

Quantum-limited estimation of the axial separation of two incoherent point sources: supplementary material

YIYU ZHOU^{1,*}, JING YANG², JEREMY D. HASSETT¹, SEYED MOHAMMAD HASHEMI RAFSANJANI³, MOHAMMAD MIRHOSSEINI⁴, A. NICK VAMIVAKAS^{1,2,5}, ANDREW N. JORDAN^{2,6}, ZHIMIN SHI^{7,†}, AND ROBERT W. BOYD^{1,2,8,‡}

¹The Institute of Optics, University of Rochester, Rochester, New York 14627, USA

²Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627, USA

³Department of Physics, University of Miami, Coral Gables, Florida 33146, USA

⁴Thomas J. Watson, Sr., Laboratory of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA

⁵Materials Science Program, University of Rochester, Rochester, New York 14627, USA

⁶Institute for Quantum Studies, Chapman University, Orange, California 92866, USA

⁷Department of Physics, University of South Florida, Tampa, Florida 33620, USA

⁸Department of Physics, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada

*Corresponding author: yzhou62@ur.rochester.edu

†email: zhiminshi@usf.edu

‡email: boyd@optics.rochester.edu

Published 26 April 2019

This document provides supplementary information to “Quantum-limited estimation of the axial separation of two incoherent point sources,” <https://doi.org/10.1364/OPTICA.6.000534>. We provide derivation of the quantum Fisher information, the construction of optimal measurements, SLM calibration and data processing, analysis of astigmatic imaging, detailed analysis of the bias of estimators, and Fisher information calculation for an Airy-disk-shaped PSF.

1. DERIVATION OF THE QUANTUM FISHER INFORMATION

Following the procedure in the supplement of Ref. [S1], the quantum Fisher information in our case can be directly calculated as

$$\mathcal{K}_s = 4[|\langle \partial_s \psi_1 | \partial_s \psi_1 \rangle - |\langle \psi_1 | \partial_s \psi_1 \rangle|^2], \quad (\text{S1})$$

and it can be readily verified that $\mathcal{K}_s = 1/4z_R^2$ for the Gaussian PSF. Here we also follow an usual method to calculate the quantum Fisher information based on the symmetric logarithmic derivative. For two point sources that are located at $z = \pm s/2$ respectively, the density matrix of a single photon can be expressed as

$$\rho = \frac{1}{2}(|\psi_1\rangle \langle \psi_1| + |\psi_2\rangle \langle \psi_2|), \quad (\text{S2})$$

where $\langle r_0 | \psi_1 \rangle = \psi(r_0; s/2)$ and $\langle r_0 | \psi_2 \rangle = \psi(r_0; -s/2)$. The orthonormal bases in which the density matrix is diagonal are found to be

$$\begin{aligned} |e_1\rangle &= A_1(|\psi_1\rangle + e^{-i\phi} |\psi_2\rangle), \\ |e_2\rangle &= A_2(|\psi_1\rangle - e^{-i\phi} |\psi_2\rangle), \\ A_1 &= \left(2 + \frac{2}{\sqrt{1+\delta^2}}\right)^{-1/2}, \\ A_2 &= \left(2 - \frac{2}{\sqrt{1+\delta^2}}\right)^{-1/2}, \\ \delta &= k\text{NA}^2 s/4, \\ \phi &= \arctan\delta. \end{aligned} \quad (\text{S3})$$

Hence the density matrix can be rewritten as

$$\rho = \frac{1}{4A_1^2} |e_1\rangle \langle e_1| + \frac{1}{4A_2^2} |e_2\rangle \langle e_2|. \quad (\text{S4})$$

Here we also define the following orthonormal bases

$$\begin{aligned} |e_3\rangle &= B_1(|m_3\rangle + e^{-i3\phi} |m_4\rangle), \\ |e_4\rangle &= B_2(|m_3\rangle - e^{-i3\phi} |m_4\rangle), \\ |m_3\rangle &= \frac{i}{c_0} |\partial_s \psi_1\rangle - c_1 |e_1\rangle - c_2 |e_2\rangle, \\ |m_4\rangle &= \frac{-i}{c_0} |\partial_s \psi_2\rangle - c_3 |e_1\rangle - c_4 |e_2\rangle, \end{aligned} \quad (\text{S5})$$

where $|\partial_s \psi_1\rangle = \partial |\psi_1\rangle / \partial s$, $|\partial_s \psi_2\rangle = \partial |\psi_2\rangle / \partial s$, and the coefficients are calculated to be

$$\begin{aligned} B_1 &= \left(\frac{\delta^2(\sqrt{\delta^2+1}-1)}{2(\delta^2+1)^{5/2}} \right)^{-1/2}, \\ B_2 &= \left(\frac{\delta^2(\sqrt{\delta^2+1}+1)}{2(\delta^2+1)^{5/2}} \right)^{-1/2}, \\ c_0 &= kNA^2/4, \\ c_1 &= A_1 \frac{-e^{i\phi} + (\delta - i)^2}{2(\delta - i)^2}, \\ c_2 &= A_2 \frac{e^{i\phi} + (\delta - i)^2}{2(\delta - i)^2}, \\ c_3 &= A_1 \frac{e^{i\phi}(\delta + i)^2 - 1}{2(\delta + i)^2}, \\ c_4 &= A_2 \frac{-1 - e^{i\phi}(\delta + i)^2}{2(\delta + i)^2}. \end{aligned} \quad (\text{S6})$$

Here the symmetric logarithmic derivative of density matrix is

$$\mathcal{L}(\rho) = \sum_{j,k; D_j+D_k \neq 0} \frac{2}{D_j+D_k} \langle e_j | \frac{\partial \rho}{\partial s} | e_k \rangle | e_j \rangle \langle e_k | \quad (\text{S7})$$

where $D_j = \langle e_j | \rho | e_j \rangle$ and the matrix elements of $\mathcal{L}(\rho)$ can be calculated as

