International Journal of Quantum Information
© World Scientific Publishing Company

Alexander S. Holevo's Researches in Quantum Information Theory in 20th Century

Masahito Hayashi
School of Data Science, The Chinese University of Hong Kong, Shenzhen, Longgang District, Shenzhen, 518172, China
International Quantum Academy (SIQA), Futian District, Shenzhen 518048, China
Graduate School of Mathematics, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan
hmasahito@cuhk.edu.cn, hayashi@iqasz.cn

Received Day Month Year
Revised Day Month Year

This paper reviews Holevo's contributions to quantum information theory during the 20 century. At that time, he mainly studied three topics, classical-quantum channel coding, quantum estimation with Craméro-Rao approach, and quantum estimation with the group covariant approach. This paper addresses these three topics.

Keywords: classical-quantum channel; quantum state estimation; group covariance.

1. Introduction

Alexander S. Holevo is undoubtedly one of the greatest researchers in quantum information theory. In particular, before quantum information theory became a popular scientific area, he led this research area as the leading person. Nevertheless, this field could not attrach much attention at that time. He is one of the greatest witnesses of the long winter experienced by the field of quantum information theory. In 1994, the author started this area with reading his textbook:[1]

A.S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory.*
North-Holland, Amsterdam (1982). Originally published in Russian in 1980.

At that time, we were still in the long winter era/period of quantum information theory. In the later part of the 1990s, the author researched many of the findings that the author will eventually write about in the 20th century. Therefore, the author is extremely grateful for Alexander S. Holevo's contributions to this area.

Looking back at his publication records, we see that his researches can be mainly divided into three streams. His first stream is the first period on quantum information, which starts in the beginning of 1970s and ends in the middle of 1980s. The above textbook summarizes the results in this period. His second stream is his researches in quantum dynamical system with quantum stochastic process and semi-group, which starts in the middle of 1980s and ends almost in the end of

20th century. The second stream is related to the mathematical description of a quantum measurement process. His third stream is the second period on quantum information, which starts in the following international conference.

Third International Conference on Quantum Communication & Measurement,
Mt. Fuji-Hakone Land, Japan, September 25-30, (1996).

At the above conference, he knew the result for a pure-state channel coding by Hausladen.[2] Inspired by this result, he obtained its generalization to the case with mixed states when he stayed in Tamagawa University after this conference.[3] This result was published as one of his most famous results:[4]

A.S. Holevo, "The capacity of the quantum channel with general signal states,"
IEEE Transactions on Information Theory **44** (1), 269-273 (1998).

His third stream still continues until today. That is, he still actively studies quantum information theory.

Unfortunately, the author is not familiar with his second stream. Since his contributions in his third stream are well known, this paper focuses on his first stream. We review his major results in his first stream, which is composed of three topics. The first topic is classical-quantum channel coding. The second topic is quantum state estimation with the Crámer-Rao approach. The third topic is quantum state estimation with the group covariant approach.

The remaining of this paper is organized as follows: Section 2 reviews his results for classical-quantum channel coding. Section 3 discusses his results for quantum state estimation with the Crámer Rao approach. Section 4 explains his results for quantum state estimation with the group covariant approach. In addition, these sections discuss how his results influenced the subsequent studies in the area of quantum information theory. Section 5 is the conclusion.

2. Classical-quantum channel coding

Classical-quantum (cq-) channel coding is the subarea in quantum information theory.

2.1. *Information quantities*

Information quantities play a central role in channel coding. Here, we introduce classical and quantum information quantities. When a random variable X is generated according to a distribution $P_X = \{P_X(x)\}_{x\in\mathcal{X}}$ on $\mathcal{X}$, its Shannon entropy $H(P_X)$ is defined as

$$H(P_X) := -\sum_{x\in\mathcal{X}} P_X(x)\log P_X(x). \tag{1}$$

When two random variables X, Y are subject to a joint distribution $P_{X,Y}$, the mutual information is given as

$$I(P_{X,Y}) := H(P_X) - H(P_Y) - H(P_{X,Y}), \tag{2}$$

where P_X and P_Y are the marginal distribution of $P_{X,Y}$ with respect to X and Y, respectively. The mutual information has another form, by using relative entropy. When two probability distributions P_X and Q_X are given, the relative entropy is defined as

$$D(P_X \| Q_X) := \sum_{x \in \mathcal{X}} P_X(x)(\log P_X(x) - \log Q_X(x)). \tag{3}$$

Then, the mutual information is rewritten as

$$I(P_{X,Y}) := \sum_{x \in \mathcal{X}} P_X(x) D\big(P_{Y|X=x} \big\| P_Y\big), \tag{4}$$

where $P_{Y|X=x}$ is the conditional distribution $P_{XY}(x,y)/P_X(x)$.

Next, we consider a quantum system $\mathcal{H}_1$. Given a density matrix ρ on $\mathcal{H}_1$, its von Neumann entropy is given as

$$S(\rho) := -\operatorname{Tr} \rho \log \rho. \tag{5}$$

When a density matrix $\rho_{1,2}$ on a joint system $\mathcal{H}_1 \otimes \mathcal{H}_2$, the mutual information is given as

$$I(\rho_{1,2}) := S(\rho_1) + S(\rho_2) + S(\rho_{1,2}), \tag{6}$$

where $\rho_1 := \operatorname{Tr}_2 \rho_{1,2}$ and $\rho_2 := \operatorname{Tr}_1 \rho_{1,2}$. The mutual information also has another form, by using relative entropy. When two densities matrices ρ and σ are given, the relative entropy is defined as

$$D(\rho \| \sigma) := \operatorname{Tr} \rho(\log \rho - \log \sigma). \tag{7}$$

Then, the mutual information is rewritten as

$$I(\rho_{1,2}) = D(\rho_{1,2} \| \rho_1 \otimes \rho_2). \tag{8}$$

Next, we assume that the state $\rho_{1,2}$ is a classical-quantum state, i.e., it has the form $\rho_{1,2} = \sum_{x \in \mathcal{X}} P_X(x) |x\rangle\langle x| \otimes \rho_{2|x}$, where P_X is a distribution on $\mathcal{X}$. In addition, $\mathcal{H}_1$ is assumed to be spanned by an orthogonal basis $\{|x\rangle\}_{x \in \mathcal{X}}$, and $\rho_{2|x}$ is a density matrix on $\mathcal{H}_2$. In this case, the formula (6) of the mutual information is rewritten as

$$I(\rho_{1,2}) = S\Big(\sum_{x \in \mathcal{X}} P_X(x) \rho_{2|x}\Big) - \sum_{x \in \mathcal{X}} P_X(x) S(\rho_{2|x}). \tag{9}$$

The formula (8) of the mutual information is rewritten as

$$I(\rho_{1,2}) = \sum_{x \in \mathcal{X}} P_X(x) D\Big(\rho_{2|x} \Big\| \sum_{x \in \mathcal{X}} P_X(x) \rho_{2|x}\Big). \tag{10}$$

2.2. *Formulation and capacity theorem*

The problem of cq-channel coding is formulated as follows:[5–9] Consider the input alphabet $\mathcal{X}$ and the output quantum system $\mathcal{H}$, which is represented as a Hilbert space. A cq-channel is given as $\boldsymbol{W} := \{W_x\}_{x\in\mathcal{X}}$, where W_x expresses the density matrix on $\mathcal{H}$ to express the output state with the input $x \in \mathcal{X}$.

