

# Quantum-inspired Multi-Parameter Adaptive Bayesian Estimation for Sensing and Imaging

Kwan Kit Lee, Christos N. Gagatsos, Saikat Guha, *Senior Member, IEEE*, Amit Ashok, *Member, IEEE*

**Abstract**—It is well known in Bayesian estimation theory that the conditional estimator attains the minimum mean squared error (MMSE) for estimating a scalar parameter of interest. In quantum, e.g., optical and atomic, imaging and sensing tasks the user has access to the quantum state that encodes the parameter. The choice of a measurement operator, i.e. a positive-operator valued measure (POVM), leads to a measurement outcome on which the aforesaid classical MMSE estimator is employed. Personick found the optimum POVM that attains the MMSE over all possible physically allowable measurements and the resulting MMSE [1]. This result from 1971 is less-widely known than the quantum Fisher information (QFI), which lower bounds the variance of an unbiased estimator over all measurements without considering any prior probability. For multi-parameter estimation, in quantum Fisher estimation theory the inverse of the QFI matrix provides an operator lower bound on the covariance of an unbiased estimator, and this bound is understood in the positive semidefinite sense. However, there has been little work on quantifying the quantum limits and measurement designs, for multi-parameter quantum estimation in a *Bayesian* setting. In this work, we build upon Personick’s result to construct a Bayesian adaptive (greedy) measurement scheme for multi-parameter estimation. We illustrate our proposed measurement scheme with the application of localizing a cluster of point emitters in a highly sub-Rayleigh angular field-of-view, an important problem in fluorescence microscopy and astronomy. Our algorithm translates to a multi-spatial-mode transformation prior to a photon-detection array, with electro-optic feedback to adapt the mode sorter. We show that this receiver performs superior to quantum-noise-limited focal-plane direct imaging.

**Index Terms**—Quantum Information, Information Theory, Bayesian Inference, Super-Resolution.

## I. INTRODUCTION

In various fields of science and engineering, multi-parameter estimation is an important problem relevant in many applications. Examples include simultaneous estimation of phase and other parameters such as loss, diffusion coefficients, uncertainty of the probe, etc. in optical/atomic interferometry [2]–[5], estimation of location of point emitters in multiple dimensions for super-resolution [6]–[8], waveform estimation [9], [10], Hamiltonian estimation [11]–[13], measuring range and velocity using lidar [14]–[16], object tracking [17], [18] and even modeling biological processes [19].

In classical estimation theory, minimizing the covariance matrix over all the unbiased estimators in the positive semidefinite sense yields the Cramer-Rao lower bound (CRLB) for multi-parameter estimation, which is the inverse of the classical Fisher information (CFI) matrix [20]. The CRLB can always be achieved asymptotically by the maximum likelihood

estimator (MLE). In quantum estimation theory, the inverse of quantum Fisher information (QFI) provides an operator lower bound on the error covariance matrix or the Quantum CRLB (QCRLB) by further minimization over all physically allowed measurements, in absence of any available priors [21]. Thus it serves as the performance benchmark of measurement design for sensing [3], [22], [23] and probe design for metrology [24], [25]. The counterpart of CFI in a Bayesian setting is referred to as the Van Trees inequality or bound [26], [27], while a Quantum Bayesian version of this QFI has been proposed by Tsang [9]. Besides using the information matrices, a lower bound of the covariance error matrix for a single-shot measurement can also be found by carrying out direct minimization over all the unbiased estimators and measurements on the covariance error matrix itself [1], [28], [29]. Although the quantum information analysis discussed above provides the ultimate precision limit of parameter estimation, the optimal measurement that achieves the limit for all parameters simultaneously may not exist in general and/or be challenging or even analytically intractable. The saturation conditions and the trade-off of the precision of estimating different parameters have been analyzed for both QFI and its Bayesian counterparts [3], [29], [30].

Apart from the theoretical analysis on the estimation precision limit, the development of practical estimation protocols is also a key concern to scientists and engineers for various real-world applications. Various approaches, ranging from the traditional methods such as maximum likelihood estimation [31]–[33], (adaptive) Bayesian estimation [34]–[37], stochastic algorithms [38], [39], to the state-of-art data-driven neural networks [40], [41], have been employed to solve a diverse range of estimation problems. Although the performance of such estimation schemes may not achieve their corresponding classical/quantum limits, they are still useful due to their tractable and efficient implementations, given that their performance meets the user/application minimum requirements.

Therefore, the theoretical challenge of seeking optimal measurements that achieve the quantum limit for multi-parameters simultaneously coupled with the practical challenge of developing multi-parameter estimation protocol(s) inspired by the quantum limits motivate our proposed sequential adaptive measurement scheme for multi-parameter estimation. We describe our sequential estimation approach within a full Bayesian inference framework by leveraging tools from the Bayesian quantum estimation theory. Sec. II outlines three different potential measurement protocols, while Sec. III provides a comprehensive review on the precision limits under different

We acknowledge support for this work by the Defense Advanced Research Projects Agency (DARPA) IAMBIC program under contract HR00112090128.

statistical frameworks (i.e., frequentist and Bayesian). The details of our proposed measurement scheme are presented in Sec. IV. In Sec. V, we employ our proposed measurement scheme to the problem of localizing an unknown number of point-emitters in a sub-Rayleigh (below diffraction-limit) field of view in an optical imaging context. This illustrative imaging application is motivated by the fact that traditional direct focal-plane imaging, which employs intensity measurements followed by electronic-domain processing, is known to be highly sub-optimal [22] in the sub-Rayleigh regime. We compare our quantum-inspired adaptive sequential measurement design with the direct imaging technique to quantify the significant optical resolution improvement obtained with our proposed scheme. In Sec. VI we briefly summarize key aspects of our work and point to future directions for further optimizations to improve upon the current results.

## II. MEASUREMENT PROTOCOLS

In classical sensing and imaging paradigm, a measurement channel is modeled by the likelihood function defined as the conditional probability  $p(\mathbf{l}|\boldsymbol{\theta})$ , where  $\mathbf{l} = [l_1, l_2, \dots, l_N]^T$  is the vector-valued measurement resulting from sensing/imaging the input  $\boldsymbol{\theta}$  (e.g., object/scene/signal). The input  $\boldsymbol{\theta}$  represents  $M$  parameters  $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_M]^T$ . Thus, the measurement channel can be expressed by the conditional probability density or likelihood  $p(\mathbf{l}|\boldsymbol{\theta})$ . This measurement model can be also applied in a more general framework of quantum sensing, where the input  $\boldsymbol{\theta}$  is replaced by a density operator  $\rho(\boldsymbol{\theta})$  describing the object being measured and the measurement channel is given by a positive-operator-valued measure (POVM)  $\{\Pi_{\mathbf{l}}\}$  operating on  $\rho(\boldsymbol{\theta})$  resulting in outcome  $\mathbf{l}$  with probability  $p(\mathbf{l}|\boldsymbol{\theta}) = \text{Tr}(\rho(\boldsymbol{\theta})\Pi_{\mathbf{l}})$  [42]. Note that a classical measurement channel can always be expressed as:  $\rho(\boldsymbol{\theta}) = \int p(\mathbf{l}|\boldsymbol{\theta})d\mathbf{l}|\alpha_{\mathbf{l}}\rangle\langle\alpha_{\mathbf{l}}|$  with the projection operator POVM  $\{\Pi_{\mathbf{l}}\} = \{|\alpha_{\mathbf{l}}\rangle\langle\alpha_{\mathbf{l}}|\}$ , where  $\{|\alpha_{\mathbf{l}}\rangle\}$  is an orthonormal complete basis (Fig. 1). Thus, in the following discussion we only consider the more general quantum formulation, and treat the classical channel as a special case.

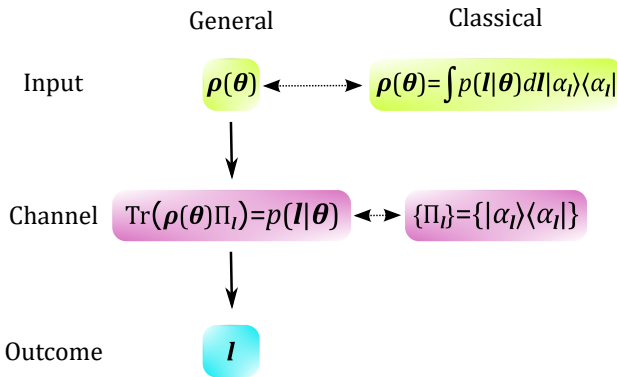


Fig. 1. A schematic diagram shows the classical and quantum channel model.

In practice, if  $N$  copies ( $N \geq 2$ ) of quantum states  $\rho(\boldsymbol{\theta})^{\otimes N}$  (or equivalently  $N$  channel uses or samples) are available, the receiver (measurement) can: (1) in the most general setting, choose a joint-measurement POVM  $\{\Pi_{\mathbf{l}_{(N)}}\}$

acting collectively on  $\rho(\boldsymbol{\theta})^{\otimes N}$ , yielding the outcome  $\mathbf{l}_{(N)}$ ; (2) employ the *local operations and classical communications* (LOCC) scheme, such that each batch of state  $\rho(\boldsymbol{\theta})^{\otimes K_{\tau}}$ , where  $K_{\tau}$  is the number of copies of state  $\rho(\boldsymbol{\theta})$  comprising the  $\tau^{\text{th}}$  measurement batch, with  $0 \leq \tau \leq S$  and  $N = \sum_{\tau=0}^S K_{\tau}$ , is measured by the POVM  $\{\Pi_{\mathbf{l}}^{(\tau)}\}$  chosen for example, based on the information available from the previous set of measurement outcomes  $\{\mathbf{l}^{(0)}, \mathbf{l}^{(1)}, \dots, \mathbf{l}^{(\tau-1)}\}$ ; or (3) use independent identical measurements on each copy of the state, described by the POVM  $\{\Pi_{\mathbf{l}}\}$ . The schematic diagram illustrating these three measurement approaches is shown in Fig. 2.

No matter the receiver strategy, after measuring all  $N$  copies, the receiver generates an estimate of  $\boldsymbol{\theta}$ , i.e.,  $\hat{\boldsymbol{\theta}}(\mathbf{l}_{\text{set}})$  where  $\mathbf{l}_{\text{set}} = \mathbf{l}_{(N)}$  for case (1) above, and  $\mathbf{l}_{\text{set}} = [\mathbf{l}^{(0)}, \mathbf{l}^{(2)}, \dots, \mathbf{l}^{(S)}]$  for cases (2) above and  $\mathbf{l}_{\text{set}} = [\mathbf{l}^{(1)}, \mathbf{l}^{(2)}, \dots, \mathbf{l}^{(N)}]$  for case (3) above. The receiver chooses the estimator to optimize a desired objective/loss function. A natural choice of the objective function associated with sensing and imaging estimation tasks is mean (expected) squared-error (MSE),  $\mathbb{E}[||\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}(\mathbf{l}_{\text{set}})||^2]$ .

