(17)「量子統計，量子情報幾何とその量子情報科学への応用」に関する研究報告

Quantum homodyne tomography as a non-parametric estimation problem

Mădălin Guţă[1] *

[1] *University of Nijmegen, Department of Mathematics*
Postbus 9010, 6500 GL Nijmegen, The Netherlands

Abstract. Quantum homodyne tomography is a quantum optical measurement scheme which can be used to estimate the state of a quantum oscillator or the associated Wigner function. The reconstruction of the state from the probability distribution of the data is an ill posed inverse problem, and various expressions of the inverse map have been found in the physics literature. However, as the density matrix may in general be infinite dimensional, the problem of controlling the statistical error of estimation becomes very important. The purpose of this paper is to introduce a few statistical concepts and methods related with non-parametric estimation, and to formulate some open problems in quantum tomography.

Keywords: Quantum homodyne tomography, Wigner function, density matrix, Non-parametric statistics, maximum likelihood estimators, minimax estimators

1 Introduction

In quantum mechanics measurements are intrinsically stochastic, the theory making predictions over the probability distributions of the outcomes. The forward map from the state ρ of the system to the probability distribution P_ρ^M of results of a measurement M, is described in the language of POVM's (positive operator valued measures). The study of the corresponding statistical inverse problem, that is of reconstructing the state ρ from the measurement results, has been initiated by Helstrom [1] and Holevo [2] and grew significantly in the last decade stimulated by technological developments in quantum engineering and measurement techniques.

An illustrative example of a "new" statistical reconstruction problem is that posed by quantum homodyne tomography, a technique proposed in [3] and realized experimentally for the first time in 1993 by Smithey *et al.* [4].

Let us consider a quantum oscillator with canonical variables $\mathbf{Q}$ and $\mathbf{P}$ satisfying the commutation relations $[\mathbf{Q}, \mathbf{P}] = i\mathbf{1}$, and let ρ be its state represented by an infinite dimensional density matrix on the space $L^2(\mathbb{R})$ on which $\mathbf{Q}$ and $\mathbf{P}$ are represented as position and momentum respectively. The main idea of quantum homodyne tomography is that ρ is in one to one correspondence with the family $\{p_\rho(x|\phi) : \phi \in [0, \pi]\}$ of probability densities of the quadratures $\mathbf{X}_\phi := \mathbf{Q}\cos\phi + \mathbf{P}\sin\phi$ for phases ϕ sweeping the interval $[0, \pi]$. This suggests the following experimental set-up for reconstructing ρ: given n systems identically prepared in state ρ, we generate random phases $\Phi_1, \ldots, \Phi_n$ which are independent, uniformly distributed over $[0, \pi]$ and then measure $\mathbf{X}_{\phi_i}$ on the $i-$th system obtaining a result X_i. If we denote the joint density of (X_i, Φ_i) by $p_\rho(x, \phi) := \frac{1}{\pi} p_\rho(x|\phi)$, then the tomography map

$$T : \rho \mapsto p_\rho(x, \phi),$$

is invertible and thus with sufficiently many data points we may hope to obtain a good estimate of the state ρ.

This is a typical statistical inverse problem where the data is distributed according to a linear transform of the parameter of interest. There are two aspects which make the problem difficult: firstly, we deal with an *ill posed inverse problem* in the sense that the map T^{-1} is unbounded with respect to the natural norms on the two spaces and thus small errors in estimating the density p_ρ from the data may lead to big errors for ρ; secondly the parameter of interest ρ is infinite dimensional, (infinite number of matrix elements), thus we deal with a more technical *non-parametric* statistical estimation problem compared with the usual case where the dimension of the density matrix is finite. The last aspect is usually disregarded in the physics literature by either assuming that the density matrix is actually finite or by making cut-off in the dimension of the matrix and neglecting the elements falling outside the finite block. However a quantitative analysis of the performance of such procedures is lacking and it is our goal to fill this gap by adopting a modern statistical perspective.

In quantum optics one often represents the state ρ by its Wigner function [5] $W_\rho(q, p)$ which is a quasi-probability distribution over $\mathbb{R}^2$ in one-to-one correspondence with the density matrix ρ, and then the problem becomes very similar to a classical inverse problem, that of positron emission tomography (PET) [6] where the role of W_ρ is played by a two dimensional probability distribution. The map from $W_\rho(q, p)$ to $p_\rho(x|\phi)$ is the Radon transform defined as

$$\mathcal{R}[f](x, \phi) = \int_{-\infty}^{\infty} f(x\cos\phi - t\sin\phi, x\sin\phi + t\cos\phi)dt.$$

Quantum tomography is thus about inverting the Radon transform which is known to be an ill posed problem [7] in the sense that we can formally write $\mathcal{R}^{-1}$ as

$$\mathcal{R}^{-1}[p](q, p) = \int_0^\pi \int K(q\cos\phi + p\sin\phi - x)p(x|\phi)d\phi dx, \tag{1}$$

where the kernel K is not a bounded function but a distribution (generalized function)

$$K(x) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} |\xi| \exp(i\xi x)d\xi. \tag{2}$$

*m.guta@math.ru.nl

Similarly, the map $T^{-1} : p(x,\phi) \mapsto \rho$ is unbounded if we consider $\|\cdot\|_1$ on both sides. However it turns out that the individual matrix elements of ρ with respect to the Fock basis $\{\psi_j(x) = (\sqrt{\pi}2^j j!)^{-1/2} H_j(x) e^{-x^2/2} : j \geq 0\}$ can be obtained by kernel integration

$$\rho_{k,j} = \frac{1}{\pi} \int_0^\pi \int f_{k,j}(x) e^{-i(j-k)\phi} p_\rho(x,\phi) d\phi dx, \quad (3)$$

where $f_{k,j}$ are bounded oscillatory functions called *pattern functions* [8, 9, 10]. The ill posedness is illustrated here by the fact that the range of $f_{k,j}$ is unbounded as we vary k, j.

2 Consistent estimators of the density matrix

The results of this section draw largely on [11] to which we refer for the proofs. Let us formulate the statistical problem in the case of estimating the density matrix. Given the data $(X_1, \Phi_1), \ldots, (X_n, \Phi_n)$ independent identically distributed with probability density $p_\rho(x,\phi)$ on $\mathbb{R} \times [0,\pi]$, we consider estimators $\hat{\rho}_n = \hat{\rho}_n((X_1,\Phi_1), \ldots (X_n,\Phi_n))$, taking values in the space of trace class operators on $L^2(\mathbb{R})$ and judge their performance at parameter ρ by the risk

$$R_{1,2}(\hat{\rho}_n, \rho) := \mathbb{E}\left(\|\hat{\rho}_n - \rho\|_{1,2}\right),$$

where the distance function may be the trace norm or the Hilbert-Schmidt norm. As a weakest desirable property we would like $\hat{\rho}_n$ to be *consistent*, i.e. $\lim_{n\to\infty} R_{1,2}(\hat{\rho}_n, \rho) = 0$ for all ρ.