To discuss a cq-channel, we discuss the mutual information when a random variable X is generated according to distribution $P = \{P(x)\}_{x\in\mathcal{X}}$ on $\mathcal{X}$. Then, the mutual information with the input distribution P is[10, 11]

$$I(P,\boldsymbol{W}) = I\Big(\sum_{x\in\mathcal{X}} P(x)|x\rangle\langle x| \otimes W_x\Big). \tag{11}$$

The aim of cq-channel coding is the transmission of classical message via n uses of the cq-channel $\boldsymbol{W}$. When we use the cq-channel $\boldsymbol{W}$ n times, the channel is given as $\boldsymbol{W}^{(n)}\{W^{(n)}_{x^n}\}_{x^n\in\mathcal{X}^n}$.[12, 13] i.e., the output state with the input $x^n = (x_1,\ldots,x_n) \in \mathcal{X}^n$ is given as

$$W^{(n)}_{x^n} := W_{x_1} \otimes \cdots W_{x_n}, \tag{12}$$

which is a density matrix on the n-fold tensor product system $\mathcal{H}^{\otimes n}$. We denote the set of messages by $\mathcal{M} := \{1,\ldots,\mathsf{M}\}$. An encoder is given as a map ϕ_n from $\mathcal{M}$ to $\mathcal{X}^n$. When $\phi(m)$ is given as $(x_1,\ldots,x_n)$, the output state is

$$W^{(n)}_{\phi(m)} = W_{x_1} \otimes \cdots W_{x_n}. \tag{13}$$

A decoder is given as a positive operator-valued measure (POVM) $\boldsymbol{\Pi} := \{\Pi_m\}_{m\in\mathcal{M}}$ over the quantum system $\mathcal{H}^{\otimes n}$, whose set of outcomes is given as $\mathcal{M}$. That is, $\boldsymbol{\Pi}$ is a resolution of the identity on $\mathcal{H}^{\otimes n}$, i.e., Π_m is a positive semi-definite matrix on $\mathcal{H}^{\otimes n}$, and the relation $\sum_{m\in\mathcal{M}} \Pi_m = I$. The pair of an encoder ϕ and a decoder $\boldsymbol{\Pi}$ is called a code and is written as Φ. The performance of a code Φ is evaluated by two parameters. One is the size $|\Phi|$ of the code Φ, which is given by M. The other is the decoding error probability $e(\Phi)$, which is given as

$$e(\Phi) := 1 - \frac{1}{\mathsf{M}} \sum_{m\in\mathcal{M}} \mathrm{Tr}\, W^{(n)}_{\phi(m)} \Pi_m. \tag{14}$$

The channel capacity is defined as

$$C(\boldsymbol{W}) := \sup_{\{\Phi_n\}_n} \Big\{ \lim_{n\to\infty} \frac{1}{n} \log|\Phi_n| \Big| e(\Phi_n) \to 0 \Big\}. \tag{15}$$

The cq-channel coding theorem, i.e., the capacity theorem is formulated as

$$C(\boldsymbol{W}) = \max_{P\in\mathcal{P}(\mathcal{X})} I(P,\boldsymbol{W}), \tag{16}$$

where $\mathcal{P}(\mathcal{X})$ expresses the set of probability distributions on $\mathcal{X}$. To show this result, we need to prove the following two inequalities.

$$C(\boldsymbol{W}) \le \max_{P\in\mathcal{P}(\mathcal{X})} I(P,\boldsymbol{W}) \tag{17}$$

$$C(\boldsymbol{W}) \ge \max_{P\in\mathcal{P}(\mathcal{X})} I(P,\boldsymbol{W}). \tag{18}$$

To show (16), he tackled the inequality (17), which is called the converse part in information theory. He proved the inequality (17) by dividing it into two steps.

In his paper[10] in 1972, as the first step, Holevo showed

$$I(\mathrm{P}[P, \boldsymbol{W}, \boldsymbol{\Pi}]) \le I(P, \boldsymbol{W}) \tag{19}$$

for any POVM $\boldsymbol{\Pi}$ on $\mathcal{Y}$, where $\mathrm{P}[P, \boldsymbol{W}, \boldsymbol{\Pi}]$ is the joint distribution on $\mathcal{X} \times \mathcal{Y}$ defined as

$$\mathrm{P}[P, \boldsymbol{W}, \boldsymbol{\Pi}](x, y) := P(x) \operatorname{Tr} W_x \Pi_y. \tag{20}$$

Nowadays, the inequality (19) can be shown by using the monotonicity of the quantum relative entropy[14, 15] for trace-preserving completely positive maps. However, at that time, the above monotonicity was not known. We can say that Holevo showed the monotonicity of the quantum relative entropy in the case of the mutual information with cq-channel as (19).

In his paper[12] in 1979, as the second step, he showed that

$$C(\boldsymbol{W}) \le \lim_{n \to \infty} \frac{1}{n} \sup_{P_n \in \mathcal{P}(\mathcal{X}^n)} \sup_{\boldsymbol{\Pi}_n} I(\mathrm{P}[P,^n, \boldsymbol{W}^{(n)}, \boldsymbol{\Pi}_n]). \tag{21}$$

In addition, he pointed out the relation

$$n \max_{P \in \mathcal{P}(\mathcal{X})} I(P, \boldsymbol{W}) = \max_{P_n \in \mathcal{P}(\mathcal{X}^n)} I(P_n, \boldsymbol{W}^{(n)}). \tag{22}$$

In fact, the above relation can be shown by using the chain rule. Combining three relations (19), (21), and (22), he derived (17).

