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

Yiyu Zhou,[1,*] Jing Yang,[2] Jeremy D. Hassett,[1] Seyed Mohammad Hashemi Rafsanjani,[3] (iD)
Mohammad Mirhosseini,[4] A. Nick Vamivakas,[1,2,5] Andrew N. Jordan,[2,6]
Zhimin Shi,[7,9] and Robert W. Boyd[1,2,8,10]

[1]The Institute of Optics, University of Rochester, Rochester, New York 14627, USA
[2]Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627, USA
[3]Department of Physics, University of Miami, Coral Gables, Florida 33146, USA
[4]Thomas J. Watson, Sr., Laboratory of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA
[5]Materials Science Program, University of Rochester, Rochester, New York 14627, USA
[6]Institute for Quantum Studies, Chapman University, Orange, California 92866, USA
[7]Department of Physics, University of South Florida, Tampa, Florida 33620, USA
[8]Department of Physics, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada
[9]e-mail: zhiminshi@usf.edu
[10]e-mail: boyd@optics.rochester.edu
*Corresponding author: yzhou62@ur.rochester.edu

Improving axial resolution is crucial for three-dimensional optical imaging systems. Here we present a scheme of axial superresolution for two incoherent point sources based on spatial mode demultiplexing. A radial mode sorter is used to losslessly decompose the optical fields into a radial mode basis set to extract the phase information associated with the axial positions of the point sources. We show theoretically and experimentally that, in the limit of a zero axial separation, our scheme allows for reaching the quantum Cramér–Rao lower bound and thus can be considered as one of the optimal measurement methods. Unlike other superresolution schemes, this scheme does not require either activation of fluorophores or sophisticated stabilization control. Moreover, it is applicable to the localization of a single point source in the axial direction. Our demonstration can be useful for a variety of applications such as far-field fluorescence microscopy.   © 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

https://doi.org/10.1364/OPTICA.6.000534

1. INTRODUCTION

Optical microscopy is one of the most important imaging modalities and has been broadly applied in various areas. One crucial metric for an optical microscope is the spatial resolution, which is typically constrained by the diffraction limit, and the Rayleigh criterion is proposed as the resolution limit of an incoherent imaging system [1–3]. In recent decades, various methods have been proposed to surpass the diffraction limit. In fluorescence microscopy, a widely used approach is to activate each fluorescence molecule individually, and therefore the overlap between neighboring molecules is avoided, and the localization precision can be improved to tens of nanometers [4–6]. This technique usually requires specially prepared samples, and the reconstruction of an image can take a long time due to the sophisticated activation and detection of individual fluorophores. Another superresolution technique is based on decomposing the optical field into the linear prolate spheroidal functions, i.e., the eigenfunctions of aperture in a coherent imaging system [7–9]. It is shown that the ultimate limit of resolution of a coherent imaging system is not determined

by diffraction but by the signal-to-noise ratio of the measurement. Therefore, a sufficiently large number of photons is needed to enable superresolution. In addition, this technique, including other approaches that require nonclassical light sources [10–15], cannot be readily applied to incoherent superresolution imaging considered here. While many other methods have been proposed to realize axial super-localization [16], such as an interferometric microscope [17–20], point-spread function (PSF) engineering [21–26], and multi-plane detection [27–30], these advances can be used to precisely measure the axial location of only a single point source, and it remains a challenge to determine a small axial separation when two incoherent, simultaneously emitting point sources overlap with each other.

To develop an efficient axial superresolution technique, we follow the procedure in Ref. [31] and formulate the estimation of axial separation in the context of quantum metrology [32–36]. The precision of a measurement method is typically quantified by the Fisher information, and the reciprocal of Fisher information is referred to as the Cramér–Rao lower bound (CRLB) and

characterizes the lower bound of measurement variance for an unbiased estimator [37,38]. To determine the axial location of point sources, the easiest and most commonly used approach, which we refer to as the direct imaging method, is to measure the size of the PSF in the image plane and then deduce the axial positions accordingly. However, our calculation in the next section shows that the corresponding Fisher information drops to zero when the axial separation of two incoherent point sources gets close to zero. This result is not surprising because the size of PSF changes slowly when point sources are almost on focus. Nonetheless, a further calculation shows that the quantum Fisher information does not vanish for an arbitrarily small axial separation. The quantum Fisher information is the upper limit of the Fisher information that cannot be exceeded by any possible types of measurement as derived in the quantum metrology theory and can be used to quantify the maximum possible amount of information that can be obtained by a measurement. Given the non-vanishing quantum Fisher information, there should exist a type of measurement that can outperform the direct imaging method and extract the maximum possible amount of information from each photon.

In the following sections, we demonstrate both theoretically and experimentally that the axial superresolution can be achieved at the single-photon level by a radial mode sorter. This radial mode sorter can losslessly project the incident photons into the radial Laguerre–Gaussian basis set. With the same amount of photons, our scheme based on the radial mode sorter can estimate the axial separation with smaller bias and standard deviation. We note that similar strategies have been studied for transverse superresolution [39–48], which are based on a Hermite–Gaussian mode sorter [40,49] or mode parity decomposition [41,42]. However, we emphasize that our radial mode sorter comes from very recent advances in spatial mode sorters [50–52] and cannot be simply realized by mode parity decomposition. Furthermore, while homodyne or heterodyne detection [12,39] provides an easier way to implement the spatial mode projective measurement, the shot noise from the reference beam fundamentally deteriorates the signal-to-noise ratio, and it has been shown that homodyne or heterodyne detection cannot even outperform the direct imaging method when the available photon number is small [53]. Moreover, the mode sorter does not require any active components, such as the local oscillator in heterodyne detection, and thus is more favorable in an experiment.

2. THEORY

The conceptual diagram for the direct imaging method and sorter-based measurement is shown in Fig. 1. The direct imaging method employs an objective to collect photons and then use a tube lens to form an image of the object as shown in Fig. 1(a). Alternatively, one can detect the optical field in a complete and orthonormal basis set as shown in Fig. 1(b), which can be realized by a spatial mode sorter and is referred to as sorted-based measurement. In the following derivations, we use the Dirac notation to represent the fields and assume a coherent state for each point source. While a semiclassical treatment is sufficient to derive these formalisms inspired by quantum metrology [46], the Dirac notation is convenient to denote the mixed state of the incoherent sources and makes it straightforward to extend the theory to other types of light sources such as single-photon state [42] and thermal state [31]. For a more tractable analysis and experiment, here we assume a Gaussian PSF, and the field distribution at the pupil
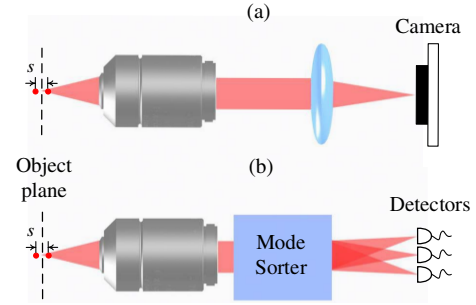

Fig. 1. Conceptual diagram for (a) direct imaging and (b) sorter-based measurement. A spatial mode sorter can direct different spatial mode components to different locations to perform spatial mode demultiplexing.

