall write

$$\mathcal{F}_{11} = \sum_{k=0}^{\infty} \sum_i \frac{\left(\frac{2\epsilon^{k+1}}{k!(1+\epsilon)^{k+2}} \text{Re}[\langle 0 | \psi_{X_R}^k \psi_{X_R}^{(1)} E(n_{k+1,i}) (\psi_{X_R}^\dagger)^{k+1} | 0 \rangle] \right)^2}{\left(\frac{\epsilon^{k+1}}{(k+1)!(1+\epsilon)^{k+2}} \langle 0 | \psi_{X_R}^{k+1} E(n_{k+1,i}) (\psi_{X_R}^\dagger)^{k+1} | 0 \rangle \right)} \leq 4\epsilon\Delta k^2, \quad (\text{H17})$$

where the equality holds when $\{E(n_k)\}$ is an arbitrary projection onto two orthonormal basis in the real span of $\{(\psi_{X_R}^\dagger)^{k-1} \psi_{X_R}^{(1)\dagger} | 0 \rangle, (\psi_{X_R}^\dagger)^k | 0 \rangle\}$ as long as $Q_0(n) \gg Q_1(n) \gg Q_2(n)$ is satisfied. For example,

$$E(n_{k,\pm}) = \frac{1}{2} \left(\frac{1}{\sqrt{k!}} (\psi_{X_R}^\dagger)^k \pm \frac{1}{\Delta k \sqrt{(k-1)!}} \psi_{X_R}^{(1)\dagger} (\psi_{X_R}^\dagger)^{k-1} \right) | 0 \rangle \langle 0 | \left(\frac{1}{\sqrt{k!}} \psi_{X_R}^k \pm \frac{1}{\Delta k \sqrt{(k-1)!}} \psi_{X_R}^{(1)} \psi_{X_R}^{k-1} \right) \quad (\text{H18})$$

is optimal. Therefore, if we want to estimate both the second moment M_2 and the centroid $\bar{X} = \tilde{M}_1 + X_R$, a straightforward method is to first use half of the whole resource to locate \bar{X} such that $\delta\bar{X} \leq s$ and then use the rest half to estimate M_2 as described in Sec. V. Note that to achieve the optimal precision wrt \tilde{M}_1 , one has to count the number of detected photons by projecting the quantum state onto

$$\tilde{B} = \left\{ \frac{1}{2} \left(\frac{1}{\sqrt{k!}} (\psi_{X_R}^\dagger)^k \pm \frac{1}{\Delta k \sqrt{(k-1)!}} \psi_{X_R}^{(1)\dagger} (\psi_{X_R}^\dagger)^{k-1} \right) | 0 \rangle \langle 0 | \left(\frac{1}{\sqrt{k!}} \psi_{X_R}^k \pm \frac{1}{\Delta k \sqrt{(k-1)!}} \psi_{X_R}^{(1)} \psi_{X_R}^{k-1} \right), k \geq 1 \right\} \quad (\text{H19})$$

unlike using Eq. (34) to estimate M_2 where we don't need to count the number of photons. Similar to the weak source scenario, this measurement scheme provides an effective FIM which is half of the optimal ones Eq. (H17) and Eq. (33),

$$\mathcal{F}(\tilde{M}_1, M_2) = \begin{pmatrix} 2\epsilon\Delta k^2 & 0 \\ 0 & 2\epsilon\Delta k^2 \end{pmatrix}. \quad (\text{H20})$$

It is only half of the QFIM

$$\mathcal{J}(\tilde{M}_1, M_2) = \begin{pmatrix} 4\epsilon\Delta k^2 & 0 \\ 0 & 4\epsilon\Delta k^2 \end{pmatrix}. \quad (\text{H21})$$

The resource required to locate \bar{X} when we want to estimate even higher order moments is still neglectable as in the weak source scenario. Now we consider the possibility of further improving Eq. (H20), here we show that above scheme is at least 93.0% efficient. According to App. (C), we have, up to $O(s^2)$,

$$P(n) = A_0(n) + A_1(n)\tilde{M}_1 + A_2(n)M_2^2 + A_3(n)\tilde{M}_1^2. \quad (\text{H22})$$