Based on formula (3) we define the unbiased estimator for the matrix element $\rho_{k,j}$:

$$\hat{\rho}^{(n)}_{k,j} = \frac{1}{n} \sum_{\ell=1}^{n} F_{k,j}(X_\ell, \Phi_\ell), \quad (4)$$

where $F_{k,j}(x,\phi) = f_{k,j}(x) e^{-i(j-k)\phi}$. The average square error for this element is bounded as follows

$$\begin{aligned}
&\mathbb{E}\left(|\hat{\rho}^{(n)}_{k,j} - \rho_{k,j}|^2\right) \leq \frac{1}{n} \int_0^\pi \int |F_{k,j}(x,\phi)|^2 p_\rho(x,\phi) d\phi dx \\
&= \frac{1}{n} \int dx f_{k,j}(x)^2 \int_0^\pi p_\rho(x,\phi) d\phi \\
&\leq \frac{C_1}{n} \|f_{k,j}\|_2^2 \leq C_2 \frac{j+k+2}{n},
\end{aligned}$$

where the constants C_1, C_2 are independent on j, k, n. Clearly the estimator $\hat{\rho}_n$ with the above matrix elements will not be consistent as the errors for estimating different matrix elements add up to infinite! We can keep the variance bounded by considering the cut-off estimator $\hat{\rho}^{(n,d)}$ whose matrix elements are given by (4) for $j, k < d$ and zero outside this $d \times d$ block. We have now a trade-off between the bias and variance of the ignored elements:

$$\begin{aligned}
\mathbb{E}(\|\rho^{(n,d)} - \rho\|_2^2) &\leq \sigma^2(n,d) + b^2(n,d) \\
&:= \frac{C_2 d^3}{n} + \sum_{\max(k,j) \geq d} |\rho_{jk}|^2, \quad (5)
\end{aligned}$$

and we still have the freedom to choose d in a advantageous way depending on n. If $d(n) \to \infty$ as $n \to \infty$ we have that the bias term converges to zero and a sufficient condition for consistency is $d = o(n^{1/3})$. Similar results can be obtained if we consider the trace norm instead of the Hilbert-Schmidt norm, and also for sieved maximum likelihood estimators [11].

However, without additional information on the state ρ we cannot say much about the *rate of convergence* or the optimal choice of d as a function of n. Thus, we will assume that ρ belongs to a class of exponentially decaying density matrices

$$\mathcal{E}(\alpha,\beta) = \{\rho : \mathrm{Tr}(e^{\alpha \mathbf{N}} \rho) \leq \beta\},$$

where $\mathbf{N}$ is the number operator $\mathbf{N}\psi_n = n\psi_n$, and $\alpha, \beta > 0$ are constants. Then by the positivity of ρ we have $|\rho_{jk}|^2 \leq \rho_{jj}\rho_{kk} \leq \beta^2 e^{-\alpha(j+k)}$. In this case the bias $b^2(n,d)$ is bounded from above by $C_3(\alpha,\beta)e^{-\alpha d}$ and

$$\mathbb{E}\left(\|\hat{\rho}^{(n,d)} - \rho\|^2\right) \leq C_3(\alpha,\beta) e^{-\alpha d} + \frac{C_2 d^3}{n}.$$

By choosing $d = \frac{1}{\alpha} \log n$ we obtain

$$\mathbb{E}\left(\|\hat{\rho}^{(n,d)} - \rho\|^2\right) \leq C_4 \frac{(\log n)^3}{n}, \quad (6)$$

a rate which is slightly worse then usual parametric rate $1/n$.

Several questions remain to be answered within this approach. Why consider exponentially decaying matrices? Surely, stronger assumptions on the class lead to faster rates of convergence but we should also worry whether the states created in the lab satisfy our assumptions. To put it differently, does there exist a natural class of states which contains the ones produced in the lab? Assuming that the state belongs to the class $\mathcal{A}(\alpha,\beta)$, what can we then say about the optimality of the above estimator? In order to avoid trivialities where an estimator is performing very good at a fixed point but very bad elsewhere, one should define optimality in such as way as to take into account the performance of the estimator over the whole class

$$R(\hat{\theta}^{(n,d)}) = \sup_{\rho \in \mathcal{E}(\alpha,\beta)} R(\hat{\theta}^{(n,d)}, \rho),$$

and compare it with the best (smallest) such maximum risk among all estimators, that is the *minimax risk*

$$R_n = \inf_{\hat{\theta}_n} R(\hat{\theta}_n) = \inf_{\hat{\theta}_n} \sup_{\rho \in \mathcal{E}(\alpha,\beta)} R(\hat{\theta}_n, \rho).$$

An estimator is called minimax if it achieves the minimax risk asymptotically. Furthermore, even if the exponential decay property seems natural for a large set of states produced in the lab, how can we know in advance to which class $\mathcal{E}(\alpha,\beta)$ the state belongs ? It is desirable to construct an *adaptive* estimator whose definition does not depend on the parameters of the class, but which still performs almost as good as the minimax estimator when the class is known.

A step in the direction of achieving this goal would be to allow the dimension d to depend on the data in such

a way that $d = d((X_1, \Phi_1), \ldots, (X_n, \Phi_n))$ is close to the optimal dimension

$$d^*(n, \rho) = \inf_d R_2(\hat{\rho}^{(n,d)}, \rho) = \inf_d \mathbb{E}\|\hat{\rho}^{(n,d)} - \rho\|_2^2.$$

Clearly, d^* depends on ρ and thus cannot be calculated from the data, but we can use an unbiased estimator of $R_2(\hat{\rho}^{(n,d)}, \rho)$ to estimate d^*:

$$\|\hat{\rho}^{(n,d)}\|_2^2 + \|\rho\|_2^2 - \frac{2}{n(n-1)} \sum_{j,k=0}^{d-1} \sum_{i \neq l}^{n} F_{j,k}(X_i, \Phi_i) F_{j,k}(X_l, \Phi_l).$$

Now we define the *cross-validation* estimator $\hat{\rho}^{(\hat{d},n)}$ whose dimension $\hat{d}$ is the minimizer of

$$d \mapsto \|\hat{\rho}^{(n,d)}\|_2^2 - \frac{2}{n(n-1)} \sum_{j,k=0}^{d-1} \sum_{i \neq l}^{n} F_{j,k}(X_i, \Phi_i) F_{j,k}(X_l, \Phi_l).$$

Although there exist general theoretical results concerning the consistency of such estimators, a complete analysis of the convergence rate is still lacking.