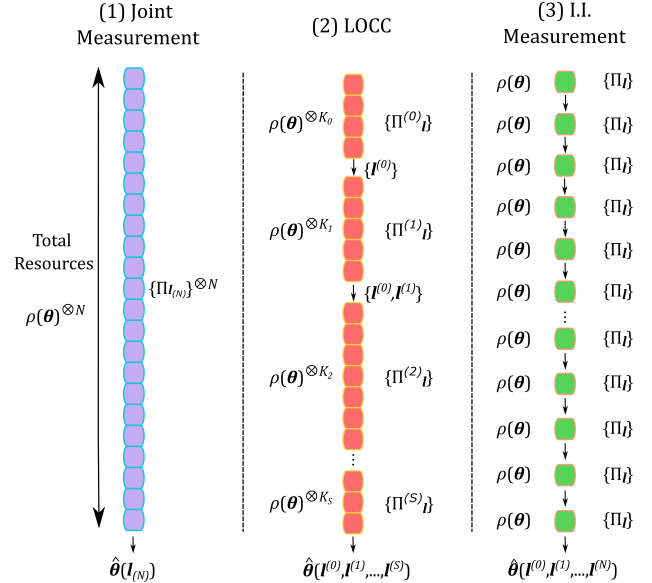


Fig. 2. A schematic diagram shows the three different measurement approaches operating on  $N$ -copies of quantum states  $\rho(\boldsymbol{\theta})^{\otimes N}$ .

## III. PRECISION LIMIT OF DIFFERENT STATISTICAL ESTIMATION FRAMEWORKS

For any *given* measurement POVM  $\{\Pi_{\mathbf{l}}\}$  choice, assuming strategy (3) above, i.e., the same measurement acts on each copy of  $\rho(\boldsymbol{\theta})$ , the problem reduces to the standard classical estimation theory problem of estimating  $\boldsymbol{\theta}$  from  $N$  i.i.d. samples of  $\mathbf{l}$ , each described by  $p(\mathbf{l}|\boldsymbol{\theta})$ . The covariance  $\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}_{\text{set}}), \boldsymbol{\theta})$  for any unbiased estimator  $\hat{\boldsymbol{\theta}}(\mathbf{l}_{\text{set}})$  of  $\boldsymbol{\theta}$  is lower bounded by  $\Sigma_C$  [29]. This means  $\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}_{\text{set}}), \boldsymbol{\theta}) - \Sigma_C$  is a semi-positive definite matrix, denoted compactly as  $\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}_{\text{set}}), \boldsymbol{\theta}) \geq \Sigma_C$ . The receiver's task is to pick the optimal estimator  $\hat{\boldsymbol{\theta}}^{\text{opt}}(\mathbf{l}_{\text{set}})$  operating on the measurement outcomes  $\mathbf{l}_{\text{set}}$ , such that  $\text{Cov}(\hat{\boldsymbol{\theta}}^{\text{opt}}(\mathbf{l}_{\text{set}}), \boldsymbol{\theta})$  saturates the bound  $\Sigma_C$  when permissible.

Tools of quantum estimation theory allow us to find a tight lower bound to  $\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}_{\text{set}}), \boldsymbol{\theta})$ , which automatically optimizes over all physically-permissible choices of a POVM  $\{\Pi_{\mathbf{l}}\}$  (again, assuming that the same measurement is used to detect each copy of  $\rho(\boldsymbol{\theta})$ ). The  $\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}_{\text{set}}), \boldsymbol{\theta})$  is lower bounded by  $\Sigma_Q$  (a *quantum* bound) [29], which itself is the infimum of a ll bounds  $\Sigma_C$  associated with a ll possible choices of  $\{\Pi_{\mathbf{l}}\}$ . For certain cases (for example when  $\boldsymbol{\theta}$  is a scalar parameter), quantum estimation theory also provides the optimal receiver POVM  $\{\Pi_{\mathbf{l}}^{\text{(opt)}}\}$  [29]. Once the optimal receiver is chosen, it uses the optimal estimator  $\hat{\boldsymbol{\theta}}^{\text{opt}}(\mathbf{l}_{\text{set}})$  using standard classical estimation tools, such that covariance  $\text{Cov}(\hat{\boldsymbol{\theta}}^{\text{opt}}(\mathbf{l}_{\text{set}}), \boldsymbol{\theta})$  saturates  $\Sigma_Q$  when permissible. Therefore, in general we can state:  $\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}_{\text{set}}), \boldsymbol{\theta}) \geq \Sigma_C \geq \Sigma_Q$ , where  $\Sigma_C$  corresponds to any choice of POVM.

The aforementioned lower bounds on the covariance of multi-parameter estimators can be defined within the statistical inference frameworks of the frequentist approach, i.e., Fisherian (with no prior), or the Bayesian (with prior  $p(\boldsymbol{\theta})$ ) inference approach. We review below some known bounds for both inference approaches.

In the Fisherian (frequentist) approach, when no prior  $p(\boldsymbol{\theta})$  is available or defined, the Cramer-Rao lower bound (CRLB)  $\Sigma_C$  on the covariance  $\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}), \boldsymbol{\theta})$  of an unbiased estimator is given by the inverse of the Fisher information (FI) matrix  $I$  [20]:

$$I_{ij} = \int \left[ \frac{\partial}{\partial \theta_i} \ln p(\mathbf{l}|\boldsymbol{\theta}) \right] \left[ \frac{\partial}{\partial \theta_j} \ln p(\mathbf{l}|\boldsymbol{\theta}) \right] p(\mathbf{l}|\boldsymbol{\theta}) d\mathbf{l}, \quad (1)$$

with  $1 \leq i, j \leq M$ , and the likelihood  $p(\mathbf{l}|\boldsymbol{\theta}) = \text{Tr}(\rho(\boldsymbol{\theta})\Pi_{\mathbf{l}})$ . The quantum version of this lower bound  $\Sigma_Q$ , which only depends on  $\rho(\boldsymbol{\theta})$  (since the measurement  $\Pi_{\mathbf{l}}$  is automatically optimized over all POVMs) is given by the inverse of the quantum Fisher information (QFI) matrix  $Q$  [43], with elements:

$$Q_{ij} = \text{Tr} \left[ \rho(\boldsymbol{\theta}) \frac{L_i L_j + L_j L_i}{2} \right], \quad (2)$$

where  $L_i$  is the symmetric logarithmic derivative (SLD) operator. The SLD operator can be determined from the following implicit relationship:

$$2 \frac{\partial}{\partial \theta_i} \rho(\boldsymbol{\theta}) = \rho(\boldsymbol{\theta}) L_i + L_i \rho(\boldsymbol{\theta}), \quad (3)$$

with  $1 \leq i \leq M$ . Thus, we have  $\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}_{\text{set}}), \boldsymbol{\theta}) \geq I^{-1} \geq Q^{-1}$  in the Fisher framework. For  $N$ -copy i.i.d. measurement of  $\rho(\boldsymbol{\theta})^{\otimes N}$ , both the classical and quantum bounds scale by a factor of  $1/N$ . The classical one is asymptotically attained by the maximum likelihood estimator (MLE). The quantum CRLB ( $Q^{-1}$ ) can not be saturated in general for  $M > 1$ .

In Bayesian setting, Personick found the minimum mean square error of a single parameter problem by optimizing first the estimator and then the measurement and [1], and the multi-parameter counterpart on the covariance  $\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}), \boldsymbol{\theta})$  of any estimator  $\hat{\boldsymbol{\theta}}(\mathbf{l})$  are found in [29]. In this work we refer to this lower bound as the Bayesian Personick lower bound (BPLB), given by,

$$\Sigma_C = \int p(\boldsymbol{\theta}) \boldsymbol{\theta} \boldsymbol{\theta}^T d\boldsymbol{\theta} - J, \quad (4)$$

where the  $M$ -by- $M$  matrix  $J$  is defined as:

$$J_{ij} = \int \frac{[\int p(\mathbf{l}, \boldsymbol{\theta}) \theta_i d\boldsymbol{\theta}] [\int p(\mathbf{l}, \boldsymbol{\theta}) \theta_j d\boldsymbol{\theta}]}{p(\mathbf{l})} d\mathbf{l}, \quad (5)$$

and  $p(\mathbf{l}, \boldsymbol{\theta}) = p(\mathbf{l}|\boldsymbol{\theta})p(\boldsymbol{\theta})$  is the joint distribution of  $\mathbf{l}$  and  $\boldsymbol{\theta}$ . This bound is analogous to the CRLB in the Fisherian setting. The posterior mean of the parameters  $\int \theta_i p(\boldsymbol{\theta}|\mathbf{l}) d\boldsymbol{\theta}$  saturates the bound in Eq. (4) and the proof can be found in Appendix A. For the quantum version of this lower bound, we first define the following operators, for  $1 \leq i \leq M$  and  $k = 0, 1, 2$  [1]:

$$\Gamma_{i,k} = \int d\boldsymbol{\theta} p(\boldsymbol{\theta}) \rho(\boldsymbol{\theta}) \theta_i^k, \quad (6)$$

and operators  $B_i$ ,  $1 \leq i \leq M$ , that satisfy:

$$2\Gamma_{i,1} = \Gamma_0 B_i + B_i \Gamma_0. \quad (7)$$

For  $k = 0$ ,  $\Gamma_{i,0} = \Gamma_{j,0}$ ,  $\forall (i, j)$ , thus we can drop the first index and denote it as  $\Gamma_0 = \int d\boldsymbol{\theta} p(\boldsymbol{\theta}) \rho(\boldsymbol{\theta})$ , the average received state. The quantum BPLB  $\Sigma_Q$  can be written as:

$$\Sigma_Q = \int p(\boldsymbol{\theta}) \boldsymbol{\theta} \boldsymbol{\theta}^T d\boldsymbol{\theta} - G, \quad (8)$$

where

$$G_{ij} = \text{Tr} \left[ \Gamma_0 \frac{B_i B_j + B_j B_i}{2} \right]. \quad (9)$$

Thus in a Bayesian inference framework, we have  $\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}), \boldsymbol{\theta}) \geq \Sigma_C \geq \Sigma_Q$ . Table I summarizes the various Fisherian and Bayesian bounds discussed here.

TABLE I  
SUMMARY OF CLASSICAL AND QUANTUM BOUNDS FOR FISHERIAN AND BAYESIAN APPROACHES.

