Resolving starlight: a quantum perspective

Mankei Tsang\textsuperscript{1, 2, *}

\textsuperscript{1}Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117583
\textsuperscript{2}Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117551

(Dated: June 6, 2019)

The wave-particle duality of light introduces two fundamental problems to imaging, namely, the diffraction limit and the photon shot noise. Quantum information theory can tackle them both in one holistic formalism: model the light as a quantum object, consider any quantum measurement, and pick the one that gives the best statistics. While Helstrom pioneered the theory half a century ago and first applied it to incoherent imaging, it was not until recently that the approach offered a genuine surprise on the age-old topic by predicting a new class of superior imaging methods. For the resolution of two sub-Rayleigh sources, the new methods have been shown theoretically and experimentally to outperform direct imaging and approach the true quantum limits. Recent efforts to generalize the theory for an arbitrary number of sources suggest that, despite the existence of harsh quantum limits, the quantum-inspired methods can still offer significant improvements over direct imaging for subdiffraction objects, potentially benefiting many applications in astronomy as well as fluorescence microscopy.

I. INGREDIENTS OF THE RESOLUTION PROBLEM: DIFFRACTION, PHOTON SHOT NOISE, STATISTICS

In 1879 Lord Rayleigh proposed a criterion of resolution for incoherent imaging in terms of two point sources \cite{Rayleigh1879}: the sources are said to be unresolvable if they are so close that their images, blurred by diffraction, overlap significantly. To quote Feynman \cite{Feynman1965}, however, “Rayleigh’s criterion is a rough idea in the first place,” and a better resolution can be achieved “if sufficiently careful measurements of the exact intensity distribution over the diffracted image spot can be made.” Thus another limiting factor is the noise in the intensity measurement, with the photon shot noise being the most fundamental source. Because of the particle nature of light, each camera pixel can record its energy in discrete quanta only, and ordinary light sources, including starlight and fluorescence, introduce further randomness to the quantum measurements \cite{Helstrom1973, Helstrom1976}.

To incorporate noise in the definition of resolution, the theory of statistical inference offers a rigorous framework \cite{Cox1961, Casella1998, Casella2002}. For example, a measure of resolution can be defined in terms of parameter estimation: given a blurry and noisy image of two point sources, how well can one estimate their separation \cite{Cox1961, Casella1998, Casella2002, Pires2012}? Or it can be framed in terms of hypothesis testing: how well can one decide from the image whether there is one or two sources \cite{Pires2012, Pires2013}? Such statistical treatments of resolution have garnered prominence in optical astronomy \cite{Cox1961, Casella1998, Casella2002, Pires2012, Pires2013, Pires2014} and fluorescence microscopy \cite{Pires2012, Pires2013, Pires2014, Pires2015}, where the number of photons is limited and shot noise is part of life.

II. QUANTUM DETECTION AND ESTIMATION THEORY

Imaging has grown into a multidisciplinary problem that straddles optics, quantum mechanics, statistics, and signal processing. In a Herculean effort that began in the 1960s, Helstrom merged the subjects into a theory of quantum detection and estimation \cite{Helstrom1973}, which marked the beginning of quantum information theory. His aim was to determine the best measurement, out of the infinite possibilities offered by quantum mechanics, that optimizes the performance of an inference task. For a given light source, the optimal performance then represents the most fundamental limit on the resolution, valid for any optics design that is allowed by quantum mechanics, as well as any computational technique in data post-processing. In setting fundamental limits, Helstrom’s theory plays a role for sensing and imaging not unlike the second law of thermodynamics for engines, ruling out unphysical superresolution methods in the same manner the second law rules out perpetual-motion machines.

The mathematics was formidable, but Helstrom managed to apply his theory to a few simple scenarios of incoherent imaging. For example, he studied the problem of locating an incoherent point source from far-field measurements \cite{Helstrom1973}, but the result was unsurprising: the quantum limit is close to the ideal performance of direct imaging, which measures the intensity on the image plane, as depicted by Fig. 1. A more intriguing problem he studied was the decision between one or two incoherent sources \cite{Helstrom1973}. Helstrom computed the mathematical form of the optimal measurement and the resulting error probabilities, but he did not propose an experimental setup or show how much improvement the optimal measurement could offer over existing imaging methods. Helstrom himself was quite pessimistic \cite{Helstrom1973}: “The optimum strategies required in order to attain the minimum error probabilities calculated here require the measurement of certain complicated quantum-mechanical projection operators, which, though possible in principle, cannot be carried out by any known apparatus.”

Unfortunately, in all the problems studied by Helstrom, the improvements predicted by his theory seemed modest at best, rendering the question of quantum limits academic. Quantum opticians turned their attention to nonclassical light sources \cite{Fry1983, Fry1984, Fry1985, Fry1986, Fry1987, Fry1988, Fry1989, Fry1990, Fry1991, Fry1992, Fry1993, Fry1994, Fry1995, Fry1996, Fry1997, Fry1998, Fry1999, Fry2000, Fry2001, Fry2002, Fry2003, Fry2004, Fry2005, Fry2006, Fry2007, Fry2008, Fry2009, Fry2010, Fry2011, Fry2012, Fry2013, Fry2014, Fry2015, Fry2016, Fry2017, Fry2018, Fry2019, Fry2020}, while classical opticians turned their attention to near-field microscopy \cite{Fry1983, Fry1984}, fluorescence control

\* mankei@nus.edu.sg; https://www.ece.nus.edu.sg/stfpage/tmk/
III. RAYLEIGH’S CURSE

With two incoherent point sources, direct imaging, and photon shot noise, many studies have shown that their separation becomes harder to estimate if they violate Rayleigh’s criterion [8–12]. The central tool used in those studies is the Fisher information, which sets general lower bounds called Cramér-Rao bounds on the parameter-estimation error [5]. The simplest Cramér-Rao bound is

\[ \text{MSE}(\theta) \geq \text{CRB}(\theta) \equiv \text{FI}(\theta)^{-1}, \]  

(3.1)

where \( \text{MSE}(\theta) \) is the mean-square error of any unbiased estimator, \( \theta \) is the unknown parameter, and \( \text{FI}(\theta) \) is the Fisher information; see Appendix A for precise definitions. The error can reach the Cramér-Rao bound in many situations, including an asymptotic limit where the sample size approaches infinity, the noise can be approximated as additive and Gaussian, and the maximum-likelihood estimator is used [79].

Assume one-dimensional paraxial imaging [80] for simplicity, as illustrated by Fig. 2, and Poisson noise, which is an excellent approximation for both optical astronomy [4, 17, 20, 21] and fluorescence microscopy [12, 22–26, 81]. The Fisher information becomes

\[ \text{FI}^{(\text{direct})}(\theta) = C(\theta)N, \]  

(3.2)

where \( \theta \) here is the separation, \( N \) is the average photon number, and \( C(\theta) \) is an \( N \)-independent prefactor that varies with \( \theta \). \( \theta \) and \( C(\theta) \) are dimensionless if \( \theta \) is normalized in Airy units (1 Airy unit is roughly \( \lambda/N.A. \) where \( \lambda \) is the wavelength and N.A. is the numerical aperture, or \( \lambda/D \) for angular resolution, where \( D \) is the aperture diameter [80, 81]). Equation (3.2) was earlier suggested by many as a fundamental measure of resolution for incoherent imaging [9–12].

IV. DISPPELLING RAYLEIGH’S CURSE

Rayleigh’s curse happens if we measure the intensity on the image plane, but what if we allow any quantum measurement that may be sensitive to the phase as well? To find the quantum limit, we can use a quantum version of the Fisher information proposed by Helstrom [27], which sets an upper bound on the Fisher information for any measurement [82, 83], as elaborated in Appendix B. We found that the Helstrom information \( (\text{HI}) \) for the separation estimation problem is given by [42]

\[ \text{FI}(\theta) \leq \text{HI}(\theta) = C(\infty)N. \]  

(4.1)

Remarkably, \( \text{HI}(\theta) \) is constant regardless of the separation and completely free of Rayleigh’s curse, as plotted in Fig. 3. The constant Helstrom information would be no surprise if it were simply a loose upper bound; the million-dollar question is whether one can find a measurement that attains the limit. Mathematical studies following Helstrom’s work have...
shown in general that a quantum-limited measurement should exist, at least in the limit of infinite sample size [84, 85]. The mathematics offers little clue to the experimental implementation, however, and finding one in quantum estimation theory is often a matter of educated guessing.

Luckily we found one. Assuming a Gaussian point-spread function, we found that sorting the light on the image plane is often a matter of educated guessing. For incoherent imaging, the wavefunction of each photon is approximated as

\[ \psi_{q}(x) = \delta(\mathbf{x} - \mathbf{x}_{q}) \]

by virtue of Born’s rule. It is also possible to measure the particle in any other orthonormal basis \( \{ \phi_{q}(x) : q \in \mathbb{N} \} \), and the probability of finding the photon in the \( q \)th spatial mode is

\[ g_{q} = \left| \int_{-\infty}^{\infty} dx \phi_{q}^{*}(x)\psi(x) \right|^{2}. \]  (5.2)

For incoherent imaging, the wavefunction of each photon is \( \psi(x-X) \), where \( \psi \) is determined by the point-spread function of a diffraction-limited imaging system and the displacement \( X \) depends on the position of the point source that emits the photon. Denoting the density of the incoherent sources as \( F(X) \), \( X \) can be regarded as a random variable with \( F(X) \) as its probability density. For direct imaging, the probability density on the image plane becomes

\[ f(x) = \int_{-\infty}^{\infty} dX |\psi(x-X)|^{2} F(X), \]  (5.3)

which agrees with the classical theory of incoherent imaging [80]. In general, the probability of finding the photon in the \( \phi_{q}(x) \) mode is

\[ g_{q} = \int_{-\infty}^{\infty} dx \phi_{q}^{*}(x)\psi(x-X) F(X). \]  (5.4)

If we treat the arrivals of the photons at the spatial modes as a temporal Poisson process, then the photon counts integrated over time are independent Poisson random variables, each with mean and variance given by \( N g_{q} \), where \( N \) is the average photon number in all modes. For direct imaging, the photon statistics should be treated as a spatial Poisson process with mean intensity \( N f(x) \) [87].

Consider two point sources, one at \( X = -\theta/2 \) and one at \( X = \theta/2 \) such that \( F(X) = [\delta(X - \theta/2) + \delta(X + \theta/2)]/2 \). If their separation is deeply sub-Rayleigh (\( \theta \ll 1 \)), the wavefunctions can be approximated as

\[ \psi \left( x \pm \frac{\theta}{2} \right) \approx \psi(x) \pm \frac{\theta}{2} \frac{\partial \psi(x)}{\partial x}, \]  (5.5)
as depicted by Fig. 5. If \( \psi(x) \) is even, \( \partial \psi(x)/\partial x \) is odd, and they can be regarded as two orthogonal modes. To the first order, the mean photon count in the fundamental \( \psi(x) \) mode is insensitive to the parameter \( \theta \), while the mean count in the derivative mode is the incoherent sum of the contributions from the two sources, or \( \propto (\theta/2)^2 + (-\theta/2)^2 = \theta^2/2 \). If the sources were coherent and in-phase instead, their contributions to the derivative mode would cancel each other, leading to much reduced signal [88]. In other words, the incoherence plays a key role in retaining a significant signal in the first order, and SPADE can extract this signal by measuring the derivative mode.

For separation estimation, as long as \( \phi_1(x) \) is orthogonal to \( \psi(x) \) and has significant overlap with the derivative mode, \( g_1(\theta) \propto \theta^2 \) for \( \theta \ll 1 \), leading to a nonzero \( (\partial g_1(\theta)/\partial \theta)^2/g_1(\theta) \) as \( \theta \to 0 \).