$$\begin{aligned} \mathcal{L}_{11} &= c_0 A_1^2 \frac{-2\delta}{(\delta^2+1)^{3/2}}, \\ \mathcal{L}_{22} &= c_0 A_2^2 \frac{2\delta}{(\delta^2+1)^{3/2}}, \\ \mathcal{L}_{21} = \mathcal{L}_{12}^* &= \frac{-2ic_0\delta}{\delta^2+1}, \\ \mathcal{L}_{31} = \mathcal{L}_{13}^* &= \frac{2iA_1c_0}{2B_1} (e^{2i\phi} - 1), \\ \mathcal{L}_{41} = \mathcal{L}_{14}^* &= \frac{-2iA_1c_0}{2B_2} (e^{2i\phi} + 1), \\ \mathcal{L}_{32} = \mathcal{L}_{23}^* &= \frac{-2iA_1c_0}{2B_2} (e^{2i\phi} + 1), \\ \mathcal{L}_{42} = \mathcal{L}_{24}^* &= \frac{2iA_2c_0}{2B_2} (e^{2i\phi} - 1), \\ \mathcal{L}_{33} = \mathcal{L}_{34} = \mathcal{L}_{43} = \mathcal{L}_{44} &= 0, \end{aligned} \quad (\text{S8})$$

where $\mathcal{L}_{jk} = \langle e_j | \mathcal{L}(\rho) | e_k \rangle$. After more algebra one can calculate that the quantum Fisher information is

$$\mathcal{K}_s(\rho) = \text{Re Tr} \mathcal{L}(\rho) \mathcal{L}(\rho) \rho = c_0^2 = \frac{1}{4z_R^2}. \quad (\text{S9})$$

2. SATURATION OF THE QUANTUM CRAMÉR-RAO BOUND

We are interested in the saturating the quantum Cramér-Rao bound in the limit of $s = 0$. Similar to the proof of Corollary 1 in Ref. [S2], one can show that a measurement consisting of projectors $\Pi_k = \sum_{\alpha} |\pi_{k\alpha}\rangle \langle \pi_{k\alpha}|$ can saturate the quantum Cramér-Rao bound at $s = 0$ if and only if for all *regular* projectors (defined as $|\langle \psi_1 | \Pi_k | \psi_1 \rangle|_{s=0} > 0$) we can have

$$\langle \partial_s^0 \psi_1 | \pi_{k\alpha} \rangle |_{s=0} = 0, \quad \forall k, \quad (\text{S10})$$

where $|\partial_s^0 \psi_1\rangle |_{s=0} = (|\partial_s \psi_1\rangle - \langle \psi_1 | \partial_s \psi_1 \rangle |\psi_1\rangle) |_{s=0}$. Such an optimal measurement can be constructed by choosing a proper trial basis $\{\pi_{k\alpha}\}$ and then follow the procedure in Ref. [S2]: (i) Identify *regular* basis vectors defined as $\langle \psi_1 | \pi_{k\alpha} \rangle |_{s=0} \neq 0$ and *null* basis vectors defined as $\langle \psi_1 | \pi_{k\alpha} \rangle |_{s=0} = 0$. (ii) Calculate $\langle \partial_s^0 \psi_1 | \pi_{k\alpha} \rangle |_{s=0}$ and check whether it vanishes or not. (iii) Assemble regular basis vectors satisfying Eq. (S10) as a regular projector $\Pi_k = \sum_{\alpha} |\pi_{k\alpha}\rangle \langle \pi_{k\alpha}|$. (iv) A null basis vector $|\pi_{k\alpha}\rangle$ is *flexible* if $\langle \partial_s \psi_1 | \pi_{k\alpha} \rangle |_{s=0} = 0$. The rank one flexible projector $\Pi_{k\alpha}$ formed by a flexible basis vector can be added to any of the previous regular projectors or the following null projectors. (v) Assemble null basis vectors that are not flexible as a null projector $\Pi_k = |\pi_{k\alpha}\rangle \langle \pi_{k\alpha}|$ defined as $\langle \psi_1 | \Pi_k | \psi_1 \rangle |_{s=0} = 0$. One can check that any measurement constructed from the previous procedure can satisfy Eq. (S10).

A. Optimal measurement basis set for the Gaussian pupil function

In the main text, we consider the case of the Gaussian pupil function, $\psi_1(r_0) = \sqrt{2/\pi} \cdot \exp(-r_0^2) \cdot \exp(-iksNA^2r_0^2/4)$. It is straightforward to find

$$\psi_1(r_0) |_{s=0} = \text{LG}_0(r_0), \quad (\text{S11})$$

$$\partial_s \psi_1(r_0) |_{s=0} = -\frac{ikNA^2}{8} [\text{LG}_0(r_0) - \text{LG}_1(r_0)]. \quad (\text{S12})$$

We take radial Laguerre-Gaussian modes $\langle r_0 | \text{LG}_p \rangle = \text{LG}_p(r_0)$ as a trial basis. So we find that for step (i) in the limit $s = 0$, the only regular basis vector is $|\text{LG}_0\rangle$ and the remaining other basis vectors are null. (ii) It can be readily shown that

$$\langle \partial_s^0 \psi_1 | \text{LG}_0 \rangle |_{s=0} = \langle \partial_s \psi_1 | \text{LG}_0 \rangle |_{s=0} - \langle \partial_s \psi_1 | \text{LG}_0 \rangle |_{s=0} \langle \text{LG}_0 | \text{LG}_0 \rangle = 0.$$

(iii) We obtain a regular projector $\Pi'_0 = |\text{LG}_0\rangle \langle \text{LG}_0|$. (iv) From Eq. (S12), we find $\langle \partial_s \psi_1 | \text{LG}_p \rangle = 0$ for $p \geq 2$. Thus the basis vectors with p higher than one are all flexible and therefore can freely added to any regular or null projector. (v) The only non-flexible null basis vector is $|\text{LG}_1\rangle$, thus we can form a null projector $\Pi'_1 = |\text{LG}_1\rangle \langle \text{LG}_1|$. We add the projector formed by flexible basis vectors of even and odd order to the previous regular projectors Π_0 and Π_1 respectively. Therefore the final optimal measurements can be

$$\Pi_0 = \sum_{p=0} |\text{LG}_{2p}\rangle \langle \text{LG}_{2p}|, \quad \Pi_1 = \sum_{p=0} |\text{LG}_{2p+1}\rangle \langle \text{LG}_{2p+1}|. \quad (\text{S13})$$

Alternatively one can add the flexible projectors to Π'_0 or Π'_1 to give rise to optimal measurements of

$$\Pi_0 = |\text{LG}_0\rangle \langle \text{LG}_0|, \quad \Pi_1 = 1 - \Pi_0, \quad (\text{S14})$$

or

$$\Pi_0 = 1 - \Pi_1, \quad \Pi_1 = |\text{LG}_1\rangle \langle \text{LG}_1|. \quad (\text{S15})$$

Any sorter that can efficiently perform the above measurements can be used to reach the quantum Fisher information when s approaches 0.

