Information Theoretical Analysis of Quantum Metrology

Yi Peng\textsuperscript{1,2} and Heng Fan\textsuperscript{1,3,4,*}

\textsuperscript{1}Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China
\textsuperscript{2}School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100190, China
\textsuperscript{3}CAS Center for Excellence in TQC, University of Chinese Academy of Sciences, Beijing 100049, China
\textsuperscript{4}Collaborative Innovation Center of Quantum Matter, Beijing 100190, China

(Dated: October 2, 2018)

We address the framework of analysing quantum metrology in the information-theoretic picture. Firstly we show how to extract the maximum amount of information in general via suitable state initialization of the probes at the beginning and a quantum measurement at the end. Our analysis can apply to both the single-parameter and the multi-parameter estimation procedures as well as to many other quantum information processing procedures. We then establish a direct connection between the information-theoretic picture of quantum metrology and its conventional variance-covariance picture, by showing that any estimation procedure achieves Heisenberg limit in variance-covariance picture can also reach the information-theoretic Heisenberg limit in the asymptotic sense. As a direct consequence, we argue that the entangled measurement is not necessary for achieving Heisenberg limit in the information-theoretic pictures, which is explicitly illustrated for the Quantum-Classical Parallel strategy of quantum metrology with a separable measurement employed and the Heisenberg limit saturated in both pictures.

Introduction.— The introduction of the concept of information to physics can trace far back to the era of Maxwell. Its intimate relation with thermodynamics has been greatly appreciated since [1–13]. It was until the seminal work of Shannon that the framework of information theory had been systematically developed and established for classical information processing devices [14, 15], after which the potential of information processing in quantum systems has been realized and thus begins the study of quantum information [9, 16–19]. The information-theoretic methodology has hence been widely employed in the study of many quantum technologies such as quantum optimal control [20], quantum error correction [21], quantum state discrimination [22–26] and even in the fundamental research of quantum mechanics [9, 27–32]. Though the temptation to build a information-theoretic picture of quantum metrology has always been common [33–37], it was addressed directly only recently [38–40]. The popularity of information-theoretic methodology in physics is partly due to the intriguing insight it provides to the study of thermodynamics [1–13]. It is natural to hope to gain similar advantage in the information-theoretic framework in other fields. Besides, an unified information-theoretic framework can provides tools to interdisciplinary researches. For instance, it could be beneficial in the study of the thermodynamic effects in aforementioned quantum technologies which is a very important issue in the development quantum technologies [41].

Information-theoretic framework of quantum metrology.— As a study of utilizing quantum effects to achieve better estimation of parameters of a physical system, quantum metrology has been among the most vigorous branches of quantum technology [42–44]. It has great applications in phase estimation [45–48], clock synchronization [49–53], gravitation wave detection [54–56], imaging of biological samples [57, 58] and many other areas [43, 59]. Generally, a metrology task is to estimate a parameter or parameters $\varphi$ using some chosen physical systems as probes. The information of $\varphi$ would be branded on the probes after sending them through the quantum channel characterized by $\varphi$. Assume that there are $n$ parameters to be estimated each of which is an entry of the vector $\varphi$. In the noiseless case, it is assumed that the information of $\varphi$ can be imprinted on a probe via the unitary process $U_\varphi = \exp(-i \sum_{i=1}^n \varphi_i \mathcal{H}_i)$ where $\varphi_i$ denotes the $i$th entry of $\varphi$ and every Hermitian operators $\mathcal{H}_i$ is within our prior knowledge. One can either employ $N$ such probes and implement one parameter-imprinting procedure $U_\varphi$ on each of the probes (parallel strategies) or employ a main probe on which the experimenter deploy $U_\varphi$ repeatedly for $N$ times (sequential strategies). The parameter-imprinting task would be assisted with some carefully designed quantum gates and even some auxiliary probes in the cases of sequential strategies. At the end of the evolution, we deploy a general POVM measurement on the probes to generate the experiment data. From the information theory point of view, the task of quantum metrology is to extract the maximum amount of information about the parameters to be estimated which is evaluated by the mutual information between the measurement outcome $m$ and $\varphi$ [22–26, 39]

\begin{equation}
H(m: \varphi) = H(m) + H(\varphi) - H(m, \varphi),
\end{equation}

where $H(m)$ and $H(\varphi)$ are Shannon information corresponding to the measurement outcome and the parameters while $H(m, \varphi)$ is the joint Shannon information of them. The search for strategies for the maximum information extraction has been addressed by many. While useful bounds has been found and widely
used [22–24, 26, 39], it has been solved for few specific cases [9, 25, 33, 34]. The development of a general formulation for nailing down the optimal information extraction strategy is not only essential for the establishment of the information-theoretic framework, but also very import by itself.