	<b>Fisher</b>	<b>Bayesian</b>
<b>Classical</b>	Information Matrix: $I$	Covariance Bound: $\Sigma_C$
<b>Quantum</b>	Information Matrix: $Q$	Covariance Bound: $\Sigma_Q$
<b>Bounds</b>	$\text{Cov}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \geq I^{-1} \geq Q^{-1}$	$\text{Cov}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) \geq \Sigma_C \geq \Sigma_Q$

To achieve the quantum bound, an optimal measurement is required (i.e. an optimal choice POVM, that acts on each copy of  $\rho(\boldsymbol{\theta})$ ). For the single parameter problem ( $M = 1$ ), the projective measurement onto the eigenvectors of the SLD operator  $L$  in Eq. (3) saturates the Fisher quantum bound, i.e., the  $I$  for the SLD measurement equals  $Q$  [44]. Likewise, the Bayesian quantum bound on the covariance is saturated (i.e.  $\Sigma_C = \Sigma_Q$ ), for the case of a single parameter ( $M = 1$ ) by a projective measurement onto the eigenvectors of the operator  $B$  in Eq. (7) [1].

For multi-parameter estimation, if the operators associated with parameter  $\theta_i$ :  $L_i$  and  $B_i$ ,  $1 \leq i \leq M$  commute with one another, for the Fisher and Bayesian frameworks respectively, the corresponding covariance bound can be saturated by the above-said measurements, calculated by evaluating the eigenvectors of  $L_i$  or  $B_i$  operators, respectively (which  $i$  does not matter as they are simultaneously diagonal). However, if the operators do not commute, which is the case in general, a measurement that is jointly optimal for all parameters may not exist and/or likely to be challenging to derive [44].

For completeness sake we also note that in the quantum case, the Holevo Cramer-Rao bound (HCRB) [45] is the most fundamental scalar lower bound on the weighted mean square error  $\text{Tr}[W\text{Cov}(\hat{\theta}(\mathbf{l}), \theta)]$ , for a positive definite  $W$ . The HCRB represents the best precision attainable with a collective measurement (discussed as case (1) above) on an asymptotically large number of identical copies of  $\rho(\theta)$ .

The aforementioned challenges relating to multi-parameter bounds and the potential to utilize quantum estimation theory to design realizable measurements for multi-parameter estimation problems motivates our proposed adaptive sequential measurement scheme described in the following section.

#### IV. ADAPTIVE SEQUENTIAL MEASUREMENT SCHEME

Consider a system or a field in the state described by the density operator:

$$\rho(\theta) = \sum_{i=1}^P b_i(\theta) |\psi_i(\theta)\rangle \langle \psi_i(\theta)|, \quad (10)$$

where  $\theta = [\theta_1, \theta_2, \dots, \theta_M]^T$  are the parameters of interest,  $|\psi_i(\theta)\rangle$  and  $b_i(\theta)$  are the parameter-dependent pure states and the corresponding weights respectively. As  $\rho(\theta)$  is unit trace, we have  $\langle \psi_i(\theta) | \psi_i(\theta) \rangle = 1, \forall i$  and  $\sum_{i=1}^P b_i(\theta) = 1$ . The states  $|\psi_i(\theta)\rangle$  are not necessarily orthogonal, i.e.  $\langle \psi_i(\theta) | \psi_j(\theta) \rangle \neq 0$  for  $i \neq j$  in general.  $P$  itself, in general, is an unknown parameter (positive integer) such that:  $P_{\min} \leq P \leq P_{\max}$ . Here we assume that  $P$  is upper bounded by  $P_{\max}$ , i.e., a prior on  $P$ . If the lower bound  $P_{\min}$  is not known/available, we can set it to 1. When  $P_{\min} \neq P_{\max}$ , both  $P$  and  $\theta$  need to be estimated. On the contrary, if  $P_{\min} = P = P_{\max}$ , i.e.,  $P$  is known *a priori* exactly, then we only need to estimate the parameters  $\theta$ .

##### A. LOCC Measurement Scheme

We propose our measurement scheme within the LOCC framework to estimate multiple parameters  $\theta$  with  $N$  independent copies of quantum state  $\rho(\theta)$  defined in Eq. (10). To illustrate our proposed scheme, we begin with the  $P$  known exactly case. In the next section, we discuss the extension where we relax this prior on  $P$ . The measurement scheme is illustrated in Fig. 3.

1) *Initialization*: The measurement is initialized by choosing the  $\{\Pi^{(0)}\}$  and  $p^{(0)}(\theta)$ , which are the POVM for measuring  $\rho(\theta)^{\otimes K_0}$  and the prior on the parameters  $\theta$  respectively. If by any means a set of pre-estimated parameters  $\hat{\theta}^{(0)}$  can be found, one may construct an estimated density operator  $\rho(\hat{\theta}^{(0)})$  and use the method described below to construct  $\{\Pi^{(0)}\}$ . Otherwise, any POVM can be used in this stage.

2) *Measurement Cycle/Step*: Let us take  $N = \sum_{\tau=0}^S K_\tau$ , such that we adapt the measurement choice  $S$  times, denoted by  $\tau$  as the iteration index,  $0 \leq \tau \leq S$ . In the  $\tau^{\text{th}}$  measurement cycle,  $K_\tau$  of copies of  $\rho(\theta)$ , the density operator of which is  $\rho(\theta)^{\otimes K_\tau}$ , are measured. The notation used here is the same as that in the previous section. In each measurement cycle/step, we employ the measurement strategy (3) introduced in Sec. I. Consider the  $\tau^{\text{th}}$  measurement cycle employing a

POVM  $\{\Pi_{l^{(\tau)}}\}$  to measure *each single copy* of  $\rho(\theta)$ . For the  $i^{\text{th}}$  copy of  $\rho(\theta)$ , where  $1 \leq i \leq K_\tau$ , the probability of obtaining the outcome  $l_i^{(\tau)}$  is  $p(l_i^{(\tau)} | \theta) = \text{Tr}[\rho(\theta) \Pi_{l_i^{(\tau)}}]$ , such that  $\Pi_{l_i^{(\tau)}} \in \{\Pi_{l^{(\tau)}}\}$ . The probability of observing the measurement outcomes  $\mathbf{l}^{(\tau)} = [l_1^{(\tau)}, l_2^{(\tau)}, \dots, l_{K_\tau}^{(\tau)}]^T$  is  $p(\mathbf{l}^{(\tau)} | \theta) = \text{Tr}[\rho(\theta)^{\otimes K_\tau} \Pi^{(\tau)}] = \prod_{i=1}^{K_\tau} \text{Tr}[\rho(\theta) \Pi_{l_i^{(\tau)}}]$ , where  $\Pi^{(\tau)} \triangleq \Pi_{l_1^{(\tau)}} \otimes \dots \otimes \Pi_{l_{K_\tau}^{(\tau)}}$ . At the end of the sequential measurement scheme, a  $N$ -copy state  $\rho(\theta)^{\otimes N}$  has been measured. Note that  $K_\tau$  can be deterministic in some situations (e.g. the number of bits being transferred in a channel), but in many sensing/imaging problems, it is likely to be a random variable. For example, in the imaging problem discussed in the next section, a single photon is described by  $\rho(\theta)$  and the number of photons (copies of  $\rho(\theta)$ )  $K_\tau$  received in a fixed time period is a random variable governed by Poisson distribution. Nevertheless, our measurement protocol works for varying  $K_\tau$  thus it is suitable for a wide range of sensing/imaging problems.

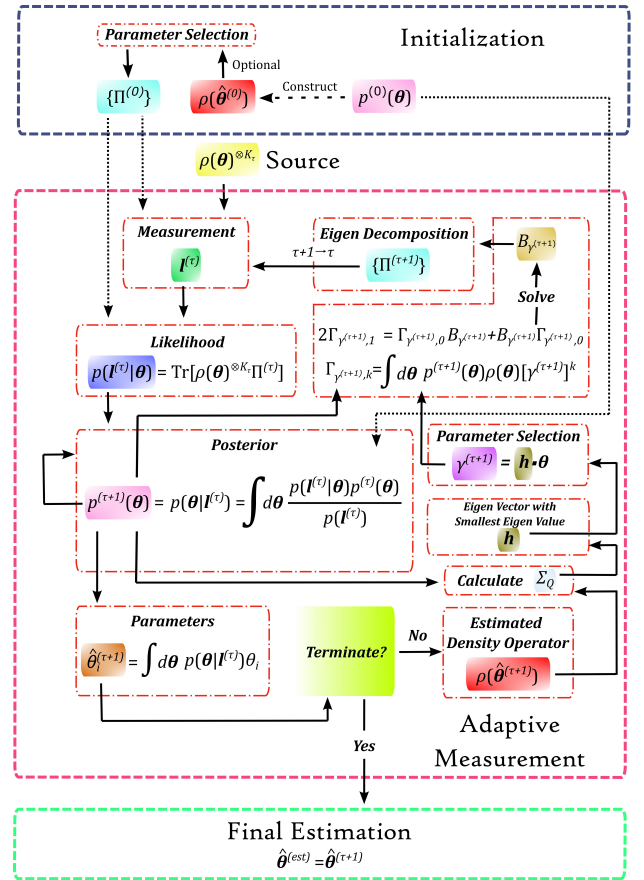


Fig. 3. A schematic diagram illustrating various steps comprising our proposed sequential adaptive measurement scheme.

The parameter estimate  $\hat{\theta}^{(\tau)}$ , available after the  $\tau^{\text{th}}$  sequential measurement is denoted by  $\hat{\theta}^{(\tau)} = [\hat{\theta}_1^{(\tau)}, \hat{\theta}_2^{(\tau)}, \dots, \hat{\theta}_M^{(\tau)}]^T$ . In a Bayesian inference setting, the parameter estimate  $\hat{\theta}^{(\tau)}$  is given by posterior mean:  $\hat{\theta}^{(\tau)} = E_{p(\theta | \mathbf{l}^{(\tau)})}[\theta]$  if we wish to minimize the objective/loss function of MSE. For other loss functions (e.g., probability of detection/classification) other

optimal estimators/detectors/classifiers can be chosen. The posterior is given by:  $p(\boldsymbol{\theta}|\mathbf{l}^{(\tau)}) = p(\mathbf{l}^{(\tau)}|\boldsymbol{\theta}) \cdot p^{(\tau)}(\boldsymbol{\theta})/p(\mathbf{l}^{(\tau)})$ , where  $p^{(\tau)}(\boldsymbol{\theta})$  is the prior at the  $\tau^{\text{th}}$  iteration. Note that the prior  $p^{(\tau)}(\boldsymbol{\theta})$  in turn equals the posterior  $p(\boldsymbol{\theta}|\mathbf{l}^{(\tau-1)})$  at the previous  $(\tau - 1)^{\text{th}}$  iteration. The density operator at the  $\tau^{\text{th}}$  iteration is represented as  $\rho(\hat{\boldsymbol{\theta}}^{(\tau)})$ . Now what remains to be determined is how we choose the POVM  $\{\Pi_{\mathbf{l}^{(\tau)}}\}$  at the  $\tau^{\text{th}}$  iteration. We use the following strategy to pick/construct the POVM  $\{\Pi_{\mathbf{l}^{(\tau)}}\}$ .