B. Optimal measurement basis set for the pupil function of a hard-edged aperture

The construction of optimal measurement for the Gaussian pupil function can be analogously done for the pupil function of hard-edged aperture. Consider the pupil function $\psi_1(r_0) = \sqrt{2/\pi} \cdot \text{circ}(\sqrt{2}r_0) \exp(-iksNA^2r_0^2/4)$, where $\text{circ}(r) = 1$ if $0 \leq r \leq 1$ and $\text{circ}(r) = 0$ if $r > 1$. The quantum Fisher information is still expressed by Eq. (S1). It is straightforward to find

$$\psi_1(r_0) |_{s=0} = Z_0(\sqrt{2}r_0), \quad (\text{S16})$$

$$\partial_s \psi_1(r_0) |_{s=0} = -\frac{ikNA^2}{16\sqrt{3}} [Z_2(\sqrt{2}r_0) + \sqrt{3}Z_0(\sqrt{2}r_0)], \quad (\text{S17})$$

where $Z_n(\sqrt{2}r_0) = \sqrt{2(n+1)/\pi} R_n(\sqrt{2}r_0) \text{circ}(\sqrt{2}r_0)$ and R_n is the Zernike polynomial. We take radial Zernike basis $\langle r_0 | Z_n \rangle = Z_n(\sqrt{2}r_0)$ as the trial basis. Following the previous procedure, we find the following optimal measurements:

$$\Pi_0 = \sum_{\text{even } n} |Z_{2n}\rangle \langle Z_{2n}|, \quad \Pi_1 = \sum_{\text{odd } n} |Z_{2n}\rangle \langle Z_{2n}|, \quad (\text{S18})$$

or

$$\Pi_0 = |Z_0\rangle \langle Z_0|, \quad \Pi_1 = 1 - \Pi_0, \quad (\text{S19})$$

or

$$\Pi_0 = 1 - \Pi_1, \quad \Pi_1 = |Z_2\rangle \langle Z_2|. \quad (\text{S20})$$

Any sorter that can efficiently perform the above measurements can be used to reach the quantum Fisher information when s approaches 0.

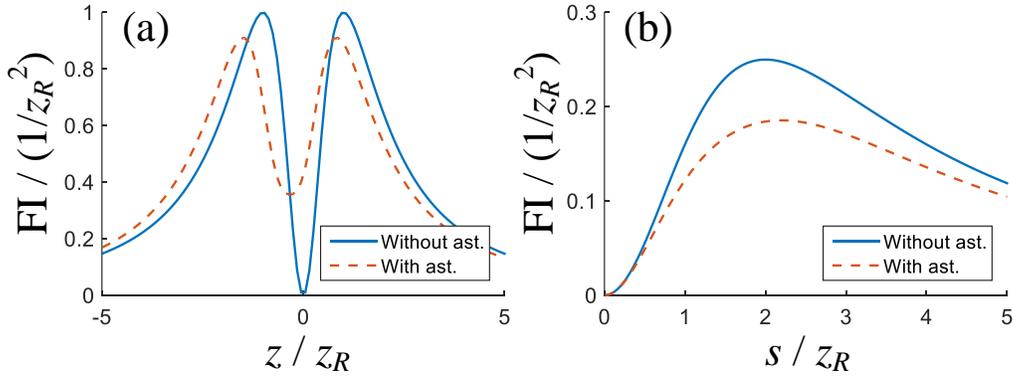


Fig. S1. Fisher information of (a) axial position z of a single point source and (b) axial separation s between a pair of point sources. The Fisher information without astigmatism (ast.) is also plotted as a reference.

3. ANALYSIS OF ASTIGMATIC IMAGING

A. Astigmatic imaging for separation estimation

In this section we calculate the Fisher information for astigmatic imaging. We introduce astigmatism to the Gaussian pupil function by adding the quadratic phase of a vertically oriented cylindrical lens, which can be expressed as

$$\psi(x_p, y_p; z) = \sqrt{2/\pi} \frac{1}{f_1 \text{NA}} \exp\left(-\frac{x_p^2 + y_p^2}{(f_1 \text{NA})^2}\right) \exp\left(-i \frac{kz}{2f_1^2} x_p^2\right) \exp\left[-i\left(\frac{kz}{2f_1^2} + \frac{k}{2f_c}\right) y_p^2\right], \quad (\text{S21})$$

where $f_c = 1.2$ m is the focal length of the astigmatic cylindrical lens and $f_1 = 4$ mm is the objective focal length. The corresponding intensity in the image plane can be calculated as

$$I(x, y; z) = \frac{2}{\pi} \frac{1}{w_x(z)w_y(z)} \exp\left(-\frac{2x^2}{w_x^2(z)}\right) \exp\left(-\frac{2y^2}{w_y^2(z)}\right), \quad (\text{S22})$$

where

$$w_x(z) = \frac{\lambda}{\pi \text{NA}} \sqrt{1 + (\pi z \text{NA}^2 / \lambda)^2}, \quad w_y(z) = \frac{\lambda}{\pi \text{NA}} \sqrt{1 + [\pi(\frac{f_1^2}{f_c} + z) \text{NA}^2 / \lambda]^2}. \quad (\text{S23})$$

The Fisher information then can be obtained as

$$\mathcal{J}_{\text{ast}}(s) = \iint_{-\infty}^{+\infty} \frac{1}{I(x, y, s/2)} \left(\frac{\partial I(x, y, s/2)}{\partial s}\right)^2 dx dy. \quad (\text{S24})$$

The Fisher information with astigmatism is plotted as the dashed line in Fig. S1. As in comparison, in Fig. S1(a) we plot the Fisher information for single point source axial localization and in Fig. S1(b) the Fisher information for measurement of separation of two

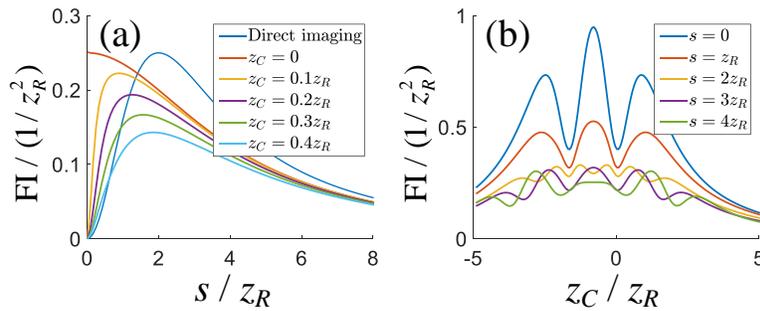


Fig. S2. (a) The Fisher information of separation estimation for sorter-based measurement with different centroid positions. The Fisher information for direct imaging with point source pair centroid $z_C = 0$ is plotted as a reference. (b) The Fisher information of centroid estimation for astigmatic imaging with different separations.

point sources. For axial localization of a single point source, astigmatism can help increase the Fisher information at $z \rightarrow 0$ as expected. However, for measuring the axial separation of a pair of point sources, the Fisher information again drops to zero, which contrasts to the improvement of radial mode sorter discussed in our manuscript.

