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 Mohammad Mirhosseini,4 A. Nick Vamivakas,1,2,5 Andrew N. Jordan,2,6 Zhimin Shi,7,9 and Robert W. Boyd1,2,8,10

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
9e-mail: zhiminshi@usf.edu
10e-mail: boyd@optics.rochester.edu
*Corresponding author: yzhou62@ur.rochester.edu
Received 14 December 2018; revised 14 March 2019; accepted 26 March 2019 (Doc. ID 355468); published 26 April 2019

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 position 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 plane for an on-axis point source is denoted by |ψ⟩, where ⟨r₀|ψ⟩ = ψ(r₀; z) and

\[ \psi(r₀; z) = \sqrt{2/\pi} \exp(-r₀^2) \exp(-ikzNA^2r₀^2/2), \]  

where NA is the numerical aperture, z is the axial position of the point source, \( k = 2\pi/\lambda \) is the wavenumber, \( \lambda \) is the wavelength, \( r₀ \) is the normalized radial coordinate in the pupil plane, and \( |r₀⟩ \) is the corresponding radial eigenstate. Here we define \( r₀ = g_j/(f_jNA) \), where \( g_j \) is the radial coordinate in the pupil plane and \( f_j \) 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

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

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 \( \rho = (|ψ₁⟩⟨ψ₁| + |ψ₂⟩⟨ψ₂|)/2 \), where \( ⟨r₀|ψ₁⟩ = ψ(r₀; s/2) \) and \( ⟨r₀|ψ₂⟩ = ψ(r₀; -s/2) \). The normalized total intensity at the image plane can be calculated as \( I(r) \equiv ⟨r|ρ|r⟩ = [I(r;s/2) + I(r; -s/2)]/2, \) where \(|r⟩\) 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 \( |LG_p⟩ \), where \( ⟨r₀|LG_p⟩ = LG_p(r₀) \) and

\[ LG_p(r₀) = \sqrt{2/\pi} \exp(-r₀^2)L_p(2r₀^2), \]  

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

$$P(p; z) = |\langle\psi| \text{LG}_p\rangle|^2 = \frac{4\pi^2 p^2}{(4\pi^2 + p^2)^{1/2}},$$  \hspace{1cm} (4)

where $z_R = \pi w_0^2/\lambda$ and $w_0 = \lambda/\pi N A$ [31]. For two equally bright sources separated by $s$, the output radial mode distribution becomes $P_s(p) \equiv |\langle LG_p|P|\text{LG}_p\rangle| = [P(p; s/2) + P(p; -s/2)]/2$. It can be noticed that for direct imaging and sorter-based measurement, the two incoherent point sources have the same response, which is independent of the magnification $K$. The quantum Fisher information, and the details are presented in Supplement 1, which gives the lower bound of classical CRLB for any possible measurements. We notice that the quantum CRLB, which gives the lower bound of classical CRLB, can be noticed that for direct imaging and sorter-based measurement by calculating the Fisher information

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

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

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

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

$$\mathcal{K}_s = 4(\partial^2 \psi_1/\partial \psi_1 - |\langle \psi_1 | \partial \psi_1 \rangle|^2),$$  \hspace{1cm} (7)

where $|\partial \psi_1 \rangle = \partial |\psi_1 \rangle/\partial s$, and it can be readily shown that $\mathcal{K}_s = 1/4\pi R^2$. 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

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

$$P^1(s) = \sum_{p=0}^{\infty} P_s(2p + 1; s) = \frac{1}{2} - \frac{4\pi^2}{8\pi R^2 + s^2}. \hspace{1cm} (8)$$