**Extraction of the maximum amount of information in quantum metrology.**— To achieve the maximum information extraction of the parameters \( \varphi \), our task is to choose the optimal combination of the state initialization \( \hat{\rho}_0 \) of the probes before applying the quantum channel characterized by \( \varphi \), and the quantum measurement at the end. The measurement can be the most general POVM measurement described by \( M \) positive Hermitian operators \( \hat{\pi}_m \) satisfying the completeness relation \( \sum_{m=1}^M \hat{\pi}_m = \mathbb{1} \). In the case of parallel strategy where \( N \) probes are employed simultaneously, the probes are initialized to \( \hat{\rho}_0 \) and pass through the parameter channel \( \hat{U}_\varphi^{\otimes N} \) and arrive at the output state \( \hat{\rho}_\varphi = (\hat{U}_\varphi^{\otimes N} \hat{\rho}_0 (\hat{U}_\varphi^{\otimes N})^\dagger \). Given \( \varphi \), the probability of obtaining the \( m \)th outcome is [19]

\[
p(m|\varphi) = \text{tr}(\hat{\pi}_m \hat{\rho}_\varphi).
\]

Let the prior probability of \( \varphi \) be \( q_\varphi \). The probability of obtaining \( m \) on average is

\[
p_m = \int d^n \varphi q_\varphi p(m|\varphi).
\]  

It turns out that we can always find a strategy where we prepare probes in pure state \( \hat{\rho}_0 = |\psi_0\rangle \langle \psi_0| \) and perform the rank-one indecomposable POVM measurement \( \hat{\pi}_m = \lambda_m |u_m\rangle \langle u_m| \) with \( \hat{\pi}_m \)'s being linearly independent, to ensure the maximum information extraction [25, 60, 61]. We remark that the POVM measurement is indecomposable if it cannot be implemented by deploying other POVM measurements according to a probability distribution [60, 61]. And the necessary conditions for the optimal information extraction are

\[
\int d^n \varphi q_\varphi \langle u_m | \hat{\rho}_\varphi | u_{m'} \rangle \ln \left[ \frac{p_m p(m'|\varphi)}{p_m p(m|\varphi)} \right] = 0 \tag{4}
\]

for \( m \neq m' \) and \( m, m' = 1, \ldots, M \), and

\[
\langle \psi_0 | \int d^n \varphi q_\varphi \sum_{m=1}^M \ln \frac{p_m}{p(m|\varphi)} (\hat{U}_\varphi^{\otimes N})^\dagger \hat{\pi}_m \hat{U}_\varphi^{\otimes N} | \psi_0 \rangle = 0 \tag{5}
\]

for any \( |\psi_0\rangle \) orthogonal to \( |\psi_0\rangle \).

Before moving to the derivation of (4) and (5), we have a few remarks on the necessary conditions (4) and (5). First of all, they are conditions that cover all the optimal information extraction strategies which employ pure initial state and rank-one indecomposable POVM measurement. Secondly, (4) and (5) can not rule out the existence of other types of optimal information extraction strategies which may utilize mixed initial state or measurement which is not rank-one indecomposable. If there is any such strategy, it can be realized by deploying the optimal strategies given by (4) and (5). This is due to the fact the mutual information \( H(m : \varphi) \) is convex in terms of the initial state and the POVM measurement employed. Thus, with (4) and (5) we can construct all the optimal information extraction strategies. Thirdly, every optimal rank-one indecomposable POVM measurement is completely orthogonal, namely \( \tilde{\pi}_m \tilde{\pi}_m' = \delta_{mm'} \) [26], when the constraints given by (4) and (5) are independent. In such circumstances, the degrees of freedom of the strategic combination of pure state and rank-one indecomposable POVM measurement with \( M \) outcomes should be \(- (M - d^2)^2\). Thus, (4) and (5) can only be satisfied if \( M = d^2 \), which implies the rank-one POVM measurement is completely orthogonal. If there are constraints which are not independent, the overcomplete rank-one indecomposable POVM should be employed for optimal information extraction [22, 25, 62].