3 Filtered back-projection estimation of the Wigner function

In this section we turn to the problem of estimating the Wigner function [12]. In [13] a similar problem is studied for the case of PET for a class of very smooth functions. Our result extends those of [13] to the class of smooth Wigner functions

$$\mathcal{W}(\beta, L) = \left\{ W_\rho : \frac{1}{4\pi^2} \int_{\mathbb{R}^2} |\widetilde{W}_\rho(w)|^2 \exp(2\beta|w|)\, dw \leq L \right\},$$

where $\widetilde{W}_\rho$ is the Fourier transform of W_ρ. By replacing W_ρ by a probability density we obtain the class $\mathcal{A}(\beta, L)$ from [13] and we notice that none of the two is contained in the other: there exist Wigner functions in $\mathcal{W}(\beta, L)$ which are not positive, and conversely there exist densities in $\mathcal{A}(\beta, L)$ which do not satisfy the bound $\|W_\rho\|_\infty \leq \frac{1}{\pi}$ characteristic for the Wigner functions [5].

If $\widehat{W}_n$ is an estimator of W_ρ, we define the pointwise risk in a fixed point $z := (q, p) \in \mathbb{R}^2$ by

$$R_z(\widehat{W}_n, W_\rho) = \mathbb{E}\left[(\widehat{W}_n(z) - W_\rho(z))^2\right].$$

The estimator called *filtered back projection* is based on the inverse Radon transform (1):

$$\widehat{W}_n(z) = \frac{1}{n} \sum_{i=1}^{n} K_\delta([z, \Phi_i] - X_i),$$

where K_δ is a regularization of the kernel (2) given by the cutoff δ

$$K_\delta(u) = \frac{1}{4\pi^2} \int_{-\delta}^{\delta} r e^{iru} dr,$$

with Fourier transform $\widetilde{K}_\delta(t) = \frac{1}{2\pi}|t| I_\delta(t)$ where I_δ is the indicator function of $\{t : |t| \leq 1/\delta\}$.

We define the dual operator $\mathcal{R}^\#$ on $L_1(\mathbb{R} \times [0, \pi])$ by

$$\mathcal{R}^\#[h](z) = \int_0^\pi h([z, \phi], \phi)\, d\phi.$$

Then

$$\mathcal{R}^\#\mathcal{R}[W](z) = \int_0^\pi \mathcal{R}[W]([z, \phi], \phi)\, d\phi$$

represents the integrals of W over all lines passing through the point x. Note that in general $\mathcal{R}^\#\mathcal{R}[W](z) \geq 0$ for all Wigner functions W and all $z \in \mathbb{R}^2$, but there exist states such that ψ_k with k odd for which $\mathcal{R}^\#\mathcal{R}[W](0) = 0$.

The parameter δ plays a similar role to that of the dimension d of the estimator of the density matrix ρ described in the previous section and the question is what is the optimal dependence $\delta = \delta(n)$? The following theorem gives an upper bound for the risk of the estimator with a specially chosen cut-off δ_n.

Theorem 1 *[12] Let $\delta_n = \log n/(2\beta)$, then for any $W \in \mathcal{W}(\beta, L)$ and any fixed $z \in \mathbb{R}^2$ such that $\mathcal{R}^\#\mathcal{R}[W](z) > 0$ we have as $n \to \infty$,*

$$\mathbb{E}\left[(\widehat{W}_n(z) - W(z))^2\right] = C^* \mathcal{R}^\#\mathcal{R}[W](z) \times \frac{(\log n)^3}{n}(1 + o(1))$$

where $C^* = 3\pi(4\pi\beta)^{-3}$.

The proof consists in bounding the risk by a sum of a bias term and variance term similarly to (5) and choosing δ_n in such a way that the two terms balance each other:

$$\mathbb{E}\left[(\widehat{W}_n(z) - W(z))^2\right] =$$
$$\left(\mathbb{E}(\widehat{W}_n(z)) - W(z)\right)^2 + \mathbb{E}\left[(\widehat{W}_n(z) - \mathbb{E}(\widehat{W}_n(z)))^2\right]$$
$$:= b_n^2(z) + \sigma_n^2(z).$$

In order to show that the estimator is minimax we need to prove that no other estimator can perform strictly better asymptotically. This amounts to proving that there exist a *lower bound* for the global risk of *any* estimator, and this lower bound is asymptotically equal to the rate of $\widehat{W}_n(z)$. For technical reasons we consider the sightly modified class of Wigner functions [13]

$$\mathcal{W}(\beta, L, \alpha_n) = \left\{ W \in \mathcal{W}(\beta, L) : \mathcal{R}^\#\mathcal{R}[W](z) \geq \alpha_n \right\},$$

for a sequence α_n such that $\lim_{n\to\infty} \alpha_n = 0$ and $\lim_{n\to\infty} (\alpha_n (\log n)^{1/3}) = \infty$. Let us denote

$$r_n(z) = \sup_{W \in \mathcal{W}(\gamma, L, \alpha_n)} \left[C^* \mathcal{R}^\#\mathcal{R}[W](z) \frac{(\log n)^3}{n} \right]^{1/2}.$$

Theorem 2 *For a fixed $z \in \mathbb{R}^2$, we have*

$$\liminf_{n\to\infty} \inf_{\widehat{T}_n} \sup_{W \in \mathcal{W}(\gamma, L, \alpha_n)} \mathbb{E}\left[\left(\frac{\widehat{T}_n(z) - W(z)}{r_n(z)} \right)^2 \right] \geq 1$$

where $\inf_{\widehat{T}_n}$ denotes the infimum over all estimators of $W(z)$.

The proof of the lower bound is based on finding a "most difficult" parametric subfamily of $\mathcal{W}(\beta, L, \alpha_n)$ such that the lower bound still holds when restricting to the subfamily and replacing the supremum with the average of the risk with respect to a specially designed distribution over the subfamily. By applying the van Trees inequality we obtain a lower bound which is an analogue of the Cramér-Rao inequality in a Bayesian set-up [14].

References

[1] C. W. Helstrom *Quantum Detection and Estimation Theory.* Academic Press, New York, 1976.

[2] A. S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory.* North-Holland, 1982.

[3] K. Vogel and H, Risken. Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase. *Phys. Rev. A*, 40, 2847–2849, 1989.

[4] D. T. Smithey, M. Beck, M. G. Raymer and A. Faridani. Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum *Phys. Rev. Lett.*, 70, 1244–1247, 1993.

[5] U. Leonhardt. *Measuring the Quantum State of Light.* Cambridge University Press, 1997.

[6] Y. Vardi, L. A. Shepp and L. Kaufman. A statistical model for positron emission tomography. *J. Am. Statist. Assoc.*, 80, 8–37, 1985.

[7] S. R. Deans. *The Radon transform and some of its applications*, John Wiley & Sons, New York 1983.