plane for an on-axis point source is denoted by $|\psi\rangle$, where $\langle r_0|\psi\rangle = \psi(r_0; z)$ and

$$\psi(r_0; z) = \sqrt{2/\pi} \exp(-r_0^2) \exp(-ikz\mathrm{NA}^2 r_0^2/2), \qquad (1)$$

where NA is the numerical aperture, z is the axial position of the point source, $k = 2\pi/\lambda$ is the wavenumber, λ is the wavelength, r_0 is the normalized radial coordinate in the pupil plane, and $|r_0\rangle$ is the corresponding radial eigenstate. Here we define $r_0 = r_p/(f_1\mathrm{NA})$, where r_p is the radial coordinate in the pupil plane and f_1 is the objective focal length. This pupil plane field distribution can be viewed as a paraxial, Gaussian approximation to the pupil function of a hard-edged circular aperture [54]. For direct imaging, a tube lens is used to perform a Fourier transform to the pupil function, and the intensity distribution on the image plane becomes

$$I(r; z) = \frac{2}{\pi} \frac{1}{w^2(z)} \exp\left(-\frac{2r^2}{w^2(z)}\right),$$
$$w(z) = \frac{M\lambda}{\pi\mathrm{NA}} \sqrt{1 + (z\pi\mathrm{NA}^2/\lambda)^2}, \qquad (2)$$

where r is the radial coordinate in the image plane, $w(z)$ denotes the Gaussian beam waist width on the image plane, and M is the magnification of the imaging system. By measuring the beam size, we can estimate the axial position z. Similar to the case of transverse superresolution [31], here we assume *a priori* knowledge of two on-axis, equally bright incoherent point sources with the centroid located at $z = 0$ plane, and the axial separation between them is s. The density matrix of these two point sources at the pupil plane can be written as $\rho = (|\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2|)/2$, 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 normalized total intensity at the image plane can be calculated as $I_s(r) \equiv \langle r|\rho|r\rangle = [I(r; s/2) + I(r; -s/2)]/2$, where $|r\rangle$ is the radial eigenstate in the image plane, and the image plane is related to the pupil plane by the Fourier transform. For sorter-based measurement, the incident field is decomposed to an orthonormal basis set, and here we consider the radial Laguerre–Gaussian (LG) basis because we notice that the axial position affects only the radial profile of pupil function. The radial LG basis in the pupil plane can be denoted as $|\mathrm{LG}_p\rangle$, where $\langle r_0|\mathrm{LG}_p\rangle = \mathrm{LG}_p(r_0)$ and

$$\mathrm{LG}_p(r_0) = \sqrt{2/\pi} \exp(-r_0^2) L_p(2r_0^2), \qquad (3)$$

where $L_p(\cdot)$ is the Laguerre polynomial. While the two-dimensional LG basis involves another azimuthal index ℓ, this

radial subset with $\ell = 0$ can still form a complete basis to describe the pupil function because of the rotational symmetry of the pupil function as shown in Eq. (1). Decomposing the pupil function $\psi(r_0; z)$ to this basis leads to the following radial mode distribution:

$$P(p; z) = |\langle\psi|LG_p\rangle|^2 = \frac{4z_R^2 z^{2p}}{(4z_R^2 + z^2)^{p+1}}, \quad (4)$$

where $z_R = \pi w_0^2/\lambda$ and $w_0 = \lambda/\pi\text{NA}$ [31]. For two equally bright sources separated by s, the output radial mode distribution becomes $P_s(p) \equiv \langle LG_p|\rho|LG_p\rangle = [P(p; s/2) + P(p; -s/2)]/2$. It can be noticed that for direct imaging and sorter-based measurement, the two incoherent point sources have the same response because Eqs. (2) and (4) are even functions of z, which suggests that the analysis presented here can also be applied to single-point localization.

We next compare the performance of direct imaging and sorter-based measurement by calculating the Fisher information for both techniques. The Fisher information for direct imaging is [31]

$$\mathcal{J}_{\text{direct}}(s) = \int_0^{2\pi} d\phi \int_0^{+\infty} \frac{1}{I_s(r)} \left(\frac{\partial I_s(r)}{\partial s}\right)^2 r dr = \frac{4s^2}{(s^2 + 4z_R^2)^2}, \quad (5)$$

which is independent of the magnification M. The Fisher information for the sorter-based measurement is

$$\mathcal{J}_{\text{sorter}}(s) = \sum_{p=0}^{\infty} \frac{1}{P_s(p)} \left(\frac{\partial P_s(p)}{\partial s}\right)^2 = \frac{4}{s^2 + 16z_R^2}. \quad (6)$$

The quantum Fisher information, i.e., the upper bound of Fisher information of any possible measurements, can be calculated as [55]

$$\mathcal{K}_s = 4[\langle\partial_s\psi_1|\partial_s\psi_1\rangle - |\langle\psi_1|\partial_s\psi_1\rangle|^2], \quad (7)$$

where $|\partial_s\psi_1\rangle = \partial|\psi_1\rangle/\partial s$, and it can be readily shown that $\mathcal{K}_s = 1/4z_R^2$. We also follow the usual way of using the symmetric logarithmic derivative to calculate the quantum Fisher information, and the details are presented in Supplement 1, which give the same result. The reciprocal of quantum Fisher information is the quantum CRLB, which gives the lower bound of classical CRLB for any possible measurements. We notice that the sorter-based measurement can reach the quantum Fisher information when the separation goes to zero, i.e., $\mathcal{J}_{\text{sorter}}(0) = \mathcal{K}_s$ [see Eq. (6)]; therefore, it can be considered to be an optimal measurement for s close to zero. However, in a realistic experiment, a mode sorter can access only a finite-dimensional Hilbert space. Therefore we follow the procedure in Ref. [56] to construct other possible optimal measurements that can reach the quantum Fisher information in the limit of $s = 0$. In Supplement 1, we show that a binary radial mode sorter is sufficient to access the quantum Fisher information. A binary sorter has two output ports, and all odd-order radial modes are directed to one output port, while all even-order modes are directed to another output port. Therefore, the photon probability distribution at two output ports is

$$P_s^0(s) = \sum_{p=0}^{\infty} P_s(2p; s) = \frac{1}{2} + \frac{4z_R^2}{8z_R^2 + s^2},$$

$$P_s^1(s) = \sum_{p=0}^{\infty} P_s(2p + 1; s) = \frac{1}{2} - \frac{4z_R^2}{8z_R^2 + s^2}. \quad (8)$$

Thus, the Fisher information for a binary sorter is

$$\mathcal{J}_{\text{binary}}(s) = \sum_{q=0}^{1} \frac{1}{P_s^q(s)} \left(\frac{\partial P_s^q(s)}{\partial s}\right)^2 = \frac{256z_R^4}{(s^2 + 8z_R^2)^2(s^2 + 16z_R^2)}. \quad (9)$$

The plot of Fisher information for different methods is shown in Fig. 2(a). It can be readily seen that the Fisher information of direct imaging begins to drop when s is smaller than $2z_R$. In incoherent imaging microscopy, the axial resolution can be expressed as $\Delta z = 2\lambda/\text{NA}^2$ [57], which can be rewritten as $\Delta z = 2\pi z_R$ with our notation. We note that the discrepancy between $2z_R$ and $2\pi z_R$ comes from our assumption of a Gaussian PSF rather than an Airy disk. However, it can be noticed that the sorter-based measurement stays nonzero and achieve the quantum Fisher information when s approaches zero, which makes it possible to break the diffraction limit. To further illustrate the improvement provided by the radial mode sorter, we calculate the Fisher information of astigmatic imaging [21], and the result is presented in Supplement 1. It is shown that while astigmatism can enhance three-dimensional localization precision of a single point source, it cannot be used directly to resolve the axial separation between two simultaneously emitting point sources without the help of photo-switchable fluorophores.

Having analyzed the performance of each method, now we need to establish the estimator of separation. For direct imaging, it can be verified that the maximum likelihood estimator is

$$\hat{w} = \sqrt{\frac{2}{N} \sum_{m=1}^{N} r_m^2}, \qquad \hat{s}_{\text{direct}} = 2z_R \sqrt{\frac{\hat{w}^2}{w_0^2} - 1}, \quad (10)$$

where r_m is the radial coordinate of m-th photon in the image plane, and N is the total detected photon number. The intuition behind this estimator is to measure the Gaussian width w by detecting the radial coordinates of photons and then use the estimated Gaussian width to calculate the separation s. However, the simple estimator of $\hat{w}$ does not take into account any experimental imperfections such as detector noise or pixelation, and thus may not be robust in a realistic experiment. Therefore, we apply
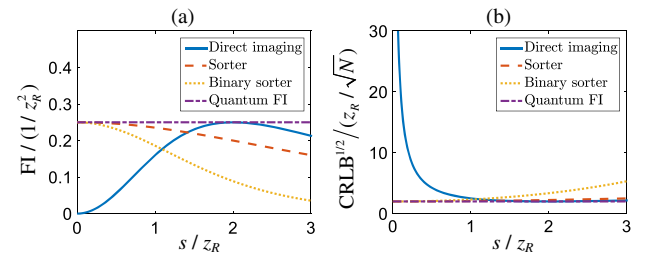