It is known that for a single parameter estimation problem, the eigen-projection measurement of  $B_1$  in Eq. (7) saturates the quantum bound  $\Sigma_Q$  [1], which reduces to a lower bound on the variance of the scalar parameter. In this case the minimum mean square error (MMSE) is given by  $\Sigma_Q = \text{Tr}[\Gamma_{1,2} - B_1\Gamma_{1,1}]$ , where  $\Gamma_{i,j}$  are defined in Eq. (6). We refer to this measurement as the Personick projection in this work. For the multi-parameter problem, the counter-part of  $\text{Tr}[B_1\Gamma_{1,1}]$  is a matrix  $G$  in Eq. (9). If all  $B_i$  operators commute, the quantum optimal measurement is given by the eigen-projections of any of the  $B_i$  operators [29]. However, there is no such guarantee that the optimal measurement for all parameters exists or can be found in general. As such for the general case, where such optimal measurements may not exist or may not be accessible, we propose to utilize Personick's result for single parameter estimation, to define a single meta parameter  $\gamma^{(\tau)}$  for estimation at the  $\tau^{\text{th}}$  iteration of the sequential measurement. This scalar parameter  $\gamma^{(\tau)}$  is defined as a linear combination of the  $M$  parameters given by the eigenvector of the matrix  $\Sigma_Q$  with the smallest eigenvalue. Our approach is optimal (in MSE sense) for estimating this single parameter (see Appendix B for proof of optimality). The BPLB  $\Sigma_Q$  on the error covariance matrix is defined

per Eqs. (8) and (9) for the density operator  $\rho(\hat{\boldsymbol{\theta}}^{(\tau-1)})$ . The scalar parameter  $\gamma^{(\tau)}$  is used to construct the operator  $B_{\gamma^{(\tau)}}$  and the corresponding Personick projection, constructed using  $B_{\gamma^{(\tau)}}$ , is chosen as the POVM  $\{\Pi_{\mathbf{l}^{(\tau)}}\}$  at the  $\tau^{\text{th}}$  iteration. The sequential measurements are terminated when all the  $N$  available copies of  $\rho(\boldsymbol{\theta})$  have been exhausted.

As a reminder, we note that in each iteration step, the Personick measurement described above is not necessarily optimal given the actual state (density operator) due to reasons stated earlier. Furthermore, as an aside we also note that our approach chooses the next optimal measurement based on the current state estimate, which can be viewed as a result of greedy optimization approach within the context of adaptive sequential estimation [46]. In contrast, a non-greedy adaptive sequential scheme selects the design for the next measurement while considering not only the current state estimate but also all possible future measurements and associated estimates within a game-theoretic framework [47], [48].

### B. Extension: $P$ not known a priori

If the scalar  $P$  in Eq. (10) i.e., the number of parameters is unknown, we can employ and initialize multiple models of density operators  $\rho(\boldsymbol{\theta}_P)$  with the corresponding prior  $p(\boldsymbol{\theta}_P)$ . Here  $\boldsymbol{\theta}_P = [\theta_1, \theta_2, \dots, \theta_{M_P}]^T$  for  $P_{\min} \leq P \leq P_{\max}$ . In such a scenario, the number of parameters, denoted by  $M_P$ , for

each model can be different in general. In  $\tau^{\text{th}}$  iteration of the sequential measurement, one model is selected and used to construct the Personick measurement. The model can be selected randomly at  $\tau = 0$ , and the one that maximizes  $p(\mathbf{l}^{(\tau-1)})$  can be used for the  $\tau^{\text{th}}$  measurement iteration. We also propose an alternative model selection method in the next section. After model selection, the measurement scheme defined in the previous section can be applied unaltered. Note that at the  $\tau^{\text{th}}$  iteration, not only the selected model but all the models are updated in a Bayesian inference setting, using the measurement outcome  $\mathbf{l}^{(\tau)}$ . When the sequential measurements eventually terminate, we simply pick a model using the same model selection criteria described above and compute the final multi-parameter estimate as the posterior mean. However, other model selection criteria and/or point estimators can also be employed, as needed.

## V. APPLICATION: MULTI POINT-EMITTER ESTIMATION

### A. Background

In traditional optical imaging, incoherent point emitters are considered to be unresolvable when their separation is smaller than the Rayleigh limit [49]. However, Tsang et al.'s recent quantum information theoretic analysis of two-point resolution, treating it as an estimation problem, has demonstrated that the Rayleigh limit is not fundamental and in fact an outcome of the sub-optimality of the diffraction-limited optical image intensity measurement, referred here to as direct imaging [22]. Furthermore, they show that the Hermite Gaussian (HG) modal projection measurement saturates the QFI, i.e., it is the optimal measurement in minimum variance unbiased estimation sense. Although this ground-breaking result has shed new light on applying modal projection measurement to achieve optical super-resolution (i.e. resolving features with sub-Rayleigh scale), the optimal measurements for more general situations are still unknown, due to the involvement of multi-parameters in such situations and the parameter dependent nature of the optimal measurement.

We employ our proposed adaptive sequential measurement scheme to illustrate one such complex multi-parameter imaging problem of optical super-resolution, specifically estimating the locations and relative brightness of incoherent point emitters comprising a cluster/constellation. Such an optical super-resolution estimation problem typically arises in many optical imaging applications such as astronomy and microscopy [50], [51]. For instance, in fluorescence microscopy, fluorophore (point-like) emitter positions reveal the structure of the specimen. The state-of-art super-resolution techniques require illumination engineering in the spatial domain, such as Stimulated Emission Depletion (STED) [52] and structured illumination [53], to reduce the effective size of the point spread function (PSF) by about 2 to 4 times. Similarly, in the temporal domain single molecule localization microscopy (SMLM) [54] yields 10 to 100 folds of resolution improvement depending on the strength of the emitters and integration time. On the other hand, astronomers exploit coherent interference of radio signals from array of telescopes, known as very

long base line interferometry (VLBI) [55], to enhance the angular resolution for imaging celestial objects. Although such long baseline interferometric techniques have a long history it only recently, due to exponential increase in available computational power, that the Event Horizon Telescope (EHT), based on the principle of VLBI, is able to capture the first image of black hole with the effective aperture size of the earth [56].

Notwithstanding the variety of existing optical super-resolution techniques, modal projection based optical super-resolution techniques inspired by quantum information theory have some inherent advantages: (1) applicability to passive imaging applications where it is not feasible to engineer object properties nor possible to deploy structured illumination (e.g., many long-standoff applications such as astronomy and long range horizontal imaging), and (2) approach (and potentially achieve) fundamental limits delivering order-of-magnitude better performance for a given aperture size and/or signal/photon budget relative to existing techniques, e.g., for applications such as exo-planet imaging and space surveillance. Moreover, we note that resolution can also be further enhanced by employing multi-baseline interferometry techniques to synthesize larger effective apertures [57], [58].

For a given optical PSF or equivalently optical aperture, spatial mode sorting required for implementing modal projection measurement can be realized using linear optics [59], [60]. Various spatial mode sorters designs, based on the principle of spatial-mode demultiplexing (SPADE) [61]–[63] or super-localization via image-inversion interferometry (SLIVER) [64], [65], have been proposed and demonstrated experimentally. However, existing mode sorters are designed to implement a fixed set of spatial modes, an adaptive mode sorter that can switch the modal basis dynamically is needed to implement the multi-parameter estimation technique proposed here. Recently, various efforts to realize such adaptive mode sorter have been reported in the literature [66]–[68].

## B. Formulation

We use the density operator formulation from [22], such that the quantum state of photons incident on the image plane viewed through an optical lens is given by the density operator  $\rho_{\text{full}}$ :

$$\rho_{\text{full}} = (1 - \epsilon)|0\rangle\langle 0| + \epsilon\rho + O(\epsilon^2), \quad (11)$$

where  $|0\rangle$  is the vacuum state,  $\rho$  is the single photon state density operator, which has the form of Eq. (10), and  $\epsilon$  is the total number of photons arriving on the image plane within the coherence time of the source. For  $\epsilon \ll 1$  (valid for weak thermal source), the photon states with order  $O(\epsilon^2)$  become negligible [69], [70]. Furthermore, in this application as we are only interested in the relative brightness of the sources, the vacuum state  $|0\rangle$  (which can be used to estimate the total source power) provides no useful information and thus we exclusively focus on  $\rho$ . The components of Eq. (10) have the following meaning in the current context:  $P$  is the number of point emitters,  $\{b_i\}_{i=1}^P$  are the relative brightness of each

point emitter or source (sum normalized to 1) and the states  $|\psi_i\rangle$  are given by:

$$|\psi_i\rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(x - x_i, y - y_i) |x, y\rangle dx dy, \quad (12)$$

such that  $(x_i, y_i)$  are the coordinates of the  $i^{\text{th}}$  point source on the image plane. Here the point spread function (PSF)  $\psi(x, y)$  of the imaging system is modeled by a 2D Gaussian function:

$$\psi(x, y) = \frac{1}{\sqrt{2\pi\sigma_x\sigma_y}} \exp\left(-\frac{x^2}{4\sigma_x^2} - \frac{y^2}{4\sigma_y^2}\right), \quad (13)$$

where  $\sigma_x$  and  $\sigma_y$  are the standard deviation (a measure of width) of the PSF in  $x$  and  $y$  direction respectively. For a given PSF,  $\sigma_x$  and  $\sigma_y$  are known parameters and we set  $\sigma_x = \sigma_y$  in our study. We define the full width at half maximum (proportional to  $\sigma_x$ ) of the PSF as Rayleigh length (rl) in our analysis.

The parameters of interest in this problem are thus the position and relative brightness of the  $P$  point emitters, i.e.  $\boldsymbol{\theta} = [x_1, \dots, x_P, y_1, \dots, y_P, b_1, \dots, b_P]^T = [\mathbf{x}, \mathbf{y}, \mathbf{b}]^T$ .