[8] G. M. D'Ariano, U. Leonhardt and H. Paul. Homodyne detection of the density matrix of the radiation field. *Phys. Rev. A*, 52, R1801–R1804, 1995.

[9] U. Leonhardt, M. Munroe, T. Kiss, Th. Richter and M. G. Raymer. Sampling of photon statistics and density matrix using homodyne detection. *Optics Communications*, 127, 144–160, 1996.

[10] U. Leonhardt, H. Paul, and G. M. D'Ariano. Tomographic reconstruction of the density matrix via pattern functions. *Phys. Rev. A* , 52, 4899–4907, 1995 .

[11] L. Artiles, R. Gill and M. Guţă. An invitation to quantum tomography. *J. Royal Statist. Soc. B: Methodology.*, 7, 109-134, 2005.

[12] L. Artiles and M. Guţă. Estimation of the Wigner function in quantum homodyne tomography with ideal detectors. in preparation.

[13] L. Cavalier. Efficient estimation of a density in a problem of tomography *Ann. Statist.*, 28, 630-647, 2000.

[14] R. D. Gill and B. Y. Levit. Applications of the van Trees inequality: a Bayesian CramerRao bound. *Bernoulli*, 1, 59-79, 1995.

Bayesian predictive density operators for the Gaussian states family

Fuyuhiko Tanaka[1] * Fumiyasu Komaki[1]

[1] *Department of Mathematical Informatics, University of Tokyo. 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.*

Abstract. Quantum state estimation has been widely investigated and there are mainly two approaches proposed: One is based on point estimation of an unknown parameter and the other is based on Bayesian method. In exchangeable quantum models with an arbitrarily chosen measurement it is shown that Bayesian predictive density operators are the best predictive density operators when we evaluate them by using the average relative entropy based on a prior [13]. We calculate the Bayesian predictive density operator for the Gaussian states family with the heterodyne measurement and confirm that their statement holds true in this model.

Keywords: quantum estimation, Gaussian states, relative entropy

1 Introduction

In classical statistics, the problem of predicting an unobserved variable y by using an observed variable x has been investigated. Suppose that a parametric model

$$\mathcal{P} = \{p(y|\theta) : \theta \in \Theta\},$$

which is a set of probability densities, is given, where Θ is a parameter space. Random variables x and y are distributed according to the same true probability density $p(\cdot|\theta)$ in $\mathcal{P}$. We predict the unobserved variable y with a predictive density $\hat{p}(y;x)$ constructed by using the observed variable x. The closeness of the true density $p(y|\theta)$ and a predicted density $\hat{p}(y;x)$ is evaluated by using the Kullback-Leibler divergence

$$D(p||\hat{p}) := \int p(y|\theta) \log \frac{p(y|\theta)}{\hat{p}(y;x)} \mathrm{d}y.$$

Aitchison [1] showed that a Bayesian predictive density $p_\pi(y|x) := \int_\Theta p(y|\theta)\pi(\theta|x)\mathrm{d}\theta$, where $\pi(\theta|x)$ is a posterior distribution, is the best predictive density when we evaluate a predictive density $\hat{p}(y;x)$ by using the average Kullback-Leibler divergence $\int \pi(\theta) \int D(p||\hat{p})p(x|\theta)\mathrm{d}x\mathrm{d}\theta$, where $\pi(\theta)$ is a probability density. Intuitively speaking, if we have some uncertainty on θ, then moderate averaged estimation from the data x is better than one based on a point estimation. We extend this result in classical statistics to the quantum setting and consider the Bayesian prediction problem of the Gaussian states family.

In quantum statistics, problems of statistical inference and state estimation has received a lot of attention over the past several years with recent developments of experimental techniques. Historically speaking, parameter estimation problem on quantum systems dates back to a quarter century, when Helstrom, Holevo, and other researchers vigorously investigated the topic and gave some extension of mathematical statistical concepts on classical probability.

*ftanaka@stat.t.u-tokyo.ac.jp

Bayesian approach for quantum statistics has also been investigated [7, 8]. Jones [9] has derived a quantum Bayes rule for pure states with the uniform prior. Later, Bužek *et al.* [3] pointed out that it can be applied to mixed states with purification ansatz. Schack *et al.* [11] extended his result to a more general framework of exchangeable states. They showed that a quantum state after a measurement can be interpreted as the state averaged over the posterior. Bužek *et al.* [3] recommended to use Bayesian technique especially when the sample size of experimental data is small. They proposed to use a posterior state corresponding to a posterior distribution in classical counterparts.

From the viewpoints of information quantity and Bayes rule, however, Bayesian estimation on quantum states has not been fully discussed. Performances of the Bayesian approach compared with other approach such as the maximal likelihood method have not been discussed theoretically. Tanaka and Komaki showed that the Bayesian method has better performance than the plug-in method when exchangeable states are considered [13]. In the present paper, we review it and calculate the Bayesian predictive density operator for the Gaussian states family with the heterodyne measurement.

2 Preliminary

We briefly summarize some notations of quantum measurement. Let $\mathcal{H}$ be a separable (possibly infinite dimensional) Hilbert space of a quantum system. An Hermitian operator ρ on $\mathcal{H}$ is called a *state* or *density operator* if it satisfies,

$$\mathrm{Tr}\rho = 1, \quad \rho \geq 0.$$

We denote the set of all states on $\mathcal{H}$ as $\mathcal{S}(\mathcal{H})$.

Let Ω be a space of all possible outcomes of an experiment (e.g., $\Omega = \mathbf{R}^n$) and suppose that a σ-algebra $\mathcal{B} := \mathcal{B}(\Omega)$ of subsets of Ω is given. An affine map μ from $\mathcal{S}(\mathcal{H})$ into a set of probability distributions on Ω, $\mathcal{P}= \{\mu(\mathrm{d}x)\}$ is called a *measurement*. There is a one-to-one correspondence between a measurement and a resolution of the identity [8]. A map from $\mathcal{B}$ into the set of positive

Hermitian operators

$$M : B \mapsto M(B),$$

where M satisfies

$$M(\phi) = O, M(\Omega) = I,$$
$$M(\cup_i B_i) = \sum_i M(B_i), \quad B_i \cap B_j = \phi, \quad \forall B_i \in \mathcal{B},$$

is called a *positive operator valued measure* (*POVM*). Any physical measurement can be represented by a POVM.

The rule describing a post-measurement state is as follows (e.g., Kraus [5], Nielsen and Chuang [10]). For simplicity, we consider only discrete outcome cases, where Ω is a countable set. Then, a family of linear operators $\{A_x\}$ satisfying

$$\sum_{x \in \Omega} A_x^* A_x = I$$

describes a measurement when considering $\{M_x = A_x^* A_x\}$ as POVM. Performing such a measurement for an arbitrarily fixed ρ yields an outcome x with probability $p_x := \mathrm{Tr}\rho M_x = \mathrm{Tr}\rho A_x^* A_x$ and the quantum state ρ changes to

$$\frac{A_x \rho A_x^*}{p_x}$$

after the outcome x is observed.