Next, we present the derivation of the necessary conditions (4) and (5) for optimal information extraction. The joint convexity of the relative entropy implies that \( H(m : \varphi) \) is a convex function of the initial state \( \hat{\rho}_0 \) and the POVM measurement \( \hat{\pi}_m \) [15, 19]. Hence there exist the pure initial state \( \hat{\rho}_0 = |\psi_0\rangle \langle \psi_0| \) and the corresponding indecomposable measurement which together accomplish the optimal information extraction [25, 35]. The rank-one POVM measurement obtained via the spectrum decomposition of an indecomposable POVM measurement is also indecomposable [60, 61]. It is also known that classifying two different outcomes as as one would not improve the information extraction (cf. Lemma 2 of Ref. [25]). Thus if the rank-one POVM measurement is obtained via spectrum decomposition of a POVM measurement, the former would outperform the latter. Hence there is always a rank-one indecomposable POVM measures ensures maximal \( H(m : \varphi) \). Further, the rank-one POVM measurement \( \{\hat{\pi}_m = \lambda_m |u_m\rangle \langle u_m|\} \) can be implemented as a completely orthogonal measurement \( \{\hat{H}_m = |v_m\rangle \langle v_m|\} \) on the composite system of the probes and an auxiliary such that \( |v_m\rangle = V^\dagger |u_m\rangle \otimes |a_m\rangle \) and \( |a_m\rangle \) belongs to an orthonormal basis of the auxiliary. \( V \) is an unitary gate acting upon the composite system

\[
V |\psi\rangle \otimes |a_1\rangle = \sum_{m=1}^M \sqrt{\lambda_m} |u_m\rangle \langle u_m| \otimes |a_m\rangle \tag{6}
\]

The optimal \( H(m : \varphi) \) considered in the subsystem of the probes is thus a conditioned optimal in the composite system of the probes and auxiliary. There are two restrictions: a) the auxiliary is initialized in the standard pure state \( |a_1\rangle \) and does not pass through the parameter channel; b) only completely orthogonal measurements on the composite system of the probes and auxiliary are considered. The set of strategy combinations of pure initial state \( |\psi_0\rangle \otimes |a_1\rangle \) and completely ortho-
nal measurement \( \{ \hat{\Pi}_m \} \) corresponds to a connected and compact hypersurface with no boundary in an Euclidean space. Hence the “maximum point” is also a “critical point”. We choose \( |\psi_0\rangle \) as a reference initial state of the probes and \( \{ |m\rangle \langle m| \} \) as a reference completely orthogonal measurement on the composite system. Our initialization and measurement setup can thus be described by two unitary gates \( \hat{U}_1 \) and \( \hat{U}_M \) such that \( |\psi\rangle = \hat{U}_1 |\psi_0\rangle \) and \( \Pi_m = \hat{U}_M |m\rangle \langle m| \hat{U}_M^\dagger \). If the strategy is optimal, the variation of the mutual information

\[
\delta H(m : \varphi) = - \int d^n \varphi |p| \sum_m \ln \left[ \frac{p_m}{p(m|\varphi)} \right] \varphi p(m|\varphi) \quad (7)
\]

will vanish in the first order of the variation of the initialization gate \( \hat{U}_1 \) and the measurement adjustment gate \( \hat{U}_M \). The variation of the conditional probability \( p(m|\varphi) \) due to the variation of the initialization and the completely orthogonal measurement is

\[
\delta p(m|\varphi) = i \text{tr} \left[ \rho \delta \hat{G}_M + \left( \hat{U}_m^{\dagger} \hat{\rho} \hat{U}_m \right) \right] \delta \hat{G}_1 \quad (8)
\]

where \( \delta \hat{G}_1 := i \hat{U}_1^\dagger \delta \hat{U}_1 \) and \( \delta \hat{G}_M := i \hat{U}_M^\dagger \delta \hat{U}_M \) are Hermitian. Notice that the freedom of choosing arbitrary pure state of the probes and completely orthogonal via the unitary gates \( \hat{U}_1 \) and \( \hat{U}_M \) is equivalent to the arbitrariness of \( \delta \hat{G}_1 \) and \( \delta \hat{G}_M \). Since the initialization gate \( \hat{U}_1 \) acts only on the probes while the measurement adjustment gate \( \hat{U}_M \) acts on the composite of the probes and the auxiliary, \( \delta \hat{G}_1 \) should only act on the probes and \( \delta \hat{G}_M \) acts on the composite system. By substituting (8) into (7), we can obtain