Fig. 2. (a) Fisher information (FI) as a function of axial separation for different methods. The sorter and binary sorter can reach the quantum Fisher information for small separation s, while the Fisher information of direct imaging drops to zero. (b) Normalized square root of Cramér–Rao lower bound (CRLB) for different methods. N is the detected photon number.

the algorithm in Ref. [58] to realize a robust, efficient Gaussian width estimator $\hat{w}$ in our experiment. For binary sorter-based measurement, the maximum likelihood estimator is

$$\hat{Q} = \frac{1}{N}\sum_{q=0}^{1} q \cdot m_q = \frac{m_1}{N}, \qquad \hat{s}_{\text{binary}} = 2z_R\sqrt{\frac{2}{1-2\hat{Q}}-2}, \tag{11}$$

where m_0 and m_1 are the photon numbers in the two output ports, and $m_0 + m_1 = N$ is the total detected photon number. The intuition behind this estimator is to use the photon probability distribution at the output ports of the sorter to estimate the separation. The lower bound of the variance of an estimator for N independent measurements is given by [38]

$$\text{Var}(\hat{s}) \geq \frac{(\partial E[\hat{s}]/\partial s)^2}{N \cdot \mathcal{J}(s)}, \tag{12}$$

where the right-hand side is referred to as the CRLB, and N is the photon number in the context of our experiment given the Poisson statistics. This formula for CRLB is also applicable to other classical photon states such as single-photon state [42] and thermal state [31], and the variance that scales as N^{-1} is referred to as the standard quantum limit [33,59]. For an unbiased estimator whose expectation is equal to the value of the estimated parameter, i.e., $E[\hat{s}] = s$, this CRLB reduces to a simpler form as $\text{Var}(\hat{s}) \geq 1/[N \cdot \mathcal{J}(s)]$, which is just the reciprocal of Fisher information, as we plot in Fig. 2(b).

3. EXPERIMENT

A schematic for the experimental setup is shown in Fig. 3. We use an attenuated laser source to illuminate the spatial light modulator (SLM) to generate the Gaussian pupil function produced by a point source. An acousto-optic modulator (AOM) is driven by a signal generator to produce 3 μs pulses, and the driving signal is also connected to an intensified charge coupled device (ICCD, PI-Max 4 1024i) for synchronization. The average detected photon number in each pulse is around 2000. We use the calibration factor provided by the manufacturer to calculate the photon number in each pixel of the camera. We emulate two incoherent sources by mixing the data for $z = \pm s/2$, which is generated by SLM separately. A computer-generated hologram is displayed on SLM 1 to generate the desired field at the first diffraction order [60]. Each time, the SLM displays the corresponding hologram to generate either $\psi(r_0; s/2)$ or $\psi(r_0; -s/2)$ to simulate a point source located at $z = s/2$ or $z = -s/2$, respectively. Since both holograms are never present at the same time, there is no

coherence between the two simulated point sources. By using a long exposure time of the camera to incoherently mix the data, we effectively generate two incoherent, simultaneously emitting point sources [61]. For the Gaussian pupil function, we use the parameters of NA = 0.1 and $f_1 = 4$ mm. The calibration data of SLM 1 are presented in Supplement 1.

To construct a binary radial mode sorter, we use two polarization-sensitive SLMs (Hamamatsu X10468-02) as shown in the schematic [50–52]. In our experiment, we use two different areas on a single SLM to act as two SLMs for reduced experimental complexity. Due to the polarization sensitivity of the SLM, this binary mode sorter is designed to work for diagonally polarized light and cannot be directly used for arbitrary polarization [52]; therefore, we use a polarizer before the SLM to filter out undesired polarization. We note that the polarization of photons is not relevant to the theory of superresolution, and thus the use of a polarization-sensitive sorter is permissible for this proof-of-principle experiment. To realize a polarization-independent sorter, one can use the previously reported interferometric scheme [50]. A quadratic phase pattern is imprinted on SLM 2 and SLM 3 as the essential ingredient of the sorter. This quadratic phase is identical to the phase of a spherical lens with a focal length of 46.5 cm, and the separation between two SLMs is 65.8 cm. Each SLM performs a fractional Fourier transform of order $\alpha = \pi/2$ to horizontally polarized light and $\alpha = \pi/4$ to vertically polarized light, respectively. One can check that even-order radial modes remain diagonally polarized and odd-order radial modes become anti-diagonally polarized after passing through both SLMs [50]. Through the use of a half-wave plate (HWP) and a polarizing beamsplitter (PBS), one can efficiently separate odd- and even-order radial modes to distinct output ports. More details about the principle of the radial mode sorter can be found in Refs. [50–52]. As mentioned earlier, this radial mode sorter cannot be realized by mode parity decomposition based on mirror reflection [41] or a 4-f system. Moreover, this radial mode sorter is in principle lossless, and the loss of our sorter comes mainly from the limited light utilization efficiency of the SLMs, which can be reduced by using other low-loss devices such as commercially available polarization directed flat lenses [51]. In our experiment, we direct the photons from different output ports to different areas of an ICCD. For direct imaging, we use a 10 cm tube lens to form the image on the ICCD detector plane. For each separation, we repeat the experiment 400 times and calculate the expectation and standard deviation from the collected data based on the maximum likelihood estimators.
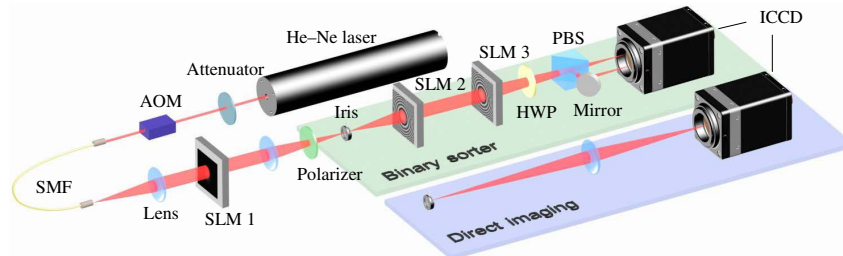

Fig. 3. Schematic of the experimental setup. A 633 nm He–Ne laser is attenuated and modulated by an acousto-optic modulator (AOM) to generate weak pulses. A computer-generated hologram is imprinted onto spatial light modulator (SLM 1) to generate the desired pupil function to simulate point sources. Two different methods, the binary sorter-based measurement and the direct imaging method, are used to estimate the separation s. In our experiment, we use a flip mirror to select the measurement method.

4. DISCUSSION

The measured separation and the standard deviation as a function of the real separation for different measurement methods are presented in Fig. 4. The average detected photon number for each measurement in our experiment is around 2000. As can be seen in Eq. (12), the loss of photons will decrease the detected photon number N and thus increase the variance of measurement and deteriorate the measurement precision. However, since all devices used in our experiment are essentially phase-only elements, the loss can always be reduced to zero by using appropriate anti-reflection coatings, and in our analysis, we assume a detection efficiency of unity. The Monte Carlo simulation results are provided as comparisons, and they agree well with the experimental data. In the simulation we set the detected photon number to be 2000, and the expectation and standard deviation of both estimators are retrieved by averaging 4000 simulations. We assume a noiseless detector with a sufficiently high spatial resolution in the simulation, and the estimators for direct imaging and binary sorter-based measurement are given by Eq. (10) and Eq. (11), respectively. One immediate observation in Fig. 4(a) is that the measured separation of direct imaging deviates from the real value when the real separation is close to zero. Another observation in Fig. 4(b) is that the measured standard deviation does not follow the CRLB and stays finite in the vicinity of $s = 0$. Neither observation is due to experimental imperfections, as they agree with the Monte Carlo simulation and should be attributed to the bias of the estimator [38,42]. The bias of an estimator is defined as the difference between the estimator's expectation value and the real value of the parameter being estimated. In Supplement 1, we provide a detailed, analytical calculation of the bias of $\hat{s}_{\text{direct}}$. The expectation value of $\hat{s}_{\text{direct}}$ at $s = 0$ can be well approximated as

$$E[\hat{s}_{\text{direct}}]|_{s=0} \approx 0.82 N^{-1/4} z_R, \tag{13}$$

which is $0.123 z_R$ for $N = 2000$ and very close to the Monte Carlo simulation $0.124 z_R$, as shown in Fig. 4(a). It can be noticed that

this bias is on the order of z_R when N is small, which qualitatively agrees with the conventional axial resolution of $2\pi z_R$. A large photon number N can lower the value of bias, which corresponds to the fact that a higher signal-to-noise ratio can enhance the resolution of direct imaging. A simple example is the deconvolution algorithm, which can be used to obtain subdiffraction resolution as long as a sufficiently high signal-to-noise ratio is available. However, the bias of direct imaging scales rather slowly with N as $N^{-1/4}$, and to reduce this bias, a sufficiently large N is needed. While a large photon number is attainable with a bright light source, in a photon-starving experiment such as fluorescence microscopy, it is usually not achievable. The slope of the estimator's expectation is calculated to be

$$\left.\frac{\partial E[\hat{s}_{\text{direct}}]}{\partial s}\right|_{s=0} \approx \frac{0.43 N^{1/4} s}{z_R}. \tag{14}$$