Now we describe our setting of state estimation. Assume that a state ρ_θ on $\mathcal{H}$ is characterized by an unknown finite-dimensional parameter $\theta \in \Theta \subset \mathbf{R}^p$.

A quantum state for n systems, $\rho^{(n)}$, is described on the n-fold tensor product Hilbert space $\mathcal{H}^{\otimes n}$. Suppose that a system composed of $n+m$ subsystems is given and that a measurement is performed only for selected n subsystems with the other m subsystems left. Then, the measurement is described by $\{A_x \otimes I\}$, where $\{A_x\}$ is a family of linear operators on $\mathcal{H}^{\otimes n}$ such that $\{M_x := A_x^* A_x\}$ is a POVM and I is the identity operator on $\mathcal{H}^{\otimes m}$.

Our aim is to estimate the true state $\sigma_\theta := \rho_\theta^{\otimes m}$ of the remaining m subsystems by using a measurement $\{M_x\}$ on the selected n subsystems $\rho_\theta^{\otimes n}$. We fix an arbitrarily chosen measurement. Note that M is given as a POVM on $\mathcal{H}^{\otimes n}$. It is not necessarily in the form of a tensor product $M_x^{\otimes n}$, which represents a repetition of the same measurement M_x for each system. Thus, all possible measurements on n subsystems, which may use entanglement, are considered.

The performance of a predictive density operator $\hat{\sigma}(x)$ is evaluated by the relative entropy $D(\sigma_\theta || \hat{\sigma}(x))$, a quantum analogue of the Kullback-Leibler divergence in classical statistics. The quantum relative entropy from ρ to σ is defined by

$$D(\rho||\sigma) := \mathrm{Tr}[\rho(\log \rho - \log \sigma)]. \tag{1}$$

It satisfies the positivity condition $D(\rho||\sigma) \geq 0$ and $D(\rho||\sigma) = 0 \Leftrightarrow \rho = \sigma$. Thus, it can be used as a measure for the goodness of state estimation.

There are mainly two approaches on inference of state σ_θ for the parametric model above. One approach is to use $\sigma_{\hat{\theta}(x)}$, where $\hat{\theta}(x)$ is an estimator of θ, depending on the observation x. The other approach corresponds to the Bayesian predictive density approach in classical statistics [9, 3]. We shall briefly review the idea. First, we assume a probability density $\pi(\theta)$ on the parameter space. In mathematical statistics $\pi(\theta)$ is usually called a *prior density*. When there is no knowledge about parameter θ, which is often called *noninformative*, several people have discussed what kind of prior should be used [12], [2]. From the data x obtained from a measurement $\{M_x\}$, a posterior distribution $\pi(\theta|x)$ is constructed as

$$\pi(\theta|x) := \frac{p^M(x|\theta)\pi(\theta)}{\int \mathrm{d}\theta \; p^M(x|\theta)\pi(\theta)},$$

where $p^M(x|\theta) = \mathrm{Tr}\rho_\theta^{\otimes n} M_x$. Next, taking an average of σ_θ with $\pi(\theta|x)$, one can obtain the Bayesian estimator

$$\sigma_\pi(x) = \int \mathrm{d}\theta \; \sigma_\theta \pi(\theta|x).$$

We call this state estimator, as in classical statistics, a *Bayesian predictive density operator*. In order to distinguish two estimators we call $\sigma_{\hat{\theta}}$, an estimator based on $\hat{\theta}$, a *plug-in predictive density operator*. In the next section, we show that Bayesian predictive density operators are better than plug-in predictive density operators.

If we assume a prior probability density $\pi(\theta)$ on the parameter space Θ, the mixture state is given by

$$\rho^{(n)} := \int \mathrm{d}\theta \; \pi(\theta) \; \rho_\theta^{\otimes n}. \tag{2}$$

A state of the form (2) is called an *exchangeable state* [11], and arises, e.g., if each subsystem is prepared in the same unknown way, as in quantum state tomography.

In a quantum exchangeable model (2), as Schack *et al.* [11] showed, a posterior distribution $\pi(\theta|x)$ naturally arises. As described above, a post-measurement state with outcome x obtained is given by

$$\rho_x^{(n+m)} = \frac{1}{p_x}[(A_x \otimes I)\left(\int \mathrm{d}\theta \; \pi(\theta)\rho_\theta^{\otimes(n+m)}\right)(A_x^* \otimes I)].$$

After the measurement of the selected n subsystems, we restrict our attention only to the remaining m subsystems. Taking a partial trace, we obtain the resulting state ρ_x^m on $\mathcal{H}^{\otimes m}$ (for partial trace, see, e.g., [10]).

The final state ρ_x^m can be rewritten using a posterior $\pi(\theta|x)$ in the form of exchangeable model [11].

$$\begin{aligned}
\rho_x^m = \mathrm{Tr}_n[\rho_x^{(n+m)}] &= \frac{1}{p_x}\int \mathrm{d}\theta \; \pi(\theta)\mathrm{Tr}_n[\rho_\theta^{\otimes(n+m)} M_x \otimes I] \\
&= \frac{1}{p_x}\int \mathrm{d}\theta \; \pi(\theta) p^M(x|\theta)\rho_\theta^{\otimes m} \\
&= \int \mathrm{d}\theta \; \pi(\theta|x)\rho_\theta^{\otimes m},
\end{aligned}$$

where $p_x := \mathrm{Tr}[\rho^{(n+m)}(M_x \otimes I)] = \int \mathrm{d}\theta \; \pi(\theta) p^M(x|\theta)$. Thus, one can interpret $\pi(\theta|x)$ as a quantum analogue of the posterior distribution in classical statistics.

Now we consider comparing two methods for estimating the true state $\sigma_\theta \in \mathcal{S}(\mathcal{H}^{\otimes m})$. Let $\hat{\sigma}(x)$ and $\tilde{\sigma}(x)$ be two predictive density operators. When the difference between two estimates $\hat{\sigma}(x)$ and $\tilde{\sigma}(x) \in \mathcal{S}(\mathcal{H}^{\otimes m})$

$$D(\sigma_\theta||\hat{\sigma}(x)) - D(\sigma_\theta||\tilde{\sigma}(x)) = \mathrm{Tr}[\sigma_\theta(\log\tilde{\sigma}(x) - \log\hat{\sigma}(x))] \tag{3}$$

is positive, $\hat{\sigma}(x)$ is better than $\tilde{\sigma}(x)$ as an estimate of the true state σ_θ. Since $\hat{\sigma}(x)$ and $\tilde{\sigma}(x)$ depend on observed data x for arbitrarily chosen measurement $\{M_x\}$ on $\mathcal{H}^{\otimes n}$, the difference (3) depends on the true parameter value θ characterizing the true state and on the data x obtained from the measurement. Thus, we take an average of (3) over $p^M(x|\theta) := \mathrm{Tr}\sigma_\theta M_x$ and $\pi(\theta)$, and evaluate $\mathrm{E}^\pi\mathrm{E}^M[D(\sigma_\theta||\hat{\sigma}(x)) - D(\sigma_\theta||\tilde{\sigma}(x))]$, where E^π and E^M denote taking an expectation with respect to $\pi(\theta)$ and $p^M(x|\theta)$. In the next section we compare plug-in predictive density operators with Bayesian predictive density operators using this quantity.