For the point emitter positions  $[\mathbf{x}, \mathbf{y}]^T$ , we use an independent Gaussian distribution  $\mathcal{N}$  prior:

$$p(\mathbf{x}, \mathbf{y}) = \prod_i^P \mathcal{N}(x_i; \bar{x}_i, \bar{\sigma}_{x_i}) \mathcal{N}(y_i; \bar{y}_i, \bar{\sigma}_{y_i}), \quad (14)$$

where for  $1 \leq i \leq P$ ,  $\bar{x}_i, \bar{y}_i, \bar{\sigma}_{x_i}, \bar{\sigma}_{y_i}$  are the mean and standard deviation of the position parameters  $x_i$  and  $y_i$  respectively.

For the brightness  $\mathbf{b}^T$  parameters a Dirichlet distribution [71] is used as a prior:  $p(\mathbf{b}) = \text{Dir}(\mathbf{b}; \mathbf{a})$ , where  $\mathbf{a} = [a_1, \dots, a_P]^T$  are the hyper-parameters of the Dirichlet distribution. Thus, the overall prior is expressed as:  $p(\mathbf{x}, \mathbf{y}, \mathbf{b}) = p(\mathbf{x}, \mathbf{y})p(\mathbf{b})$ .

This specifies all the relevant details (i.e., photon state density operator, prior distribution) for the proposed adaptive sequential measurement scheme described in the previous section. Note that as  $p(\mathbf{x}, \mathbf{y}, \mathbf{b})$  is not a conjugate prior for the Poisson likelihood, we update the hyper-parameters of the prior distribution at  $\tau^{\text{th}}$  iteration to derive the posterior, which assumes the role of the prior in the next  $(\tau + 1)^{\text{th}}$  iteration. The prior hyper-parameters are:  $\mathbf{h} = [\bar{x}_1, \dots, \bar{x}_P, \bar{y}_1, \dots, \bar{y}_P, \bar{\sigma}_{x_1}, \dots, \bar{\sigma}_{x_P}, \bar{\sigma}_{y_1}, \dots, \bar{\sigma}_{y_P}, a_1, \dots, a_P, \delta]^T = [\bar{\mathbf{x}}, \bar{\mathbf{y}}, \bar{\boldsymbol{\sigma}}_x, \bar{\boldsymbol{\sigma}}_y, \mathbf{a}, \delta]^T$ . Here,  $\delta$  is another hyper-parameter associated with the brightness prior distribution which is explained later.

To update the hyper-parameters of the position prior at the  $(\tau + 1)^{\text{th}}$  iteration, we use the first- and the second-moments of the posterior distribution at the  $\tau^{\text{th}}$  iteration:

$$\bar{\alpha}_i^{(\tau+1)} = \int \alpha_i p(\boldsymbol{\theta} | \mathbf{l}^{(\tau)}; \mathbf{h}^{(\tau)}) d\boldsymbol{\theta}, \quad (15)$$

$$\bar{\sigma}_{\alpha_i}^{(\tau+1)} = \int [\alpha_i - \bar{\alpha}_i^{(\tau+1)}]^2 p(\boldsymbol{\theta} | \mathbf{l}^{(\tau)}; \mathbf{h}^{(\tau)}) d\boldsymbol{\theta}, \quad (16)$$

where  $\alpha$  represents  $x$  or  $y$  co-ordinate.

For the hyper-parameters  $\mathbf{a}^T$  of the brightness prior, an expectation maximization (EM) approach is used. We first find the mean of the brightness vector as:

$$\hat{\mathbf{b}}_i^{(\tau+1)} = \int b_i p(\boldsymbol{\theta} | \mathbf{l}^{(\tau)}; \mathbf{h}^{(\tau)}) d\boldsymbol{\theta}. \quad (17)$$

Then,  $\mathbf{a}^T$  is updated such that  $\hat{\mathbf{b}}^{(\tau+1)}$  becomes the mode of the distribution:

$$\begin{aligned} \mathbf{a}^{(\tau+1)} &= \hat{\mathbf{b}}^{(\tau+1)} [a_0^{(\tau)} + \delta^{(\tau)} - P] + 1 \\ &= \hat{\mathbf{b}}^{(\tau+1)} [a_0^{(\tau+1)} - P] + 1, \end{aligned} \quad (18)$$

where  $a_0^{(\tau)} = \sum_i^P a_i^{(\tau)}$  and  $a_0^{(\tau+1)} = a_0^{(\tau)} + \delta^{(\tau)}$ . Qualitatively the larger the  $a_0^{(\tau)}$ , the smaller the total variance of the Dirichlet distribution. Adding  $\delta^{(\tau)} \geq 0$  leads to  $a_0^{(\tau+1)} \geq a_0^{(\tau)}$ , such that the variance reduces monotonically with each iteration  $\tau$ . Note that the introduction of  $\delta^{(\tau)}$  does not change the position of the mode in the distribution. We set  $\delta^{(\tau)}$  to a constant for all  $\tau$ .

When  $P$  (i.e., number of point emitters) is unknown *a priori*, we select the model in each measurement cycle as follows. Let  $p_P(\mathbf{l}^{(\tau)})$  denote the likelihood of the model consisting of  $P$  point emitters in the  $\tau^{\text{th}}$  cycle. We calculate the following weighted log likelihood  $Z_P^{(\tau)}$ :

$$Z_P^{(\tau)} = \sum_{t=1}^{\tau} \exp \left[ -\kappa \left( 1 - \frac{t}{\tau} \right) \right] \ln p_P(\mathbf{l}^{(t)}), \quad (19)$$

and pick the model with largest  $Z_P^{(\tau)}$  as the estimate in the  $(\tau + 1)^{\text{th}}$  measurement cycle. Here we set  $\kappa = 10$  for this illustrative application. The selection of  $\kappa$  affects the performance of the estimation scheme and its optimization is outside the scope of this work. We use this model selection method instead of simply picking the model that maximizes  $p_P(\mathbf{l}^{(\tau)})$  so that the estimation of  $P$  is not purely determined by the single measurement result at iteration  $\tau$ . As measurements proceed, the more recent likelihoods successively improve due to improving precision of parameter estimates and thus larger weight are assigned to them. It should be noted that this model selection method is only illustrative and more appropriate and efficient model selection methods may be applicable in other applications.

### C. Simulation Results

We demonstrate the performance of the proposed adaptive sequential measurement scheme for 100 distinct randomized realizations of 3-point emitter constellations.

The position of the 1<sup>st</sup> point emitter is uniformly distributed inside a circle with radius of 0.375 rl (Rayleigh length). The position of the  $i^{\text{th}}$  emitter,  $i > 1$ , is  $[x_i, y_i] = [x_{i-1}, y_{i-1}] + [(d + \delta d)\cos\phi, (d + \delta d)\sin\phi]$ , where  $d$  is a constant,  $\delta d$  and  $\phi$  are uniformly distributed random variables over the intervals  $[-\delta d_0/2, \delta d_0/2]$  and  $[0, 2\pi)$  respectively, for some constant  $\delta d_0$ , such that  $0 \leq \delta d_0 < 2d$ . The position of the  $i^{\text{th}}$  emitter  $[x_i, y_i]$  is re-selected if it falls outside the 0.375 rl circle (field of view) or the separation of any pair of sources is smaller than  $d - \delta d_0/2$ . By doing so, for each emitter, the closest neighbour is located around  $d$  and minimum separation of any pair of point emitters is guaranteed to be no less than  $d - \delta d_0/2$ . In

the simulation below we set  $d = 0.1$  (rl) and  $\delta d_0 = 0.1d$ . The relative brightness of point emitters set to be equal/uniform. The average total photon budget  $N$  is set to  $5 \times 10^5$  and each adaptive sequential step utilizes around  $10^4$  photons (i.e., the mean of  $K_\tau$  is  $10^4$  for  $\tau \leq 1$ ). The adaptive sequential scheme is initialized by employing 1000 photons for a direct imaging measurement (i.e., the mean of  $K_0$  is 1000) followed by using an expectation maximization (EM) algorithm to estimate the initial model parameters. The remaining photons are detected by using Personick projection measurement in each adaptive sequential step.

For the traditional direct imaging (serving as a baseline), which uses direct focal plane intensity measurements of all available  $N$  photon copies, the Richardson-Lucy deconvolution algorithm [72] is first used to deconvolve the blurred image followed by the k-mean clustering algorithm [73] to find the position and relative brightness of identified point emitters<sup>1</sup>.

### D. Estimation with $P$ known exactly

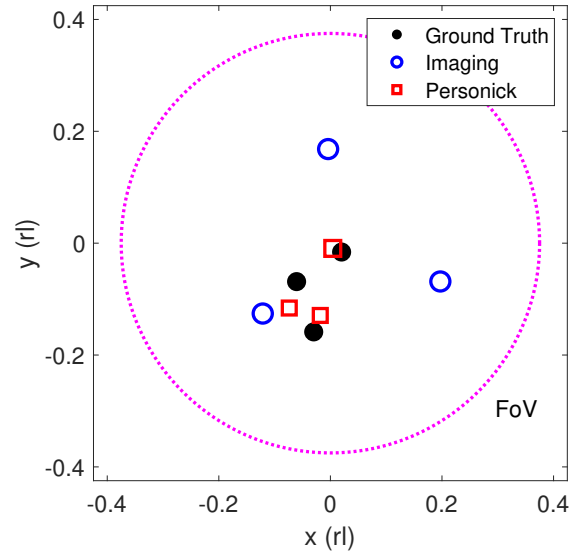


Fig. 4. An illustrative example of a cluster of point emitters estimated with direct imaging (baseline) and Personick projection measurements (proposed adaptive measurement), when  $P$  is known exactly. The black dots, blue circles and red squares correspond to the ground truth, estimates obtained from direct imaging and Personick projection measurements respectively. The marker size is proportional to the point emitter brightness.

For each of the 100 constellations, we employ 10 Monte Carlo simulation (i.e., different noise realizations). Fig. 4 shows an illustrative realization of the point emitter cluster and estimated emitter locations and brightness using the two measurement schemes.

To obtain the average performance of the proposed adaptive measurement scheme, for each point emitter realization, we first pair the ground truth point emitter location with the estimated locations, such that the sum of the position errors

<sup>1</sup>We would like to acknowledge the DARPA IAMBIC JHU-APL team for providing this algorithm.

defined as:  $\sum_{i=1}^P \sqrt{(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2}$ , over all point-source matched pairs is minimized. The average (over all emitters) position error distribution of the point emitters is shown in Fig. 5. We observe that the proposed adaptive scheme outperforms the direct imaging. More specifically, the mean position error obtained by our adaptive scheme is six-fold lower than that of the direct imaging. Also, the position error distribution of the Personick measurement is more concentrated and position errors for all estimates is less than  $d = 0.1(\text{rl})$ .