Together with Eq. (12), it immediately follows that the CRLB becomes $\text{Var}(\hat{s}_{\text{direct}})|_{s=0} \geq 0.74 z_R^2 / \sqrt{N}$. In contrast to the diverging CRLB solely predicted by the reciprocal of Fisher information, the CRLB calculated here takes into account the bias and explains the non-diverging standard deviation as shown in the experiment and Monte Carlo simulation. The scaled CRLB is calculated to be $\sqrt{\text{Var}(\hat{s}_{\text{direct}})}/(z_R/\sqrt{N}) \geq 5.8$, which is close to the standard deviation in the Monte Carlo simulation result 6.6 as shown in Fig. 4(b). It should be noted that Eq. (12) is an inequality instead of an equality, which causes the discrepancy between 5.8 and 6.6. Furthermore, $\partial E[\hat{s}_{\text{direct}}]/\partial s|_{s=0} = 0$ implies that the expectation value has a slope of zero when $s = 0$, as can be seen in Fig. 4(a). Hence, despite of a finite standard deviation, it is intrinsically unrealistic to use the measured s to recover the real s in the vicinity of $s = 0$ for the direct imaging method, and any attempt to construct an unbiased estimator will lead to a diverging standard deviation. Another observation is that the variance of the estimator for direct imaging scales as $\text{Var}(\hat{s}_{\text{direct}})|_{s=0} \propto N^{-1/2}$; therefore, this estimator cannot reach the standard quantum limit when s is small [33,59].

For sorter-based measurement, it can also be noticed that the standard deviation in both simulation and experiment deviates from the reciprocal of Fisher information and drops to zero when s is small. In Supplement 1, we show that $\partial E[\hat{s}_{\text{binary}}]/\partial s|_{s=0} = 0$ and thus $\text{Var}(\hat{s}_{\text{binary}})|_{s=0} \geq 0$, which explains the zero standard deviation that violates the reciprocal of Fisher information as shown in Fig. 4(d). It has been pointed out that this so-called superefficiency exists only on a set of points with zero measure, and the region of superefficiency reduces for more samples [31,62]. In addition, we have also shown in Supplement 1 that $E[\hat{s}_{\text{binary}}]|_{s=0} = 0$, which coincides with the Monte Carlo simulation and suggests that the sorter-based measurement can provide more precise, less biased measurement when s is small. However, we still observe a small, nonzero separation at $s = 0$ in our experiment, and the zero standard deviation is not visible either. We attribute this inconsistency to experimental imperfections, including dark noise of the detector and misalignment of the sorter. At the point of $s = 0$, all photons are supposed to be sorted to the output port of even-order radial modes, and no photons should be detected at the other output port. Nevertheless, when we experimentally characterize our sorter, we observe that 0.28% of detected photons are routed to the wrong output port on average when $s = 0$. In the data processing, we have subtracted this averaged crosstalk before estimating the separation, but the
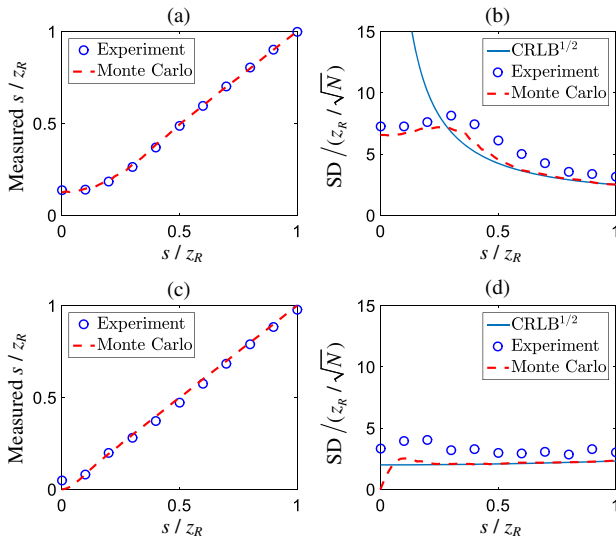

Fig. 4. (a) Measured separation and (b) standard deviation (SD) of s as a function of actual separation for direct imaging method. (c) Measured separation and (d) SD of s as a function of actual separation for binary sorter-based measurement. The Monte Carlo simulation results and the square root of corresponding CRLB are provided as comparisons. N is the detected photon number.

associated shot noise cannot be simply eliminated and thus leads to the experimental inconsistency, as we describe above. In Supplement 1, we quantify the effect of crosstalk on data processing, and the analytical calculation shows that $E[\hat{s}_{\text{binary}}]|_{s=0} = 0.043z_R$, which is very close to the measured value of $0.049z_R$. There are several ways to further mitigate the crosstalk, such as aligning the sorter more carefully, replacing the ICCD by low-noise single-pixel detectors, and developing a more robust estimator [58]. Despite these experimental imperfections, it is apparent that sorter-based measurement can outperform direct imaging, given the strong bias and higher standard deviation of direct imaging compared to that of sorter-based measurement. In addition, the advantage of sorter-based measurement is supposed to be more obvious with a larger photon number, because the variance of direct imaging scales as $N^{-1/2}$ rather than N^{-1}.

To further provide a quantitative description of the improvement of our experiment, we compare the two methods in terms of bias and variance. For direct imaging, if we want to reduce its bias to $0.049z_R$, which is obtained in sorter-based measurement, the photon number needs to be on the order of 10^5 to satisfy $0.82N^{-1/4}z_R = 0.049z_R$. Furthermore, in Fig. 4, it can be seen that the measured standard deviation of direct imaging is approximately twice that of sorter-based measurement. Remember that the variance of direct imaging scales as $N^{-1/2}$; thus, the standard deviation scales as $N^{-1/4}$, and 16 times more photons are needed to reduce the standard deviation of direct imaging to the level of sorter-based measurement. We note that here we are comparing the experimental data rather than noise-free theoretical predictions. The Monte Carlo simulation shows that the sorter-based measurement has zero bias and zero standard deviation at $s = 0$, and thus the direct imaging needs infinite photons to beat the sorter-based measurement.