For the opposite direction (18), Hausladen et. al.[2] showed the inequality (18) when all W_x are pure states. Their key idea is typical subspaces.[16, 17] Inspired by this result,[2] using the idea of typical subspaces, Holevo showed the inequality (18) with general mixed states W_x. Later, independently, Schumacher and Westmoreland[18] showed the same result. In addition, with Burnashev, Holevo derived an exponential decay date of the decoding error probability when the transmission rate is smaller than the capacity $C(\boldsymbol{W})$ and all W_x are pure states.[19]

Since the result (16) forms the foundation of quantum communication, it motivated many studies in quantum information theory. For example, Ogawa and Nagaoka[20] showed the strong converse theorem for cq-channel coding. To prove (16) via information spectrum method,[21] the reference[22] invented a useful matrix inequality, which is often called Hayashi-Nagaoka inequality. Bennet et al[23] considered cq-channel coding with entanglement-assistance, which leads to reverse Shannon theorem,[24] which is known as a new topic in Shannon theory. Holevo[25] provided a more elegant proof. Also using the result (16), Devetak[26] showed the coding theorem for sending a quantum state via a noisy quantum channel. In this way, the result (16) yields various results in quantum information theory.

3. Quantum estimation with Crámer Rao type bounds

Holevo made great contributions to quantum state estimation as well. In this problem, we consider a parameterized family of density operators $\mathcal{S} = \{\rho_\theta\}_{\theta\in\Theta}$ on a Hilbert space $\mathcal{H}$. When Θ is a continuous subset of $\mathbb{R}^d$, an estimator is formulated as a POVM $\mathbf{\Pi}$ on $\mathcal{H}$ whose data set is $\mathbb{R}^d$. Since $\mathbb{R}^d$ is a continuous set, its rigorous formulation requires measure theory. A POVM Π is defined as a map from the set $\mathcal{B}(\mathbb{R}^d)$ of Borel sets to the set of positive semi-definite operators on $\mathcal{H}$. The map Π should satisfy the following conditions.

$$\Pi(\emptyset) = 0, \quad \Pi(\mathbb{R}^d) = I, \quad \Pi(\cup_j B_j) = \sum_j \mathbf{\Pi}(B_j), \tag{23}$$

where $B_j \in \mathcal{B}(\mathbb{R}^d)$ are disjoint sets. We often consider the *unbiased* condition, which is formulated as

$$\int_{\mathbb{R}^d} \hat{\theta} \operatorname{Tr} \rho_\theta \Pi(d\hat{\theta}) = \theta \tag{24}$$

for $\theta \in \Theta$. Since the unbiased condition is too restrictive, taking the Taylor expansion in (24) at $\theta_0 \in \Theta$, we often consider its relaxed version, the locally unbiased condition at $\theta_0 \in \Theta$, which is formulated as

$$\int_{\mathbb{R}^d} \hat{\theta} \operatorname{Tr} \rho_{\theta_0} \Pi(d\hat{\theta}) = \theta_0 \tag{25}$$

$$\frac{\partial}{\partial \theta^k} \int_{\mathbb{R}^d} \hat{\theta}^l \operatorname{Tr} \rho_{\theta_0} \Pi(d\hat{\theta}) = \delta_{k,l}. \tag{26}$$

To measure the precision of an estimator Π, we focus on the mean square matrix $V(\Pi)$ as

$$V_\theta^{k,l}(\Pi) := \int_{\mathbb{R}^d} (\hat{\theta}^k - \theta^k)(\hat{\theta}^l - \theta^l) \operatorname{Tr} \rho_\theta \Pi(d\hat{\theta}). \tag{27}$$

When Π is an unbiased estimator, the mean square matrix $V_\theta(\Pi)$ coincides with the covariance matrix. For $d = 1$, we simply call $V_\theta(\Pi)$ the mean square.

Helstrom defined the SLD $L_{\theta,k}$ as a Hermitian (self-adjoint) operator satisfying[8, 27]

$$L_{\theta,k} \circ \rho_\theta = \frac{\partial \rho_\theta}{\partial \theta^k}. \tag{28}$$

where $X \circ Y := \frac{1}{2}(XY + YX)$ Then, Helstrom defined the SLD Fisher information matrix $J_{\theta,k,l}$ as

$$J_{\theta,k,l} := \operatorname{Tr} L_{\theta,k}(L_{\theta,l} \circ \rho_\theta). \tag{29}$$

For a locally unbiased estimator Π at θ, using Schwarz inequality, we obtain the matrix inequality based on positive semi-definite matrices.

$$V_\theta(\Pi) \geq J_\theta^{-1}. \tag{30}$$

In the following, we use matrix inequalities in the above sense. For $d = 1$, the following locally unbiased estimator achieves the equality in (30). When Π is the spectral decomposition of the operator $\frac{1}{J_\theta}(L_\theta + \theta)$, the equality in (30) holds.

However, the equality in (30) does not hold in the multiple-parameter case. This difficulty is caused by the non-commutativity of SLDs $L_{\theta,k}$. The first attempt to tackle this problem was done by Yuen and Lax.[28] They considered complex parameters to identify the state, i.e., they assumed $\Theta \subset \mathbb{C}^d$. They focused on the unbiased condition (24) by replacing $\mathbb{R}^d$ by $\mathbb{C}^d$, and considered right logarithmic derivatives based on complex parameters. Their method works well for the quantum Gaussian family, whose mathematical formulation is given in the references[29–31] [1, Chapter 5], which has crucial non-commutativity related to the canonical observables Q and P. But, it cannot be applied to a general real-multiple-parameter model. At that time, several Russian researchers[32, 34] focused this topic.