For different measurement outcome n , there are only two situations:

- $P(n) = O(s^2)$, then

$$A_0(n) = \sum_{k=0}^{\infty} \frac{\epsilon^{k+1}}{(k+1)!(1+\epsilon)^{k+2}} \langle 0 | \psi_{X_R}^{k+1} E(n) (\psi_{X_R}^\dagger)^{k+1} | 0 \rangle, \quad (\text{H23})$$

$$A_1(n) = \sum_{k=0}^{\infty} \frac{2\epsilon^{k+1}}{k!(1+\epsilon)^{k+2}} \text{Re}[\langle 0 | \psi_{X_R}^k \psi_{X_R}^{(1)} E(n) (\psi_{X_R}^\dagger)^{k+1} | 0 \rangle], \quad (\text{H24})$$

$$A_2(n) = \sum_{k=0}^{\infty} \frac{\epsilon^{k+1}}{k!(1+\epsilon)^{k+1}} \langle 0 | \psi_{X_R}^k \psi_{X_R}^{(1)} E(n) \psi_{X_R}^{(1)\dagger} (\psi_{X_R}^\dagger)^k | 0 \rangle, \quad (\text{H25})$$

$$A_3(n) = \sum_{k=0}^{\infty} \frac{(k+1)\epsilon^{k+1}}{k!(1+\epsilon)^{k+2}} \langle 0 | \psi_{X_R}^k \psi_{X_R}^{(1)} E(n) \psi_{X_R}^{(1)\dagger} (\psi_{X_R}^\dagger)^k | 0 \rangle. \quad (\text{H26})$$

Other terms can be ignored in the subdiffraction limit.

- $P(n) = O(1)$, then

$$A_0(n) = \sum_{k=0}^{\infty} \frac{\epsilon^{k+1}}{(k+1)!(1+\epsilon)^{k+2}} \langle 0 | \psi_{X_R}^{k+1} E(n) (\psi_{X_R}^\dagger)^{k+1} | 0 \rangle, \quad (\text{H27})$$

$$A_1(n) = \sum_{k=0}^{\infty} \frac{2\epsilon^{k+1}}{k!(1+\epsilon)^{k+2}} \text{Re}[\langle 0 | \psi_{X_R}^k \psi_{X_R}^{(1)} E(n) (\psi_{X_R}^\dagger)^{k+1} | 0 \rangle], \quad (\text{H28})$$

and $A_2(n)$ and $A_3(n)$ can be ignored in the subdiffraction limit. For simplicity we can assume Eq. (H25) and Eq. (H26) are also true.

One important property derived from this relation is that

$$\sum_n A_2(n) = \sum_n A_3(n) = \epsilon \Delta k^2. \quad (\text{H29})$$

The entries of the FIM are

$$\mathcal{F}_{11} = \sum_n \frac{(2A_3(n)M_1')^2}{A_0'(n) + A_2(n)M_2^2 + A_3(n)M_1'^2}; \quad (\text{H30})$$

$$\mathcal{F}_{12} = \mathcal{F}_{21} = \sum_n \frac{(2A_3(n)M_1')(2A_2(n)M_2)}{A_0'(n) + A_2(n)M_2^2 + A_3(n)M_1'^2}; \quad (\text{H31})$$

$$\mathcal{F}_{22} = \sum_n \frac{(2A_2(n)M_2)^2}{A_0'(n) + A_2(n)M_2^2 + A_3(n)M_1'^2}. \quad (\text{H32})$$

where $M_1' = \tilde{M}_1 + A_1(n)/(2A_3(n))$ and $A_0'(n) = A_0(n) - A_1(n)^2/(4A_3(n)) \geq 0$. We define another 2-by-2 matrix \mathcal{F}' by replacing all $A_0(n)$ above with 0. Clearly, $\text{tr}(\mathcal{F}^{-1}) \geq \text{tr}(\mathcal{F}'^{-1})$ because $\mathcal{F}' \succeq \mathcal{F}$. Using Eq. (H29), we have