3 Bayesian predictive density operators

In classical statistics, Aitchison [1] showed that the Bayesian predictive density $p_\pi(y|x)$ has better performance under the Kullback-Leibler divergence than any plug-in predictive density $p(y|\hat{\theta})$ when a proper prior $\pi(\theta)$ is given. Tanaka and Komaki derived the corresponding result for quantum predictive density operators.[13]

Theorem 1 *Suppose that we perform a measurement for selected n subsystems $\rho_\theta^{\otimes n}$ of a system $\rho_\theta^{\otimes(n+m)}$ composed of $n+m$ subsystems in order to estimate the remaining m subsystems $\sigma_\theta = \rho_\theta^{\otimes m}$. The true parameter value θ is unknown and a prior probability density $\pi(\theta)$ is assumed. Let $\hat{\sigma}(x)$ be any predictive density operator, where x is an outcome of a measurement $\{M_x\}$ for the n subsystems. Performance of a predictive density operator $\hat{\sigma}(x)$ is measured with the average relative entropy*

$$\mathrm{E}^\pi\mathrm{E}^M[D(\sigma_\theta||\hat{\sigma}(x))] = \int \mathrm{d}\theta\ \pi(\theta)\int \mathrm{d}x\ p^M(x|\theta)D(\sigma_\theta||\hat{\sigma}(x))$$

from the true state σ_θ. Then, the Bayesian predictive density operator $\sigma_\pi(x)$ based on the observation x and the prior $\pi(\theta)$ is the best predictive density operator.

Proof.

First of all, for arbitrary $\hat{\sigma}(x), \tilde{\sigma}(x)$, we rewrite the difference of two average relative entropy.

$$\begin{aligned}
&\mathrm{E}^\pi\mathrm{E}^M[D(\sigma_\theta||\hat{\sigma}(x)) - D(\sigma_\theta||\tilde{\sigma}(x))] \\
&= \int \mathrm{d}\theta\ \pi(\theta)\int \mathrm{d}x\ p^M(x|\theta)\mathrm{Tr}[\sigma_\theta(\log\tilde{\sigma}(x) - \log\hat{\sigma}(x))] \\
&= \int \mathrm{d}x\ p_x\int \mathrm{d}\theta\ \pi(\theta|x)\mathrm{Tr}[\sigma_\theta(\log\tilde{\sigma}(x) - \log\hat{\sigma}(x))] \\
&= \int \mathrm{d}x\ p_x\mathrm{Tr}[\sigma_\pi(x)(\log\tilde{\sigma}(x) - \log\hat{\sigma}(x))].
\end{aligned}$$

The positivity of the above form indicates that $\tilde{\sigma}(x)$ is better than $\hat{\sigma}(x)$. We set $\tilde{\sigma}(x) = \sigma_\pi(x)$,then we obtain

$$\begin{aligned}
&\mathrm{E}^\pi\mathrm{E}^M[D(\sigma_\theta||\hat{\sigma}(x)) - D(\sigma_\theta||\sigma_\pi(x))] \\
&= \int \mathrm{d}x\ p_x D(\sigma_\pi(x)||\hat{\sigma}(x)) \geq 0.
\end{aligned}$$

The last inequality holds due to the positivity of the relative entropy $D(\sigma||\sigma') \geq 0$ and $p_x \geq 0$. Since $\hat{\sigma}(x)$ is arbitrarily chosen, it is shown that $\sigma_\pi(x)$ is better than any other $\hat{\sigma}(x)$.

Remark. Our argument holds when a prepared state is described as $\rho_\theta \otimes \sigma_\theta$, where $\rho_\theta \in \mathcal{S}(\mathcal{H})$ and $\sigma_\theta \in \mathcal{S}(\mathcal{K})$, and $\mathcal{H}$ and $\mathcal{K}$ are distinct Hilbert spaces. This setting is a generalization of that introduced in Section 2.

In different setting, Krattenthaler and Slater obtained similar result [4]. While they consider a prior density $\pi(\theta)$ with respect to an unknown state, we consider a posterior density $\pi(\theta|x)$ with respect to a post-measurement state.

4 Prediction of unknown Gaussian state from one sample

We consider the prediction problem of the Gaussian states family below (See, e.g., Holevo [8] for the Gaussian states family).

$$\begin{aligned}
\mathcal{M} &:= \{\rho_{\theta,N} : \theta \in \mathbf{C}\}, \\
\rho_{\theta,N} &:= \frac{1}{\pi N}\int_{\mathbf{C}} \exp\left(-\frac{|\alpha-\theta|^2}{N}\right)|\alpha\rangle\langle\alpha|\mathrm{d}^2\alpha
\end{aligned} \tag{4}$$

and assuming that the photon expectation parameter $N(>0)$ is known. We omit N unless otherwise necessary.

The parameter estimation problem of the model (4) was investigated by Yuen and Lax [14] and Holevo [8]. They obtain the Cramér-Rao type bound, i.e., the lower bound of the trace of the mean square error matrix with an arbitrary weight matrix, based on the RLD Fisher information matrix. They showed that the heterodyne measurement $\{\frac{|\alpha\rangle\langle\alpha|}{\pi}\}$ achieves the bound and it is optimal. This measurement is optimal also in an asymptotic sense, which was shown by Hayashi [6].

Here, we consider the prediction problem in the Bayesian framework. Assume that unknown parameter θ is distributed subject to $\pi(\theta) = \frac{1}{2\pi\tau^2}\exp\left(-\frac{|\theta-\xi|^2}{2\tau^2}\right)$, where $\xi \in \mathbf{C}, \tau^2 > 0$ are so-called hyperparameter.