B. Effect of centroid misalignment and astigmatic imaging for centroid estimation

Assume that the centroid of the point source pair is z_C and the separation is s , so the positions of two point sources are $z_1 = z_C + s/2$ and $z_2 = z_C - s/2$, respectively. Therefore, the output radial mode distribution becomes $P'(p; z_C, s) \equiv \langle \text{LG}_p | \rho | \text{LG}_p \rangle = [P(p; z_C + s/2) + P(p; z_C - s/2)]/2$. Consequently, the Fisher information of separation in the presence of centroid misalignment can be written as

$$\mathcal{J}'(s) = \sum_{p=0}^{\infty} \frac{1}{P'(p; z_C, s)} \left(\frac{\partial P'(p; z_C, s)}{\partial s} \right)^2, \quad (\text{S25})$$

where the result is shown in Fig.S2(a).

As the next step, we analyze the Fisher information of centroid estimation for astigmatic imaging. The intensity at the image plane can be rewritten as

$$I'(x, y; z_C, s) = \frac{1}{2} I(x, y; z_C + s/2) + \frac{1}{2} I(x, y; z_C - s/2), \quad (\text{S26})$$

where the definition of $I(x, y; z)$ follows Eq. (S22), and the Fisher information of centroid estimation is thus

$$\mathcal{J}'_{\text{ast}}(z_C, s) = \iint_{-\infty}^{+\infty} \frac{1}{I(x, y; z_C, s)} \left(\frac{\partial I(x, y, z_C, s)}{\partial z_C} \right)^2 dx dy. \quad (\text{S27})$$

Here we use the value of $f_C = 0.5$ m and the result is shown in Fig.S2(b).

4. SLM CALIBRATION AND DATA PROCESSING

In our experiment we use SLM 1 to generate Gaussian point spread function, and we calibrate SLM 1 to compensate the aberration and experimental imperfection. The pupil function we want to generate is

$$\psi(r_0; z) = \sqrt{2/\pi} \exp(-r_0^2) \exp(-ikzNA^2r_0^2/2), \quad (\text{S28})$$

where $z = \pm s/2$ and this pupil function is generated at the first diffraction order of the computer-generated hologram imprinted on the SLM. We apply Seidel aberrations to the computer-generated hologram to improve mode quality and tune the parameter z to fit the expected curve. For direct imaging, the measured width w_0 of the Gaussian PSF should increase with z as expressed in Eq. (2) in the manuscript, and the corrected experimental result is shown in Fig. S3(a). For binary sorter-based measurement, the value of Q calculated from the output photon numbers of the binary sorter defined in Eq. (11) can be written as a function of z as

$$Q = \frac{1}{2} \left(1 - \frac{1}{z^2/32z_R^2 + 1} \right). \quad (\text{S29})$$

We correct the SLM to fit this curve and the calibrated data is shown in Fig. S3(b). Due to the detector noise and misalignment of radial mode sorter, the output Q has a small, nonzero value of 0.28% at $z = 0$, i.e. $Q(z = 0) = 0.28\%$, and we treat this experimentally measured nonzero value as a constant and subtract it before estimating the axial position. In other words, we use a new quantity $\bar{Q} = Q - 0.28\%$ in the estimator to calculate s . However, the shot noise associated with this crosstalk cannot be simply removed by this subtraction. At $z = 0$, Q is in fact a random variable with an average of 0.28%. For a specific measurement, if the measured Q is lower than 0.28%, then \bar{Q} becomes negative and the above equation results in an imaginary z . In our experiment, once a negative \bar{Q} is obtained we force it to be zero to guarantee a real z . Given that \bar{Q} is either zero or positive at $z = 0$, the expectation of \bar{Q} becomes positive rather than 0, and consequently the expectation of \hat{z} becomes positive. At $z = 0$ the two point sources coincide and thus this is why we measured a positive s when the s is zero. This treatment is also used in the estimator of direct imaging as will be analyzed in detail in the next section.

To quantify our treatment, we model the crosstalk of 0.28% as binomial distribution. Specifically, we assume a probability of $p = 0.28\%$ for a photon to be detected in the output port for odd-order radial modes. For N photons, the number of photons that appear in the output port for odd-order radial modes follows the binomial distribution as

$$B(k; N, p) = \frac{N!}{k!(N-k)!} p^k (1-p)^{N-k}. \quad (\text{S30})$$

Here the estimator is defined as

$$\hat{Q} = k/N, \quad \hat{s} = 2z_R \sqrt{\frac{2}{1 - 2\hat{Q}} - 2}. \quad (\text{S31})$$

The expectation of \hat{Q} is $E[\hat{Q}] = p = 0.28\%$. As we mentioned above, we subtract this number and use $\hat{\bar{Q}} = \hat{Q} - 0.28\%$ as the new estimator to mitigate the effect of crosstalk. However, it is possible to obtain a negative $\hat{\bar{Q}}$, and we force it to be zero whenever a negative value is obtained. Then the expectation of \hat{s} can be calculated as

$$E[\hat{s}] = \sum_{k > Np}^N 2z_R B(k; N, p) \sqrt{\frac{2}{1 - 2\frac{k - Np}{N}} - 2}. \quad (\text{S32})$$

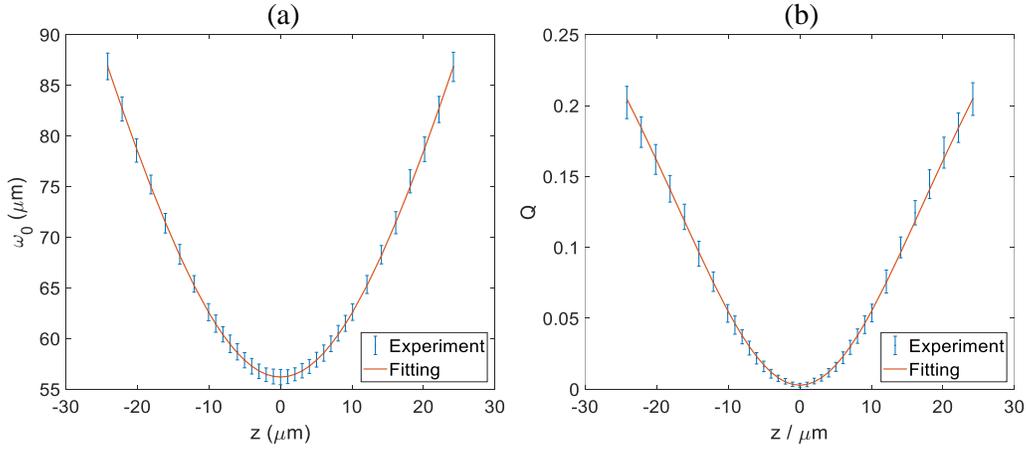


Fig. S3. (a) Measured Gaussian width w_0 as a function of z . (b) Measured Q of binary sorter as a function of z .