To resolve this problem, for a general real-multiple-parameter model, Holevo[1, 35] defined the right logarithmic derivative $L^R_{\theta,k}$ as

$$\rho_\theta L^R_{\theta,k} = \frac{\partial \rho_\theta}{\partial \theta^k}. \tag{31}$$

Here, $L^R_{\theta,k}$ is not necessarily self-adjoint operator. Holevo then defined the RLD Fisher information matrix J^R_θ as

$$J^R_{\theta,k,l} := \operatorname{Tr} \rho L^R_{\theta,k} (L^R_{\theta,l})^\dagger. \tag{32}$$

For a locally unbiased estimator Π at θ, using Schwarz inequality, we obtain the matrix inequality

$$V_\theta(\Pi) \geq (J^R_\theta)^{-1}. \tag{33}$$

In the one-parameter case, the inequality $J^R_\theta \geq J_\theta$ holds so that (33) does not give a better bound than (30). However, in the multiple parameter case, due to the effect of the off-diagonal part, there are several cases where (33) gives a better bound than (30). To clarify such a case, he introduced the D operator $\mathcal{D}_\theta$ on the set of self-adjoint operators as

$$\rho_\theta \circ \mathcal{D}_\theta(X) = i[X, \rho_\theta]. \tag{34}$$

When the space spanned by $\{L_{\theta,k}\}_k$ is invariant with respect to the D operator $\mathcal{D}_\theta$, which is called the D invariant property, Holevo[1, 35] showed that

$$(J^R_\theta)^{-1} = J^{-1}_\theta + \frac{i}{2} J^{-1}_\theta D_\theta J^{-1}_\theta \tag{35}$$

where the antisymmetric matrix D_θ is defined as

$$D_{\theta,k,l} := \operatorname{Tr} \mathcal{D}_\theta(L_{\theta,k})(L_{\theta,l} \circ \rho_\theta). \tag{36}$$

This fact means that (33) is a stricter matrix inequality than (30) in the above case.

Furthermore, to utilize the imaginary part of the inequality (33), Holevo[1, 35] considered the weighted sum of the components of the mean square error matrix $V_\theta(\Pi)$. We choose a $d \times d$ positive semi-definite matrix G, and focus on

$$\operatorname{Tr} GV_\theta(\Pi). \tag{37}$$

In fact, since it is impossible to simultaneously minimize all diagonal components in $V_\theta(\Pi)$, we need to handle their trade-off. The above strategy enables us to discuss their trade-off. To discuss this problem, Holevo focused on the following lemma.

Lemma 1. *[1, Lemma 6.6.1], [33, (2.9)], [34, (8.1 1)] Given a Hermitian matrix R, we have*

$$\min_{V:Hermitian} \{\operatorname{Tr} GV | V \geq \pm R\} = \operatorname{Tr} |\sqrt{G} R \sqrt{G}|. \tag{38}$$

The minimum holds when $V = \sqrt{G}^{-1} |\sqrt{G} R \sqrt{G}| \sqrt{G}^{-1}$.

This lemma can be easily shown by diagonalizing the Hermitian matrix $\sqrt{G} R \sqrt{G}$.

Combining the matrix inequality (33) and Lemma 1, Holevo[1, 35] showed that

$$\operatorname{Tr} GV_\theta(\Pi) \geq \operatorname{Tr} G\mathbf{Re}(J_\theta^R)^{-1} + \operatorname{Tr} |\sqrt{G} \mathbf{Im}(J_\theta^R)^{-1} \sqrt{G}| \tag{39}$$

for a locally unbiased estimator Π. Hence, the right hand side of (39) is called the RLD bound. When the D invariant property holds, the relation (35) simplifies (39) to

$$\operatorname{Tr} GV_\theta(\Pi) \geq \operatorname{Tr} GJ_\theta^{-1} + \frac{1}{2} \operatorname{Tr} |\sqrt{G} J_\theta^{-1} D_\theta J_\theta^{-1} \sqrt{G}|. \tag{40}$$

When we employ (30) instead of (33), the lower bound is composed only of the first term in (40). Also, Holevo[36] also showed the equality in (40).

However, the above approach works only when the D invariant property holds. To resolve this problem, we consider a more general approach. When Π is a locally unbiased estimator, we choose d self-adjoint operators $\boldsymbol{X}(\Pi) = (X^1(\Pi), \ldots, X^d(\Pi))$ as

$$X^k(\Pi) := \int_{\mathbb{R}^d} (\hat{\theta}^k - \theta^k) \Pi(d\hat{\theta}). \tag{41}$$

Then, we have the condition

$$\operatorname{Tr} X^k(\Pi) \frac{\partial \rho_\theta}{\partial \theta^l} = \delta_{k,l}. \tag{42}$$

Also, Holevo [1, (6.7.73)] showed the matrix inequality

$$V_\theta(\Pi) \geq Z(\boldsymbol{X}(\Pi)), \tag{43}$$

where the Hermitian matrix $Z(\boldsymbol{X})$ is defined as

$$Z^{k,l}(\boldsymbol{X}) := \operatorname{Tr} X^k X^l \rho_\theta. \tag{44}$$

Then, we define

$$C_\theta(G,\boldsymbol{X}) := \inf_{V:\text{Symmetric}} \{\operatorname{Tr} GV | V \geq Z(\boldsymbol{X})\}. \tag{45}$$

Since the relation (43) guarantees that the matrix $V_\theta(\Pi)$ satisfies the condition in (45), we find the inequality

$$\operatorname{Tr} GV_\theta(\Pi) \geq C_\theta(G,\boldsymbol{X}(\Pi)). \tag{46}$$

Therefore, for a locally unbiased estimator Π, we obtain

$$\operatorname{Tr} GV_\theta(\Pi) \geq C_\theta^{HN}(G) := \inf_{\boldsymbol{X}} C_\theta(G,\boldsymbol{X}), \tag{47}$$

where the above infimum is taken for d self-adjoint operators $\boldsymbol{X} = (X^1,\ldots,X^d)$ satisfying the condition (42).

When the D invariant property holds, Holevo [1, Section 6.7] showed that the lower bound in (47) equals the right hand side of (40). He also showed that the infimum in (47) is attained when the d self-adjoint operators $\boldsymbol{X}$ is given as

$$X^k = \sum_{l=1}^d (J_\theta^{-1})^{k,l} L_{\theta,l}. \tag{48}$$

Using this fact, Holevo[1,36] constructed an estimator to attain the lower bound (47) under the quantum Gaussian family.

Later, using Lemma 1, Nagaoka solved the minimization (45) as

$$C_\theta(G,\boldsymbol{X}) = \operatorname{Tr}\sqrt{G}\mathbf{Re}Z(\boldsymbol{X})\sqrt{G} + \operatorname{Tr}|\sqrt{G}\mathbf{Im}Z(\boldsymbol{X})\sqrt{G}|. \tag{49}$$

Nagaoka[37] wrote that Holevo introduced the lower bound $C_\theta^{HN}(G)$. In fact, the aim of the paper[37] was to present Nagaoka's new bound, which is the two-parameter case of the bound called Nagaoka-Hayashi bound in the recent paper,[38] as well as to present its advantage over the lower bound $C_\theta^{HN}(G)$. Also, since the inequality (47) can be obtained with a few steps and a combination of the matrix inequality (42) and Lemma 1, Nagaoka considered that the inequality (47) should be attributed to the contribution by Holevo. Following to Nagaoka's idea, many subsequent studies[38–46] called it the Holevo bound. In these studies, one of the authors was directly suggested by Nagaoka or his collaborators about the use of the terminology "the Holevo bound". Recently, due to these studies, other studies[47–49] followed this terminology without direct suggestion from Nagaoka or his collaborator.