In this work, we focus mainly on the superresolution of axial separation for two point sources, but we note that the theory presented above can be directly applied to the localization of axial position z of a single point source as long as the separation s is replaced by $s \rightarrow z/2$, which presents an alternative to sophisticated interferometric microscopy [17–20]. In contrast to the interferometric detection scheme, which requires nanometer-scale stabilization over a path length on the order of 1 m [16], the common-path radial mode sorter used in our experiment is robust, and no additional stabilization control is needed. We assume that the two point sources are on axis and their center position is known. In Supplement 1, we analyze the effect of misaligned centroid, i.e., centroid of point source pair $z_C \neq 0$. As shown in Fig. 5(a), while the Fisher information drops in the presence

of a misaligned centroid, the radial mode sorter can provide improved precision for a small separation with $|z_C| < 0.3z_R$. In a realistic scenario, an adaptive measurement can be used to estimate both the centroid and separation, as discussed in Ref. [31]. However, unlike the case of transverse centroid estimation, the direct imaging does not provide sufficient Fisher information for measuring the axial centroid of a point source pair. Here we notice that astigmatic imaging [21] presents an effective method to overcome this difficulty. The analysis is included in Supplement 1, and the Fisher information of centroid estimation for astigmatic imaging is shown in Fig. 5(b). It can be seen that astigmatic imaging provides appreciable Fisher information over a broad range of centroid and separation. Hence, a hybrid measurement consisting of both the radial mode sorter and astigmatic imaging can be a practical scheme for axial superresolution. Another assumption we make is that a Gaussian PSF is used for more tractable analysis and experiment. While the Gaussian PSF is a widely adopted approximation [57,63], for a high-NA imaging system, a more accurate PSF model may be needed [64]. In this case, one can always establish a complete and orthonormal basis based on the PSF model and construct a sorter accordingly to achieve superresolution accordingly [46]. Very recently, it has been pointed out theoretically that for the pupil function of a hard-edged aperture, the optimal measurement basis turns out to be the Zernike basis [55,56], and we discuss other optimal measurements that are easier to implement in Supplement 1. Therefore, based on our result, it can be anticipated that three-dimensional superresolution can be realized as long as a Zernike mode sorter is available. In Supplement 1, we calculate the Fisher information of various measurements for an Airy-disk-shaped PSF, and the result is shown in Fig. 6. It can be seen that while the Zernike mode sorter provides optimal measurement [55], the LG mode sorter as a sub-optimal measurement can still provide nonzero Fisher information at near-zero separation, outperforming the direct imaging measurement. Here we take into account only the zeroth- and first-order radial modes in calculating the Fisher information of the LG mode sorter and Zernike mode sorter, which should be reasonably achievable in an experiment. Moreover, given the widely used Gaussian-to-tophat laser beam shaper [65,66], it is possible to convert the pupil function of a hard-edged aperture to a Gaussian and then apply the radial mode sorter subsequently. Recent advances in multi-plane light conversion [67] also provide an alternative possible solution for building a Zernike mode sorter. Finally, despite the classical light source used in our experiment, our method can be used for other light sources such as single-photon emitters [31,42], because the quantum state of
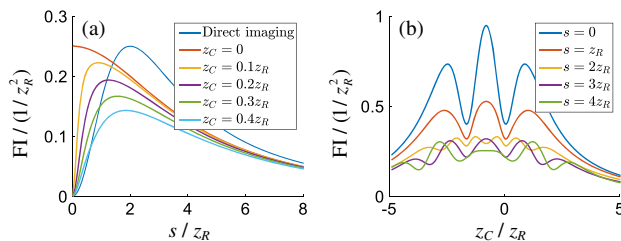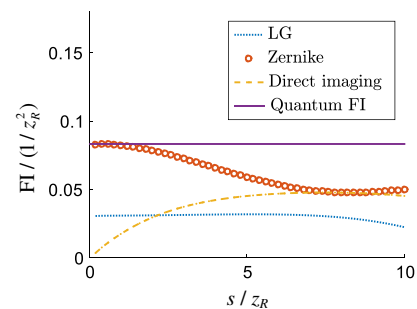

Fig. 5. (a) Fisher information of separation estimation for sorter-based measurement with different centroid positions. The Fisher information for direct imaging with point source pair centroid $z_C = 0$ is plotted as a reference. (b) Fisher information of centroid estimation for astigmatic imaging with different separations.

Fig. 6. Fisher information of different measurements for an Airy-disk-shaped PSF model.

photons represents the temporal coherence of light and is generally independent of the spatial degree of freedom. Therefore, it is possible to combine the radial mode sorter and intensity correlation measurement to further increase the resolution for single-photon sources [68].

In conclusion, we theoretically and experimentally demonstrate the axial superresolution based on a radial mode sorter. The binary radial mode sorter employed in out experiment can reach the quantum CRLB for arbitrarily small axial separation. Our method makes three-dimensional superresolution imaging promising and can be potentially useful for enhancing the resolution of optical microscopes.

See Supplement 1 for supporting content.

REFERENCES

1. Rayleigh, "Xxxi. investigations in optics, with special reference to the spectroscope," London Edinburgh Dublin Philos. Mag. J. Sci. **8**(49), 261–274 (1879).
2. J. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, 2008).
3. M. Born and E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light* (Elsevier, 2013).
4. E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging intracellular fluorescent proteins at nanometer resolution," Science **313**, 1642–1645 (2006).
5. M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)," Nat. Methods **3**, 793–796 (2006).
6. S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," Opt. Lett. **19**, 780–782 (1994).
7. V. N. Beskrovnyy and M. I. Kolobov, "Quantum limits of super-resolution in reconstruction of optical objects," Phys. Rev. A **71**, 043802 (2005).
8. M. I. Kolobov, *Quantum Imaging* (Springer, 2007).
9. M. I. Kolobov and C. Fabre, "Quantum limits on optical resolution," Phys. Rev. Lett. **85**, 3789–3792 (2000).
10. L. A. Rozema, J. D. Bateman, D. H. Mahler, R. Okamoto, A. Feizpour, A. Hayat, and A. M. Steinberg, "Scalable spatial superresolution using entangled photons," Phys. Rev. Lett. **112**, 223602 (2014).
11. N. Treps, U. Andersen, B. Buchler, P. K. Lam, A. Maitre, H.-A. Bachor, and C. Fabre, "Surpassing the standard quantum limit for optical imaging using nonclassical multimode light," Phys. Rev. Lett. **88**, 203601 (2002).
12. M. A. Taylor, J. Janousek, V. Daria, J. Knittel, B. Hage, H.-A. Bachor, and W. P. Bowen, "Subdiffraction-limited quantum imaging within a living cell," Phys. Rev. X **4**, 011017 (2014).
13. M. A. Taylor, J. Janousek, V. Daria, J. Knittel, B. Hage, H.-A. Bachor, and W. P. Bowen, "Biological measurement beyond the quantum limit," Nat. Photonics **7**, 229–233 (2013).
14. H. Shin, K. W. C. Chan, H. J. Chang, and R. W. Boyd, "Quantum spatial superresolution by optical centroid measurements," Phys. Rev. Lett. **107**, 083603 (2011).
15. O. Schwartz, J. M. Levitt, R. Tenne, S. Itzhakov, Z. Deutsch, and D. Oron, "Superresolution microscopy with quantum emitters," Nano Lett. **13**, 5832–5836 (2013).
16. A. von Diezmann, Y. Shechtman, and W. Moerner, "Three-dimensional localization of single molecules for super-resolution imaging and single-particle tracking," Chem. Rev. **117**, 7244–7275 (2017).
17. M. P. Backlund, Y. Shechtman, and R. L. Walsworth, "Fundamental precision bounds for three-dimensional optical localization microscopy with Poisson statistics," Phys. Rev. Lett. **121**, 023904 (2018).
18. G. Shtengel, J. A. Galbraith, C. G. Galbraith, J. Lippincott-Schwartz, J. M. Gillette, S. Manley, R. Sougrat, C. M. Waterman, P. Kanchanawong, M. W. Davidson, R. D. Fetter, and H. F. Hess, "Interferometric fluorescent super-resolution microscopy resolves 3d cellular ultrastructure," Proc. Natl. Acad. Sci. USA **106**, 3125–3130 (2009).
19. M. Schrader and S. Hell, "4pi-confocal images with axial superresolution," J. Microsc. **183**, 110–115 (1996).
20. J. Bewersdorf, R. Schmidt, and S. Hell, "Comparison of i5m and 4pi-microscopy," J. Microsc. **222**, 105–117 (2006).
21. B. Huang, W. Wang, M. Bates, and X. Zhuang, "Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy," Science **319**, 810–813 (2008).
22. S. R. P. Pavani, M. A. Thompson, J. S. Biteen, S. J. Lord, N. Liu, R. J. Twieg, R. Piestun, and W. Moerner, "Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function," Proc. Natl. Acad. Sci. USA **106**, 2995–2999 (2009).
23. S. Jia, J. C. Vaughan, and X. Zhuang, "Isotropic three-dimensional super-resolution imaging with a self-bending point spread function," Nat. Photonics **8**, 302–306 (2014).
24. M. Martínez-Corral, P. Andres, J. Ojeda-Castaneda, and G. Saavedra, "Tunable axial superresolution by annular binary filters. Application to confocal microscopy," Opt. Commun. **119**, 491–498 (1995).
25. T. R. Sales and G. M. Morris, "Fundamental limits of optical superresolution," Opt. Lett. **22**, 582–584 (1997).
26. F. Tamburini, G. Anzolin, G. Umbriaco, A. Bianchini, and C. Barbieri, "Overcoming the Rayleigh criterion limit with optical vortices," Phys. Rev. Lett. **97**, 163903 (2006).
27. M. F. Juette, T. J. Gould, M. D. Lessard, M. J. Mlodzianoski, B. S. Nagpure, B. T. Bennett, S. T. Hess, and J. Bewersdorf, "Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples," Nat. Methods **5**, 527–529 (2008).
28. E. Toprak, H. Balci, B. H. Blehm, and P. R. Selvin, "Three-dimensional particle tracking via bifocal imaging," Nano Lett. **7**, 2043–2045 (2007).
29. P. A. Dalgarno, H. I. Dalgarno, A. Putoud, R. Lambert, L. Paterson, D. C. Logan, D. P. Towers, R. J. Warburton, and A. H. Greenaway, "Multiplane imaging and three dimensional nanoscale particle tracking in biological microscopy," Opt. Express **18**, 877–884 (2010).
30. S. Abrahamsson, J. Chen, B. Hajj, S. Stallinga, A. Y. Katsov, J. Wisniewski, G. Mizuguchi, P. Soule, F. Mueller, C. D. Darzacq, X. Darzacq, C. Wu, C. I. Bargmann, D. A. Agard, M. Dahan, and M. G. L. Gustafsson, "Fast multicolor 3d imaging using aberration-corrected multifocus microscopy," Nat. Methods **10**, 60–63 (2012).
31. M. Tsang, R. Nair, and X.-M. Lu, "Quantum theory of superresolution for two incoherent optical point sources," Phys. Rev. X **6**, 031033 (2016).
32. C. W. Helstrom, "Quantum detection and estimation theory," J. Stat. Phys. **1**, 231–252 (1969).
33. V. Giovannetti, S. Lloyd, and L. Maccone, "Advances in quantum metrology," Nat. Photonics **5**, 222–229 (2011).
34. A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (Springer, 2011), vol. **1**.
35. C. L. Degen, F. Reinhard, and P. Cappellaro, "Quantum sensing," Rev. Mod. Phys. **89**, 035002 (2017).
36. M. G. Paris, "Quantum estimation for quantum technology," Int. J. Quantum. Inform. **7**, 125–137 (2009).
37. H. L. Van, *Trees, Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory* (Wiley, 2004).
38. S. M. Kay, *Fundamentals of Statistical Signal Processing*, 1st ed., Vol. I: Estimation theory (Prentice-Hall, 1993),.
39. F. Yang, A. Tashchilina, E. S. Moiseev, C. Simon, and A. I. Lvovsky, "Far-field linear optical superresolution via heterodyne detection in a higher-order local oscillator mode," Optica **3**, 1148–1152 (2016).
40. M. Paúr, B. Stoklasa, Z. Hradil, L. L. Sánchez-Soto, and J. Rehacek, "Achieving the ultimate optical resolution," Optica **3**, 1144–1147 (2016).
41. R. Nair and M. Tsang, "Far-field superresolution of thermal electromagnetic sources at the quantum limit," Phys. Rev. Lett. **117**, 190801 (2016).