For $N = 2000$ and $p = 0.28\%$ the above equation can be numerically calculated to be $E[\hat{s}] = 0.043z_R$, which is very close to the experimentally measured value of $0.049z_R$.

5. BIAS ANALYSIS OF THE ESTIMATOR

A. Direct imaging method

The intensity distribution on the image plane for the direct imaging method can be written as [see Eq. (2) in the manuscript]

$$I(r; z) = \frac{2}{\pi} \frac{1}{w^2(z)} \exp\left(-\frac{2r^2}{w^2(z)}\right), \quad (\text{S33})$$

$$w(z) = \frac{M\lambda}{\pi\text{NA}} \sqrt{1 + (z\pi\text{NA}^2/\lambda)^2},$$

which can be interpreted as the probability density function (PDF) on the image plane for single photons. The magnification M is assumed to be 1 to simplify the calculation, and as we show in the manuscript this magnification does not influence the Fisher information. Since the point sources are located at $z = \pm s/2$, in the following we rewrite $w(z)$ as $w(s) = w(z = s/2)$. The maximum likelihood estimator of direct imaging for N measurements is

$$\hat{s}_{\text{direct}} = 2z_R \sqrt{\frac{2}{Nw_0^2} \sum_{m=1}^N r_m^2 - 1}, \quad (\text{S34})$$

and our goal here is to analyze the bias of this estimator at the limit of $s \rightarrow 0$. The CRLB of an estimator is [S3]

$$\text{Var}(\hat{s}) \geq \frac{(\partial E[\hat{s}]/\partial s)^2}{N \cdot I(s)}, \quad (\text{S35})$$

where $\text{Var}(\hat{s})$ is the variance of the estimator, $E[\hat{s}]$ is the expectation of estimator, and $I(s) = 4s^2/(s^2 + 4z_R^2)^2$ is the Fisher information. For an unbiased estimator, i.e. $E[\hat{s}] = s$ and thus $\partial E[\hat{s}]/\partial s = 1$, the CRLB is simply given by the inverse of Fisher information. In the manuscript we show that the Fisher information of direct imaging vanishes when $s \rightarrow 0$, but the variance of direct imaging in both experiment and simulation does not diverge to infinity. This is because of the bias of the estimator as we analyze in the following. In the polar coordinate, the PDF becomes

$$f(r) = \frac{4}{w^2(s)} \exp\left(-\frac{2r^2}{w^2(s)}\right), \quad r \geq 0. \quad (\text{S36})$$

Since the estimator is defined in terms of r^2 , here we define $\bar{r} = r^2$ and the PDF of \bar{r} can be calculated as

$$\bar{f}(\bar{r}) = f(r(\bar{r})) \frac{dr}{d\bar{r}} = \frac{2}{w^2(s)} \exp\left(-\frac{2\bar{r}}{w^2(s)}\right), \quad r \geq 0, \quad (\text{S37})$$

where $r(\bar{r}) = \sqrt{\bar{r}}$. For N independent measurements (r_1, r_2, \dots, r_m) , the PDF of the sum of random variables is the convolution of each PDF. Therefore, the PDF of $x = \sum_{m=1}^N r_m^2$ can be expressed as $\bar{f}(\bar{r}) \otimes \bar{f}(\bar{r}) \otimes \dots \otimes \bar{f}(\bar{r})$, where \otimes denotes the operation of convolution. This PDF turns to be the Erlang distribution, which is a special case of the gamma distribution, and can be given as

$$g(x; N, k') = \frac{1}{(N-1)!} k'^N x^{N-1} \exp(-k'x), \quad x \geq 0, \quad (\text{S38})$$

where $k' = 2/w^2(s)$. Therefore, the PDF of $\bar{x} = 2x/Nw_0^2 - 1$, which is the term under the square root of Eq. (S34), can be readily obtained as

$$P(\bar{x}) = g(\bar{x} + 1; N, k), \quad \bar{x} \geq -1, \quad (\text{S39})$$

where $k = Nw_0^2k'/2 = N/[1 + (s\pi NA^2/2\lambda)^2]$. It can be noticed that the estimator, which can be written as $\hat{s}_{\text{direct}} = 2z_R\sqrt{\bar{x}}$, become ill-posed when $\bar{x} < 0$. The intuition behind this behavior is that the estimator calculates the separation s based on the measured $w(s)$. By definition we know that $w(s) \geq w_0$, but in an experiment it is possible to get a value of $w(s)$ that is even less than w_0 , especially when the available photon number is small. One way to proceed is to force $\bar{x} = 0$ whenever a negative $\bar{x} = 0$ is measured, and this is how we perform our experiment as well as the Monte Carlo simulation. Hence, the integration region $-1 \leq \bar{x} \leq 0$ can be ignored because $\sqrt{\bar{x}}$ is forced to be 0, and consequently the expectation becomes

$$\begin{aligned} E[\hat{s}_{\text{direct}}] &= E[2z_R\sqrt{\bar{x}}] = 2z_R \int_0^\infty \sqrt{\bar{x}}g(\bar{x} + 1; N, k)d\bar{x} \\ &= 2z_R \int_1^\infty \sqrt{\bar{x} - 1}g(\bar{x}; N, k)d\bar{x}. \end{aligned} \quad (\text{S40})$$