However, Holevo [1, Section 6.7] considered the infimum (47) only when the D invariant property holds. Nagaoka[37] wrote down the infimum (47) for the general case at the first time. While Nagaoka's paper[37] was written as a technical report, the reference[39] provided the full derivation of (47) with the form (49) as a journal publication. In the reference [39, Remark 2], the author mentioned that this bound was essentially introduced in [1, Sec. VI-7] while the notation of[39] is based on the reference.[37] At that time, the author was overconfident about the explanation in the reference[37] for the reference.[1] The main issue of the reference [1, Sec. VI-7] is a statement different from (47). Since Nagaoka wrote down the infimum $C_\theta(G, \boldsymbol{X})$ as (49) and the expression (49) has been widely used for the calculation of this bound, it is suitable to call it the Holevo-Nagaoka bound as has been done in the reference.[50] Indeed, some of the recent papers cited only the reference[1] to discuss this bound whereas it is hard to derive the bound (47) with the form (49) only from the reference.[1] When a paper uses this bound, it might be better to cite paper[37] (and reference[39]) in addition to reference.[1]

At a later time, the importance of the lower bound $C_\theta^{HN}(G)$ was revealed. In the following, we discuss its importance in subsequent studies. Nagaoka[51] introduced the concept of the n-fold independent and identical distributed (i.i.d.) extension of a given state family $\mathcal{S} = \{\rho_\theta\}$ as $\mathcal{S}_n = \{\rho_\theta^{\otimes n}\}$. At that time, many strong objections to this extension were presented from the viewpoint of physics due to the no-cloning theorem. Therefore, this research direction was not well accepted at that time. The paper[39] showed the asymptotic attainability of the lower bound $C_\theta^{HN}(G)$ under the i.i.d. extension of any submodel of the qubit system including the full model. To show this fact, the paper[39] showed the asymptotic approximation of the i.i.d. extension by the qubit system by the Gaussian state family, which is called the local asymptotic normality. The paper[52] showed the local asymptotic normality of full parameter model in the qudit system, which implies the asymptotic attainability of the RLD bound under the i.i.d. extension of the full parameter model of the qubit system. Finally, the papers[40,43] showed the asymptotic attainability of the lower bound $C_\theta^{HN}(G)$ under the i.i.d. extension of any submodel of the qubit system in various settings. In this way, the lower bound $C_\theta^{HN}(G)$ plays the key role among subsequent works. Nowadays, it is widely considered that the lower bound $C_\theta^{HN}(G)$ gives the ultimate bound in state estimation.

4. Quantum estimation with group covariant estimator

Quantum system has useful group symmetry, which is an essential difference from classical system. It is not impossible to consider a group symmetry in a classical system, but the irreducible decomposition does not have a simple form in a classical system. In contrast, Since a quantum system often has a very elegant irreducible decomposition, we can expect that group symmetry greatly simplifies the problem for quantum state estimation. Based on this idea, Holevo formulated quantum state estimation theory when the state family $\mathcal{S} = \{\rho_\theta\}_{\theta\in\Theta}$ has the symmetry with respect to a representation f of a group G on a Hilbert space $\mathcal{H}$.[1,53] When the parameter space Θ is closed with respect to an action of the group G and the state family $\mathcal{S}$ satisfies the following condition, the state family is called a covariant family with respect to the representation f.

$$f(g)\rho_\theta f(g)^\dagger = \rho_{g\theta} \tag{50}$$

for $g \in G$ and $\theta \in \Theta$. In this section, we assume the above condition for our state family $\mathcal{S}$. Next, we consider a POVM Π whose data set is Θ. We say that a POVM

Π is covariant with respect to the representation f when the following condition holds:

$$f(g)M(B)f(g)^\dagger = M(gB) \tag{51}$$

for any Borel set B of Θ, where $gB := \{g\theta | \theta \in B\}$. The concept for the group covariant measurement can be attributed back to Mackey.[54]

To discuss the precision of our estimate, we need to employ an error function $R(\theta, \hat{\theta})$. Then, the average error is given as

$$R_\theta(\Pi) := \int_\Theta R(\theta, \hat{\theta}) \operatorname{Tr} \Pi(d\hat{\theta})\rho_\theta. \tag{52}$$

When we focus on the average for θ, using a prior distribution ν on Θ, we discuss the Bayesian risk

$$R_\nu(\Pi) := \int_\Theta R_\theta(\Pi)\nu(d\theta). \tag{53}$$

When G is a compact group, the invariant probability measure μ on G can be defined. We often employ μ as our prior distribution. When we are interested in the worst case θ, we discuss the minimax risk

$$R(\Pi) := \max_{\theta \in \Theta} R_\theta(\Pi). \tag{54}$$

In this problem setting, it is natural to impose the group invariance to the error function $R(\theta, \hat{\theta})$, which is formulated as

$$R(g\theta, g\hat{\theta}) = R(\theta, \hat{\theta}). \tag{55}$$

When the error function R satisfies (55) and the estimator Π is covariant, we have

$$R_\theta(\Pi) = R_{\theta'}(\Pi) = R(\Pi) = R_\nu(\Pi) \tag{56}$$

for $\theta', \theta \in \Theta$.

When the error function R satisfies (55) and G is a compact group, as a quantum version of Hunt-Stein theorem,[55] Holevo[1,53] showed the following relations.

$$\min_{\Pi \in \mathcal{M}} R(\Pi) = \min_{\Pi \in \mathcal{M}} R_\mu(\Pi) = \min_{\Pi \in \mathcal{M}_{cov}} R(\Pi) = \min_{\Pi \in \mathcal{M}_{cov}} R_\mu(\Pi) = \min_{\Pi \in \mathcal{M}_{cov}} R_\theta(\Pi), \tag{57}$$

where $\mathcal{M}$ expresses the set of POVM whose data set is Θ and $\mathcal{M}_{cov}$ expresses the set of covariant POVM whose data set is Θ.