Although one can perform a measurement for n subsystems $\rho_\theta^{\otimes n}$ and predict the remaining m subsystems, we consider $n = m = 1$ case for simplicity. When $n = 1$, it is natural to adopt the heterodyne measurement above. Then the estimator of θ is given by $\hat{\theta}(\alpha) = \alpha$, where the measurement outcome α is distributed by

$$\alpha \sim p^M(\alpha|\theta) = \frac{1}{\pi(N+1)}\exp\left(-\frac{|\alpha-\theta|^2}{N+1}\right).$$

We calculate the average relative entropy for two predictive density operator $\rho_{\hat{\theta}}$ and $\hat{\rho}_\pi$. Straightforward cal-

culation yields

$$\rho_{\hat{\theta}(\alpha)} = \frac{1}{\pi N}\int_{\mathbf{C}} \exp\left(-\frac{|\beta-\alpha|^2}{N}\right)|\beta\rangle\langle\beta|\mathrm{d}^2\beta,$$

$$\hat{\rho}_\pi(\alpha) = \frac{1}{\pi(N+2\Delta^2)}\int_{\mathbf{C}} \exp\left(-\frac{|\beta-\bar{\theta}|^2}{N+2\Delta^2}\right)|\beta\rangle\langle\beta|\mathrm{d}^2\beta,$$

where

$$\bar{\theta} := \frac{\left(\frac{N+1}{2}\right)^{-1}\alpha + (\tau^2)^{-1}\xi}{\left(\frac{N+1}{2}\right)^{-1}+(\tau^2)^{-1}},\ (\Delta^2)^{-1} := \left(\frac{N+1}{2}\right)^{-1}+(\tau^2)^{-1}. \tag{5}$$

The average relative entropy for them is also obtained by

$$\mathcal{R}_p := E^\pi E^M[D(\rho_\theta||\rho_{\hat{\theta}})] = (N+1)\log\left(\frac{N+1}{N}\right),$$

$$\begin{aligned}\mathcal{R}_\pi &:= E^\pi E^M[D(\rho_\theta||\hat{\rho}_\pi)] \\ &= \log\frac{1}{N+1} + N\log\frac{N}{N+1} \\ &\quad - \log\frac{1}{N+2\Delta^2+1} - (N+2\Delta^2)\log\frac{N+2\Delta^2}{N+2\Delta^2+1},\end{aligned}$$

where we used the formula for the Gaussian states family

$$\begin{aligned}D(\rho_{\zeta,N}||\rho_{\zeta',M}) &= \log\left(\frac{M+1}{N+1}\right) + N\log\left(\frac{N}{N+1}\frac{M+1}{M}\right) \\ &\quad + \log\left(\frac{M+1}{M}\right)|\zeta-\zeta'|^2.\end{aligned}$$

Since $\mathcal{R}_\pi$ is monotone increasing with τ^2,

$$\begin{aligned}\sup_{\tau^2>0}\mathcal{R}_\pi &= \lim_{\tau^2\to\infty}\mathcal{R}_\pi \\ &= \log\frac{1}{N+1} + N\log\frac{N}{N+1} \\ &\quad - \log\frac{1}{2N+2} - (2N+1)\log\frac{2N+1}{2N+2} \\ &\equiv \mathcal{R}_*.\end{aligned}$$

In addition, from the straightforward calculation we can show $\mathcal{R}_p > \mathcal{R}_* \geq \mathcal{R}_\pi$. Thus, it is shown that the Bayesian predictive density operator $\hat{\rho}_\pi$ is better than the plug-in density operator $\rho_{\hat{\theta}}$ based on $\hat{\theta}(\alpha)$.

5 Concluding remarks

Since the model $\mathcal{M}$ is translation invariant, it seems natural to adopt the Lebesgue measure $\pi_J(\theta)\mathrm{d}^2\theta \propto \mathrm{d}^2\theta$ as a noninformative prior. Although $\int \pi_J(\theta)\mathrm{d}^2\theta = \infty$, as classical statistics, various quantities are obtained by taking the limit $\tau^2 \to \infty$. Since $2\Delta^2 = N+1$, Bayesian predictive density operator is given by

$$\hat{\rho}_{\pi_J} = \frac{1}{\pi(2N+1)}\int_{\mathbf{C}} \exp\left(-\frac{|\beta-\bar{\theta}|^2}{2N+1}\right)|\beta\rangle\langle\beta|\mathrm{d}^2\beta,$$

and the average relative entropy is equal to $\mathcal{R}_*(<\infty)$.

Strictly speaking, the proof of theorem 1 is valid only for finite-dimensional cases (i.e., $\dim\mathcal{H} < \infty$). However, from the result above, we expect that it holds even when $\dim\mathcal{H} = \infty$ under some regularity conditions such as the exchangeability of the order of Tr and $\int \mathrm{d}\theta\ \pi(\theta)$ and integrability of $\rho_\pi(x) = \int \mathrm{d}\theta\ \pi(\theta|x)\rho_\theta$.

Acknowledgments F.T. was supported by the JSPS Research Fellowships for Young Scientists.

References

[1] J. Aitchison. Goodness of prediction fit. *Biometrika*, 62: 547–554, 1975.

[2] S .L. Braunstein and C. M. Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72: 3439–3443, 1994.

[3] V. Bužek and R. Derka and G. Adam and P. L. Knight. Reconstruction of quantum states of spin systems: from quantum Bayesian inference to quantum tomography. *Ann. Phys. (N.Y.)*, 266: 454–496, 1998.

[4] C. Krattenthaler and P. B. Slater. Asymptotic Redundancies for Universal Quantum Coding. *IEEE Trans. Info. Theor.*, 46: 801, 2000.

[5] K. Kraus. *States, Effects, and Operations. Fundamental Notions of Quantum Theory.* Springer-Verlag, 1983.

[6] M. Hayashi. Asymptotic quantum theory for the thermal states family. In *Quantum Communication, Computing and Measurement 2.* Kluwer/Plenum, 2000, 99–104.

[7] C. W. Helstrom. *Quantum Detection Theory.* Academic Press, 1976.

[8] S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory.* North-Holland, 1982.

[9] K. R. W. Jones. Principles of quantum inference. *Ann. Phys. (N.Y.)*, 207: 140–170, 1991.

[10] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

[11] R. Schack, T. Brun, and C. Caves. Quantum Bayes rule. *Phys. Rev. A*, 64: 014305, 2001.

[12] P. B. Slater. Information gains expected from separate and joint measurements of N identical spin-1/2 systems: Noninformative Bayesian analyses. *J. Math. Phys.*, 38: 2274, 1997.

[13] F. Tanaka and F. Komaki. Bayesian predictive density operators for exchangeable quantum-statistical models. *Phys. Rev. A*, 71: 052323, 2005.

[14] H. P. Yuen and M. Lax: Multiple-parameter quantum estimation and measurement of nonselfadjoint observables. *IEEE Trans. Inform. Theory*, 19: 740–750, 1973.

Differential geometrical aspects of quantum estimation theory — On the geometry of generalized RLD metric —

Hiroshi Nagaoka[1] *

[1] *Graduate School of Information Systems,*
The University of Electro-Communications,
1-5-1, Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan.