When s is small, we can use the Taylor expansion of k as $k \approx N(1 - c_0s^2)$ with $c_0 = \pi^2NA^4/4\lambda^2 = 1/4z_R^2$. In addition, when N is relatively large, such as 2000 used in our experiment, we can use the Stirling approximation $N! \approx \sqrt{2\pi N}N^{N+1/2}e^{-N}$ and thus $g(\bar{x}; N, k)$ can be approximated as

$$g(\bar{x}; N, k) \approx g(\bar{x}; N, N(1 - c_0s^2)) \approx \sqrt{\frac{N}{2\pi}}(1 - c_0s^2)^N x^{N-1} e^{(1-x)N} e^{Nxc_0s^2}. \quad (\text{S41})$$

Using $e^{Nxc_0s^2} \approx 1 + Nxc_0s^2$ and $(1 - c_0s^2)^N \approx 1 - Nc_0s^2$, we have $e^{Nxc_0s^2}(1 - c_0s^2)^N \approx 1 + N(x - 1)c_0s^2$ and thus

$$\begin{aligned} g(\bar{x}; N, k) &\approx \sqrt{\frac{N}{2\pi}} x^{N-1} e^{(1-x)N} [1 + N(x - 1)c_0s^2], \\ \frac{\partial g(\bar{x}; N, k)}{\partial s} &\approx \sqrt{\frac{2N}{\pi}} x^{N-1} e^{(1-x)N} N(x - 1)c_0s, \end{aligned} \quad (\text{S42})$$

therefore

$$E[\hat{s}_{\text{direct}}]_{s=0} = \sqrt{\frac{2N}{\pi}} z_R \int_1^\infty \sqrt{\bar{x} - 1} \bar{x}^{N-1} e^{-(\bar{x}-1)N} d\bar{x}. \quad (\text{S43})$$

For CRLB, we have

$$\begin{aligned} \left. \frac{\partial E[\hat{s}_{\text{direct}}]}{\partial s} \right|_{s=0} &= 2z_R \int_1^\infty \sqrt{\bar{x} - 1} \frac{\partial g(\bar{x}; N, k)}{\partial s} d\bar{x} \\ &= \sqrt{\frac{N}{2\pi}} \frac{Ns}{z_R} \int_1^\infty (\bar{x} - 1)^{3/2} \bar{x}^{N-1} e^{-(\bar{x}-1)N} d\bar{x}. \end{aligned} \quad (\text{S44})$$

In the manuscript we evaluate scaled standard deviation as $\text{Var}(\hat{s})^{1/2}/(z_R/\sqrt{N})$, which can be expressed as

$$\begin{aligned} \left. \frac{\text{Var}(\hat{s})^{1/2}}{z_R/\sqrt{N}} \right|_{s=0} &\geq \frac{\partial E[\hat{s}_{\text{direct}}]/\partial s}{z_R I(s)} = \frac{2z_R}{s} \frac{\partial E[\hat{s}_{\text{direct}}]}{\partial s} \\ &= \sqrt{\frac{2N}{\pi}} N \int_1^\infty (\bar{x} - 1)^{3/2} \bar{x}^{N-1} e^{-(\bar{x}-1)N} d\bar{x}. \end{aligned} \quad (\text{S45})$$

The integral of above equations can be analytically calculated by the following approximations. Eq. (S44) can be rewritten as

$$\left. \frac{\partial E[\hat{s}_{\text{direct}}]}{\partial s} \right|_{s=0} = \frac{s}{\sqrt{2\pi}z_R} N^{3/2} \int_1^\infty (\bar{x} - 1)^{3/2} \bar{x}^{N-1} e^{-(\bar{x}-1)N} d\bar{x}. \quad (\text{S46})$$

Then we set $\bar{x} = 1 + t/\sqrt{N}$ and therefore the integral becomes

$$\left. \frac{\partial E[\hat{s}_{\text{direct}}]}{\partial s} \right|_{s=0} = \frac{s}{\sqrt{2\pi}z_R} N^{1/4} \int_0^\infty t^{3/2} \left(1 + \frac{t}{\sqrt{N}}\right)^{N-1} e^{-\sqrt{N}t} dt. \quad (\text{S47})$$

Now we can use the following approximation

$$\begin{aligned} \left(1 + \frac{t}{\sqrt{N}}\right)^{N-1} &= \exp\left((N-1)\ln\left(1 + \frac{t}{\sqrt{N}}\right)\right) \\ &= \exp\left((N-1)\left(\frac{t}{\sqrt{N}} - \frac{t^2}{2N} + \dots\right)\right) \\ &\approx \exp\left(\sqrt{N}t - \frac{t^2}{2} - \frac{t}{\sqrt{N}} + \frac{t^2}{2N}\right) \\ &= \exp(\sqrt{N}t) \exp\left(-\frac{t^2}{2}\right) \exp\left(-\frac{t}{\sqrt{N}} + \frac{t^2}{2N}\right) \\ &\approx \exp(\sqrt{N}t) \exp\left(-\frac{t^2}{2}\right) \left(1 - \frac{t}{\sqrt{N}} + \frac{t^2}{2N}\right), \end{aligned} \quad (\text{S48})$$

therefore

$$\begin{aligned}
\left. \frac{\partial E[\hat{s}_{\text{direct}}]}{\partial s} \right|_{s=0} &\approx \frac{s}{\sqrt{2\pi z_R}} \propto N^{1/4} \int_0^\infty t^{3/2} e^{-t^2/2} \left(1 - \frac{t}{\sqrt{N}} + \frac{t^2}{2N}\right) dt \\
&= \frac{s}{\sqrt{2\pi z_R}} N^{1/4} 2^{1/4} \int_0^\infty e^{-t_1} \left(t_1^{1/4} - \sqrt{\frac{2}{N}} t_1^{3/4} + \frac{1}{N} t_1^{5/4}\right) dt \\
&\approx \frac{s}{\sqrt{2\pi z_R}} N^{1/4} 2^{1/4} \Gamma\left(\frac{5}{4}\right) \\
&\approx \frac{0.43 N^{1/4} s}{z_R},
\end{aligned} \tag{S49}$$

where $t_1 = t^2/2$ and $\Gamma(z) = \int_0^\infty x^{z-1} \exp(-x) dx$ is the Gamma function. Immediately we can get

$$\left. \frac{\text{Var}(\hat{s})^{1/2}}{z_R/\sqrt{N}} \right|_{s=0} \geq \frac{\partial E[\hat{s}_{\text{direct}}]/\partial s}{z_R I(s)} = \frac{2z_R}{s} \frac{\partial E[\hat{s}_{\text{direct}}]}{\partial s} \approx 0.86 N^{1/4}. \tag{S50}$$