\[
\int d^n \varphi |p| \sum_m \ln \left[ \frac{p_m}{p(m|\varphi)} \right] \left[ \rho \otimes |a_0\rangle \langle a_0| \right] \varphi p(m|\varphi) = 0 \quad (9)
\]

and

\[
\int d^n \varphi |p| \sum_m \ln \left[ \frac{p_m}{p(m|\varphi)} \right] \left( \hat{U}_m^{\dagger} \right)^\dagger \left[ \rho \otimes |v_m\rangle \langle v_m| \right] = 0. \quad (10)
\]

Expand (9) to different blocks by the orthogonal projections \( \mathbb{1} \otimes |a_m\rangle \langle a_m| \) with \( m = 1, \ldots, M \). Then it is clear that (9) is equivalent to (4). If we expand (10) in the orthonormal basis which contains \( |\psi_0\rangle \), we can realize immediately that (10) is equivalent to (5).

**Connection with the conventional variance-covariance picture.**—Quantum metrology has conventionally been studied using the covariance matrix for assessing a estimation which reduces to the variance in the single-parameter estimation case [42, 43, 59, 63]

\[
\Delta_\varphi := \left( \langle \varphi_{\text{est}} | / \partial \varphi \langle \varphi_{\text{est}} \rangle \right)_\text{av} - \varphi^2 \quad (11)
\]

where \( \varphi_{\text{est}} \) is an estimation to the parameter \( \varphi \). The optimal strategy in the conventional variance-covariance picture [64–67] is not exactly the same as its counterpart (4) and (5) in the information-theoretic picture. There is a intimate relation between the two pictures of quantum metrology in the single-parameter estimation scenario. In a single-parameter estimation, \( \Delta_\varphi \) is bounded from below by

\[
\Delta_\varphi \geq 1/\sqrt{sWN}^\alpha, \quad (12)
\]

where \( W \) is the width of the spectrum of the probe Hamiltonian \( \hat{H} \) such that \( \hat{U}_M = e^{-i\varphi \hat{H}} \). The estimation \( \varphi_{\text{est}} \) is extracted from \( s \) experiment data points. If an entangled initial state is employed in the parallel strategy or one uses a sequential strategy, the Heisenberg limit \( \alpha = 1 \) can be obtained. Otherwise, only the standard quantum limit \( \alpha = 1/2 \) is available. In the information-theoretic picture, the standard quantum limit of the mutual information \( H(m : \varphi) \) is \( \frac{1}{2} \ln N \) and its Heisenberg is limit \( \ln N \) [39]. It can be shown that the mutual information between the estimation \( \varphi_{\text{est}} \) and the parameter \( \varphi \) is asymptotically upper bounded as

\[
H(\varphi_{\text{est}} : \varphi) \leq \ln(\sqrt{s} WN^\alpha) + H(\varphi) \quad (13)
\]

with \( \epsilon \ll 1 \), if the estimation error \( \Delta_\varphi \) is lower bounded as (12). Further, if (12) is asymptotically saturated so would be (13). As a consequence, any estimation strategy achieves the asymptotic Heisenberg limit (standard quantum limit) in the variance description (12) would also obtain the information-theoretic Heisenberg limit (standard quantum limit). It confirms the general connection between information-theoretic picture and the variance-covariance description of quantum metrology.

Next, we will prove that the information-theoretic bound (13) can be obtained from (12) in the asymptotic sense. Given \( r \) times of repetition of the aforementioned estimation which produces \( s \) data points, one can obtain a final estimation

\[
\bar{\varphi}_{\text{est}} = \frac{1}{r} \sum_{j=1}^r \varphi_{\text{est}}^{(j)} \quad (14)
\]