To summarize, SPADE relies on the subtle interplay between the coherence induced by diffraction, the incoherence of the sources, and the signal-dependent nature of photon shot noise. It would have been difficult to discover such a fortuitous possibility via conventional wisdom alone, but quantum estimation theory—and quantum information theory in general—have the advantage of being oblivious to conventional wisdom. The mathematics may look daunting, but it can sometimes give rise to new physics beyond our imagination.

VI. IMPLEMENTATIONS OF SPADE

To implement SPADE, different spatial modes should be coupled into physically separate channels before detection. This in principle requires only linear optics [89], but the most efficient implementation remains unclear. Many methods have been proposed and demonstrated, particularly for the purpose of mode-division multiplexing in optical communication [90]. Here we highlight a few methods that have been experimentally demonstrated for the two-point resolution problem.

A. Interferometry

Nair proposed an interferometer called SLIVER (super-localization via image-inversion interferometry) that can in principle achieve a quantum-limited Fisher information for \( \theta \to 0 \) and any even point-spread function [48]. Although image-inversion interferometry has earlier been proposed and demonstrated [91–94] to achieve a modest resolution improvement for general confocal microscopy, its extraordinary precision for two-point resolution was hitherto not recognized.

The setup, depicted by Fig. 6, consists of a two-arm interferometer with spatial inversion in one arm. The inversion can be implemented via mirrors, lenses, or a Dove prism for example. As a result of the inversion and the interference at the second beamsplitter, all the even modes on the image plane are routed to one output port while the odd modes are routed to the other port. Hence the fundamental mode \( \psi(x) \), as long as it is even, is separated from the odd derivative mode, which is detected at the other port. Tang, Durak, and Ling reported a proof-of-concept demonstration of SLIVER [70], although their reported errors were not close to the quantum limit.

SLIVER works best for sub-Rayleigh separations but is suboptimal for larger separations. A variant of SLIVER called pix-SLIVER replaces the detectors by detector arrays and can work better for larger separations [50]. Another way to generalize SLIVER is to think of image inversion as a special case of fractional Fourier transform (FRFT). A tree of FRFT interferometers, with the image-inversion interferometer at its root,
FIG. 6. An image-inversion interferometer. Through the inversion and the interference, the even modes are coupled to one port while the odd modes are coupled to the other port.

can sort the Hermite-Gaussian modes and implement SPADE [95]. The interferometer-tree concept can be generalized to sort in any other basis if appropriate mode-dependent phases can be introduced [96, 97].

Along this direction, Hassett and co-workers demonstrated a Michelson interferometer with variable FRFT in one arm and used it to infer the Hermite-Gaussian-mode spectrum $g_q$ of a shifted Gaussian beam [77]. They suggested that the setup could be useful for estimating sub-Rayleigh separations, although its statistical performance remains to be studied. In another work, Zhou and co-workers demonstrated a binary radial-mode sorter that is also based on FRFT interferometry and used it to enhance the estimation of the axial separation between two sources [78].

B. SPLICE

Tham, Ferretti, and Steinberg proposed an elegant setup called SPLICE (super-resolved position localization by inversion of coherence along an edge) to capture the derivative mode [71]. SPLICE consists of a phase plate that introduces a $\pi$ phase shift to half of the image plane and a single-mode fiber, as illustrated by Fig. 7. An odd mode on the image plane is thus coupled into the fiber and detected, while all other modes orthogonal to it are rejected by the fiber. Despite the imperfect match between the odd mode and the derivative mode, Tham and co-workers were still able to demonstrate a mean-square error around five times the quantum bound and a significant improvement over direct imaging [71].

C. Holograms

A hologram is capable of performing a spatial matched filter, and it can be designed such that the diffracted intensities at specific points in the far field are proportional to the modal spectrum $g_q$ [80, 98]. The use of such a hologram for separation estimation was demonstrated by Paúr and co-workers [72]. Their reported mean-square errors were around twice the quantum bound, but it is important to note that they scaled the quantum bound with respect to the diffracted photon number, not the photon number before the hologram, meaning that the result did not take into account the low diffraction efficiency of their hologram. Efficient SPADE is possible with multiple holograms however [89, 99, 100].

D. Point-spread-function shaping

In the context of direct imaging, the approximation given by Eq. (5.7) for $\theta \ll 1$ leads to

$$\text{FI}^{(\text{direct})} \approx \frac{N \theta^2}{16} \int_{-\infty}^{\infty} dx \frac{[\partial^2 \psi(x)]^2}{\psi(x)^2}$$

(6.1)

It is often assumed [10, 11] that this can be approximated by

$$\text{FI}^{(\text{direct})} \approx \frac{N \theta^2}{16} \int_{-\infty}^{\infty} dx \frac{1}{\psi(x)^2} \left[ \frac{\partial^2 \psi(x)^2}{\partial x^2} \right]^2,$$

(6.2)

which scales quadratically with $\theta$. This is indeed true if $|\psi(x)|^2$ is Gaussian, but it turns out that the integral in Eq. (6.2) may not converge if $|\psi(x)|^2$ has zeros, and one must go back to Eq. (6.1), which can give a linear scaling of $\text{FI}^{(\text{direct})}$ with $\theta$ instead. Paúr and co-workers exploited this phenomenon by introducing a signum phase mask at the pupil plane of a direct-imaging system, changing $\psi(x)$ from a Gaussian to an odd function with a zero in the middle [76]. Although the resulting Fisher information still approaches zero for $\theta \to 0$, they were able to demonstrate a significant improvement of the estimation accuracy with a simple change.
E. Heterodyne

Given the experimental difficulties of performing efficient SPADE, a seemingly appealing alternative is to perform heterodyne detection of the derivative mode by interfering the light with a shaped reference beam on a detector, as demonstrated by Yang and co-workers [73]. It was later found however that the homodyne or heterodyne Fisher information still suffers from Rayleigh’s curse for weak thermal light [56]. This can be attributed to the constant vacuum noise that plagues a heterodyne or homodyne detection regardless of the signal, compared with the Poisson variance that reduces with the signal for photon counting. A similar problem was discovered earlier in the context of stellar interferometry [101, 102]. The surprisingly poor performance of heterodyne detection demonstrates the importance of analyzing a measurement using rigorous quantum optics as well as statistics, even when dealing with classical light, to ensure an acceptable statistical performance.

F. Sum-frequency generation

Donohue and co-workers implemented SPADE in the time or frequency domain for estimating the separation between optical pulses via an interesting nonlinear-optical technique: sum-frequency generation [74]. If the light is combined with a strong local-oscillator pulse in a second-order nonlinear medium with the right phase matching, the Hamiltonian of the sum-frequency generation is the same as that of linear optics [103], and a temporal or spectral mode projection can be implemented if the local oscillator has the desired mode shape and the up-converted signal is measured. While the efficiency of their measurement was only 0.7%, the principle was clearly demonstrated in their experiment.

G. Two-photon measurement

Last but not the least, we should mention an even more radical proposal by Parniak and co-workers, which uses a two-photon measurement to estimate the centroid and the separation of two sources simultaneously near the quantum limit [75]. Its principle would be too difficult to explain here and its applicability to usual light sources is questionable, but it demonstrates the fact that our model of linear optics and photon counting does not encompass all the possibilities offered by quantum mechanics, and there exist multiphoton measurements that can offer advantages in multiparameter estimation, at least in principle.

VII. EXTENDED SOURCES

A. Estimation of the second moment

While the two-point problem is historic and significant, it has rather limited applications, and the important next step is to apply the concepts developed so far to more general objects. Suppose now that the number of point sources is arbitrary, and the object intensity is given in general by $F(X)$. Similar to the sub-Rayleigh approximation earlier, here we focus on a subdiffraction regime where the object width around $X = 0$, defined as $\Delta$, is much smaller than the width of the point-spread function, or $\Delta \ll 1$. Then, similar to Eq. (5.5), the photon wavefunction due to each point $X$ within the object can be approximated as

$$\psi(x - X) \approx \psi(x) - X \frac{\partial \psi(x)}{\partial x}, \quad (7.1)$$

Summing the incoherent contributions from all the points via Eq. (5.4), the mean photon count in the derivative mode $\phi_1(x) \propto \partial \psi(x)/\partial x$ is

$$N_{g_1} \approx Nc_1^2 \int_{-\infty}^{\infty} dXX^2 F(X), \quad (7.2)$$

where $c_1$ is a constant and $\int_{-\infty}^{\infty} dXX^2 F(X)$ is the second moment of $F(X)$. Thus we can expect SPADE to enhance the estimation of the second moment for any subdiffraction object in the same way it enhances the two-point resolution. As the second moment can be related to the width of $F(X)$, it should not be surprising that SPADE can also enhance the estimation of the object size [43, 45].

B. Even moments

To go another step further, let us expand $\psi(x - X)$ up to the $q$th order. It is more convenient to work in the spatial frequency domain, as defined by

$$\psi(x) \rightarrow \Psi(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx \psi(x) \exp(-ikx), \quad (7.3)$$

which leads to

$$\psi(x - X) \rightarrow \exp(-ikX)\Psi(k) \approx \sum_{p=0}^{q} \frac{(-ikX)^p}{p!} \Psi(k). \quad (7.4)$$

A natural orthonormal basis that includes the fundamental mode $\psi(x) \rightarrow \Psi(k)$ and the derivative mode $-\partial \psi(x)/\partial x \rightarrow -ik\Psi(k)$ can be defined as [55]

$$\{ \phi_q(x) \rightarrow \Phi_q(k) = (-i)^q b_q(k) \Psi(k) : q \in \mathbb{N}_0 \}, \quad (7.5)$$

where $\{b_q(k)\}$ are the orthogonal polynomials obtained by applying the Gram-Schmidt procedure to monomials $\{1, k, k^2, \ldots\}$ with respect to the weighted inner product [104]

$$\langle u(k), v(k) \rangle \equiv \int_{-\infty}^{\infty} dk |\Psi(k)|^2 u^*(k)v(k), \quad (7.6)$$

leading to $\{b_q(k), b_p(k)\} = \int_{-\infty}^{\infty} dk \Phi_q^*(k)\Phi_p(k) = \delta_{qp}$. The basis is called the point-spread-function-adapted basis [55], or
the PAD basis for short [44]. For example, if $|\Psi(k)|^2$ is Gaussian, then $\{b_q(k)\}$ are the Hermite polynomials. An important property of $b_q(k)$ that follows from the Gram-Schmidt procedure is that $\langle b_q(k), k^p \rangle = 0$ if $p < q$. The overlap function in Eq. (5.4) becomes

$$\int_{-\infty}^{\infty} dx \phi_q^*(x) \psi(x - X)$$

$$= \sum_{p=0}^{q} \frac{(-iX)^p}{p!} \int_{-\infty}^{\infty} dk \Phi_q^p(k) \Psi(k) k^p \quad (7.7)$$

$$= \sum_{p=0}^{q} \frac{(-iX)^p}{p!} q^q (b_q(k), k^p) = c_q X^q, \quad (7.8)$$

where $c_q$ is a real constant. In other words, $\Phi_q(k)$ is orthogonal to all the terms in Eq. (7.4) except the last term (and the neglected higher-order terms). The mean photon count given by Eq. (5.4) becomes

$$Ng_q \approx Nc_q^2 \int_{-\infty}^{\infty} dX X^{2q} F(X). \quad (7.9)$$

Similar to the relation between the derivative mode and the second moment, each PAD mode can access an even moment while rejecting the background noise from all the lower moments [44]. Hence SPADE with respect to the PAD basis can be expected to enhance the estimation of all even moments.