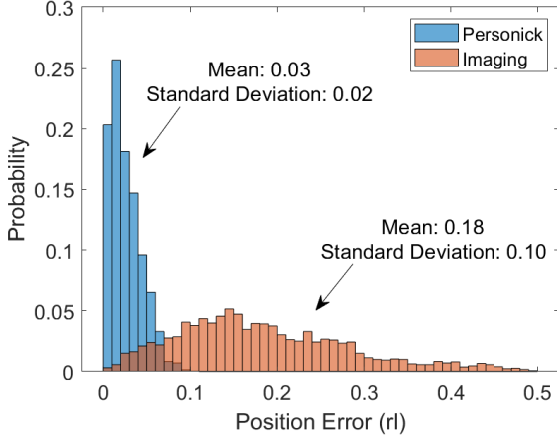


Fig. 5. Distribution of the point emitter position errors obtained with the two measurement schemes, when  $P$  is known exactly.

### E. Estimation with unknown $P$

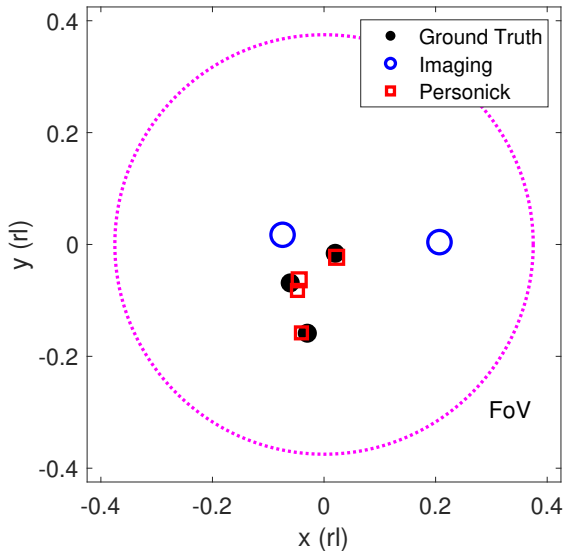


Fig. 6. An illustrative example of a cluster of point emitters estimated using direct imaging and Personick projection measurement schemes, when  $P$  is unknown. The marker definition is same as in Fig. 4.

When  $P_{max} = 6$  is used as a prior, recall that the estimation algorithm has to also estimate  $P$ . One such illustrative example is shown in Fig. 6. We observe that even if the Personick

measurement predicts the wrong number of sources (i.e., four instead of three), the reconstructed point emitter distribution is closer to the ground truth compared to the reconstruction obtained with the direct imaging measurement, which underestimates the number of point emitters as two in this particular instance. To analyze the performance quantitatively, using the same set of constellations and same number of simulations, the distribution of number of point emitters estimated by the two measurement schemes is shown in Fig. 7. We observe that the adaptive Personick projective scheme estimates the correct number of point emitters with a 50% success rate relative to only 10% for direct imaging. The  $P$  estimated by our proposed Personick projective measurement scheme is more concentrated around the true value  $P = 3$  while that of direct imaging is more spread out across the range of possible  $P$ . Fig. 8 shows the corresponding position error distribution, computed only for cases where  $P \geq 3$  such that none of the estimated point emitters sources are merged. We observe that when the  $P$  is estimated correctly the proposed adaptive scheme maintains the significant performance advantage over direct imaging in terms of significantly lower point emitter localization error.

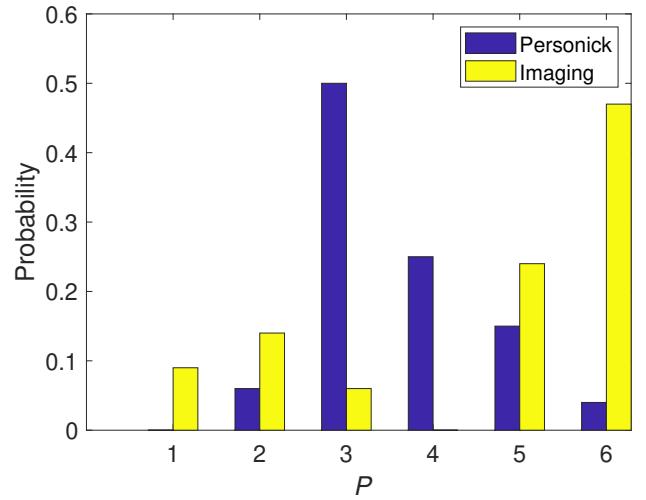


Fig. 7. Distribution of the number of point emitters estimated by the two measurement schemes.

## VI. CONCLUSIONS

Based on quantum estimation theory, we propose an adaptive sequential Bayesian multi-parameter estimation scheme with applications in sensing and imaging. Using an illustrative example task of optical super-resolution of point emitters embedded in a constellation, relevant in many optical imaging applications e.g., astronomy and microscopy, we demonstrate its superior performance relative to the direct diffraction-limited imaging approach. Our simulation study results show a nearly six-fold lower point emitter localization error achieved by our proposed measurement/estimation scheme relative to the state-of-the-art direct imaging scheme in the sub-Rayleigh regime. It is also worth noting that our POVM choice i.e., measurement design used in each sequential measurement step is provably optimal (in the MSE sense) when estimating

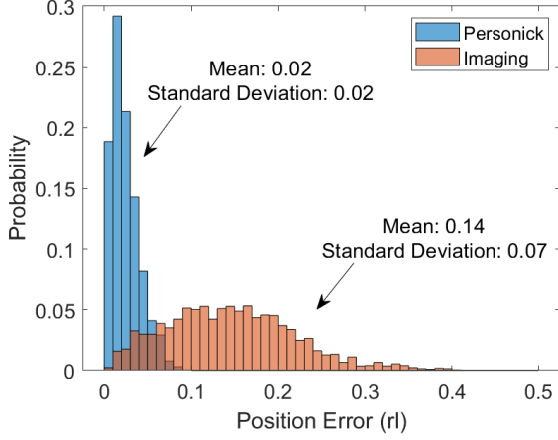


Fig. 8. Distribution of the point emitter position errors obtained with the two measurement schemes, when  $P$  is not known *a priori*.

a scalar parameter, which is a linear combination of the multiple parameters of interest in a given estimation task. We believe that our proposed multi-parameter estimation scheme is an important step towards achieving quantum optimal performance for several practical sensing and imaging tasks, especially for sources/objects/signals which are fully described by semi-classical models that span a wide-range of real-world applications, such as exo-planet search, fluorescence microscopy and space situational awareness.

We further note that in this work we illustrated the proposed sequential adaptive measurement scheme, applicable to a general multi-parameter estimation problem, for a specific choice of priors, adaptation batch size and other relevant implementation parameters. However, the optimization of the implementation details, such as the choice of prior, model selection method, the number of adaptation steps, number of photons in each adaptation steps etc., require future work in context of a given problem/application. Moreover, with regard to the scalar parameter choice  $\gamma^{(\tau)}$  in each adaption step, one may also consider a more general function of the parameters of interest instead of their linear combination as the single parameter as chosen in this work. In such a case the optimal choice of such a scalar function remains an open problem that requires further analysis. We are also currently exploring extensions of proposed multi-parameter estimation scheme along various directions; including optimal or near-optimal measurement design for more than one parameters in each adaptive sequential step as well as non-greedy adaptive sequential scheme(s) within a game-theoretic framework inspired by our prior work [47].

#### APPENDIX A

##### POSTERIOR MEAN SATURATES QUANTUM BOUND

For any parameters  $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_M]^T$  and their estimators  $\hat{\boldsymbol{\theta}}(\mathbf{l}) = [\hat{\theta}_1(\mathbf{l}), \hat{\theta}_2(\mathbf{l}), \dots, \hat{\theta}_M(\mathbf{l})]^T$ , where  $\mathbf{l} = [l_1, l_2, \dots, l_N]^T$  are the measurement through a channel represented by POVM  $\{\Pi\}$ , the covariance matrix elements are:  $[\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}), \boldsymbol{\theta})]_{ij} = \text{E}[(\theta_i - \hat{\theta}_i(\mathbf{l}))(\theta_j - \hat{\theta}_j(\mathbf{l}))]$ . In Bayesian setting, the expectation is taken over the joint distribution

$p(\boldsymbol{\theta}, \mathbf{l})$  of both  $\boldsymbol{\theta}$  and  $\mathbf{l}$ . Upon expanding, the covariance matrix elements can be also written as:

$$[\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}), \boldsymbol{\theta})]_{ij} = \text{E}[\theta_i \theta_j] - \text{E}[\theta_i \hat{\theta}_j(\mathbf{l})] - \text{E}[\theta_j \hat{\theta}_i(\mathbf{l})] + \text{E}[\hat{\theta}_i(\mathbf{l}) \hat{\theta}_j(\mathbf{l})]. \quad (20)$$

The first term  $\text{E}[\theta_i \theta_j]$  in Eq. (20) matches the first term in Eq. (4). Now, substitute  $\hat{\theta}_i(\mathbf{l}) = \int \theta_i p(\boldsymbol{\theta}|\mathbf{l}) d\boldsymbol{\theta}$ , the second term of Eq. (20) becomes:

$$\begin{aligned} \text{E}[\theta_i \hat{\theta}_j(\mathbf{l})] &= \int \int \theta_i \left[ \int \theta'_j p(\boldsymbol{\theta}'|\mathbf{l}) d\boldsymbol{\theta}' \right] p(\boldsymbol{\theta}, \mathbf{l}) d\boldsymbol{\theta} d\mathbf{l} \\ &= \int \left[ \int \theta_i p(\boldsymbol{\theta}|\mathbf{l}) d\boldsymbol{\theta} \right] \left[ \int \theta'_j p(\boldsymbol{\theta}'|\mathbf{l}) d\boldsymbol{\theta}' \right] p(\mathbf{l}) d\mathbf{l} \\ &= J_{ij}. \end{aligned} \quad (21)$$

Similarly, the third and the fourth terms in Eq. (20) equal  $J_{ij}$ , and thus  $\text{Cov}(\hat{\boldsymbol{\theta}}(\mathbf{l}), \boldsymbol{\theta}) = \Sigma_C$  exactly. If  $M = 1$ , i.e.  $\boldsymbol{\theta}$  is a single parameter,  $\Sigma_C$  gives the minimum mean square error with the posterior mean as the estimator.