by averaging the \( r \) estimations \( \varphi_{\text{est}} \). We use the superscript \( (j) \) to indicate from which group of data is the estimation \( \bar{\varphi}_{\text{est}} \) made. For a reasonable estimator one should expect the average of the estimation \( \langle \varphi_{\text{est}} \rangle_\text{av} \) asymptotically approaches the parameter \( \varphi \) and \( \partial \langle \varphi_{\text{est}} \rangle_\text{av}/\partial \varphi \) approaches unity, when \( N \) grows bigger [43, 48, 68]. In the case of large \( N \), we should have \( \langle \varphi_{\text{est}} \rangle_\text{av} = \varphi \) with \( \sigma_\varphi = \Delta_\varphi/\sqrt{r} \) being the standard deviation of \( \varphi_{\text{est}} \). We presume the number of repetition \( r \) being sufficiently large and \( \varphi_{\text{est}}^{(j)} \)s are mutually independent. Then the central limit theorem indicates that the probability of \( \varphi_{\text{est}} \) conditional on \( \varphi \) should be Gaussian [69]

\[
p(\varphi_{\text{est}} | \varphi) \approx \exp \left[ - \left( \varphi_{\text{est}} - \varphi \right)^2 / 2\sigma_\varphi^2 \right] / \sqrt{2\pi\sigma_\varphi^2}. \quad (15)
\]
Given the prior probability distribution \(q_\varphi\) of the parameter \(\varphi\), we can calculate the information of \(\tilde{\varphi}_{\text{est}}\) conditioned on \(\varphi\)

\[
H(\tilde{\varphi}_{\text{est}}|\varphi) = -\int \text{d}\varphi \varphi p(\tilde{\varphi}_{\text{est}}|\varphi) \ln p(\tilde{\varphi}_{\text{est}}|\varphi) \\
= \frac{1}{2} \ln (2\pi e/r) + \int \text{d}\varphi q_\varphi \ln \Delta_\varphi.
\]

(16)

With \(\sigma_\varphi\) being small enough, the conditional probability \(p(\tilde{\varphi}_{\text{est}}|\varphi)\) would be mainly concentrated within a small area of \(|\tilde{\varphi}_{\text{est}} - \varphi| < k\sigma_\varphi\) such that it can be approximated by a delta function \(\delta(\tilde{\varphi}_{\text{est}} - \varphi)\) [70]. We can choose \(k = \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}} \sqrt{\ln(6/\sigma_{\text{max}})}\) with \(\sigma_{\text{min}} := \min_\varphi \sigma_\varphi\) and \(\sigma_{\text{max}} := \max_\varphi \sigma_\varphi\) such that \(k\sigma_\varphi\) is still very small and approaches 0 very fast as \(\sigma_{\text{max}} \to 0\). As a result the Shannon information of \(\tilde{\varphi}_{\text{est}}\) would approach to that of \(\varphi\)

\[
H(\tilde{\varphi}_{\text{est}}) = [1 + O(1/r)] H(\varphi) + o(\sigma_{\text{max}}/\sqrt{r}),
\]

(17)

provided that \(q_\varphi\) is fairly smooth, namely its first derivative \(q^{(1)}(\varphi)\) and the second derivative \(q^{(2)}(\varphi)\) with respect to \(\varphi\) exist and are finite. Here we presume the complexity of estimating \(\varphi\) of different values is of the same order, namely \(\sigma_\varphi\) is of the same order for different values of \(\varphi\). For the convenience of mathematical analysis we also assume the first derivative \(\sigma^{(1)}\) and the second derivative \(\sigma^{(2)}\) of \(\sigma_\varphi\) with respect to \(\varphi\) exist and are finite. \(O(1/r)\) means terms of the same or higher order than \(1/r\) and \(O(\sigma_{\text{max}}/\sqrt{r})\) denotes the terms much smaller than \(\sigma_{\text{max}}/\sqrt{r}\). One can obtain the mutual information \(H(\tilde{\varphi}_{\text{est}} : \varphi) = H(\tilde{\varphi}_{\text{est}}) - H(\tilde{\varphi}_{\text{est}}|\varphi)\) between \(\tilde{\varphi}_{\text{est}}\) and \(\varphi\) by subtracting (16) from (17). Then we obtain the upper bound of the mutual information

\[
H(\tilde{\varphi}_{\text{est}} : \varphi) \leq \ln \left( \sqrt{\pi} W N^\alpha \right) + H(\varphi).
\]