If $\psi(x)$ is Gaussian, the PAD basis becomes the Hermite-Gaussian basis, and its sensitivity to even moments is noted in Refs. [43, 73]. The general PAD basis is proposed in Refs. [55, 57] for the two-point problem and applied to general imaging in Refs. [44, 47]. The use of SPlICE for moment estimation was recently proposed by Bonsma-Fisher and co-workers [69].

C. Error analysis

Define the moment parameters as

$$\theta_\mu = \int_{-\infty}^{\infty} dX X^{2q} F(X), \quad (7.10)$$

where $\mu \in \mathbb{N}$ denotes the moment order. Appendix D introduces the multiparameter-estimation theory in more detail. The mean and variance of the photon count $n_q$ in each PAD mode is

$$Ng_q \approx Nc_q^2 \theta_{2q}, \quad (7.11)$$

so the estimator $\hat{\theta}_{2q} = n_q / (Nc_q^2)$ is approximately unbiased, and the mean-square error is [43, 44]

$$\text{MSE}_{2q}^{(\text{SPADE})} \approx \frac{\theta_{2q}}{Nc_q^2} = \frac{O(\Delta^{2q})}{N}, \quad (7.12)$$

where the subscript $2q$ denotes the error for the $\theta_{2q}$ parameter, the big-O notation denotes terms on the order of the argument, and $\theta_\mu = O(\Delta^\mu)$. For direct imaging on the other hand, we showed that the Cramér-Rao bound for any moment is [43, 44]

$$\text{MSE}_{\mu}^{(\text{direct})} \geq \text{CRB}_{\mu}^{(\text{direct})} = \frac{O(1)}{N}, \quad (7.13)$$

so SPADE can achieve much lower errors for the even moments in the $\Delta \ll 1$ subdiffraction regime.

As large as the enhancement seems, the signal-to-noise ratio (SNR), defined as

$$\text{SNR}_\mu \equiv \frac{\theta_\mu^2}{\text{MSE}_\mu}, \quad (7.14)$$

offers a more sobering perspective, as the signal $\theta_\mu^2 = O(\Delta^{2q})$ is an even smaller number. For SPADE and even moments, the SNR turns out to be equal to the mean photon count in a PAD mode, or

$$\text{SNR}_q^{(\text{SPADE})} \approx Ng_q = NO(\Delta^{2q}), \quad (7.15)$$

which decreases for smaller $\Delta$ and higher moments. The degradation of the SNR can be attributed to the inherently low efficiency of a subdiffraction source coupling into a higher-order mode. While this shows that SPADE has its own limitations, the fact remains that direct imaging is even worse, with a SNR given by

$$\text{SNR}_q^{(\text{direct})} = NO(\Delta^{2q}), \quad (7.16)$$

which is $NO(\Delta^{4q})$ for $\mu = 2q$. With enough photons, the enhancements offered by SPADE can still be useful, especially for the lower moments.

D. Odd moments

To estimate an odd moment, consider projections into the pair of so-called iPAD modes

$$\phi_q^\pm(x) = \frac{\phi_q(x) \pm \phi_{q+1}(x)}{\sqrt{2}}, \quad (7.17)$$

which result from the interference of two adjacent PAD modes [44]. It makes intuitive sense that, if each $\phi_q$ mode is sensitive to the $2q$th moment, then a superposition of two adjacent PAD modes should be sensitive to an odd moment in-between. Expanding $\psi(x - X)$ up to the $(q + 1)$th order and following the same steps as Eqs. (7.7) and (7.8), the overlap function becomes

$$\int dx \phi_q^\pm \psi(x - X) \approx \frac{1}{\sqrt{2}} (c_q X^q \pm c_{q+1} X^{q+1}), \quad (7.18)$$

where $|\Psi(k)|^2$ is assumed to be even such that $\{b_q(k)\}$ are alternatively even and odd, leading to $\langle b_q(k), k^{q+1} \rangle = 0$. Let the output counts be $n_q^\pm$. The mean counts are

$$Ng_q^\pm \approx \frac{N}{2} \int dX \left( c_q X^q \pm c_{q+1} X^{q+1} \right)^2 F(X). \quad (7.19)$$
Subtracting one count by the other, the mean is
\[
N(g_q^{(+)} - g_q^{(-)}) \approx 2N c_q c_{q+1} \theta_{2q+1},
\]
so an estimator of the odd moment \(\theta_{2q+1}\) can be constructed as \(\hat{\theta}_{2q+1} = (n_q^{(+)} - n_q^{(-)})/(2N c_q c_{q+1})\). The variance of \(n_q^{(+)} - n_q^{(-)}\) is \(N(g_q^{(+)} + g_q^{(-)}) \approx N(c_q^2 \theta_{2q} + c_{q+1}^2 \theta_{2q+2})\), so the mean-square error becomes \([43, 44]\),
\[
\text{MSE}_{2q+1}^{\text{(SPADE)}} \approx \frac{1}{4N} \left( \frac{\theta_{2q}}{c_q^2} + \frac{\theta_{2q+2}}{c_{q+1}^2} \right) = \frac{O(\Delta^{2q})}{N},
\]
and the SNR becomes
\[
\text{SNR}_{2q+1}^{\text{(SPADE)}} \approx \frac{N(g_q^{(+)} - g_q^{(-)})^2}{g_q^{(+)} + g_q^{(-)}} = NO(\Delta^{2q+2}).
\]

For the first moment \((q = 0)\), the error is the same as the well known \(O(1)/N\) error for point-source localization \([17, 22–27]\). For the third and higher moments, however, there is significant enhancement over direct imaging. Note also that \(n_q^{(+)} + n_q^{(-)}\) can give information about the even moments as well.

For readers who wonder how the concept of Rayleigh’s curse should be generalized for moment estimation, see Appendix E.

### E. Fourier object analysis via moments

The moments can be used in a (generalized) Fourier analysis that may be more familiar to opticians \([7]\). Suppose that \(F(X)\) can be expanded as
\[
F(X) = \sum_{\mu=0}^{\infty} \hat{F}_\mu h_\mu(X) G(X),
\]
where \(G(X)\) is a nonnegative reference density, \(\{h_\mu(X) = \sum_{\nu=0}^\infty H_{\mu\nu} X^\nu : \mu \in \mathbb{N}_0\}\) are orthogonal polynomials that satisfy
\[
\int_{-\infty}^{\infty} dX G(X) h_\mu(X) h_\nu(X) = \delta_{\mu\nu},
\]
and \(\{\hat{F}_\mu\}\) are Fourier coefficients. Each \(h_\mu(X)\) has \(\mu\) distinct zeros on the support of \(G(X)\) \([104]\), so each \(h_\mu(X) G(X)\) can be regarded as a wavelet that exhibits localized oscillations. The Fourier coefficients can be expressed as
\[
\hat{F}_\mu = \int dX h_\mu(X) F(X) = \sum_{\nu=0}^\mu H_{\mu\nu} \theta_\nu.
\]
In other words, each Fourier coefficient of order \(\mu\) can be reconstructed from moments up to order \(\mu\). Thus the number of accurately estimated moments can be regarded as a measure of resolution, and SPADE can help by bringing in more accurate moments and increasing the number of obtainable Fourier coefficients for a subdiffraction object.

With a finite number of moments or Fourier coefficients and no other prior information, the reconstruction of \(F(X)\) is ill-posed and requires regularization \([7]\). Many linear or nonlinear algorithms can be used, depending on the application \([7, 105–107]\).

### F. Quantum limits

Through the Helstrom information, we have learned earlier that SPADE is optimal for estimating the separation of two point sources. References \([28, 43]\) show that direct imaging is close to optimal for locating a subdiffraction object with a known shape, while Ref. \([43]\) also shows that SPADE is close to optimal for estimating its size. Generalizing such results for arbitrary moments is much more difficult, as there are now an infinite number of parameters and an infinite number of spatial modes. Zhou and Jiang showed essentially \([47]\) that any measurement should give a Fisher information that scales with \(\Delta\)
\[
\text{FI}_\mu = NO(\Delta^{-\mu_1}), \quad \mu_1 \leq \mu,
\]
where \(\mu_1\) is an integer. With the Cramér-Rao bound \(\text{MSE}_\mu \geq 1/\text{FI}_\mu\), the SNR should scale as
\[
\text{SNR}_\mu \leq \theta_\mu^2 \text{HI}_\mu = NO(\Delta^{\mu_2}), \quad \mu_2 \geq \mu,
\]
where \(\mu_2\) is another integer. This means that, for a given \(\mu\), the SNR must decrease for smaller \(\Delta\), and the decrease is faster for higher \(\mu\). The best scaling with \(\Delta\) is achieved at \(\mu_1 = \mu_2 = \mu\), matching the scaling of the SPADE error given by Eq. (7.12) for the even moments. Zhou and Jiang did not provide a tractable bound on the prefactor of Eq. (7.26) however, so it remains a question whether SPADE is at all close to the quantum limit in absolute terms, or there may yet be superior measurements.

Using more standard quantum estimation theory, Ref. \([46]\) proves a quantum limit given by
\[
\text{FI}_\mu \leq H_{1\mu} \leq H_{1\mu}' = NO(\Delta^{-2(\mu/2)}),
\]
where \(H_{1\mu}'\) is an absolute limit that does not depend on the measurement and can be approximated analytically or numerically. The scaling of \(1/H_{1\mu}'\) with \(\Delta\) matches the scaling of the SPADE error given by Eqs. (7.12) and (7.21), suggesting that SPADE is close to quantum-optimal for both even and odd moments, but a more quantitative comparison of the quantum limit with the SPADE performance remains to be done. A limit on the SNR is
\[
\text{SNR}_\mu \leq \theta_\mu^2 H_{1\mu}' = NO(\Delta^{2(\mu/2)}).
\]
For a given subdiffraction object, Ref. \([46]\) also shows that \(\theta_\mu^2 H_{1\mu}'\) must decay quickly with higher \(\mu\), meaning that higher moments are fundamentally more difficult to estimate.

For the motivated readers, Appendices A–J introduce the mathematical concepts that underpin the recent studies of quantum limits to incoherent imaging.
VIII. OTHER GENERALIZATIONS

A. Unknown centroid

A crucial assumption in the preceding discussion is that the object is highly concentrated near a known coordinate \( X = 0 \), and the SPADE device is ideally aligned with \( X = 0 \). To put it the other way, \( \Delta \) should be regarded as the object width plus any misalignment of SPADE with the object centroid, and misalignment can reduce the enhancement by increasing the effective \( \Delta \). As direct imaging can locate the centroid accurately, the misalignment can be minimized if the object of interest has been imaged before and its centroid is already known accurately, as is often the case in astronomy. Otherwise some overhead photons should be used to locate the centroid first. References [42, 68] find that, despite the overhead, SPADE can still offer significant enhancements of the two-point resolution over direct imaging with the same total photon number.

In principle, it turns out to be possible to estimate the centroid and the separation simultaneously at the quantum limit if a multiphoton measurement is performed, as demonstrated by Parniak and co-workers [58, 75], but the applicability of their measurement to usual light sources is questionable.

B. Strong thermal light

While the model of weak thermal light and Poisson statistics works well for astronomical or fluorescent sources at optical frequencies [4, 12, 17, 20–26, 81], thermal sources at lower frequencies or scattered laser sources can exhibit super-Poisson statistics [3]. Nair computed the Helstrom information for separation estimation with the exact thermal state and also proposed variations of SPADE and SLIVER to approach it [50]. Lupo and Pirandola computed the quantum limit for the same problem but assumed arbitrary quantum states, including the thermal state as a special case [51]. Yang and co-workers studied the use of mode homodyne or heterodyne detection for the two-point problem and found that, although it is not competitive for weak thermal light, it can offer an enhancement over direct imaging for strong thermal light [56]. Appendices C and I summarize some useful bounds on the Fisher information for strong thermal light.