When G is not compact, the invariant prior μ does not exist. Instead of (57), we have[56,57]

$$\min_{\Pi \in \mathcal{M}} R(\Pi) = \min_{\Pi \in \mathcal{M}_{cov}} R(\Pi) = \min_{\Pi \in \mathcal{M}_{cov}} R_\theta(\Pi). \tag{58}$$

Now, we assume that there exists an invariant measure μ on Θ, which is a weaker condition than the compactness of Θ. For a covariant POVM Π, there exists a positive semi-definite operator T_0 such that

$$M(d\theta) = f(g_\theta)T_0 f(g_\theta)^\dagger \mu(d\theta), \tag{59}$$

where g_θ is chosen as $g_\theta \theta_0 = \theta$, and

$$\int_\Theta f(g_\theta) T_0 f(g_\theta)^\dagger \mu(d\theta) = I. \tag{60}$$

Conversely, given a positive semi-definite operator T_0 satisfying (60), we can construct a covariant POVM Π with (59). Therefore, our optimization in (57) and (58) is simplified to the optimization for the choice of a positive semi-definite operator T_0 satisfying (60).

Furthermore, condition (60) can be simplified into the case with irreducible representation and the case of a commutative group G without multiplicity. In the case with irreducible representation, the left hand side of (60) is a constant multiplication of I. Hence, the condition (60) is replaced by the normalization of the trace of T_0. In the case of a commutative compact group G without multiplicity, each irreducible representation is one-dimensional. The set of irreducible representations is denoted by $\hat{G}$. We denote the vector in the irreducible representation space identified by $\lambda \in \hat{G}$ by $|e_\lambda\rangle$. The condition "without multiplicity" means that the uniqueness of the vector $|e_\lambda\rangle$. We denote the set of irreducible representations appearing in $\mathcal{H}$ by $\hat{S}$. Then, any vector in $\mathcal{H}$ is written as $\sum_{\lambda \in \hat{S}} a_\lambda |e_\lambda\rangle$. The condition (60) is simplified as $\|a_\lambda\|^2 = 1$ for $\lambda \in \hat{S}$.

Holevo[1, 53] applied this approach to several covariant models. The first example is the model with the group U(1), in which Helstrom[58] studied the case when the error function is the delta function. In this model, a state ρ and a representation of the group U(1) are fixed. This model can be considered as the estimation of unknown phase operation. Holevo derived the optimal estimator under proper error functions. The second example is a model with Weyl-Heinsenberg group. He discusses the estimation in quantum Gaussian family [1, Chapter 5]. The third example is a model with group SU(2).[53]

Later, this approach diverged into various directions. The first one is the asymptotic estimation on the full pure state family.[59] The i.i.d. extension of the full pure state family is written as a state family on the symmetric tensor product space, which is an irreducible space. Hence, this approach exactly derives the optimal estimator with a finite size. In contrast, the Cramér-Rao approach gives only an asymptotically approximately optimal estimator. In this way, while the group covariant approach can be applied limited cases, this approach can drive a stronger conclusion in this way.

The second one is the estimation of unitary action. When we estimate the unknown applied unitary and the set of possible unitaries forms a group, the group covariant approach works well. In this case, we can optimize the input state as well as the measurement. To consider this problem, many researchers[60–67] used covariant measurements. Thanks to the covariant state estimation theory, the problem is simplified to the optimization of the input state. As results, they could derive Heisenberg scaling in this problem setting. In this approach, some of them exactly derived the optimal error with the finite size. Indeed, many researchers[68–72] claimed the achievement of Heisenberg scaling by using the Cramér-Rao approach. However, the Cramér-Rao approach has serious drawback in the estimation of unitary action. Basically, Cramér-Rao approach realizes the local optimal estimator. Using the two-step adaptive method,[73] this approach realizes the global optimal estimator in the case of state estimation. However, the two-step adaptive method works only with the usual scaling.[74] In addition, the bound obtained by the Cramér-Rao approach is smaller than the limit of the minimum error with a finite size.[74] This relation shows that it is impossible to show the achievability of the bound obtained by the Cramér-Rao approach. That is, if it is possible, we can derive an asymptotic bound that contradicts with the limit of the minimum error with a finite size. In this way, Holevo's covariant state estimation theory greatly supports many researches for Heisenberg scaling.

The key idea of Holevo's group symmetric approach is reducing the number of free parameters by using the symmetry. Nowadays, this idea is widely accepted in quantum information theory, and his result is the first successful example in this direction. For example, approximate cloning employs the group symmetry.[75] That is, there exists this philosophy behind many recent results of quantum information theory with group covariance.[76]

5. Conclusion

We have reviewed Holevo's researches in quantum information theory in the 20th century. Although majority of the research were done in the 1970s, they have strongly influenced our current studies in quantum information theory. They form the foundations of many topics in quantum information theory. That is to say, he contributed many tools for these areas.

Interestingly, he contributed quantum estimation via two different approaches, the Craméro-Rao approach and the group covariant approach - he addressed. That is, he addressed quantum estimation without sticking to one approach. This fact shows that he has the ability to tackle key problem without being constrained to a specific strategy. The author would like to conclude this article by expressing his sincere respect for Holevo's significant contributions to quantum information theory.

References

1. A. S. Holevo, *Probabilistic and statistical aspects of quantum theory*, Translated from the Russian by the author, North-Holland Series in Statistics and Probability, 1, North-Holland Publishing Co., Amsterdam, 1982 , xii+312 pp.
2. P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, W. Wooters, "Classical information capacity of a quantum channel," *Phys. Rev. A* **54**, 1869-1876 (1996).
3. O. Hirota, A. S. Holevo, and C. M. Caves, eds. *Quantum Communication, Computing, and Measurement*, Plenum Publishing, (1997).
4. A.S. Holevo, "The capacity of the quantum channel with general signal states," *IEEE Trans. Inf. Theory* 44, 269 (1998); arXiv:quant-ph/9611023.