(18)

from (12). We can show that the contribution from \(r\) in (18) is superficial. First of all, the mere effect of the \(r\) repetition of estimation is increasing the amount of data from \(s\) to \(sr\). Since \(\tilde{\varphi}_{\text{est}}\) is the average of \(r\) estimations each of which is constructed from \(s\) data point, \(\tilde{\varphi}_{\text{est}}\) cannot be a better estimation than a general estimation \(\varphi_{\text{est}}\) evaluated directly from the \(sr\) experiment data. On the other hand, the scaling behavior of such a general estimation with respect to \(sr\) in the variance picture (12) can never be better than \(1/\sqrt{sr}\). It means that the two data processing algorithms are equally good asymptotically when both \(r\) and \(s\) are sufficiently large. Therefore, \(r\) can be incorporated to \(s\) while \(\tilde{\varphi}_{\text{est}}\) can be replace by an estimation \(\varphi_{\text{est}}\) constructed from \(sr\) data points. Secondly, by using the jointly convex property of the mutual information which is a result of the joint convexity of the relative entropy [15, 19], we can show that \(H(\tilde{\varphi}_{\text{est}} : \varphi) \leq H(\tilde{\varphi}_{\text{est}} : \varphi)\). One can then chose \(r = \ln s\) and (18) is equivalent to (13) with \(\epsilon = \ln (\ln s)/\ln s \ll 1\) given that \(s\) is big.

**Information-theoretic Heisenberg limit of the Quantum-Classical parallel strategy.**— We know that the entangled measurement is not necessary to achieve the conventional Heisenberg limit (12) [59]. A direct consequence of (13) is that entangled measurement is not necessary to accomplish the information-theoretic Heisenberg limit either. We can verify this explicitly in a general Quantum-Classical parallel strategy of quantum metrology where the probes are initialized in a GHZ state while a separate measurement is employed at the end [59]. The scheme employed here (cf. Fig. 1) is similar to the Parallel QPEP presented in Ref. [39] except for several differences. The first difference is that the Hamiltonian \(\hat{H}\) is of arbitrary form as long as its is known and the dimension \(d\) of the Hilbert space of a single probe is finite. Secondly, we employ neither CNOT gate nor any other kind of multi-partite gate at the measurement stage. Lastly, we replace the inverse Quantum Fourier Transformation with a modified Semiclassical Fourier Transformation [71]. The \(N\) probes are divided into \(L = \ln(N + 1)\) groups and label them with \(0, 1, \ldots, L - 1\). The \(\ell\)th group has \(2^{L-\ell-1}\) probes as shown in Fig. 1(b). The multi-probe gates \(\hat{V}\) would initialize the \(2^{L-\ell-1}\) probes to the GHZ state which would be transformed by the parameter-imprinting channel \(\hat{U}_\varphi\) with \(\hat{U}_\varphi = e^{-i\varphi \hat{H}}\)

\[
|0\rangle^\otimes 2^{L-\ell-1} + e^{-i2^{L-\ell-1}W\varphi} |1\rangle^\otimes 2^{L-\ell-1} \big/ \sqrt{2}.
\]

(19)

Here \(|0\rangle\) is the ground state of a probe while \(|1\rangle\) is its highest excited state. Notice that \(W\) is the width of the spectrum of \(\hat{H}\) and we ignored the unimportant global phase in (19). The effect of the Semiclassical Fourier Transformation shown in Fig. 1(a) plus separable local measurement of the probe Hamiltonian is equivalent to the following procedure: Firstly we deploy a separate measurement of \(X^\otimes 2^{L-1}\) on the 0th group with \(\hat{X} := |0\rangle \langle 1| + |1\rangle \langle 0|\). If the measurement tells us that the number of probes having the physical quantity \(X\) of value \(-1\) in the 0th group is odd, we record the measurement result by \(m_0 = 1\), otherwise \(m_0 = 0\). We mark that the eigenvalue \(-1\) of \(\hat{X}\) corresponds to the signal 1 in the quantum circuit shown of Fig. 1 which signifies the highest level of the probe Hamiltonian \(\hat{H}\). We move forward by measuring \(\hat{A}^{2^{L-1-\ell}}\) on the \(\ell\)th group of \(2^{L-1-\ell}\) probes according to the measurement results \(m_0, \ldots, m_{\ell-1}\) of former groups by carefully choosing

\[
\hat{A}_\ell := \hat{X} \cos \left( 2\pi \sum_{j=0}^{\ell-1} \frac{m_2^j}{2^j} \right) - \hat{Y} \sin \left( 2\pi \sum_{j=0}^{\ell-1} \frac{m_2^j}{2^j} \right),
\]