C. Two point sources with unequal brightnesses

Řeháček and co-workers studied the quantum limits and the optimal measurements for two point sources with unequal brightnesses in Refs. [59, 60]. They found that, while significant enhancements over direct imaging remain possible, the performance gets worse for unequal sources. With hindsight this is perhaps not surprising, as moments up to the third are needed to fully parametrize unequal sources and the SNR for the third moment is fundamentally poorer. The use of SPLICE for this case was also studied by Bonsma-Fisher and co-workers [69].

D. Partially coherent sources

Larson and Saleh studied the separation estimation problem for two partially coherent sources and suggested that Rayleigh’s curse would recur [65]. Their work has been challenged by Ref. [66], however, which points out a few problems with Ref. [65] and also shows that SPADE can overcome the curse as long as the sources are not highly correlated, contrary to the claim by Larson and Saleh; see also their reply [67]. In any case, the debate is irrelevant to observational astronomy and fluorescence microscopy, where there is no sound reason to doubt the established model of spatially incoherent sources [4, 12, 22–26, 81].

E. Two-dimensional imaging

Although we have so far focused on imaging in one dimension for simplicity, the same principles carry over to two dimensions. For two point sources, there are now two parameters for their vectorial separation. The quantum limits for the two parameters are the same as that for the one-dimensional case, and SPADE with respect to the transverse-electromagnetic (TEM) modes or a pair of SLIVER devices can still estimate the vectoral separation near the quantum limit [53]. For extended sources in two dimensions, a generalization of the PAD and iPAD modes have been studied in Refs. [43, 44, 47], and quantum limits have been studied in Refs. [43, 47].

F. Three-dimensional imaging

Reference [88] studies quantum limits to the three-dimensional localization of one point source as well as two coherent sources using the full vectoral electromagnetics model (the discussion of incoherent sources there is flawed and superseded by Ref. [42]). In the context of the paraxial model on the other hand, the axial dimension requires special treatment [80]. Backlund, Shechtman, and Walsworth computed the quantum limit to the three-dimensional localization of a point source in paraxial imaging and proposed special interferometers to achieve it [61]. Yu and Prasad [63, 64] and Napoli and co-workers [62] studied the same problem but for two incoherent sources. Zhou and co-workers recently demonstrated a FRFT interferometer to enhance the estimation of the axial separation between two sources [78].

G. Biased estimators

The simplest form of the Cramér-Rao bound is applicable to unbiased estimators only, and it turns out that biased estimators may violate it significantly [79]. For example, the Cramér-Rao bound for separation estimation with direct imaging blows up to infinity as \( \theta \to 0 \), but the maximum-likelihood estimator, being biased for this problem, can still achieve a finite error for all \( \theta \) [70, 71, 108]. For SPADE, the maximum-
likelihood estimator can also violate the Cramér-Rao bound and give a vanishing error as \( \theta \to 0 \) \cite{42}. Given these violations, one may wonder if the Cramér-Rao bound is meaningful outside the theoretical construct of asymptotic statistics \cite{79} after all. The loophole can be fixed by using a Bayesian version of the Cramér-Rao bound \cite{5, 109} that is valid for any biased or unbiased estimator, as introduced in Appendix J. Reference \cite{52} shows that, from the Bayesian and minimax perspectives, there remains a significant performance gap between direct imaging and SPADE for separation estimation, even if biased estimators are permitted.

H. One-versus-two hypothesis testing

Another way of defining the two-point resolution is to consider the error probabilities of deciding whether there is one point source or two point sources with the same total brightness. As mentioned in Sec. II, Helstrom performed a pioneering study of this problem using his quantum detection theory \cite{29}, but his proposed measurement depends on the separation in the two-source hypothesis, he did not suggest any experimental setup to realize it, and he did not show how much improvement it could offer. In the context of direct imaging, the problem was also studied in Refs. \cite{13–16}.

Coming in full circle, Lu and co-workers recently showed that the quantum limit to the hypothesis-testing problem is indeed a substantial improvement over direct imaging, and both SPADE and SLIVER can reach the quantum limit in the sub-Rayleigh regime, without knowing the separation in advance \cite{54}.

IX. COMPARISON WITH OTHER IMAGING TECHNIQUES

In the wider context of imaging research, SPADE is but one of the countless superresolution proposals in the literature. It nonetheless possesses many unique advantages and avoids some common pitfalls of prior ideas, thanks to its firm footing in quantum optics and statistics. Its advantages over direct imaging and computational techniques have already been emphasized in previous sections, and here we highlight some other important or popular ideas in imaging and how SPADE compares.

A. Stellar interferometry

SPADE perhaps bears the most resemblance to stellar interferometry \cite{4, 20, 110}, as they are both examples of applying coherent optical processing to incoherent imaging. Conventional wisdom suggests, however, that the advantage of stellar interferometry over direct imaging lies in its robustness against atmospheric turbulence, and it cannot compete with direct imaging if the imaging system is diffraction-limited \cite{4, 20}. To quote Goodman \cite{4}, “the reader may well wonder why the Fizeau stellar interferometer, which uses only a portion of the telescope aperture, is in any way preferred to the full telescope aperture in this task of measuring the angular diameter of a distant object. The answer lies in the effects of the random spatial and temporal fluctuations of the earth’s atmosphere (‘atmospheric seeing’)...” Furthermore, to quote Zmuidzinas \cite{20}, “it is important to remember that the imperfect beam patterns of sparse-aperture interferometers extract a sensitivity penalty as compared with filled-aperture telescopes, even after accounting for the differences in collecting areas.” Another idea that sounds similar to SLIVER is nulling interferometry \cite{110}, which was proposed for the specific purpose of exoplanet detection. The idea there is to remove the light from a bright star via destructive interference while leaving the light from a nearby planet intact.

Our theory offers the novel insight that, even for a diffraction-limited system, image-plane optical processing can provide further enhancements for imaging subdiffraction objects. It remains open questions whether nulling interferometry or similar ideas turn out to perform similarly to SLIVER or SPADE in the subdiffraction regime, and how these techniques and the quantum limits \cite{46, 47, 59, 60} may impact important astronomical applications, such as exoplanet detection.

B. Multiphoton coincidence

While modern stellar interferometers all rely on amplitude interference \cite{110}, also called a \( g^{(1)} \) measurement in quantum optics, the intensity interferometer by Hanbury Brown and Twiss—a \( g^{(2)} \) measurement—deserves a mention as well, since it inspired the foundation of quantum optics \cite{3} and is still being held in high regard by quantum opticians. In astronomy, however, the intensity interferometer has in fact been obsolete for decades because of its poor SNR \cite{4, 110}. It relies on the postselection of two-photon-coincidence events, which are much rarer than the one-photon events used in amplitude interferometry and therefore must give much less information in principle. For example, Davis and Tango reported an amplitude interferometer that obtained similar results to those from the intensity interferometer, using only \( \sim 2\% \) of the observation time \cite{111}. For microscopy, the use of multiphoton coincidence has recently been demonstrated in some heroic experiments \cite{112–115}, but again its statistical performance needs to be studied more carefully. SPADE, on the other hand, is a \( g^{(1)} \) measurement that relies on the much more abundant one-photon events without the need for coincidence detection and its statistical performance has been proved rigorously.

C. Electron microscopy and near-field microscopy

If the object is on a surface and accessible, then no technique can compete with electron microscopy, atomic force microscopy, and scanning-tunneling microscopy in terms of
resolution. Those techniques impose stringent requirements on the sample however, and that is why optical microscopy remains useful, especially for biological imaging, as it is able to image biological samples in a more natural environment and provide protein-specific contrast via fluorophore tagging.

In terms of optics, near-field techniques have not been successful because of the short depth of focus and other technical challenges [37]. In recent years, the use of plasmonics and metamaterials to enhance the near field [38] has also attracted immense interest in the academy, but the requirement of close proximity to the object and the impact of loss remain showstoppers in practice [116].

Being a far-field technique, SPADE is more compatible with biological imaging, not to mention its unique capability for astronomy and remote sensing. Unlike metamaterials, SPADE requires only low-loss optical components and there is no stringent requirement on their feature size, so fabrication is more straightforward.

D. Superresolution fluorescence microscopy

Far-field superresolution techniques such as PALM and STED have been hugely successful in biological fluorescence microscopy [37, 39, 40], but many of them rely on sophisticated control of the source emission, which introduces many other problems, such as the need for special fluorophores, slow speed in the case of PALM, and phototoxicity in the case of STED. SPADE, on the other hand, is a passive far-field measurement that can complement or supersede the superresolution techniques by extracting more information from the light or alleviating the need for source control. The combination of SPADE with microscope configurations, such as confocal and structured illumination [81], awaits further research.

E. Nonclassical light

The application of nonclassical light to sensing and imaging has been an active research topic in quantum optics for many decades [30–36, 117]. It is now well known, however, that nonclassical light is extremely fragile against loss and decoherence [117], and any theoretical advantage can be easily lost in practice, not to mention that the efficient generation and detection of nonclassical light remain challenging. More recent proposals, such as quantum illumination and quantum reading [35], apply to high-noise scenarios, but the achievable improvement turns out to be quite modest even in theory [118].

As SPADE works with classical light, linear optics, and photon counting, loss and other imperfections are not nearly as detrimental. If we are to believe that the second quantum revolution is near and applications using nonclassical resources will soon be widespread [119], then SPADE should be an even surer bet.

For astronomy, obviously the light sources cannot be controlled, but the use of entangled photons and quantum repeaters has been proposed to teleport photons in stellar interferometry and increase its baseline [120–122]. Unfortunately, quantum repeaters are nowhere near practical yet, and conventional linear optical devices remain the best option in the foreseeable future.

F. Superoscillation, amplification, postselection

There are so many other superresolution ideas that going through them all would not be feasible. We list here only a few more: superoscillation [123], amplification [124], and postselection [125]. They either require steep trade-offs with the SNR or have questionable statistics [126, 127]. These examples once again demonstrate the importance of a rigorous analysis using quantum optics and statistics. It is important to keep in mind that superresolution is possible even with direct imaging and data processing, and it is ultimately limited by the SNR [7]. A superresolution technique is viable only if it can beat direct imaging on statistical terms.

X. CONCLUSION

Just as the design of engines must go beyond mechanics and consult thermodynamics, the design of optical sensing and imaging systems must go beyond electromagnetics and consult statistics. With the increasingly dominant role of photon shot noise in modern applications, quantum mechanics is also relevant. Quantum information theory can tackle all these subjects in one unified formalism, setting limits to what we can do, and also telling us how much further we can go. For incoherent imaging, it gives us the pleasant surprise that there is still plenty of room for improvement, and we just need to find a way to achieve it. We found one in the form of SPADE, which requires only low-loss linear optics and photon counting. While we started with the simple model of two point sources, we have since generalized the theory to deal with any subdiffraction object, showing that substantial improvements remain possible. The theoretical groundwork has been laid, proof-of-principle experiments have been done, and applications in astronomy and fluorescence microscopy can now be envisioned. Special-purpose applications that require only the low-order moments, such as two-point resolution and object-size estimation, should be the first to benefit, while more general imaging protocols will require further research.

Many open problems still remain. On the theoretical side, the exact quantum limits to general imaging and the optimal measurements to achieve them remain unclear. The theory for three-dimensional imaging and spectroscopy remains underdeveloped. On the practical side, an efficient implementation of SPADE at the right wavelengths is needed for applications. The performance of SPADE in the presence of atmospheric turbulence and other technical noises also needs to be assessed. Fortunately, adaptive optics [128] and photodetectors [129] have become so good in recent years that we can be a lot more optimistic than Helstrom about reaching the quantum limits in the near future.
ACKNOWLEDGMENTS

We are grateful to the seminal contributions of the authors of Refs. [42–78], especially the crucial roles of Ranjith Nair, Xiao-Ming Lu, and Shan Zheng Ang in our early papers. We acknowledge the financial support from the Singapore National Research Foundation under Project No. QEP-P7.