42. W.-K. Tham, H. Ferretti, and A. M. Steinberg, "Beating Rayleigh's curse by imaging using phase information," Phys. Rev. Lett. **118**, 070801 (2017).

43. Z. S. Tang, K. Durak, and A. Ling, "Fault-tolerant and finite-error localization for point emitters within the diffraction limit," Opt. Express **24**, 22004–22012 (2016).

44. J. Řeháček, Z. Hradil, D. Koutný, J. Grover, A. Krzic, and L. L. Sánchez-Soto, "Optimal measurements for quantum spatial superresolution," Phys. Rev. A **98**, 012103 (2018).

45. J. Řeháček, Z. Hradil, B. Stoklasa, M. Paúr, J. Grover, A. Krzic, and L. L. Sánchez-Soto, "Multiparameter quantum metrology of incoherent point sources: towards realistic superresolution," Phys. Rev. A **96**, 062107 (2017).

46. M. Tsang, "Subdiffraction incoherent optical imaging via spatial-mode demultiplexing: semiclassical treatment," Phys. Rev. A **97**, 023830 (2018).

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

48. S. Zhou and L. Jiang, "Modern description of Rayleigh's criterion," Phys. Rev. A **99**, 013808 (2019).

49. Y. Zhou, J. Zhao, Z. Shi, S. M. H. Rafsanjani, M. Mirhosseini, Z. Zhu, A. E. Willner, and R. W. Boyd, "Hermite-Gaussian mode sorter," Opt. Lett. **43**, 5263–5266 (2018).

50. Y. Zhou, M. Mirhosseini, D. Fu, J. Zhao, S. M. Hashemi Rafsanjani, A. E. Willner, and R. W. Boyd, "Sorting photons by radial quantum number," Phys. Rev. Lett. **119**, 263602 (2017).

51. D. Fu, Y. Zhou, R. Qi, S. Oliver, Y. Wang, S. M. H. Rafsanjani, J. Zhao, M. Mirhosseini, Z. Shi, P. Zhang, and R. W. Boyd, "Realization of a scalable Laguerre-Gaussian mode sorter based on a robust radial mode sorter," Opt. Express **26**, 33057–33065 (2018).

52. Y. Zhou, M. Mirhosseini, S. Oliver, J. Zhao, S. M. H. Rafsanjani, M. P. J. Lavery, A. E. Willner, and R. W. Boyd, "Using all transverse degrees of freedom in quantum communications based on a generic mode sorter," Opt. Express **27**, 10383–10394(2019).

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

54. P. N. Petrov, Y. Shechtman, and W. Moerner, "Measurement-based estimation of global pupil functions in 3d localization microscopy," Opt. Express **25**, 7945–7959 (2017).

55. Z. Yu and S. Prasad, "Quantum limited superresolution of an incoherent source pair in three dimensions," Phys. Rev. Lett. **121**, 180504 (2018).

56. J. Yang, S. Pang, Y. Zhou, and A. N. Jordan, "Optimal measurements for quantum multi-parameter estimation with general states," arXiv:1806.07337 (2018).

57. J. B. Pawley, *Handbook of Biological Confocal Microscopy* (Springer, 2010).

58. C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke, "Fast, single-molecule localization that achieves theoretically minimum uncertainty," Nat. Methods **7**, 373–375 (2010).

59. V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum-enhanced measurements: beating the standard quantum limit," Science **306**, 1330–1336 (2004).

60. M. Mirhosseini, O. S. Magana-Loaiza, C. Chen, B. Rodenburg, M. Malik, and R. W. Boyd, "Rapid generation of light beams carrying orbital angular momentum," Opt. Express **21**, 30196–30203 (2013).

61. B. Rodenburg, M. Mirhosseini, O. S. Magaña-Loaiza, and R. W. Boyd, "Experimental generation of an optical field with arbitrary spatial coherence properties," J. Opt. Soc. Am. B **31**, A51–A55 (2014).

62. D. V. Hinkley and D. Cox, *Theoretical Statistics* (Chapman and Hall/CRC, 1979).

63. B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "Gaussian approximations of fluorescence microscope point-spread function models," Appl. Opt. **46**, 1819–1829 (2007).

64. S. F. Gibson and F. Lanni, "Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy," J. Opt. Soc. Am. A **9**, 154–166 (1992).

65. W. Cheng, W. Han, and Q. Zhan, "Compact flattop laser beam shaper using vectorial vortex," Appl. Opt. **52**, 4608–4612 (2013).

66. O. Homburg and T. Mitra, "Gaussian-to-top-hat beam shaping: an overview of parameters, methods, and applications," Proc. SPIE **8236**, 82360A (2012).