$$\mathcal{F}'_{11} + \frac{M_2}{M_1'} \mathcal{F}'_{12} = \mathcal{F}'_{22} + \frac{M_1'}{M_2} \mathcal{F}'_{12} = 4\epsilon \Delta k^2, \quad (\text{H33})$$

and

$$\begin{aligned} \mathcal{F}'_{11} + \mathcal{F}'_{22} &\leq 4 \sum_n \max\{A_2(n), A_3(n)\} \\ &\leq 4\Delta k^2 \left(\sum_{k=0}^{\lfloor \epsilon \rfloor} \frac{\epsilon^{k+1}}{(1+\epsilon)^{k+1}} + \sum_{k=\lfloor \epsilon \rfloor+1}^{\infty} \frac{(k+1)\epsilon^{k+1}}{(1+\epsilon)^{k+2}} \right) \\ &= 4\Delta k^2 \left(\epsilon + \lfloor \epsilon + 1 \rfloor \left(\frac{\epsilon}{1+\epsilon} \right)^{\lfloor \epsilon + 2 \rfloor} \right) \leq 4 \left(1 + \frac{1}{e} \right) \epsilon \Delta k^2. \end{aligned} \quad (\text{H34})$$

Therefore

$$\begin{aligned}
(\delta\tilde{M}_1)^2 + (\delta\tilde{M}_2)^2 &\geq \text{tr}(\mathcal{F}^{-1}) \geq \text{tr}(\mathcal{F}'^{-1}) = \frac{\mathcal{F}'_{11} + \mathcal{F}'_{22}}{\mathcal{F}'_{11}\mathcal{F}'_{22} - \mathcal{F}'_{12}^2} \\
&= \frac{1}{4\epsilon\Delta k^2 \left(1 - \frac{4\epsilon\Delta k^2}{\mathcal{F}'_{11} + \mathcal{F}'_{22}}\right)} \geq \frac{1+e}{4} \frac{1}{\epsilon\Delta k^2}.
\end{aligned} \tag{H35}$$

We conclude that our measurement scheme is at least $\frac{1+e}{4} \approx 93.0\%$ efficient for arbitrary ϵ in the sense that if one achieve certain estimation precision $(\delta\tilde{M}_1)^2 + (\delta\tilde{M}_2)^2$ by repeating our measurement N times, the optimal measurement scheme requires at least $93.0\% \cdot N$ times to achieve such precision.

We can easily generalize above measurement scheme to 2D imaging when the PSF is separable. $\psi_{X_R Y_R}^{(10)\dagger}$ and $\psi_{X_R Y_R}^{(01)\dagger}$ are orthogonal. As in 1D imaging,

$$\tilde{M}_{10} = \bar{X} - X_R, \quad \tilde{M}_{01} = \bar{Y} - Y_R \tag{H36}$$

are estimated by

$$\frac{1}{k!} (\psi_{X_R Y_R}^\dagger)^k \frac{\psi_{X_R Y_R}^\dagger \pm \frac{1}{\Delta k_x} \psi_{X_R Y_R}^{(10)\dagger}}{\sqrt{2}} |0\rangle \langle 0| \frac{\psi_{X_R Y_R} \pm \frac{1}{\Delta k_x} \psi_{X_R Y_R}^{(10)}}{\sqrt{2}} (\psi_{X_R Y_R})^k \tag{H37}$$

and

$$\frac{1}{k!} (\psi_{X_R Y_R}^\dagger)^k \frac{\psi_{X_R Y_R}^\dagger \pm \frac{1}{\Delta k_y} \psi_{X_R Y_R}^{(01)\dagger}}{\sqrt{2}} |0\rangle \langle 0| \frac{\psi_{X_R Y_R} \pm \frac{1}{\Delta k_y} \psi_{X_R Y_R}^{(01)}}{\sqrt{2}} (\psi_{X_R Y_R})^k \tag{H38}$$

with optimal FIs equal to

$$\mathcal{F}_{1010} = 4\epsilon\Delta k_x^2, \quad \mathcal{F}_{0101} = 4\epsilon\Delta k_y^2. \tag{H39}$$

We won't discuss the optimality of simultaneously estimating $\tilde{M}_{10}, \tilde{M}_{01}$ and the second moments M_{20}, M_{11}, M_{02} here.