Appendix A: Cramér-Rao bound and Fisher information

Let \( \{ P_Y(y|\theta) > 0 : y \in \Omega, \theta \in \Theta \subset \mathbb{R} \} \) be a family of probability distributions for an observed random variable \( Y \), where \( \theta \) is an unknown scalar parameter and the support \( \Omega \) is assumed to be countable and common to all distributions for simplicity. Let \( \hat{\theta}(Y) \) be an estimator of \( \theta \). Define the mean-square error as

\[
\text{MSE}(\theta) \equiv \mathbb{E} \left[ (\hat{\theta}(Y) - \theta)^2 \right] = \sum_y P_Y(y|\theta) \left[ \hat{\theta}(y) - \theta \right]^2 ,
\]

(A1)

where \( \mathbb{E} \) denotes the expectation. The unbiased condition is

\[
\mathbb{E} \left[ \hat{\theta}(Y) \right] = \theta .
\]

(A2)

Under certain regularity conditions on the distributions \([130]\), the Cramér-Rao bound given by Eq. (3.1) holds for any unbiased estimator, where the Fisher information is \([5, 79]\)

\[
\text{FI}(\theta) \equiv \sum_y \frac{1}{P_Y(y|\theta)} \left[ \frac{\partial P_Y(y|\theta)}{\partial \theta} \right]^2 .
\]

(A3)

Generalization for probability densities is straightforward \([5, 79]\).

Appendix B: Helstrom information

Let \( \{ \rho(\theta) : \theta \in \Theta \subset \mathbb{R} \} \) be a family of density operators for a quantum object. Under a quantum measurement, the generalized Born’s rule is given by

\[
P_Y(y|\theta) = \text{tr} \ E_Y(y) \rho(\theta),
\]

(B1)

where \( \text{tr} \) denotes the operator trace and \( E_Y(y) \) is called the positive operator-valued measure (POVM), which models the measurement statistics \([131]\). Define the Helstrom information as \([27]\)

\[
\text{HI} = \text{tr} \ L^2 = \text{tr} \ \frac{\partial \rho}{\partial \theta} \ L,
\]

(B2)

where \( L \) is a solution to

\[
\frac{\partial \rho}{\partial \theta} = \frac{1}{2} \left( \rho L + L \rho \right).
\]

(B3)

For any POVM, Helstrom proved \( \text{MSE} \geq \text{HI}^{-1} \) \([27]\), while Nagaoka \([82]\) and Braunstein and Caves \([83]\) proved

\[
\text{FI}(\theta) \leq \text{HI}(\theta) .
\]

(B4)

Although they also proved that \( \max_{E_Y} \text{FI}(\theta) = \text{HI}(\theta) \), and a projection in the eigenstates of \( L \) gives an optimal POVM, it is important to keep in mind that \( L \) is a function of \( \theta \), and the optimal POVM derived from it at one value of \( \theta \) may be suboptimal at other values. In practice, obviously \( \theta \) is unknown, and there is no guarantee that one can find a POVM that is optimal across a range of \( \theta \). A solution, proposed by Nagaoka and refined by Hayashi and Matsumoto \([84]\) and Fujiiwara \([85]\), is to consider repeated adaptive measurements, and they showed that the total Fisher information of such measurements can approach the Helstrom information in the limit of infinitely many measurements under certain technical conditions.

The Helstrom information has more tractable expressions in the following cases:

1. Express \( \rho \) in the diagonal form

\[
\rho = \sum_j \lambda_j |e_j\rangle \langle e_j| ,
\]

(B5)

where \( \{ \lambda_j \geq 0 \} \) are the eigenvalues of \( \rho \) and \( \{|e_j\} \) are the orthonormal eigenkets, including all the ones that give \( D_{jk} \equiv \langle e_j| \partial \rho / \partial \theta | e_k \rangle \neq 0 \). Then

\[
L = \sum_{j,k} \frac{2D_{jk}}{\lambda_j + \lambda_k} |e_j\rangle \langle e_k| , \quad \text{HI} = \sum_{j,k} \frac{2|D_{jk}|^2}{\lambda_j + \lambda_k} .
\]

(B6)

For example, Ref. \([42]\) uses this brute-force method to compute Eq. (4.1) for the two-point problem, since the dimension of the problem is manageable.

2. If \( \rho = |e\rangle \langle e| \) is a pure state, \( \rho^2 = \rho \), and

\[
L = 2 \frac{\partial \rho}{\partial \theta} = 2 \left( \frac{\partial |e\rangle \langle e|}{\partial \theta} + |e\rangle \langle e| \frac{\partial |e\rangle \langle e|}{\partial \theta} \right) ,
\]

(B7)

\[
\text{HI} = 4 \left[ \frac{\partial |e\rangle \langle e|}{\partial \theta} \frac{\partial |e\rangle \langle e|}{\partial \theta} - \left( i \langle e| \frac{\partial |e\rangle \langle e|}{\partial \theta} \right)^2 \right] .
\]

(B8)

Note that \( i \langle e| \frac{\partial |e\rangle}{\partial \theta} \) is real because \( 2 \text{Re}(\langle e| \frac{\partial |e\rangle}{\partial \theta}) = \langle e| \frac{\partial |e\rangle}{\partial \theta} + (\frac{\partial |e\rangle}{\partial \theta}) |e\rangle = \partial(\langle e| e\rangle)/\partial \theta = 0 \), where \( \text{Re} \) denotes the real part.

Equation (B8) is useful for computing upper bounds on the Helstrom information, as elaborated in Appendix F.

3. For a tensor product of density operators \( \rho = \otimes_j \rho_j \),

\[
\text{HI}^{(o)} = \sum_j \text{HI}^{(\rho_j)} ,
\]

(B9)

where \( \text{HI}^{(\rho)} \) denotes the Helstrom information for density operator \( \rho \). This is useful for dealing with multiple temporal modes.

4. Let

\[
\mathcal{F}(\rho, \sigma) \equiv \text{tr} \sqrt{\sigma \rho \sqrt{\sigma}}
\]

(B10)
be the fidelity between two states $\rho$ and $\sigma$ [131, 132]. It can be shown [83, 131, 133, 134] that

$$\text{HI}(\theta) = \lim_{\delta \to 0} \frac{8}{\delta^2} [1 - \mathcal{F}(\rho(\theta), \rho(\theta + \delta))], \tag{B11}$$

which leads to

$$\mathcal{F}(\rho(\theta), \rho(\theta + \delta)) = 1 - \frac{\delta^2}{8} \text{HI}(\theta) + o(\delta^2), \tag{B12}$$

$$\text{HI}(\theta) = -4 \left. \frac{\partial^2}{\partial q^2} \mathcal{F}(\rho(\theta), \rho(\theta')) \right|_{\theta' = \theta}, \tag{B13}$$

where $o(\delta^2)$ denotes terms asymptotically smaller than $\delta^2$. These formulas are convenient if the fidelity is easier to compute; see Ref. [50] for an example. The relation also means that the Helstrom information inherits many properties of the fidelity; some important ones are discussed in Appendices F–H.

5. If $\rho$ is a thermal state with mutual coherence matrix $\Gamma$ [3], Helstrom showed that [27]

$$\text{HI} = \text{tr} \frac{\partial}{\partial \theta} \bar{\Upsilon}, \tag{B14}$$

where $\bar{\Upsilon}$ is a solution to

$$\frac{\partial}{\partial \theta} = \frac{1}{2} \{\Gamma \bar{\Upsilon}(I + \Gamma) + (I + \Gamma)\bar{\Upsilon}\Gamma\}, \tag{B15}$$

$I$ being the identity matrix. Reference [46] uses this to prove bounds for thermal states, as elaborated in Appendix C.

6. If $\rho$ is a Gaussian state, a general formula has been derived by Monras [135].

### Appendix C: Thermal state in the ultraviolet limit

Consider thermal light in one temporal mode and multiple spatial modes, and let $\{q_{\lambda}\}$ be the annihilation operators for the spatial modes. As first proposed by Glauber [136], the thermal state is [3, 27]

$$\sigma = \mathbb{E}(|\alpha\rangle \langle\alpha|), \tag{C1}$$

where $\alpha$ is a column vector of zero-mean complex Gaussian random variables and $|\alpha\rangle$ is a multimode coherent state that obeys $a_q |\alpha\rangle = \alpha_q |\alpha\rangle$ [3]. The statistics of $\alpha$ are completely determined by the mutual coherence matrix $\Gamma$ in statistical optics [3, 4], viz.,

$$\mathbb{E}(\alpha) = 0, \quad \mathbb{E}(\alpha\alpha^\dagger) = \Gamma, \tag{C2}$$

where $\dagger$ denotes the transpose and $\dagger$ denotes the Hermitian transpose. The photon-counting distribution is

$$P(n) = \langle n | \sigma | n \rangle = \mathbb{E}(|\langle n | \alpha \rangle|^2), \tag{C3}$$

$$|n\rangle = \prod_q (a_q^\dagger)^{n_q} |\text{vac}\rangle, \tag{C4}$$

$$|\langle n | \alpha \rangle|^2 = \exp(-\alpha^\dagger \alpha) \prod_q \frac{[\alpha_q]^{2n_q}}{n_q!}, \tag{C5}$$

where $|n\rangle$ is a Fock state and $|\text{vac}\rangle$ is the vacuum state. Equation (C3) agrees with the semiclassical theory by Mandel [3]. With $M$ temporal modes, the density operator can be modeled as $M$ copies of $\sigma$, or

$$\rho = \sigma^{\otimes M}. \tag{C6}$$

To simplify the thermal state for optical frequencies, let

$$\epsilon \equiv \text{tr} \Gamma \tag{C7}$$

be the average photon number per temporal mode and

$$g \equiv \frac{\Gamma}{\text{tr} \Gamma} \tag{C8}$$

be the normalized mutual coherence matrix. Define the ultraviolet limit as $\epsilon \to 0$ while holding $N = M \epsilon$ constant. The zero-photon probability per temporal mode is

$$P(0, \ldots, 0) = \mathbb{E} \left[ \exp(-\alpha^\dagger \alpha) \right] = 1 - \epsilon + O(\epsilon^2), \tag{C9}$$

the one-photon probability is

$$P(0, \ldots, n_q = 1, 0, \ldots) = \mathbb{E} \left[ \exp(-\alpha^\dagger \alpha)|\alpha_q|^2 \right] = \epsilon g_q + O(\epsilon^2), \tag{C10}$$

where the diagonal entries of a matrix are abbreviated as $g_{qq} = g_q$, and the probability of two or more photons is $O(\epsilon^2)$. The photon counts summed over $M$ temporal modes hence become Poisson in the ultraviolet limit [4, 20]. A simplified quantum model in this limit is [42, 102, 120]

$$\sigma = (1 - \epsilon) |\text{vac}\rangle \langle\text{vac}| + \epsilon \rho_1 + O(\epsilon^2), \tag{C11}$$

where the one-photon density operator is

$$\rho_1 = \sum_{q, p} g_{qp} |\phi_q\rangle \langle\phi_p|, \quad |\phi_q\rangle = a_q^\dagger |\text{vac}\rangle. \tag{C12}$$

For paraxial incoherent imaging in particular [43, 46],

$$\rho_1 = \int_{-\infty}^{\infty} dX F(X)e^{-i\hat{k}X} |\psi\rangle \langle\psi| e^{i\hat{k}X}, \tag{C13}$$

where $\hat{k}$ is the spatial-frequency or momentum operator, $|\psi\rangle$ is the one-photon state with spatial wavefunction $\langle x |\psi\rangle = \psi(x)$, and $|x\rangle$ is the one-photon position eigenket that obeys $\langle x |\psi\rangle = \delta(x - x')$. $f(x) = \langle x |\rho_1 |x\rangle$ gives Eq. (5.3), while $g_q = \langle \phi_q |\rho_1 |\phi_q\rangle$ gives Eq. (5.4). If $f$ and $g$ depend on $\theta$ (but $\epsilon$ does not), the Fisher information for the Poisson processes is given by Eqs. (5.6) and (5.8).