67. G. Labroille, B. Denolle, P. Jian, P. Genevaux, N. Treps, and J.-F. Morizur, "Efficient and mode selective spatial mode multiplexer based on multi-plane light conversion," Opt. Express **22**, 15599–15607 (2014).

68. R. Tenne, U. Rossman, B. Rephael, Y. Israel, A. Krupinski-Ptaszek, R. Lapkiewicz, Y. Silberberg, and D. Oron, "Super-resolution enhancement by quantum image scanning microscopy," Nat. Photonics **13**, 116–122 (2019).

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

YIYU ZHOU[1,*], **JING YANG**[2], **JEREMY D. HASSETT**[1], **SEYED MOHAMMAD HASHEMI RAFSANJANI**[3], **MOHAMMAD MIRHOSSEINI**[4], **A. NICK VAMIVAKAS**[1,2,5], **ANDREW N. JORDAN**[2,6], **ZHIMIN SHI**[7,†], **AND ROBERT W. BOYD**[1,2,8,‡]

[1] *The Institute of Optics, University of Rochester, Rochester, New York 14627, USA*
[2] *Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627, USA*
[3] *Department of Physics, University of Miami, Coral Gables, Florida 33146, USA*
[4] *Thomas J. Watson, Sr., Laboratory of Applied Physics, California Institute of Technology, Pasadena, California 91125, USA*
[5] *Materials Science Program, University of Rochester, Rochester, New York 14627, USA*
[6] *Institute for Quantum Studies, Chapman University, Orange, California 92866, USA*
[7] *Department of Physics, University of South Florida, Tampa, Florida 33620, USA*
[8] *Department 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*

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

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

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

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

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

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

Hence the density matrix can be rewritten as

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

Here we also define the following orthonormal bases

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

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

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

Here the symmetric logarithmic derivative of density matrix is

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

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

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

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

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

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

$$\left\langle \partial_s^0 \psi_1 \middle| \pi_{k\alpha} \right\rangle \big|_{s=0} = 0, \quad \forall k, \tag{S10}$$

where $|\partial_s^0 \psi_1\rangle\big|_{s=0} = (|\partial_s \psi_1\rangle - \langle \psi_1|\partial_s \psi_1\rangle |\psi_1\rangle)|_{s=0}$. 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 $\langle \psi_1|\pi_{k\alpha}\rangle\big|_{s=0} \neq 0$ and *null* basis vectors defined as $\langle \psi_1|\pi_{k\alpha}\rangle\big|_{s=0} = 0$. (ii) Calculate $\langle \partial_s^0 \psi_1|\pi_{k\alpha}\rangle\big|_{s=0}$ 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 $\langle \partial_s \psi_1|\pi_{k\alpha}\rangle\big|_{s=0} = 0$. 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 $\langle \psi_1|\Pi_k|\psi_1\rangle\big|_{s=0} = 0$. 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, $\psi_1(r_0) = \sqrt{2/\pi} \cdot \exp(-r_0^2) \cdot \exp\left(-iks\text{NA}^2 r_0^2/4\right)$. It is straightforward to find

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

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

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

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

(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

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

Alternatively one can add the flexible projectors to Π_0' or Π_1' to give rise to optimal measurements of

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

or

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

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 $\psi_1(r_0) = \sqrt{2/\pi} \cdot \text{circ}(\sqrt{2}r_0) \exp\left(-iks\text{NA}^2 r_0^2/4\right)$, 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

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

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

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:

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

or

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

or

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

Any sorter that can efficiently perform the above measurements can be used to reach the quantum Fisher information when s approaches 0.
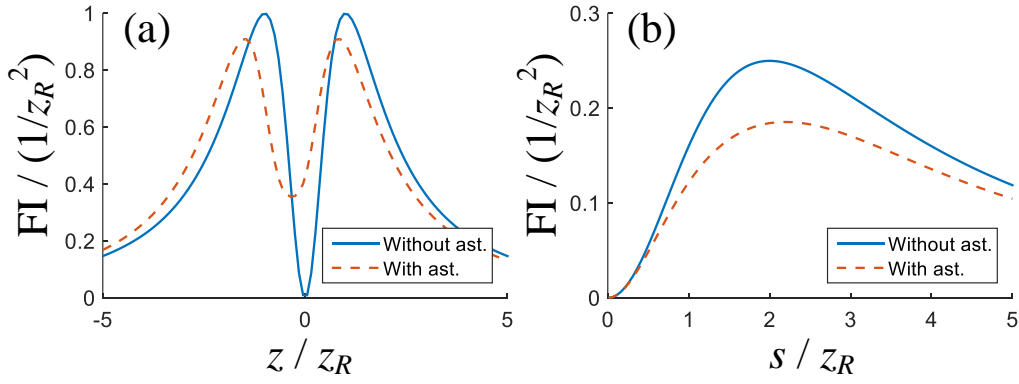
Fig. S1. Fisher information of (a) axial position z of a single point source and (b) axial separation s between a pair of point sources. The Fisher information without astigmatism (ast.) is also plotted as a reference.

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

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

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

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

where

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

The Fisher information then can be obtained as

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

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
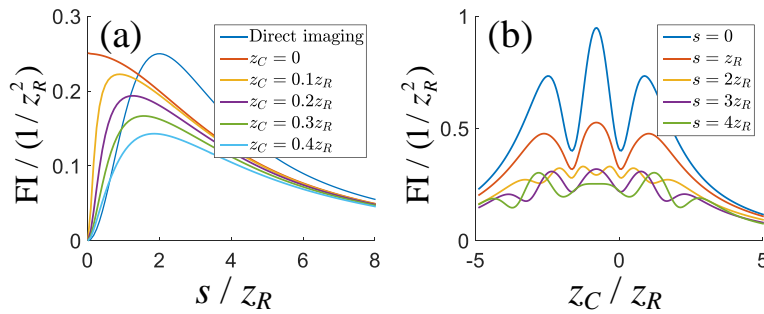

Fig. S2. (a) The Fisher information of separation estimation for sorter-based measurement with different centroid positions. The Fisher information for direct imaging with point source pair centroid $z_C = 0$ is plotted as a reference. (b) The Fisher information of centroid estimation for astigmatic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

point sources. For axial localization of a single point source, astigmatism can help increase the Fisher information at $z \to 0$ 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 $z_1 = z_C - s/2$, respectively. Therefore, the output radial mode distribution becomes $P'(p; z_C, s) \equiv \langle \mathrm{LG}_p | \rho | \mathrm{LG}_p \rangle = [P(p; z_C + s/2) + P(p; z_C - s/2)]/2$. Consequently, the Fisher information of separation in the presence of centroid misalignment can be written as

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

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

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

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

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

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

$$\psi(r_0; z) = \sqrt{2/\pi} \exp\left(-r_0^2\right) \exp\left(-ikz\mathrm{NA}^2 r_0^2/2\right), \tag{S28}$$

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 abberations 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

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

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

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

Here the estimator is defined as

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

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 $\bar{Q}$, and we force it to be zero whenever a negative value is obtained. Then the expectation of $\hat{s}$ can be calculated as

$$E[\hat{s}] = \sum_{k > Np}^{N} 2z_R B(k; H, p) \sqrt{\frac{2}{1 - 2\frac{k - Np}{N}} - 2}. \tag{S32}$$
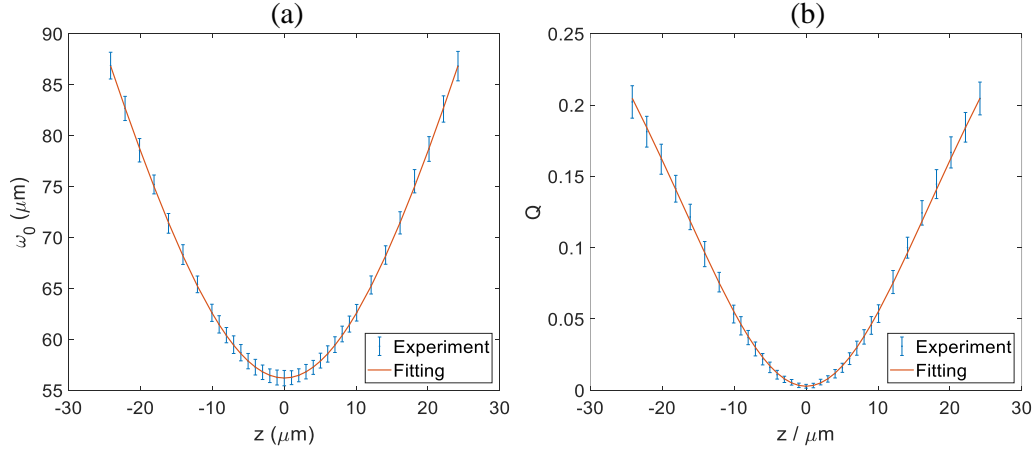
Fig. S3. (a) Measured Gaussian width w_0 as a function of z. (b) Measured Q of binary sorter as a function of z.