Thus, the Fisher information for a binary sorter is

$$\mathcal{J}_\text{binary}(s) = \sum_{q=0}^{\infty} \frac{1}{P^q(s)} \left( \frac{\partial P^q(s)}{\partial s} \right)^2 = \frac{256\pi^4}{(s^2 + 8\pi R^2)^2(s^2 + 16\pi R^2)^2}.$$  \hspace{1cm} (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 $2\pi R$. In incoherent imaging microscopy, the axial resolution can be expressed as $\Delta z = 2\lambda/NA^2$ [57], which can be rewritten as $\Delta z = 2\pi R$ with our notation. We note that the discrepancy between $2\pi R$ and $2\pi \Delta z$ 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

$$\hat{s} = \sqrt{\frac{2N}{\sum m^2}} - \hat{s}_\text{direct} = 2\pi R \sqrt{\frac{\hat{w}^2}{w_0^2} - 1},$$  \hspace{1cm} (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 $\hat{w}$ 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.](image-url)

(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 \( \hat{\vartheta} \) in our experiment. For binary sorter-based measurement, the maximum likelihood estimator is

\[
\hat{Q} = \frac{1}{N} \sum_{q=0}^{1} q \cdot m_q = \frac{m_0}{N},
\]

\[
\hat{s}_{\text{Binary}} = 2\varepsilon R \sqrt{\frac{2}{1 - 2Q} - 2},
\]

(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 \[38\]

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

(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 heralded 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 \( \text{Var}(\hat{s}) \geq 1/[N \cdot J(s)] \), 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 \( \text{NA} = 0.1 \) 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.](image-url)
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

$$\frac{\partial E[\hat{\theta}_{\text{direct}}]}{\partial \theta} = \frac{0.43N^{1/4}}{z_R}. \quad (14)$$

Together with Eq. (12), it immediately follows that the CRLB becomes $\text{Var}(\hat{\theta}_{\text{direct}}) \geq 0.74z_R^2/\sqrt{N}$ 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 $\sqrt{\text{Var}(\hat{\theta}_{\text{direct}})/(z_R^2/\sqrt{N})} \geq 5.8$, 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, $\partial E[\hat{\theta}_{\text{direct}}]/\partial \theta|_{\theta=0} = 0$ implies that the expectation value has a slope of zero when $\theta = 0$, as can be seen in Fig. 4(a). Hence, despite of a finite standard deviation, it is intrinsically unrealistic to use the measured $\theta$ to recover the real $\theta$ in the vicinity of $\theta = 0$ 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 $\text{Var}(\hat{\theta}_{\text{direct}})|_{\theta=0} \propto N^{-1/2}$; 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 $\partial E[\hat{\theta}_{\text{binary}}]/\partial \theta|_{\theta=0} = 0$ and thus $\text{Var}(\hat{\theta}_{\text{binary}})|_{\theta=0} = 0$, 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 $E[\hat{\theta}_{\text{binary}}]|_{\theta=0} = 0$, 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
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 $E[F_{\text{binary}}]_{s=0} = 0.043 z_R$, which is very close to the measured value of $0.049 z_R$. 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 0.049$z_R$, which is obtained in sorter-based measurement, the photon number needs to be on the order of $10^7$ to satisfy $0.82 N^{-1/4} z_R = 0.049 z_R$. 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 $s \rightarrow s/2$, 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 $|z_C| < 0.3 z_R$. 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 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 $E[F_{\text{binary}}]_{s=0} = 0.043 z_R$, which is very close to the measured value of $0.049 z_R$. 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 0.049$z_R$, which is obtained in sorter-based measurement, the photon number needs to be on the order of $10^7$ to satisfy $0.82 N^{-1/4} z_R = 0.049 z_R$. 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 $s \rightarrow s/2$, 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 $|z_C| < 0.3 z_R$. 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 the
photon super-resolution is achievable only if the temporal coherence of light is long. There is a trade-off: longer coherence length gives better spatial resolution, but shorter coherence length is needed for better temporal resolution. This suggests that a long coherence length is desirable for interferometric super-resolution microscopy.

In conclusion, we theoretically and experimentally demonstrate the axially super-resolution based on a radial mode sorter. The binary radial mode sorter employed in our experiment can reach the quantum CRRLB 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.

**Acknowledgments.** We thank Mankei Tsang for helpful discussions.

See Supplement 1 for supporting content.

**REFERENCES**