The ultraviolet limit and the negligence of $O(\epsilon^2)$ terms mean that multiphoton coincidence events and bunching effects are ignored [3, 4]. Besides thermal sources, the model here also applies to any incoherent sources, such as fluorescent sources [22–26, 81] or even electrons [10, 11], as long as they obey an incoherent-imaging model with Poisson counting statistics.
Reference [46] shows that the information for the thermal \( \sigma \) given by Eqs. (B14) and (B15) on a per-photon basis is upper-bounded by its ultraviolet limit, which coincides with the information for the \( \rho_1 \) given by Eq. (C12), viz.,

\[
\frac{H^I(\sigma)}{\epsilon} \leq \lim_{\epsilon \to 0} \frac{H^I(\sigma)}{\epsilon} = H^I(\rho_1). \quad (C14)
\]

With \( H^I(\rho) = M H^I(\sigma) \) for \( M \) temporal modes, the total information in the ultraviolet limit becomes

\[
H^I(\rho) \leq \lim_{\epsilon \to 0} H^I(\rho) = NH^I(\rho_1), \quad (C15)
\]

which means that \( H^I(\rho_1) \) also serves as a limit for thermal states with arbitrary \( \epsilon \).

**Appendix D: Multiparameter estimation**

Now suppose that \( \theta \in \Theta \subset \mathbb{R}^K \) is a column vector of parameters, and the estimator is also a vector. Define the mean-square error covariance matrix as

\[
\text{MSE}_{\mu\nu}(\theta) \equiv \mathbb{E} \left[ \left( \hat{\theta}_\mu(Y) - \theta_\mu \right) \left( \hat{\theta}_\nu(Y) - \theta_\nu \right) \right]. \quad (D1)
\]

Diagonal entries of a matrix are again abbreviated as \( \text{MSE}_{\mu\mu} = \text{MSE}_\mu \). The multiparameter Cramér-Rao bound [5] can be expressed as the matrix inequality [131]

\[
\text{MSE} \geq \text{CRB} \equiv F^{-1}, \quad (D2)
\]

\[
\text{FI}_{\mu\nu}(\theta) \equiv \sum_y \frac{1}{P_Y(y|\theta)} \frac{\partial P_Y(y|\theta)}{\partial \theta_\mu} \frac{\partial P_Y(y|\theta)}{\partial \theta_\nu}. \quad (D3)
\]

The matrix inequality means that \( \text{MSE} - \text{CRB} \) is positive-semidefinite, or equivalently \( u^\top (\text{MSE} - \text{CRB}) u \geq 0 \) for any real column vector \( u \). For example, explicit expressions of the multiparameter Cramér-Rao bound for moment estimation with direct imaging have been derived in Refs. [43, 44].

The Helstrom information matrix is defined as

\[
H_{\mu\nu} = \text{Re} \text{tr} \rho L_{\mu\nu} L_{\nu} = \text{tr} \frac{\partial \rho}{\partial \theta_\mu} L_{\nu}, \quad (D4)
\]

\[
\frac{\partial \rho}{\partial \theta_\mu} = \frac{1}{2} (\rho L_{\mu} + L_{\mu} \rho). \quad (D5)
\]

The matrices can be shown to inherit all the properties of their scalar version by substituting the directional derivative \( \partial/\partial \theta = \sum_{\mu} u_\mu \partial/\partial \theta_\mu \) and \( L = \sum_{\mu} u_\mu L_{\mu} \) for an arbitrary real vector \( u \). For example, upon the substitutions, the scalar Fisher information becomes \( u^\top \text{FI} u \) and the scalar Helstrom information becomes

\[
\text{tr} \rho L^2 = u^\top H u = u^\top H_{\mu\nu} u = \text{tr} \rho L_{\mu\nu} L_{\nu}, \quad (D6)
\]

where we have used the fact that, since \( u^\top H u \) and \( u \) are real, \( u^\top H u = \text{Re} \sum_{\mu} u_\mu \bar{u}_\mu H_{\mu\nu} u_\nu = \sum_{\mu} u_\mu \text{Re}(H_{\mu\nu}) u_\nu = \sum_{\mu} u_\mu H_{\mu\nu} u_\nu \). The Nagaoka bound given by Eq. (B4) becomes \( u^\top \text{FI} u \leq u^\top H u \), meaning that Eq. (B4) still holds as a matrix inequality. A consequence of the matrix inequality is that the inverses obey the reverse relation [131], so the Nagaoka bound leads to

\[
\text{MSE} \geq G^{-1} \geq H^{-1}. \quad (D7)
\]

For multiple parameters, it may not be possible for the Fisher information matrix of any measurement to achieve the Helstrom value, and there exist tighter quantum bounds [137].

An important property of the Fisher information is that it transforms like a metric tensor upon reparametrization. Define a new set of parameters \( \xi \) via an invertible function of \( \theta \). The Fisher information matrix with respect to the new parameters becomes

\[
\text{FI}[\xi] = J^\top \text{FI}[\theta] J, \quad J_{\mu \nu} = \frac{\partial \theta_\mu}{\partial \xi_\nu}, \quad (D8)
\]

and the Cramér-Rao bound becomes

\[
\text{CRB}[\xi] = J^{-1} \text{CRB}[\theta] (J^{-1})^\top, \quad (J^{-1})_{\mu \nu} = \frac{\partial \xi_\nu}{\partial \theta_\mu}. \quad (D9)
\]

The Helstrom information and the Helstrom bound transform in the same way.

**Appendix E: Rayleigh’s curse for moment estimation**

It must be stressed that a zero of the Fisher information at one parameter value may be removed simply by reparametrization. For example, if \( \theta \) is the separation between two sources, the direct-imaging information is given by Eq. (3.2), where \( C(\theta) \propto \theta^2 \) for \( \theta \ll 1 \) if the point-spread function is Gaussian. If the parameter is redefined as \( \xi = \theta^2 \), then \( (d\theta/d\xi)^2 = 1/(4\theta^2) \), and the transformed information \( \text{FI}[\xi] = (d\theta/d\xi)^2 \text{FI}[\theta] \) no longer vanishes as \( \theta \to 0 \). Thus a sound definition of Rayleigh’s curse should not be based on the zeros or the absolute value of the Fisher information for a particular parametrization.

A better definition is to compare the Fisher information with a reference value that assumes sparse sources. For two point sources, an obvious reference is

\[
\text{RI}(\theta) = C(\infty) N, \quad (E1)
\]

which is the information assuming that the sources can be located individually as if they were well separated. Rayleigh’s curse can therefore be defined as a reduction in the ratio of the Fisher information to Eq. (E1) for small \( \theta \).

To define the reference more precisely and generally for multiple sources and multiple parameters, suppose that the object consists of discrete point sources, each with relative intensity \( p_s \) and position \( \xi_s \). The object intensity becomes

\[
F(X) = \sum_s p_s \delta(X - \xi_s). \quad (E2)
\]

Let \( \{\xi_0, \xi_1, \ldots, p_1, p_2, \ldots\} \) be the unknown parameters and \( p_0 = 1 - \sum_{s=1}^\infty p_s \). Define the reference information matrix
It can be shown that Eq. (E4) gives Eq. (E1) for two equally bright point sources upon reparametrization. The reference bound becomes

\[
\begin{align*}
\mathcal{R}^B_{\xi,p} &= \left( \mathcal{R}^{[\xi,p]}_{\ell} \right)^{-1} = \left( \mathcal{R}^{[\xi]}_{\ell} 0 \right), \\
\mathcal{R}^B_{\ell} &= \frac{\delta_{st}}{C(\infty)N p_s}, \\
\mathcal{R}^B_{\xi} &= \frac{1}{N} \left( p_s \delta_{st} - p_s \mu \right).
\end{align*}
\]

If \( \psi(x) \) is a real function multiplied by any constant phase,

\[
\begin{align*}
\langle \psi | \hat{k} | \psi \rangle &= 0, \\
\langle \psi | \hat{k}^2 | \psi \rangle &= \frac{1}{4} \int_{-\infty}^{\infty} dx \frac{1}{|\psi(x)|^2} \left| \frac{\partial |\psi(x)|^2}{\partial x} \right|^2, \quad (E10)
\end{align*}
\]

and Eq. (E8) becomes the well known Cramér-Rao bound on the localization errors for direct imaging of sparse point sources [17, 22–26, 28]. Stochastic localization microscopy [37], for example, can achieve this bound by making different subsets of sparse sources emit in each image. A generalized Rayleigh’s curse can therefore be defined as any substantial increase in the errors with respect to this reference bound due to sub-Rayleigh separations. As shown in Appendix H, the Helstrom information is below this reference value, viz.,

\[
\mathcal{H}^{[\xi,p]} \leq \mathcal{R}^{[\xi,p]},
\]

so the reference is indeed a quantum limit.

In terms of the moment parameters \( \theta_{\mu} = \sum_s p_s \xi_{\mu}^s \), the reparametrized reference is now

\[
\begin{align*}
\mathcal{R}_{\mu \nu} &= \sum_{s,t} \frac{\partial \theta_{\mu}}{\partial \xi_{s}} \frac{\partial \theta_{\nu}}{\partial \xi_{t}} \mathcal{R}^B_{\ell} + \sum_{s,t} \frac{\partial \theta_{\mu}}{\partial p_s} \frac{\partial \theta_{\nu}}{\partial p_t} \mathcal{R}^B_{\xi} \\
&= \frac{1}{N} \left[ \mu \nu \theta_{\mu+\nu} - \theta_{\mu} \theta_{\nu} \right], \\
\mathcal{R}^B_{\mu} &= \frac{1}{N} \left[ \mu^2 \theta_{2\mu} - 2 \mu \theta_{\mu} - \theta_{\mu}^2 \right] = \frac{O(\Delta^{2\mu-2})}{N}. \quad (E13)
\end{align*}
\]

One can see that Eqs. (7.12) and (7.21) for SPADE and the quantum limit given by the inverse of Eq. (7.28) all obey worse scalings with \( \Delta \) than Eq. (E15) for \( \Delta \ll 1 \) and \( \mu \geq 3 \), so the generalized Rayleigh’s curse is fundamental starting from the third moment. This conclusion is similar to that by Zhou and Jiang [47], at least on a superficial level.

**Appendix F: Ignorance inequalities**

The Fisher and Helstrom information quantities obey a “no-free-lunch” principle analogous to the second law of thermodynamics: they cannot increase under any operation that does not depend on the parameter. The principle is technically called monotonicity and underlies most of the techniques for computing upper bounds on the information.