#### APPENDIX B PARAMETER SELECTION

For a single parameter  $\gamma = \mathbf{h} \cdot \boldsymbol{\theta}$ , where  $\mathbf{h}$  is some unit vector, the variance of estimating  $\gamma$  is:

$$\begin{aligned} \text{Var}(\gamma) &= \int \int d\mathbf{l} [\mathbf{h} \cdot \hat{\boldsymbol{\theta}}(\mathbf{l}) - \mathbf{h} \cdot \boldsymbol{\theta}]^2 \\ &= \mathbf{h}^T \text{Cov}[\hat{\boldsymbol{\theta}}(\mathbf{l}), \boldsymbol{\theta}] \mathbf{h} \\ &\geq \mathbf{h}^T \Sigma_Q \mathbf{h}. \end{aligned} \quad (22)$$

Applying Eq. (6) for  $k = 1$  to  $\gamma$ , we have:

$$\begin{aligned} \Gamma_{1,\gamma} &= \int d\boldsymbol{\theta} p(\boldsymbol{\theta}) \rho(\boldsymbol{\theta}) (\mathbf{h} \cdot \boldsymbol{\theta}) \\ &= \sum_i h_i \Gamma_{1,i}. \end{aligned} \quad (23)$$

Thus, if we set  $B_\gamma = \sum_i h_i B_i$ , Eq. (7) can be satisfied. Under the constrain  $|\mathbf{h}|^2 = 1$ , we search for the  $\mathbf{h}$  that minimizes the MMSE =  $\text{Tr}[\Gamma_{2,\gamma} - B_\gamma \Gamma_{1,\gamma}]$  using Lagrange multiplier:

$$\begin{aligned} L &= \int d\boldsymbol{\theta} p(\boldsymbol{\theta}) (\mathbf{h} \cdot \boldsymbol{\theta})^2 - \text{Tr} \left[ \left( \sum_j h_j B_j \right) \left( \sum_k h_k \Gamma_{1,k} \right) \right] \\ &\quad - \lambda \left( \sum_i h_i^2 - 1 \right). \end{aligned} \quad (24)$$

Taking the derivative with respect to  $\mathbf{h}$  and  $\lambda$ , we have:

$$\begin{aligned} \frac{\partial L}{\partial h_i} &= 2 \int d\boldsymbol{\theta} p(\boldsymbol{\theta}) (\mathbf{h} \cdot \boldsymbol{\theta}) \theta_i - 2\lambda h_i \\ &\quad - \text{Tr} \left[ B_i \left( \sum_k h_k \Gamma_{1,k} \right) + \left( \sum_j h_j B_j \right) \Gamma_{1,i} \right] = 0, \end{aligned} \quad (25)$$

$$\frac{\partial L}{\partial \lambda} = \sum_i h_i^2 - 1 = 0. \quad (26)$$

Using Eq. (25), we have:

$$\begin{aligned}
 h_i &= \frac{1}{\lambda} \sum_j h_j \left[ \int d\theta p(\theta) \theta_i \theta_j - \frac{1}{2} \text{Tr} \left( B_i \Gamma_{1,j} + B_j \Gamma_{1,i} \right) \right] \\
 &= \frac{1}{\lambda} \sum_j h_j \left[ \int d\theta p(\theta) \theta_i \theta_j - \text{Tr} \left( \Gamma_0 \frac{B_i B_j + B_j B_i}{2} \right) \right] \\
 &= \frac{1}{\lambda} \sum_j h_j \left[ \int d\theta p(\theta) \theta_i \theta_j - G_{ij} \right]. \quad (27)
 \end{aligned}$$

Substituting Eq. (27) into Eq. (26), we have:

$$\lambda = \sqrt{\sum_i \left\{ \left[ \sum_j h_j \left[ \int d\theta p(\theta) \theta_i \theta_j - G_{ij} \right] \right]^2 \right\}}. \quad (28)$$

It can be easily seen that Eq. (27) is precisely the equations to determine the eigenvector of  $\Sigma_Q$ , with the corresponding eigenvalue  $\lambda$ . Thus, if we pick the eigenvector  $\mathbf{h}_m$  with the smallest eigenvalue  $\lambda_m$ , the MMSE would be:  $\mathbf{h}_m^T \Sigma_Q \mathbf{h}_m = \lambda_m$ , which can be saturated by the eigen-projection measurement of  $B_\gamma$ . Since the MMSE is lower bounded by 0,  $\lambda_m$  is the global minimum. In other words,  $\lambda_m$  is the minimum MMSE one can get for a single parameter which is the linear combination of the original parameters  $\theta$ .

#### REFERENCES

- [1] S. D. Personick, "Application of quantum estimation theory to analog communication over quantum channels," *IEEE Trans. Inf. Theory*, vol. 17, pp. 240–246, 1971.
- [2] P. J. D. Crowley, A. Datta, M. Barbieri, and I. A. Walmsley, "Tradeoff in simultaneous quantum-limited phase and loss estimation in interferometry," *Phys. Rev. A*, vol. 89, p. 023845, Feb 2014. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.89.023845>
- [3] M. D. Vidrighin, G. Donati, M. G. Genoni, X.-M. Jin, W. S. Kolthammer, M. S. Kim, A. Datta, M. Barbieri, and I. A. Walmsley, "Joint estimation of phase and phase diffusion for quantum metrology," *Nature Communications*, vol. 5, no. 1, p. 3532, Apr 2014. [Online]. Available: <https://doi.org/10.1038/ncomms4532>
- [4] E. Roccia, V. Cimini, M. Sbroscia, I. Gianani, L. Ruggiero, L. Mancino, M. G. Genoni, M. A. Ricci, and M. Barbieri, "Multiparameter approach to quantum phase estimation with limited visibility," *Optica*, vol. 5, no. 10, pp. 1171–1176, Oct 2018. [Online]. Available: <http://opg.optica.org/optica/abstract.cfm?URI=optica-5-10-1171>
- [5] I. Gianani, F. Albarelli, A. Verna, V. Cimini, R. Demkowicz-Dobrzanski, and M. Barbieri, "Kramers–kronig relations and precision limits in quantum phase estimation," *Optica*, vol. 8, no. 12, pp. 1642–1645, Dec 2021. [Online]. Available: <http://opg.optica.org/optica/abstract.cfm?URI=optica-8-12-1642>
- [6] S. Z. Ang, R. Nair, and M. Tsang, "Quantum limit for two-dimensional resolution of two incoherent optical point sources," *Phys. Rev. A*, vol. 95, p. 063847, Jun 2017. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.95.063847>
- [7] C. Napoli, S. Piano, R. Leach, G. Adesso, and T. Tufarelli, "Towards superresolution surface metrology: Quantum estimation of angular and axial separations," *Phys. Rev. Lett.*, vol. 122, p. 140505, Apr 2019. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.122.140505>
- [8] S. Prasad and Z. Yu, "Quantum-limited superlocalization and superresolution of a source pair in three dimensions," *Phys. Rev. A*, vol. 99, p. 022116, Feb 2019. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.99.022116>
- [9] M. Tsang, H. M. Wiseman, and C. M. Caves, "Fundamental quantum limit to waveform estimation," *Phys. Rev. Lett.*, vol. 106, p. 090401, Mar 2011. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.106.090401>
- [10] S. Boulebnane, M. P. Woods, and J. M. Renes, "Waveform estimation from approximate quantum nondemolition measurements," *Phys. Rev. Lett.*, vol. 127, p. 010502, Jun 2021. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.127.010502>
- [11] H. Yuan, "Sequential feedback scheme outperforms the parallel scheme for hamiltonian parameter estimation," *Phys. Rev. Lett.*, vol. 117, p. 160801, Oct 2016. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.117.160801>
- [12] H. Yuan and C.-H. F. Fung, "Quantum parameter estimation with general dynamics," *npj Quantum Information*, vol. 3, no. 1, p. 14, Apr 2017. [Online]. Available: <https://doi.org/10.1038/s41534-017-0014-6>
- [13] S. Pang and T. A. Brun, "Quantum metrology for a general hamiltonian parameter," *Phys. Rev. A*, vol. 90, p. 022117, Aug 2014. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.90.022117>
- [14] S. Gao, M. O'Sullivan, and R. Hui, "Complex-optical-field lidar system for range and vector velocity measurement," *Optics express*, vol. 20, no. 23, pp. 25 867–25 875, 2012.
- [15] Q. Zhuang, Z. Zhang, and J. H. Shapiro, "Entanglement-enhanced lidars for simultaneous range and velocity measurements," *Phys. Rev. A*, vol. 96, p. 040304, Oct 2017. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.96.040304>
- [16] D. Bruneau and J. Pelon, "Simultaneous measurements of particle backscattering and extinction coefficients and wind velocity by lidar with a mach–zehnder interferometer: principle of operation and performance assessment," *Appl. Opt.*, vol. 42, no. 6, pp. 1101–1114, Feb 2003. [Online]. Available: <http://opg.optica.org/ao/abstract.cfm?URI=ao-42-6-1101>
- [17] B. Xu, Y. Ren, P. Zhu, and M. Lu, "A pso-based approach for multi-cell multi-parameter estimation," in *The 2014 International Conference on Control, Automation and Information Sciences (ICCAIS 2014)*, 2014, pp. 65–70.
- [18] C. Hellriegel and E. Gratton, "Real-time multi-parameter spectroscopy and localization in three-dimensional single-particle tracking," *Journal of The Royal Society Interface*, vol. 6, no. suppl\_1, pp. S3–S14, 2009. [Online]. Available: <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2008.0313.focus>
- [19] E. Van Derlinden, K. Bernaerts, and J. Van Impe, "Simultaneous versus sequential optimal experiment design for the identification of multi-parameter microbial growth kinetics as a function of temperature," *Journal of Theoretical Biology*, vol. 264, no. 2, pp. 347–355, 2010. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S002251931000007X>
- [20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1997.
- [21] C. W. Helstrom, "Quantum detection and estimation theory," *Journal of Statistical Physics*, vol. 1, no. 2, pp. 231–252, Jun 1969. [Online]. Available: <https://doi.org/10.1007/BF01007479>
- [22] M. Tsang, R. Nair, and X.-M. Lu, "Quantum theory of superresolution for two incoherent optical point sources," *Phys. Rev. X*, vol. 6, p. 031033, Aug 2016. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevX.6.031033>
- [23] C. Lupo, Z. Huang, and P. Kok, "Quantum limits to incoherent imaging are achieved by linear interferometry," *Phys. Rev. Lett.*, vol. 124, p. 080503, Feb 2020. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.124.080503>
- [24] M. D. Lang and C. M. Caves, "Optimal quantum-enhanced interferometry using a laser power source," *Phys. Rev. Lett.*, vol. 111, p. 173601, Oct 2013. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.111.173601>
- [25] R. Demkowicz-Dobrzański and L. Maccone, "Using entanglement against noise in quantum metrology," *Phys. Rev. Lett.*, vol. 113, p. 250801, Dec 2014. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.113.250801>
- [26] H. L. Van Trees, *Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory*. John Wiley & Sons, 2004.
- [27] R. D. Gill and B. Y. Levit, "Applications of the van trees inequality: A bayesian cramer–rao bound," *Bernoulli*, vol. 1, no. 1/2, pp. 59–79, 1995. [Online]. Available: <http://www.jstor.org/stable/3318681>
- [28] J. Rubio and J. Dunningham, "Quantum metrology in the presence of limited data," *New Journal of Physics*, vol. 21, no. 4, p. 043037, apr 2019. [Online]. Available: <https://doi.org/10.1088/1367-2630/ab098b>
- [29] —, "Bayesian multiparameter quantum metrology with limited data," *Physical Review A*, vol. 101, no. 3, p. 032114, 2020.
- [30] S. Ragy, M. Jarzyna, and R. Demkowicz-Dobrzański, "Compatibility in multiparameter quantum metrology," *Phys. Rev. A*, vol. 94, p. 052108, Nov 2016. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.94.052108>
- [31] X.-L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ecm algorithm: A general framework," *Biometrika*, vol. 80, no. 2, pp. 267–278, 1993.