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]

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

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

(S33)

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

$$\hat{s}_{\mathrm{direct}} = 2z_R \sqrt{\frac{2}{Nw_0^2} \sum_{m=1}^{N} r_m^2 - 1},$$

(S34)

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

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

(S35)

where $\mathrm{Var}(\hat{s})$ 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 $s \to 0$, 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

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

(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

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

(S37)

where $r(\bar{r}) = \sqrt{\bar{r}}$. For N independent measurements $(r_1, r_2, \cdots, r_m)$, 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 $\bar{f}(\bar{r}) \otimes \bar{f}(\bar{r}) \otimes \cdots \otimes \bar{f}(\bar{r})$, 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

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

(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

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

where $k = Nw_0^2 k'/2 = N/[1 + (s\pi NA^2/2\lambda)^2]$. 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

$$E[\hat{s}_{\text{direct}}] = E[2z_R\sqrt{\bar{x}}] = 2z_R \int_0^\infty \sqrt{\bar{x}} g(\bar{x} + 1; N, k) d\bar{x}$$

$$= 2z_R \int_1^\infty \sqrt{\bar{x} - 1} g(\bar{x}; N, k) d\bar{x}. \tag{S40}$$

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

$$g(\bar{x}; N, k) \approx g(\bar{x}; N, N(1 - c_0 s^2)) \approx \sqrt{\frac{N}{2\pi}}(1 - c_0 s^2)^N x^{N-1} e^{(1-x)N} e^{Nxc_0 s^2}. \tag{S41}$$

Using $e^{Nxc_0 s^2} \approx 1 + Nxc_0 s^2$ and $(1 - c_0 s^2)^N \approx 1 - Nc_0 s^2$, we have $e^{Nxc_0 s^2}(1 - c_0 s^2)^N \approx 1 + N(x - 1)c_0 s^2$ and thus

$$g(\bar{x}; N, k) \approx \sqrt{\frac{N}{2\pi}} x^{N-1} e^{(1-x)N} [1 + N(x - 1)c_0 s^2],$$

$$\frac{\partial g(\bar{x}; N, k)}{\partial s} \approx \sqrt{\frac{2N}{\pi}} x^{N-1} e^{(1-x)N} N(x - 1)c_0 s, \tag{S42}$$

therefore

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

For CRLB, we have

$$\frac{\partial E[\hat{s}_{\text{direct}}]}{\partial s}\bigg|_{s=0} = 2z_R \int_1^\infty \sqrt{\bar{x} - 1} \frac{\partial g(\bar{x}; N, k)}{\partial s} d\bar{x}$$

$$= \sqrt{\frac{N}{2\pi}} \frac{Ns}{z_R} \int_1^\infty (\bar{x} - 1)^{3/2} \bar{x}^{N-1} e^{-(\bar{x}-1)N} d\bar{x}. \tag{S44}$$

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

$$\frac{\text{Var}(\hat{s})^{1/2}}{z_R/\sqrt{N}}\bigg|_{s=0} \geq \frac{\partial E[\hat{s}_{\text{direct}}]/\partial s}{z_R I(s)} = \frac{2z_R}{s} \frac{\partial E[\hat{s}_{\text{direct}}]}{\partial s}$$

$$= \sqrt{\frac{2N}{\pi}} N \int_1^\infty (\bar{x} - 1)^{3/2} \bar{x}^{N-1} e^{-(\bar{x}-1)N} d\bar{x}. \tag{S45}$$

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

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

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

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

Now we can use the following approximation

$$(1 + \frac{t}{\sqrt{N}})^{N-1} = \exp\left((N - 1)\ln\left(1 + \frac{t}{\sqrt{N}}\right)\right)$$

$$= \exp\left((N - 1)(\frac{t}{\sqrt{N}} - \frac{t^2}{2N} + \cdots)\right)$$

$$\approx \exp\left(\sqrt{N}t - \frac{t^2}{2} - \frac{t}{\sqrt{N}} + \frac{t^2}{2N}\right) \tag{S48}$$

$$= \exp\left(\sqrt{N}t\right) \exp\left(-\frac{t^2}{2}\right) \exp\left(-\frac{t}{\sqrt{N}} + \frac{t^2}{2N}\right)$$

$$\approx \exp\left(\sqrt{N}t\right) \exp\left(-\frac{t^2}{2}\right)(1 - \frac{t}{\sqrt{N}} + \frac{t^2}{2N}),$$

therefore

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

$$= \frac{s}{\sqrt{2\pi}z_R} N^{1/4} 2^{1/4} \int_0^\infty e^{-t_1}(t_1^{1/4} - \sqrt{\frac{2}{N}}t_1^{3/4} + \frac{1}{N}t_1^{5/4})dt$$

$$\approx \frac{s}{\sqrt{2\pi}z_R} N^{1/4} 2^{1/4} \Gamma\left(\frac{5}{4}\right)$$

$$\approx \frac{0.43 N^{1/4} s}{z_R},$$

(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

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

(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

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

(S51)

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

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

(S52)

which is $E[\hat{s}_{\text{direct}}]\big|_{s=0} = 0.1226 z_R$ 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]

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

(S53)

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

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

(S54)

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

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

(S55)

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

$$E[\hat{s}_{\text{binary}}] = \sum_{k=0}^N 2z_R P(k; N, p)\sqrt{\frac{2}{\frac{2k}{N}-1} - 2}$$

$$\approx \sum_{k=0}^N 2z_R \frac{N!}{k!(N-k)!}(1 - \frac{s^2}{16z_R^2})^k(\frac{s^2}{16z_R^2})^{N-k}\sqrt{\frac{2}{\frac{2k}{N}-1} - 2}$$

$$= \sum_{k=0}^N 2z_R \frac{N!}{k!(N-k)!}(1 - \frac{s^2}{16z_R^2})^{N-k}(\frac{s^2}{16z_R^2})^k\sqrt{\frac{2}{\frac{2(N-k)}{N}-1} - 2}.$$

(S56)

In the last step we make the substitution $k \to N - k$. 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

$$E[\hat{s}_{\text{binary}}] \approx 0 + 4z_R\sqrt{N}(1 - \frac{s^2}{16z_R^2})^{N-1}(\frac{s^2}{16z_R^2}) + 2z_R\sqrt{2N}(N-1)(1 - \frac{s^2}{16z_R^2})^{N-2}(\frac{s^2}{16z_R^2})^2 + \cdots$$

$$\approx \frac{\sqrt{N}s^2}{4z_R} + 2\sqrt{2N}z_R(N-1)(\frac{s^2}{16z_R^2})^2 + \cdots$$
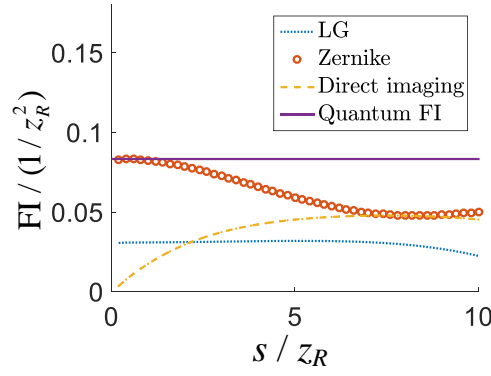
(S57)

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

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

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

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

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

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

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]

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

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 as 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

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

where $p_{max} + 1$ is the number of Zernike modes used for superresolution. In our simulation we choose $p_{max} = 1$, 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

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

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

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

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

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

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

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

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).