In the classical case, a version of monotonicity can be derived by considering a pair of random variables \( Y \) and \( Z \) [138]. It is straightforward to show that the Fisher information for the joint distribution \( P_{YZ}(y,z|\theta) = P_{Y|Z}(y|z,\theta)P_{Z}(z|\theta) = P_{Z|Y}(z|y,\theta)P_{Y}(y|\theta) \) is

\[
\begin{align*}
\mathcal{F}^{(Y)}(\theta) &= \sum_{y} P_{Y}(y|\theta) \mathcal{F}^{(Y|Z)}(y,\theta) + \mathcal{F}^{(Z)}(\theta) \\
&= \sum_{y} P_{Y}(y|\theta) \mathcal{F}^{(Z|Y)}(y,\theta) + \mathcal{F}^{(Y)}(\theta), \quad (F1)
\end{align*}
\]

where \( \mathcal{F}^{(O)} \) denotes the Fisher information for \( P_{O} \). Since \( \mathcal{F} \geq 0 \), we obtain

\[
\mathcal{F}^{(Z)} \leq \mathcal{F}^{(Y)}, \quad \mathcal{F}^{(Y)} \leq \mathcal{F}^{(Y|Z)} \cdot \quad (F3)
\]

These inequalities mean that ignoring any part of the data cannot increase the information.

In the quantum case, let \( \sigma \) be a density operator in Hilbert space \( \mathcal{H}_A \otimes \mathcal{H}_B \) and

\[
\rho = \text{tr}_B \sigma, \quad \text{tr}_B \text{ denotes the partial trace over } \mathcal{H}_B \text{ and means that the } \text{degrees of freedom in } \mathcal{H}_B \text{ are ignored. It can be proved that } [131]
\]

\[
\mathcal{H}^{(\sigma)} \leq \mathcal{H}^{(\sigma)} \cdot \quad (F5)
\]

A simple proof is as follows: Let \( \tau \) be a purification of \( \sigma \) in \( \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C \), which satisfies \( \sigma = \text{tr}_C \tau \). Then it can be shown via Uhlmann’s theorem for the fidelity [131, 132] and the relation between the Helstrom information and the fidelity that

\[
\mathcal{H}^{(\sigma)} = \inf_{\tau} \mathcal{H}^{(\tau)}. \quad \text{(F6)}
\]

Since purifications of \( \sigma \) are also purifications of \( \rho \) with \( \rho = \text{tr}_B \sigma \), \( \mathcal{H}^{(\sigma)} \leq \mathcal{H}^{(\tau)} \) for all \( \tau \), and we obtain \( \mathcal{H}^{(\sigma)} \leq \inf_{\tau} \mathcal{H}^{(\tau)} = \mathcal{H}^{(\sigma)} \).

Reference [46] uses the ignorance inequality, or more specifically Eq. (F6), to prove Eq. (7.28). The trick is to find a judicious purification \( \sigma \) of \( \rho \) such that \( \mathcal{H}^{(\sigma)} \) is easier to compute via Eq. (B8) and also offers a reasonably tight bound.

**Appendix G: Data-processing inequalities**

Another version of monotonicity in the classical case can be derived by assuming that \( P_{Y|Z} \) does not depend on \( \theta \), such
that $F_{\mid Z}^{(V)} = 0$ and $F_{\mid Z}^{(V)} = F_{\mid Z}^{(Z)}$ from Eq. (F1). It then follows from Eq. (F3) that
\[
F_{\mid Y}^{(V)} \leq F_{\mid Z}^{(Z)}.
\]
This is called the data-processing inequality (DPI), as $Y$ follows from $Z$ in a Markov chain that can model any data processing. Given Eq. (G1), it is also possible to prove Eq. (F3), as ignoring data is a form of processing.

A practical consequence of the DPI is that the direct-imaging information given by Eq. (5.6), which assumes infinitesimally small pixels, is the highest that can be obtained by any processing of the spatial intensity on the image plane. For example, in direct imaging with large pixels, each pixel measurement can be modeled as the sum of the photon counts from the infinitesimal pixels that cover the area, so the information must be less than Eq. (5.6) by virtue of the DPI. A more complicated example is the intensity masks proposed by Bertero and co-workers for incoherent imaging \[7, 139\], which also process the intensity only and must lead to an information lower than Eq. (5.6).

Appendix J: Bayesian Cramér-Rao bound

We have so far adopted the perspective of asymptotic statistics \[79\], but it has many conceptual and technical issues, especially when the sample size is finite. These issues can be

The upper bound is a quantum analog of Eqs. (F1) and (F3). If $P_{\theta}$ does not depend on $\theta$, then $F_{\mid Z}^{(Z)} = 0$, and the property is simply known as convexity \[144\]. If $\tau$ does not depend on $\theta$ on the other hand, $H_{\theta}^{(\tau)} = 0$, and the bound is known as a classical-simulation bound \[145\].

For example, for a distribution of discrete point sources, combining Eqs. (C13) and (E2) yields
\[
\rho_1 = \sum_s p_s \tau(s, \xi),
\]
and extended convexity leads to Eq. (E12). In other words, the right-hand side of Eq. (E12) is the information upon observing each photon and also the random variable $S$ with $P_S(s) = p_s$. Knowing $S$ reveals the point source that emits the photon and enables photometry and localization of each point source using only the photons from it. This can be done with direct imaging of sparse sources and is in fact the principle behind stochastic localization microscopy \[37\], which can achieve Eq. (E12). Without any control of the sources, Eq. (7.28) is a tighter quantum limit for subdiffraction objects. See also Refs. \[43, 46\] for derivations of quantum bounds for location, size, and moment parameters using convexity and classical simulation.
overcome by a Bayesian approach. For parameter estimation, the approach assumes that the parameter $\theta$ is a random variable with a certain prior probability density $f_\theta(\vartheta)$. Let the support of $f_\theta$ be $\Theta' \subseteq \Theta$. The Bayesian mean-square error is

$$BMSE \equiv \mathbb{E}[\text{MSE}(\theta)] = \int_{\Theta'} d^K \theta f_\theta(\vartheta) \text{MSE}(\vartheta). \quad (J1)$$

For any estimator—biased or unbiased—and a $f_\theta(\vartheta)$ that vanishes on the boundary of $\Theta'$, a Bayesian Cramér-Rao bound is [5, 109, 147]

$$BMSE \geq \left\{ \mathbb{E}[\text{FI}(\theta)] + \text{FI}(\theta) \right\}^{-1}, \quad (J2)$$

where

$$\text{FI}(\theta) = \int_{\Theta'} d^K \theta \frac{1}{f_\theta(\vartheta)} \frac{\partial f_\theta(\vartheta)}{\partial \theta_\mu} \frac{\partial f_\theta(\vartheta)}{\partial \theta_\nu} \quad (J3)$$

is the prior Fisher information. The bound was first proposed by Schützenberger [109] and rediscovered by Van Trees [5].

A quantum version in terms of the Helstrom information is straightforward given Eq. (B4).

The Bayesian bound resembles the frequentist version given by Eq. (3.1) when $\text{FI}(\theta)$ is constant over $\Theta'$ and the prior information $\text{FI}(\theta)$ is negligibly small, but note that the Bayesian bound applies to the “global” error $BMSE$, not the error function $\text{MSE}(\theta)$, and it is valid for any biased or unbiased estimator.

The Bayesian bound is also useful for minimax estimation, the goal of which is to minimize the worst-case error $\sup_{\theta \in \Theta} \text{MSE}(\theta)$ [79]. Through the relation

$$\sup_{\theta \in \Theta} \text{MSE}(\theta) \geq BMSE \quad (J4)$$

for any $f_\theta$, lower bounds on $BMSE$ also apply to the worst-case error, and the bounds can be tightened by choosing $f_\theta$ judiciously. See Ref. [52] for an application of the Bayesian bound and the minimax perspective to the two-point problem.

[1] Lord Rayleigh, “XXXI. Investigations in optics, with special reference to the spectroscope,” Philosophical Magazine Series 5 8, 261–274 (1879).
[2] Richard P. Feynman, Robert B. Leighton, and Matthew Sands, The Feynman Lectures on Physics, Vol. 1 (California Institute of Technology, Pasadena, 2013) Chap. 30. Diffraction.
[3] Leonard Mandel and Emil Wolf, Optical Coherence and Quantum Optics (Cambridge University Press, Cambridge, 1995).
[4] Joseph W. Goodman, Statistical Optics (Wiley, New York, 1985).
[5] Harry L. Van Trees, Detection, Estimation, and Modulation Theory, Part I. (John Wiley & Sons, New York, 2001).
[6] A. J. den Dekker and A. van den Bos, “Resolution: a survey,” Journal of the Optical Society of America A 14, 547–557 (1997).
[7] Geoffroy de Villiers and E. Roy Pike, The Limits of Resolution (CRC Press, Boca Raton, 2016).
[8] Oscar Falconi, “Limits to which Double Lines, Double Stars, and Disks can be Resolved and Measured,” J. Opt. Soc. Am. 57, 987–993 (1967).
[9] Ming-Jer Tsai and Keh-Ping Dunn, Performance Limitations on Parameter Estimation of Closely Spaced Optical Targets Using Shot-Noise Detector Model, Tech. Rep. ADA073462 (Lincoln Laboratory, MIT, 1979).
[10] E. Bettens, D. Van Dyck, A. J. den Dekker, J. Sijbers, and A. van den Bos, “Model-based two-object resolution from observations having counting statistics,” Ultramicroscopy 77, 37–48 (1999).
[11] S. Van Aert, A. J. den Dekker, D. Van Dyck, and A. van den Bos, “High-resolution electron microscopy and electron tomography: resolution versus precision,” Journal of Structural Biology 138, 21–33 (2002).
[12] Sripad Ram, E. Sally Ward, and Raimund J. Ober, “Beyond Rayleigh’s criterion: A resolution measure with application to single-molecule microscopy,” Proceedings of the National Academy of Sciences of the United States of America 103, 4457–4462 (2006).
[13] J. L. Harris, “Resolving Power and Decision Theory,” Journal of the Optical Society of America 54, 606–611 (1964).
[14] Carmen O. Acuna and Joseph Horowitz, “A statistical approach to the resolution of point sources,” Journal of Applied Statistics 24, 421–436 (1997).
[15] M. Shahram and P. Milanfar, “Imaging below the diffraction limit: a statistical analysis,” IEEE Transactions on Image Processing 13, 677–689 (2004).
[16] M. Shahram and P. Milanfar, “Statistical and Information-Theoretic Analysis of Resolution in Imaging,” IEEE Transactions on Information Theory 52, 3411–3437 (2006).
[17] Edward J. Farrell, “Information Content of Photoelectric Star Images,” Journal of the Optical Society of America 56, 578–587 (1966).
[18] L. B. Lucy, “Statistical Limits to Super Resolution,” Astronomy and Astrophysics 261, 706 (1992).
[19] L. B. Lucy, “Resolution limits for deconvolved images,” The Astronomical Journal 104, 1260–1265 (1992).
[20] Jonas Zmuidzinas, “Cramér–Raosensitivity limits for astronomical instruments: implications for interferometer design,” Journal of the Optical Society of America A 20, 218–233 (2003).
[21] Eric D. Feigelson and G. Jogesh Babu, Modern Statistical Methods for Astronomy (Cambridge University Press, Cambridge, 2012).
[22] Kim L. Mortensen, L. Stirling Churchman, James A. Spudich, and Henrik Flyvbjerg, “Optimized localization analysis for single-molecule tracking and super-resolution microscopy,” Nature Methods 7, 377–381 (2010).
[23] Alex Small and Shane Stahlheber, “Fluorophore localization algorithms for super-resolution microscopy,” Nature Methods 11, 267–279 (2014).
[24] Hendrik Deschout, Francesca Cella Zanacchi, Michael Mlodzianoski, Alberto Diaspro, Joerg Bewersdorf, Samuel T. Hess, and Kevin Braeckmans, “Precisely and accurately localizing single emitters in fluorescence microscopy,” Nature
Mikhail I. Kolobov, ed., Quantum Imaging (Academic Press, New York, 1976).