The right-hand side of the above inequality is calculated to be 5.8 for $N = 2000$, which is close to the Monte Carlo simulation result 6.6 as shown in Fig. 4(b) in the manuscript. We note that the above equation is an inequality instead of an equation, thus our analytic result is reasonable since $5.8 < 6.6$. Another apparent observation is that the CRLB of direct imaging scales with N as

$$\text{Var}(\hat{s}) \geq \frac{(\partial E[\hat{s}]/\partial s)^2}{N \cdot I(s)} \propto \frac{(N^{1/4})^2}{N} = \frac{1}{\sqrt{N}} \tag{S51}$$

which means that this estimator cannot reach the standard quantum limit [S4]. Similar math tricks can be applied to the evaluation of $E[\hat{s}_{\text{direct}}]_{s=0}$ and the result is

$$E[\hat{s}_{\text{direct}}]_{s=0} = \left(\frac{2}{N}\right)^{1/4} \frac{z_R}{\sqrt{\pi}} \Gamma\left(\frac{3}{4}\right) \approx 0.82 N^{-1/4} z_R, \tag{S52}$$

which is $E[\hat{s}_{\text{direct}}]_{s=0} = 0.1226 z_R$ for $N = 2000$. Thus this approximated result is very accurate. It can be noticed that this bias term is proportional to $N^{-1/4}$, which scales rather slowly with N and a sufficiently large $N \approx 10^5$ is needed to reduce the bias to $0.05 z_R$.

B. Sorter-based measurement

For the binary mode sorter used in our experiment, the probability distribution is a binomial distribution which can be described as [see Eq. (8) in the manuscript]

$$P_0(s) = \frac{1}{2} + \frac{4z_R^2}{8z_R^2 + s^2}, \quad P_1(s) = \frac{1}{2} - \frac{4z_R^2}{8z_R^2 + s^2}. \tag{S53}$$

The sum of N independent binomial distribution becomes another binomial distribution and the probability distribution is

$$P(k; N, p) = \binom{N}{k} p^k (1-p)^{N-k}, \tag{S54}$$

where $p = P_0(s)$. The estimator used for sorter-based measurement can be written as

$$\hat{Q} = \frac{N-k}{N}, \quad \hat{s}_{\text{binary}} = 2z_R \sqrt{\frac{2}{1-2\hat{Q}} - 2}. \tag{S55}$$

In the limit of $s \rightarrow 0$, we have $p \approx 1 - s^2/16z_R^2$. Therefore

$$\begin{aligned}
E[\hat{s}_{\text{binary}}] &= \sum_{k=0}^N 2z_R P(k; N, p) \sqrt{\frac{2}{\frac{2k}{N} - 1} - 2} \\
&\approx \sum_{k=0}^N 2z_R \frac{N!}{k!(N-k)!} \left(1 - \frac{s^2}{16z_R^2}\right)^k \left(\frac{s^2}{16z_R^2}\right)^{N-k} \sqrt{\frac{2}{\frac{2k}{N} - 1} - 2} \\
&= \sum_{k=0}^N 2z_R \frac{N!}{k!(N-k)!} \left(1 - \frac{s^2}{16z_R^2}\right)^{N-k} \left(\frac{s^2}{16z_R^2}\right)^k \sqrt{\frac{2}{\frac{2(N-k)}{N} - 1} - 2}.
\end{aligned} \tag{S56}$$

In the last step we make the substitution $k \rightarrow N-k$. When s is small, we can discard all higher-order terms and only keep the term with a small k . We keep the terms with $k = 0, 1, 2$ and the above equation becomes

$$\begin{aligned}
E[\hat{s}_{\text{binary}}] &\approx 0 + 4z_R \sqrt{N} \left(1 - \frac{s^2}{16z_R^2}\right)^{N-1} \left(\frac{s^2}{16z_R^2}\right) + 2z_R \sqrt{2N(N-1)} \left(1 - \frac{s^2}{16z_R^2}\right)^{N-2} \left(\frac{s^2}{16z_R^2}\right)^2 + \dots \\
&\approx \frac{\sqrt{N} s^2}{4z_R} + 2\sqrt{2N} z_R (N-1) \left(\frac{s^2}{16z_R^2}\right)^2 + \dots
\end{aligned} \tag{S57}$$

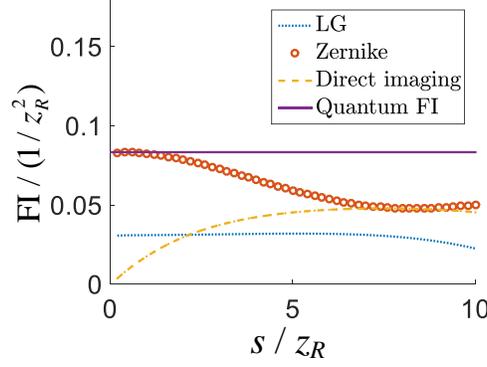


Fig. S4. Fisher information of different measurements for an Airy-disk-shaped PSF.

If we assume that $Ns^2/16z_R^2 \ll 1$, then all higher-order terms can be neglected and the result can be simplified to

$$E[\hat{s}_{\text{binary}}] = \frac{\sqrt{N}s^2}{4z_R}. \quad (\text{S58})$$

At the first sight this result seems problematic as the expectation value increases with N . This is because we make a strong assumption $Ns^2/16z_R^2 \ll 1$. Therefore when N is sufficiently large, the expectation value will not increase because other terms need to be taken into account. In addition, the point we want to make with this result is that $E[\hat{s}_{\text{binary}}]|_{s=0} = 0$ as well as $\partial E[\hat{s}_{\text{binary}}]/\partial s|_{s=0} = 0$. However, unlike the direct imaging, the Fisher information of sorter-based measurement is $I_{\text{sorter}}(s) = 4/(s^2 + 16z_R^2)$ and is nonzero at $s = 0$ as $I_{\text{sorter}}(0) = 1/4z_R^2$. Therefore we have

$$\text{Var}(\hat{s}_{\text{sorter}})|_{s=0} \geq \frac{(\partial E[\hat{s}_{\text{sorter}}]/\partial s)^2}{N \cdot I_{\text{sorter}}(s)} \Big|_{s=0} = 0. \quad (\text{S59})$$

Both the expectation and variance agrees with the Monte Carlo simulation.

6. FISHER INFORMATION CALCULATION FOR AN AIRY-DISK-SHAPED PSF

In our manuscript we mainly discuss the Gaussian PSF model which is valid only in the paraxial regime. In this section, we calculate the Fisher information for an Airy-disk-shaped PSF. The main conclusion of this section is that: (1) For an Airy-disk-shaped PSF, the Zernike mode sorter provides the optimal estimation for the axial separation of two point sources. (2) Although the Laguerre-Gaussian mode sorter is sub-optimal, it can still provide persistent nonzero Fisher information and thus outperform the direct imaging in the near-zero separation regime.