5. A. S. Holevo, “Towards the mathematical theory of quantum communication channels,” *Probl. Pered. Inform.,* vol. 8, no. 1, pp. 63–71, 1972. (English translation: Probl. Inform. Transm., vol. 8, no. 1, pp. 47–56).
6. A. S. Holevo, “Some estimates of the information transmitted by quantum communication channel,” *Probl. Pered. Inform.,* vol. 9, no. 3, pp. 3–11, 1973. (English translation: Probl. Inform. Transm., vol. 9, no. 3, pp. 177–183).
7. A. S. Holevo, “Problems in the mathematical theory of quantum communication channels,” *Rep. Math. Phys.,* vol. 12, no. 2, pp. 273–278, 1977.
8. C.W. Helstrom, *Quantum detection and estimation theory*, Academic Press, New York, 1976.
9. C.W. Helstrom, “Detection theory and quantum mechanics,” *Inform. and Control* **10** (3) (1967), 254 – 291.
10. A. S. Holevo, “Bounds for the Quantity of Information Transmitted by a Quantum Communication Channel,” *Problems Information Transmission*, **9**, 3 (1973) 177 – 183.
11. L. B. Levitin, “On quantum measure of information,” in: *Proc. IV All-Union Conference on Information Transmission and Coding Theory*, Tashkent, 1969, pp. 111–115. (Russian).
12. A. S. Holevo, “On Capacity of a Quantum Communications Channel,” *Problems Information Transmission*, **15**, 4 (1979) 247– 253.
13. R. L. Stratonovich and A. G. Vantsyan, “Asymptotically error-free decoding in pure quantum channels”, *Probl. Control Inform. Theory* **7** 3 (1978), 161–174.
14. G. Lindblad, “Completely positive maps and entropy inequalities,” *Communications in Mathematical Physics*, 40(2):147–151, June 1975.
15. A. Uhlmann, “Relative entropy and the Wigner-Yanase-Dyson-Lieb concavity in an interpolation theory,” *Communications in Mathematical Physics*, 54(1) (1977) 21 – 32.
16. B. Schumacher, “Quantum coding,” *Phys. Rev. A* 51, 2738 (1995)
17. R. Jozsa, B. Schumacher, A new proof of the quantum noiseless coding theorem, *J. Modern Optics* **41**, (1994) 2343 – 2349.
18. B. Schumacher, M.D. Westmoreland, “Sending classical information via noisy quantum channels,” *Phys. Rev. A* 56, 131 (1997)
19. M. V. Burnashev and A. S. Holevo, “On reliability function of quantum communication channel,” *Problems Inform. Transmission,* **34**, 2 (1998) 97–107; LANL Rep, quant-ph/9703013, Mar. 10, 1997.
20. T. Ogawa and H. Nagaoka, “Strong converse to the quantum channel coding theorem,” *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2486-2489 (1999).
21. T. S. Han, *Information-Spectrum Methods in Information Theory.* Berlin, Germany: Springer-Verlag, 2003, (The original Japanese edition was published by Baifukan-Press, Tokyo, Japan, in 1998).
22. M. Hayashi and H. Nagaoka, “General formulas for capacity of classical-quantum channels,” *IEEE Transactions on Information Theory*, vol. 49, no. 7, (2003) 1753-1768.
23. C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, “Entanglement-Assisted Classical Capacity of Noisy Quantum Channels,” *Phys. Rev. Lett.* **83**, 3081 (1999);
24. C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor and A. Winter, “The Quantum Reverse Shannon Theorem and Resource Tradeoffs for Simulating Quantum Channels,” *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2926-2959, (2014).
25. A.S. Holevo, “On entanglement-assisted classical capacity,” *J.Math. Phys.* **43**, (2002) 4326 – 4333.
26. I. Devetak, “The private classical capacity and quantum capacity of a quantum chan-

nel," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 44-55 (2005).
27. C.W. Helstrom, "Minimum mean-square error estimation in quantum statistics," *Phys. Lett.,* 25A, 101-102 (1967).
28. H.P. Yuen and M. Lax, "Multiple-parameter quantum estimation and measurement of non-selfadjoint observables," *IEEE Trans. Inform. Theory,* **19**, 740–750 (1973).
29. A. S. Holevo, "On the quasiequivalence of quasifree states on the C^*-algebra of CAR," *Theoret. and Math. Phys.*, **14**, 2 (1973) 107–111.
30. A. S. Holevo, "Quasifree states of the C^* algebra of CCR. II," *Theoret. and Math. Phys.,* 6:2 (1971) 103–107.
31. A. S. Holevo, "Quasi-free states on the C^* algebra of CCR," *Theoret. and Math. Phys.*, **6** (1971) 1–12.
32. V. P. Belavkin, "On generalised uncertainty relations and efficients measurements on quantum systems," *Teoreticheskaya i Matematichescheskaya Fizika,* 26, 316-319 (1976). (In Russian); English translation, "Generalized uncertainty relations and efficient measurements in quantum systems," *Theoretical and Mathematical Physics,* 26, 213–222 (1976).
33. V. P. Belavkin and B. A. Grishanin, "Optimum estimation of quantum channels by the generalized Heisenberg inequality method," *Probl. Peredachi Inf.*, **9** 3 (1973), 44–52; Problems Inform. Transmission, 9:3 (1973), 209– 215.
34. R. L. Stratonovich, "The quantum generalization of optimal statistical estimation and testing hypothesis," *J. Stoch.* 1 (1973) 87–126.
35. A. S. Holevo, "Noncommutative analogues of the Cramér-Rao inequality in the quantum measurement theory", *Proceedings of the Third Japan–USSR Symposium on Probability Theory*, (Tashkent, 1975), Lecture Notes in Math., 550, Springer, Berlin, 1976, 194 – 222
36. A. S. Holevo, "Some statistical problems for quantum Gaussian states," *IEEE Trans. Information Theory*, **21** 5 (1975) 533–543
37. H. Nagaoka, "A new approach to Cramér-Rao bounds for quantum state estimation," *IEICE Technical Report*, **89**, 228, IT 89-42, 9-14, (1989); Asymptotic Theory of Quantum Statistical Inference: Selected Papers, edited by M. Hayashi (World Scientific, Singapore, 2005), pp. 100–112.
38. L. O. Conlon, J. Suzuki, P. K. Lam, and S. M. Assad, "Efficient computation of the nagaoka–hayashi bound for multiparameter estimation with separable measurements," *npj Quantum Information* **7**, 1 (2021).
39. M. Hayashi and K. Matsumoto, "Asymptotic performance of optimal state estimation in qubit system," *Journal of Mathematical Physics*, Vol. 49, 102101 (2008): "Asymptotic performance of optimal state estimation in quantum two level system," arXiv:quant-ph/0411073 (2004).
40. L. Yamagata, A. Fujiwara, and R.D. Gill, "Quantum local asymptotic normality based on a new quantum likelihood ratio," *Ann. Stat.* **41**(4), (2013) 2197– 2217
41. J. Suzuki, "Parameter estimation of qubit states with unknown phase parameter," *International Journal of Quantum Information*,**13** 1450044 (2015)
42. J. Suzuki, "Explicit formula for the Holevo bound for two-parameter qubit-state estimation problem featured," *J. Math. Phys.* **57**, 042201 (2016).
43. Y. Yang, G. Chiribella, and M. Hayashi, "Attaining the Ultimate Precision Limit in Quantum State Estimation," *Communications in Mathematical Physics,* vol. 368(1), 223 – 293 (2019).
44. J. Suzuki, "Information Geometrical Characterization of Quantum Statistical Models in Quantum Estimation Theory," *Entropy*, 21(7), 703 (2019).
45. J. Suzuki, Y. Yang, and M. Hayashi, "Quantum state estimation with nuisance pa-

rameters," *Journal of Physics A: Mathematical and Theoretical*, vol. 53 453001 (2020).