- [32] N. T. Longford, "A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects," *Biometrika*, vol. 74, no. 4, pp. 817–827, 1987.
- [33] S. Guindon and O. Gascuel, "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood," *Systematic biology*, vol. 52, no. 5, pp. 696–704, 2003.
- [34] N. Wiebe and C. Granade, "Efficient bayesian phase estimation," *Phys. Rev. Lett.*, vol. 117, p. 010503, Jun 2016. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.117.010503>
- [35] A. M. Ellison, "Bayesian inference in ecology," *Ecology Letters*, vol. 7, no. 6, pp. 509–520, 2004. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1461-0248.2004.00603.x>
- [36] M. P. V. Stenberg, Y. R. Sanders, and F. K. Wilhelm, "Efficient estimation of resonant coupling between quantum systems," *Phys. Rev. Lett.*, vol. 113, p. 210404, Nov 2014. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.113.210404>
- [37] F. Huszár and N. M. T. Houlsby, "Adaptive bayesian quantum tomography," *Phys. Rev. A*, vol. 85, p. 052120, May 2012. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.85.052120>
- [38] X.-L. Luo, J. Feng, and H.-H. Zhang, "A genetic algorithm for astroparticle physics studies," *Computer Physics Communications*, vol. 250, p. 106818, 2020.
- [39] P. Carnevali, L. Coletti, and S. Patarnello, "Image processing by simulated annealing," in *Readings in Computer Vision*. Elsevier, 1987, pp. 551–561.
- [40] G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," *Science*, vol. 355, no. 6325, pp. 602–606, 2017.
- [41] L. Ge, H. Liang, J. Yuan, and D. Thalmann, "3d convolutional neural networks for efficient and robust hand pose estimation from single depth images," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5679–5688.
- [42] D. F. V. James, "Quantum computation and quantum information," *Physics Today*, vol. 54, no. 11, pp. 60–62, 2001. [Online]. Available: <https://doi.org/10.1063/1.1428442>
- [43] J. Liu, H. Yuan, X.-M. Lu, and X. Wang, "Quantum fisher information matrix and multiparameter estimation," *Journal of Physics A: Mathematical and Theoretical*, vol. 53, no. 2, p. 023001, dec 2019. [Online]. Available: <https://doi.org/10.1088/1751-8121/ab5d4d>
- [44] J. S. Sidhu and P. Kok, "Geometric perspective on quantum parameter estimation," *AVS Quantum Science*, vol. 2, no. 1, p. 014701, 2020. [Online]. Available: <https://doi.org/10.1116/1.5119961>
- [45] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*. Edizioni della Normale, 2011.
- [46] A. Vince, "A framework for the greedy algorithm," *Discrete Applied Mathematics*, vol. 121, no. 1-3, pp. 247–260, 2002.
- [47] L.-C. Huang, M. A. Neifeld, and A. Ashok, "Face recognition with non-greedy information-optimal adaptive compressive imaging," *Appl. Opt.*, vol. 55, no. 34, pp. 9744–9755, Dec 2016. [Online]. Available: <http://opg.optica.org/ao/abstract.cfm?URI=ao-55-34-9744>
- [48] S. Bharathi, D. Kempe, and M. Salek, "Competitive influence maximization in social networks," in *Internet and Network Economics*, X. Deng and F. C. Graham, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 306–311.
- [49] L. Rayleigh, "Xxxi. investigations in optics, with special reference to the spectroscope," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 8, no. 49, pp. 261–274, 1879.
- [50] M. Yamanaka, N. I. Smith, and K. Fujita, "Introduction to super-resolution microscopy," *Microscopy*, vol. 63, no. 3, pp. 177–192, 03 2014. [Online]. Available: <https://doi.org/10.1093/jmicro/dfu007>
- [51] M. Honma, K. Akiyama, M. Uemura, and S. Ikeda, "Super-resolution imaging with radio interferometry using sparse modeling," *Publications of the Astronomical Society of Japan*, vol. 66, no. 5, 09 2014, 95. [Online]. Available: <https://doi.org/10.1093/pasj/psu070>
- [52] B. Hein, K. I. Willig, and S. W. Hell, "Stimulated emission depletion (sted) nanoscopy of a fluorescent protein-labeled organelle inside a living cell," *Proceedings of the National Academy of Sciences*, vol. 105, no. 38, pp. 14 271–14 276, 2008.
- [53] M. G. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," *Journal of microscopy*, vol. 198, no. 2, pp. 82–87, 2000.
- [54] J. R. Allen, S. T. Ross, and M. W. Davidson, "Single molecule localization microscopy for superresolution," *Journal of Optics*, vol. 15, no. 9, p. 094001, 2013.
- [55] M. H. Cohen, "Introduction to very-long-baseline interferometry," *Proceedings of the IEEE*, vol. 61, no. 9, pp. 1192–1197, 1973.
- [56] K. Akiyama, K. Bouman, and D. Woody, "First m87 event horizon telescope results. i. the shadow of the supermassive black hole," *Astrophysical Journal Letters*, vol. 875, 04 2019.
- [57] C. G. Ebeling, A. Meiri, J. Martineau, Z. Zalevsky, J. M. Gerton, and R. Menon, "Increased localization precision by interference fringe analysis," *Nanoscale*, vol. 7, no. 23, pp. 10430–10437, 2015.
- [58] A. Meiri, C. G. Ebeling, J. Martineau, Z. Zalevsky, J. M. Gerton, and R. Menon, "Interference based localization of single emitters," *Optics Express*, vol. 25, no. 15, pp. 17 174–17 191, 2017.
- [59] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discrete unitary operator," *Phys. Rev. Lett.*, vol. 73, pp. 58–61, Jul 1994. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.73.58>
- [60] J.-F. Morizur, L. Nicholls, P. Jian, S. Armstrong, N. Treps, B. Hage, M. Hsu, W. Bowen, J. Janousek, and H.-A. Bachor, "Programmable unitary spatial mode manipulation," *J. Opt. Soc. Am. A*, vol. 27, no. 11, pp. 2524–2531, Nov 2010. [Online]. Available: <http://opg.optica.org/josaa/abstract.cfm?URI=josaa-27-11-2524>
- [61] Y. Zhou, J. Zhao, Z. Shi, S. M. H. Rafsanjani, M. Mirhosseini, Z. Zhu, A. E. Willner, and R. W. Boyd, "Hermite gaussian mode sorter," *Opt. Lett.*, vol. 43, no. 21, pp. 5263–5266, Nov 2018. [Online]. Available: <http://opg.optica.org/ol/abstract.cfm?URI=ol-43-21-5263>
- [62] Y. Zhou, J. Yang, J. D. Hassett, S. M. H. Rafsanjani, M. Mirhosseini, A. N. Vamivakas, A. N. Jordan, Z. Shi, and R. W. Boyd, "Quantum-limited estimation of the axial separation of two incoherent point sources," *Optica*, vol. 6, no. 5, pp. 534–541, May 2019. [Online]. Available: <http://opg.optica.org/optica/abstract.cfm?URI=optica-6-5-534>
- [63] N. K. Fontaine, R. Ryf, H. Chen, D. T. Neilson, K. Kim, and J. Carpenter, "Laguerre-gaussian mode sorter," *Nature Communications*, vol. 10, no. 1, p. 1865, Apr 2019. [Online]. Available: <https://doi.org/10.1038/s41467-019-09840-4>
- [64] Z. S. Tang, K. Durak, and A. Ling, "Fault-tolerant and finite-error localization for point emitters within the diffraction limit," *Opt. Express*, vol. 24, no. 19, pp. 22 004–22 012, Sep 2016. [Online]. Available: <http://opg.optica.org/oe/abstract.cfm?URI=oe-24-19-22004>
- [65] W. Larson, N. V. Tabiryan, and B. E. A. Saleh, "A common-path polarization-based image-inversion interferometer," *Opt. Express*, vol. 27, no. 4, pp. 5685–5695, Feb 2019. [Online]. Available: <http://opg.optica.org/oe/abstract.cfm?URI=oe-27-4-5685>
- [66] M. R. Grace, Z. Dutton, A. Ashok, and S. Guha, "Approaching quantum-limited imaging resolution without prior knowledge of the object location," *JOSA A*, vol. 37, no. 8, pp. 1288–1299, 2020.
- [67] D. Fu, Y. Zhou, R. Qi, S. Oliver, Y. Wang, S. M. H. Rafsanjani, J. Zhao, M. Mirhosseini, Z. Shi, P. Zhang *et al.*, "Realization of a scalable laguerre-gaussian mode sorter based on a robust radial mode sorter," *Optics express*, vol. 26, no. 25, pp. 33 057–33 065, 2018.
- [68] H. Defienne and D. Faccio, "Arbitrary spatial mode sorting in a multimode fiber," *Physical Review A*, vol. 101, no. 6, jun 2020. [Online]. Available: <https://doi.org/10.1103/PhysRevA.101.063830>
- [69] L. Mandel, "Fluctuations of photon beams: the distribution of the photoelectrons," *Proceedings of the Physical Society (1958-1967)*, vol. 74, no. 3, p. 233, 1959.
- [70] J. W. Goodman, *Statistical optics*. John Wiley & Sons, 2015.
- [71] M. H. DeGroot, *Optimal Statistical Decisions*. John Wiley & Sons, 2005.
- [72] W. H. Richardson, "Bayesian-based iterative method of image restoration\*," *J. Opt. Soc. Am.*, vol. 62, no. 1, pp. 55–59, Jan 1972. [Online]. Available: <http://www.osapublishing.org/abstract.cfm?URI=josa-62-1-55>
- [73] J. Wu, "Cluster analysis and k-means clustering: an introduction," in *Advances in K-means Clustering*. Springer, 2012, pp. 1–16.