In the pupil plane, the normalized pupil function for an Airy-disk-shaped PSF can be written as

$$\psi_H(r_p; z) = \frac{1}{\sqrt{\pi}(f_1 \text{NA})^2} \text{circ}\left(\frac{r_p}{f_1 \text{NA}}\right) \exp\left(-ikzr_p^2/2f_1^2\right), \quad (\text{S60})$$

where r_p is the radial coordinate in the pupil plane, f_1 is the objective focal length, and $\text{circ}(x) = 1$ for $x < 1$ and $\text{circ}(x) = 0$ for $x \geq 1$. The density matrix for two point sources can be expressed as $\rho_H = (|\psi_H^1\rangle\langle\psi_H^1| + |\psi_H^2\rangle\langle\psi_H^2|)/2$, where $\langle r_p | \psi_H^1 \rangle = \psi_H(r_p; s/2)$ and $\langle r_p | \psi_H^2 \rangle = \psi_H(r_p; -s/2)$. The quantum Fisher information (QFI) can be directly calculated as [S1]

$$\mathcal{K}_H = 4(\langle(\partial_s \Psi)^2\rangle - \langle\partial_s \Psi\rangle^2) = \frac{\pi^2 \text{NA}^4}{12\lambda^2}, \quad (\text{S61})$$

where $\Psi(r_p; s) = -ksr_p^2/4f_1^2$ is the phase gradient of $\psi_H(r_p; s/2)$ and the angular bracket denotes $\langle\Phi\rangle = \int_0^\infty \Phi(r_p) |\psi_H(r_p; z)|^2 2\pi r_p dr_p$. It is clear that the QFI is a constant as shown in Fig. S4.

The Fisher information calculation for Zernike sorter is similar to that of the radial mode sorter. Here we use the well-known Zernike polynomials as $\langle r_p | Z_n^m \rangle = Z_n^m(r_p/f_1 \text{NA})$. The Zernike mode is normalized in such a way that $\int_0^{f_1 \text{NA}} |\langle r_p | Z_n^m \rangle|^2 2\pi r_p dr_p = 1$. Since the azimuthal part of the Zernike basis is ignored, we have $m = 0$ and n has to be an even number following the definition of Zernike polynomials. The Fisher information can thus be computed as

$$\mathcal{J}_{\text{Zernike}}(s) = \sum_{p=0}^{p_{\text{max}}} \left| \partial_s \langle Z_{2p}^0 | \rho_H | Z_{2p}^0 \rangle \right|^2 / \langle Z_{2p}^0 | \rho_H | Z_{2p}^0 \rangle, \quad (\text{S62})$$

where $p_{\text{max}} + 1$ is the number of Zernike modes used for superresolution. In our simulation we choose $p_{\text{max}} = 1$, which means that we only use two modes of lowest orders, which should be reasonably achievable in an experiment. We plot $\mathcal{J}_{\text{Zernike}}(s)$ in Fig. S4 as orange circles. It can be seen that the Zernike mode sorter can reach the QFI at a near-zero separation s .

As a comparison, we also calculate the Fisher information for direct imaging measurement. The intensity distribution in the image plane related to the pupil function by a Fourier transform as

$$I(r, s) = \frac{1}{2} |\mathcal{F}[\psi_H(r_p, s/2)](r, s/2)|^2 + \frac{1}{2} |\mathcal{F}[\psi_H(r_p, -s/2)](r, -s/2)|^2, \quad (\text{S63})$$

where $\mathcal{F}[\cdot]$ denotes the Fourier transform with coordinate transformation from r_p (in the pupil plane) to r (in the image plane). Unlike the case of the Gaussian PSF, the Fourier transform of the hard-edged pupil function does not have an analytic form, so the above equation has to be computed numerically. Then the Fisher information for direct imaging measurement can be written as

$$\mathcal{J}_{\text{Direct}}(s) = \int_0^\infty \frac{1}{I(r, s)} \left(\frac{\partial I(r, s)}{\partial s} \right)^2 2\pi r dr, \quad (\text{S64})$$

and the result is plotted as the yellow dashed line in Fig. S4. It is clear that the Fisher information drops to zero for a near-zero separation s .

As the next step, we calculate the Fisher information for the use of Laguerre-Gaussian (LG) mode sorter for an Airy-disk-shaped PSF instead of a Gaussian PSF. Here we use the LG basis in the pupil plane as

$$\langle r_p | \text{LG}_p \rangle = \sqrt{\frac{2}{\pi}} \frac{1}{w_p} \exp\left(-\frac{r_p^2}{w_p^2}\right) L_p\left(\frac{2r_p^2}{w_p^2}\right), \quad (\text{S65})$$

where w_p is the beam waist radius. Here we use the value of $w_p = f_1 \text{NA} / 1.121$ which is obtained by numerically maximizing the inner product $\langle \text{LG}_0 | Z_0^0 \rangle$. The intuition behind is to make $|\text{LG}_0\rangle$ be as close as possible to $|Z_0^0\rangle$. Then the Fisher information can be calculated as

$$\mathcal{J}_{\text{LG}}(s) = \sum_{p=0}^{p_{\max}} |\partial_s \langle \text{LG}_p | \rho_H | \text{LG}_p \rangle|^2 / \langle \text{LG}_p | \rho_H | \text{LG}_p \rangle. \quad (\text{S66})$$

The result is shown as the blue dotted line in Fig. S4. It can be seen that although LG mode sorter has a lower Fisher information than the optimal Zernike mode sorter, its value at near-zero separation persists and does not drop to zero, outperforming the direct imaging.

REFERENCES

- [S1] Z. Yu and S. Prasad, "Quantum limited superresolution of an incoherent source pair in three dimensions," *Phys. Rev. Lett.* **121**, 180504 (2018).
- [S2] J. Yang, S. Pang, Y. Zhou, and A. N. Jordan, "Optimal measurements for quantum multi-parameter estimation with general states," arXiv:1806.07337 (2018).
- [S3] S. M. Kay, *Fundamentals of statistical signal processing, Volume I: Estimation theory (1st ed.)* (Prentice-Hall, Englewood Cliffs, 1993).
- [S4] V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum-enhanced measurements: beating the standard quantum limit," *Science* **306**, 1330–1336 (2004).