Carl W. Helstrom, Quantum Detection and Estimation Theory (Springer, New York, 2007).

Jonathan P. Dowling, Quantum optical metrology - the low-dimensional limit of resolving two point sources limited by an arbitrary point spread function,” in 2017 IEEE International Symposium on Information Theory (ISIT) (IEEE, Aachen, Germany, 2017) pp. 441–445.

Andrzej Chrostowski, Rafal Demkowicz-Dobrzański, Marcin Jarzyna, and Konrad Banaszek, “On super-resolution imaging as a multiparameter estimation problem,” International Journal of Quantum Information 15, 1740005 (2017).

J. Reháček, Z. Hradil, B. Stoklasa, M. Pařír, J. Grover, A. Krzic, and L. L. Sánchez-Soto, “Multiparameter quantum metrology of incoherent point sources: Towards realistic superresolution,” Physical Review A 96, 062107 (2017).

J. Reháček, Z. Hradil, D. Koutný, J. Grover, A. Krzic, and L. L. Sánchez-Soto, “Optimal measurements for quantum spatial superresolution,” Physical Review A 98, 012103 (2018).

Mikael P. Backlund, Yoav Shechtman, and Ronald L. Walsworth, “Fundamental Precision Bounds for Three-Dimensional Optical Localisation Microscopy with Poisson Statistics,” Physical Review Letters 121, 023904 (2018).

Carmine Napoli, Samanta Piana, Richard Leach, Gerardo Adesso, and Tommaso Tufarelli, “Towards Superresolution Surface Metrology: Quantum Estimation of Angular and Axial Separations,” Physical Review Letters 122, 140505 (2019).

Zhixian Yu and Sudhakar Prasad, “Quantum Limited Superresolution of an Incoherent Source Pair in Three Dimensions,” Physical Review Letters 121, 180504 (2018).

Sudhakar Prasad and Zhixian Yu, “Quantum-limited superlocalization and superresolution of a source pair in three dimensions,” Physical Review A 99, 022116 (2019).
[65] Walker Larson and Bahaa E. A. Saleh, “Resurgence of Rayleigh’s curse in the presence of partial coherence,” Optica 5, 1382–1389 (2018).
[66] Mankei Tsang and Ranjith Nair, “Resurgence of Rayleigh’s curse in the presence of partial coherence: comment,” Optica 6, 400–401 (2019).
[67] Walker Larson and Bahaa E. A. Saleh, “Resurgence of Rayleigh’s curse in the presence of partial coherence: reply,” Optica 6, 402–403 (2019).
[68] Michael R. Grace and Saikat Guha, private communication (2019).
[69] Kent Bonsma-Fisher, Weng-Kian Tham, Hugo Ferretti, and Aephraim M. Steinberg, “Beating Rayleigh’s Curve by Using Imaging with Phase-sensitive measurements,” arXiv:1906.00081 [physics, physics:quant-ph] (2019).
[70] Zong Sheng Tang, Kadir Durak, and Alexander Ling, “Fault-tolerant and finite-error localization for point emitters within the diffraction limit,” Optics Express 24, 22004 (2016).
[71] Weng-Kian Tham, Hugo Ferretti, and Aephraim M. Steinberg, “Beating Rayleigh’s curve by imaging with phase-sensitive measurements,” Physical Review Letters 118, 070801 (2017).
[72] Martin Pátr, Bohumil Stoklasa, Zdeněk Hradil, Luís L. Sánchez-Soto, and Jaroslav Reháček, “Achieving the ultimate optical resolution,” Optica 3, 1144 (2016).
[73] Fan Yang, Arina Tashchilina, E. S. Moiseev, Christoph Simon, and A. I. Lvovsky, “Far-field linear optical superresolution via heterodyne detection in a higher-order local oscillator mode,” Optica 3, 1148 (2016).
[74] J. M. Donohue, V. Ansari, J. Řeháček, Z. Hradil, B. Stoklasa, M. Pátr, L. L. Sánchez-Soto, and C. Silberhorn, “Quantum-Limited Time-Frequency Estimation through Mode-Selective Photon Measurement,” Physical Review Letters 121, 090501 (2018).
[75] Michal Parniak, Sebastian Borównka, Kajetan Boroszko, Wojciech Wasilewski, Konrad Banaśek, and Rafał Demkowicz-Dobrzański, “Beating the Rayleigh Limit Using Two-Photon Interference,” Physical Review Letters 121, 250503 (2018).
[76] Martin Pátr, Bohumil Stoklasa, Jai Grover, Andrej Krzic, Luís L. Sánchez-Soto, Zdeněk Hradil, and Jaroslav Reháček, “Tempering Rayleigh’s curse with PSF shaping,” Optica 5, 1177–1180 (2018).
[77] J. Hassett, T. Malhorta, M. A. Alonso, R. W. Boyd, S. M. Hashemi Rafsanjani, and A. N. Vamivakas, “Sub-Rayleigh Limit Localization with a Spatial Mode Analyzer,” in Frontiers in Optics / Laser Science (2018), paper JWA4.124 (Optical Society of America, 2018) p. JWA4.124.
[78] Yiyu Zhou, Jing Yang, Jeremy D. Hassett, Seyed Mohammad Hashemi Rafsanjani, Mohammad Mirhosseini, A. Nick Vamivakas, Andrew N. Jordan, Zhimin Shi, and Robert W. Boyd, “Quantum-limited estimation of the axial separation of two incoherent point sources,” Optica 6, 534–541 (2019).
[79] E. L. Lehmann and George Casella, The Theory of Point Estimation (Springer, New York, 1998).
[80] Joseph W. Goodman, Introduction to Fourier Optics (McGraw-Hill, New York, 2004).
[81] James B. Pawley, ed., Handbook of Biological Confocal Microscopy (Springer, New York, 2006).
[82] Hiroshi Nagaoka, “A new approach to Cramér-Rao bounds for quantum state estimation,” IEICE Technical Report IT 89-42, 9–14 (1989), reprinted in Asymptotic Theory of Quantum Statistical Inference, edited by Masahito Hayashi (World Scientific, Singapore, 2005) Chap. 8, 100–112.
[83] Samuel L. Braunstein and Carlton M. Caves, “Statistical distance and the geometry of quantum states,” Physical Review Letters 72, 3439–3443 (1994).
[84] Masahito Hayashi, ed., Asymptotic Theory of Quantum Statistical Inference: Selected Papers (World Scientific, Singapore, 2005).
[85] Akio Fujiwara, “Strong consistency and asymptotic efficiency for adaptive quantum estimation problems,” Journal of Physics A: Mathematical and General 39, 12489 (2006).
[86] Horace P. Yuen and Jeffrey H. Shapiro, “Optical communication with two-photon coherent states—Part I: Quantum-state propagation and quantum-noise,” IEEE Transactions on Information Theory 24, 657–668 (1978).
[87] Donald L. Snyder and Michael I. Miller, Random Point Processes in Time and Space (Springer-Verlag, New York, 1991).
[88] Mankei Tsang, “Quantum limits to optical point-source localization,” Optica 2, 646–653 (2015).
[89] Jean-François Morizur, Lachlan Nicholls, Pu Jian, Seiji Armstrong, Nicolas Treps, Boris Hage, Magnus Hsu, Warwick Bowen, Jiri Janousek, and Hans-A. Bachor, “Programmable unitary spatial mode manipulation,” Journal of the Optical Society of America A 27, 2524–2531 (2010).
[90] Gaifang Li, Neng Bai, Ningbo Zhao, and Cen Xia, “Space-division multiplexing: the next frontier in optical communication,” Advances in Optics and Photonics 6, 413–487 (2014).
[91] Kai Wicker and Rainer Heintzmann, “Interferometric resolution improvement for confocal microscopes,” Optics Express 15, 12206 (2007).
[92] Kai Wicker, Simon Sindbert, and Rainer Heintzmann, “Characterisation of a resolution enhancing image inversion interferometer,” Optics Express 17, 15491 (2009).
[93] D. Weigel, R. Foerster, H. Babovsky, A. Kiessling, and R. Kowarschik, “Enhanced resolution of microscopic objects by image inversion interferometry,” Optics Express 19, 26451 (2011).
[94] D. Weigel, H. Babovsky, A. Kiessling, and R. Kowarschik, “Investigation of the resolution ability of an image inversion interferometer,” Optics Communications 284, 2273–2277 (2011).
[95] Xin Xue, Haiqing Wei, and Andrew G. Kirk, “Beam analysis by fractional Fourier transform,” Optics Letters 26, 1746–1748 (2001).
[96] Ayman F. Abouraddy, Timothy M. Yarnall, and Bahaa E. A. Saleh, “Generalized optical interferometry for modal analysis in arbitrary degrees of freedom,” Optics Letters 37, 2889–2891 (2012).
[97] Lane Martin, Davood Mardani, H. Esat Kondakci, Walker D. Larson, Soroush Shabahang, Ali K. Jahromi, Tanya Malhotra, A. Nick Vamivakas, George K. Atia, and Ayman F. Abouraddy, “Basis-neutral Hilbert-space analyzers,” Scientific Reports 7, 44995 (2017).
[98] Andrew Forbes, Angela Dudley, and Melanie McLaren, “Creation and detection of optical modes with spatial light modulators,” Advances in Optics and Photonics 8, 200–227 (2016).
[99] Guillaume Labroille, Bertrand Denolle, Pu Jian, Philippe Genevaux, Nicolas Treps, and Jean-François Morizur, “Efficient and mode selective spatial mode multiplexer based on multi-plane light conversion,” Optics Express 22, 15599–15607 (2014).
[100] Nicolas K. Fontaine, Roland Ryf, Haoshuo Chen, David T. Neilson, Kwangwoong Kim, and Joel Carpenter, “Laguerre-Gaussian mode sorter,” Nature Communications 10, 1865 (2019).
[101] Charles H. Townes, “Noise and sensitivity in interferometry,” in Principles of Long Baseline Stellar Interferometry, edited by Peter R. Lawson (Jet Propulsion Laboratory, Pasadena, CA: National Aeronautics and Space Administration, 1981).
The incoherent case,” Inverse Problems 5, 441–461 (1989).

[140] Alexander S. Holevo, Quantum Systems, Channels, Information (de Gruyter, Berlin, 2012).

[141] D. Petz and C. Ghinea, “Introduction to quantum Fisher information,” in Quantum Probability and Related Topics, QP-PQ: Quantum Probability and White Noise Analysis, Vol. 27, edited by Rolando Rebolledo and Miguel Orszag (World Scientific, Singapore, 2011) pp. 261–281.

[142] S. Alipour and A. T. Rezakhani, “Extended convexity of quantum fisher information in quantum metrology,” Physical Review A 91, 042104 (2015).

[143] Shilin Ng, Shan Zheng Ang, Trevor A. Wheatley, Hidehiro Yonezawa, Akira Furusawa, Elanor H. Huntington, and Mankei Tsang, “Spectrum analysis with quantum dynamical systems,” Physical Review A 93, 042121 (2016).

[144] Akio Fujiwara, “Quantum channel identification problem,” Physical Review A 63, 042304 (2001).

[145] R. Demkowicz-Dobrzański, J. Kolodyński, and M. Guţă, “The elusive Heisenberg limit in quantum-enhanced metrology,” Nature Communications 3, 1063 (2012).

[146] M. Stein, A. Mezghani, and J. A. Nossek, “A Lower Bound for the Fisher Information Measure,” IEEE Signal Processing Letters 21, 796–799 (2014).

[147] Richard D. Gill and Boris Y. Levit, “Applications of the Van Trees inequality: A Bayesian Cramér-Rao bound,” Bernoulli 1, 59–79 (1995).