(20)

where \(\hat{Y} := i(-|0\rangle \langle 1| + |1\rangle \langle 0|)\). As one may have realized that \(\hat{X} (\hat{Y})\) is the spin along the \(x (y)\) direction in the spin-1/2 system. Similarly, we record \(m_\ell = 0, 1\) according to whether we obtain even or odd number of
probes having the physical quantity $\hat{A}_\ell$ of value $-1$ in the $\ell$th group. The conditional probability of obtaining the measurement result $m = \sum_{\ell=0}^{L-1} m_\ell 2^\ell$ conditioned on the parameter $\varphi$ is

$$p(m|\varphi) = \frac{\sin^2 [(N + 1)W\varphi/2]}{(N + 1)^2 \sin^2 [(W\varphi - 2\pi m)/(N + 1)]/2}. \quad (21)$$

With the energy spectrum width $W$ of $\hat{H}$ set to 1, we obtain the mutual information between $m$ and $\varphi$

$$H(m : \varphi) \simeq \ln N - 1.22, \quad (22)$$

given an uniform prior probability distribution in the interval $0 \leq \varphi \leq 2\pi$. It is the exactly the same as the result presented for QPEP in Ref. [39]. Note that though we do not employ any entangled measurement, the measurement employed here is made possible with the assistant of classical communication [19].

The Width of energy spectrum and the Hilbert space dimension.—Just as in the variance-covariance picture (12), it is the energy spectrum width not the dimension of the Hilbert space of the probe that enters the standard quantum limit $\alpha = 1/2$ and the Heisenberg limit $\alpha = 1$ in the information-theoretic picture (13). This can also be seen explicitly in (22) of a general quantum-classical parallel strategy of estimation. The dimension $d$ of the probe Hilbert space can be involved in some special cases. One typical example is the case where the probe energy levels are equally separated $\hat{H} = -\sum_{k=1}^{d-1} 2k\pi$ as presented in Ref. [39]. The probe energy spectrum width and Hilbert space dimension would then be equivalent $W = 2\pi(d-1)$. In such case, the effect of the probe Hilbert space dimension $d$ would manifest itself as $\ln(d-1)$ for both the standard quantum limit and the Heisenberg limit.

Conclusion.—We have investigated the optimal information extraction strategy in quantum metrology. The general method to obtain such optimal strategy is outlined via (4) and (5). Our result can be generalized to information extraction task such as the discrimination of quantum states and quantum channels. We have further established a direct connection between the conventional variance-covariance picture and the information-theoretic picture of quantum metrology in the single-parameter estimation scenario. This connection indicates that the usefulness of entanglement at different stages of a metrology task is the same in both pictures. We explicitly illustrate this by showing a slightly modified Quantum-Classical Parallel strategy which can achieve the Heisenberg limit in both pictures. Our results also indicate that it is the energy spectrum width rather than the dimension of the probe which plays a nontrivial role in quantum metrology in the information-theoretic pictures.

We would like to thank M. J. W. Hall for invaluable comment on this work. This work was supported by Ministry of Science and Technology of China (Grants No. 2016YFA0302104 and 2016YFA0300600), National Natural Science Foundation of China (Grant Nos. 91536108 and 11774406) and Chinese Academy of Sciences (XDPB08-3).

---

[1] J. C. Maxwell, *Theory of Heat* (Longmans, London, 1902) pp. 338–339.
[2] L. Sizilard, *Zeitschrift für Physik*, 53, 840 (1929).
[3] L. Brillouin, *Science and information theory* (Academic Press, New York, 1962).
[4] R. Landauer, *IBM J. Res. Dev.*, 5, 183 (1961).
[5] O. Penrose, *Foundations of Statistical Mechanics* (Pergamon Press, London, 1970) pp. 225–238.
[6] K. Maruyama, F. Nori, and V. Vedral, *Rev. Mod. Phys.*, 81, 1 (2009).
[7] E. T. Jaynes, *Phys. Rev.*, 106, 620 (1957).
[8] E. T. Jaynes, *Phys. Rev.*, 108, 171 (1957).
[9] L. B. Levitin, in *Information Complexity and Control in Quantum Physics*, Vol. 294, edited by A. Blaquiere, S. Diner, and G. Lochak, International Centre for Me-