46. K. Yamagata, "Maximum logarithmic derivative bound on quantum state estimation as a dual of the Holevo bound," *J. Math. Phys.* **62**, 062203 (2021).
47. F. Albarelli, J. F. Friel, and A. Datta, "Evaluating the Holevo Cramér-Rao Bound for Multiparameter Quantum Metrology," *Phys. Rev. Lett.* 123, 200503 (2019).
48. M. Tsang, F. Albarelli, and A. Datta "Quantum Semiparametric Estimation," *Phys. Rev. X* 10, 031023 (2020)
49. J. Friel, P. Palittapongarnpim, F. Albarelli, and A. Datta "Attainability of the Holevo-Cramér-Rao bound for two-qubit 3D magnetometry," arXiv.2008.01502
50. J. Suzuki, "Quantum-state estimation problem via optimal design of experiments," *International Journal of Quantum Information* **19**, No. 08, 2040007 (2021)
51. H. Nagaoka, "On the parameter estimation problem for quantum statistical models," In *Poceeding of 12th Symposium on Information Theory and Its Applications (SITA)*, Inuyama, Tottori, Japan, December 6–9, 1989, p 577 – 582; Asymptotic Theory of Quantum Statistical Inference: Selected Papers, edited by M. Hayashi (World Scientific, Singapore, 2005), pp. 120–127.
52. J. Kahn and M. Guţă, "Local asymptotic normality for finite dimensional quantum systems," *Commun. Math. Phys.* 289(2), 597 – 652 (2009)
53. A. S. Holevo, "Covariant measurements and uncertainty relations," *Rep. Math. Phys.*, **16**, 3 (1979) 385 – 400.
54. G. W. Mackey, "Imprimitivity for representations of locally compact groups I," *Proc. Nat. Acad. Sci. U.S.A.*, **35** (1949), 537–545.
55. T. S. Ferguson, *Mathematical statistics: a decision theoretic approach*, Acad Press, NY, 1967.
56. N.A. Bogomolov, "Minimax measurements in a general statistical decision theory," *Theor. Probl. Appl.* 26, 787 (1982)
57. M. Ozawa, "On the noncommutative theory of statistical decision," *Research Reports on Information Sciences*, Report number: A-74, Tokyo Institute of Technology (1980).
58. C. W. Helstrom, "Estimation of a displacement parameter of a quantum system," *International Journal of Theoretical Physics*, **11** (1974) 357–378
59. M. Hayashi, "Asymptotic estimation theory for a finite dimensional pure state model," Journal of Physics A: Mathematical and General **31**, 4633 – 4655 (1998)
60. G. Chiribella, G.M.D'Ariano, and M.F. Sacchi, "Optimal estimation of group transformations using entanglement," Phys. Rev. A **72**, 042338 (2005)
61. M. Hayashi, "Fourier Analytic Approach to Quantum Estimation of Group Action," *Communications in Mathematical Physics*, **347**, 3 – 82 (2016).
62. A. Luis, and J. Perina, "Optimum phase-shift estimation and the quantum description of the phase difference," *Phys. Rev. A* **54**, 4564 (1996)
63. V. Bužek, R.Derka, and S. Massar, "Optimal quantum clocks," *Phys. Rev. Lett.*, **82**, 2207 (1999).
64. M. Hayashi, "Parallel treatment of estimation of SU(2) and phase estimation," *Phys. Lett. A*, **354**(3), 183–189 (2006).
65. H. Imai and M. Hayashi, "Fourier analytic approach to phase estimation in quantum systems," *New Journal of Physics*, **11**, 043034 (2009).
66. E. Bagan, M. Baig, and R. Munoz-Tapia, "Quantum reverse-engineering and reference-frame alignment without nonlocal correlations," *Phys. Rev. A* **70**, 030301(R) (2004).
67. G. Chiribella, G.M. D'Ariano, P. Perinotti, and M.F. Sacchi, "Efficient use of quantum resources for the transmission of a reference frame," *Phys. Rev. Lett.* **93**, 180503 (2004).

68. V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum-enhanced measurements: beating the standard quantum limit," *Science*, **306**, 1330–1336 (2004).
69. V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum-enhanced "Quantum metrology," *Phys. Rev. Lett.*, **96**, 010401 (2006).
70. T. Nagata, R. Okamoto, J. O'Brien, K. Sasaki, and S. Takeuchi, "Beating the standard quantum limit with four-entangled photons," *Science*, **316**(5825), 726 (2007).
71. R. Okamoto, H.F. Hofmann, T. Nagata, J.L. O'Brien, K. Sasaki, and S. Takeuchi, "Beating the standard quantum limit: phase super-sensitivity of N-photon interferometers," *N. J. Phys.* **10**, 073033 (2008).
72. J.A. Jones, S.D.Karlen, J. Fitzsimons, A. Ardavan, S.C.Benjamin, G.A.D. Briggs, J.J.L. Morton, "Magnetic field sensing beyond the standard quantum limit using 10-spin NOON states," *Science* **324**, 1166–1168 (2009).
73. M. Hayashi and K. Matsumoto, "Statistical model with measurement degree of freedom and quantum physics," RIMS koukyuroku No 1055 (Kyoto: Kyoto University) p 96 (1998) (In Japanese); *Asymptotic Theory of Quantum Statistical Inference.* ed M Hayashi, Singapore: World Scientific, 2005, p. 162 (reprinted, English translation).
74. M. Hayashi, "Comparison between the Cramer-Rao and the mini-max approaches in quantum channel estimation," *Commun. Math. Phys.* **304**(3), 689–709 (2011).
75. R.F. Werner, "Optimal cloning of pure states," *Phys. Rev. A* **58**, 1827 (1998)
76. M. Hayashi, *A Group Theoretic Approach to Quantum Information*, Springer (2017).