Abstract. The SLD (symmetric logarithmic derivative) and RLD (right logarithmic derivative) are very important in quantum estimation theory. They are also interesting in that they give two extremes of monotone Riemannian metrics. Compared with the SLD, however, the geometrical position of RLD is not so clear. The aim of the present article is to provide a general framework to treat the geometry of RLD, where the complexified cotangent space plays an essential role rather than the (real) tangent space.

Keywords: quantum estimation, differential geometry, SLD, RLD, monotone metric

1 Introduction

Let $M = \{\rho_\theta \,|\, \theta = (\theta^1, \cdots, \theta^m) \in \Theta\}$ be a family of faithful (i.e., strictly positive) density operators smoothly parametrized by an m-dimensional parameter θ ranging over an open set Θ in $\mathbb{R}^m$. We consider M as a smooth manifold with the coordinate system $\theta = [\theta^i]$. For each point ρ_θ, the symmetric logarithmic derivatives (SLDs) $L_i^{\rm S} = L_{i,\theta}^{\rm S}$ $(i = 1, \ldots, m)$ are defined by

$$\partial_i \rho_\theta = \frac{1}{2}(\rho_\theta L_i^{\rm S} + L_i^{\rm S} \rho_\theta), \tag{1}$$

where $\partial_i := \frac{\partial}{\partial \theta^i}$, and the SLD Fisher information matrix $G_\theta^{\rm S} = [g_{ij}^{\rm S}(\theta)]$ is defined by

$$g_{ij}^{\rm S}(\theta) := {\rm Re\ tr}[\rho_\theta L_i^{\rm S} L_j^{\rm S}]. \tag{2}$$

Similarly, the right logarithmic derivatives (RLDs) $\tilde{L}_i^{\rm R} = \tilde{L}_{i,\theta}^{\rm R}$ and the RLD Fisher information matrix $\tilde{G}_\theta^{\rm R} = [\tilde{g}_{ij}^{\rm R}(\theta)]$ are defined by

$$\partial_i \rho_\theta = \rho_\theta \tilde{L}_i^{\rm R}, \tag{3}$$

$$\tilde{g}_{ij}^{\rm R}(\theta) = {\rm tr}[\rho_\theta \tilde{L}_j^{\rm R} (\tilde{L}_i^{\rm R})^*]. \tag{4}$$

The inverse matrices of G_θ^S and $\tilde{G}_\theta^{\rm R}$ give lower bounds for the variance-covariance matrix of any unbiased estimator (unbiased measurement) of the parameter θ (: quantum versions of the Cramér-Rao inequality [1, 2, 3]). We also introduce the real part of the RLD Fisher information matrix:

$$G_\theta^{\rm R} = [g_{ij}^{\rm R}(\theta)], \quad g_{ij}^{\rm R}(\theta) = {\rm Re}\ \tilde{g}_{ij}^{\rm R}(\theta).$$

From a differential geometrical viewpoint, these quantities define two Riemannian metrics $g^{\rm S}$ and $g^{\rm R}$ on the manifold M by $g^{\rm S}(\partial_i, \partial_j) = g_{ij}^{\rm S}$ and $g^{\rm R}(\partial_i, \partial_j) = g_{ij}^{\rm R}$. Moreover, these metrics are known to be the minimum and maximum, respectively, among the class of normalized monotone Riemannian metrics [4]. Here, the monotonicity of a Riemannian metric[1] g on the quantum state space means that the length of a tangent vector measured by g does not increase under any quantum operation (TPCP map), and the normality of g means that it coincides with the classical Fisher information on a manifold of mutually commutative states[2]. The original RLD Fisher information matrix also defines a "complex monotone metric" $\tilde{g}^{\rm R}$ by $\tilde{g}^{\rm R}(\partial_i, \partial_j) = \tilde{g}_{ij}^{\rm R}$, although it is not a Riemannian metric in the usual sense.

A remarkable difference between SLD and RLD is that $L_i^{\rm S}$ and $G_\theta^{\rm S}$ are definable for a manifold of non-faithful states, including pure states in particular, while $\tilde{L}_i^{\rm R}$ and $\tilde{G}_\theta^{\rm R}$ are not. Nevertheless, a substitute for $(\tilde{G}_\theta^{\rm R})^{-1}$ is properly defined even when $\tilde{G}_\theta^{\rm R}$ does not exist for which a quantum Cramér-Rao inequality holds(; see [5] for coherent models).

The aim of the present article is to provide a general framework to treat the geometry of RLD, where the complexified cotangent space plays an essential role rather than the (real) tangent space.

2 Basic definitions

Let $\mathcal{H}$ be a $d(< \infty)$-dimensional Hilbert space, and denote the totality of linear (hermitian, resp.) operators on $\mathcal{H}$ by $\mathcal{L} = \mathcal{L}(\mathcal{H})$ ($\mathcal{L}_{\rm h} = \mathcal{L}_{\rm h}(\mathcal{H})$, resp.). The set of density operators is denoted by $\bar{\mathcal{S}} = \bar{\mathcal{S}}(\mathcal{H})$, which is decomposed as

$$\bar{\mathcal{S}} = \bigcup_{r=1}^{d} \mathcal{S}_r,$$

where $\mathcal{S}_r = \{\rho \in \bar{\mathcal{S}} \,|\, {\rm rank}\rho = r\}$. Note that $\mathcal{S}_r$ is naturally regarded as a smooth manifold with $\dim \mathcal{S}_r = 2dr - r^2 - 1$. In particular, $\mathcal{S}_1$ is the set of pure states, which can be identified with the complex projective space $P^{d-1}(\mathbb{C})$, and $\mathcal{S}_d$ is the set of faithful states.

Let a state $\rho \in \mathcal{S}_r$ be arbitrarily fixed, and let $T_\rho = T_\rho(\mathcal{S}_r)$ be the tangent space of $\mathcal{S}_r$ at ρ, which is a real linear space of dimension $n := 2dr - r^2 - 1$. Given a local coordinate system $[\theta^i]$ of $\mathcal{S}_r$ around ρ, we have $T_\rho = {\rm span}_{\mathbb{R}}\{(\frac{\partial}{\partial \theta^i})_\rho\}_{i=1}^n$.

*nagaoka@is.uec.ac.jp

[1] More precisely, the monotonicity is meaningful not for a Riemannian metric on a single manifold but for a way of assigning a Riemannian metric to an arbitrary manifold of quantum states.

[2] Every monotone metric is shown to be a constant multiple of the classical Fisher information on a manifold of commutative states.

The complexification of T_ρ is denoted by $T_\rho^{\mathbb{C}} = T_\rho \otimes \mathbb{C} = T_\rho \oplus \sqrt{-1}T_\rho = \mathrm{span}_{\mathbb{C}}\{(\frac{\partial}{\partial\theta^i})_\rho\}_{i=1}^n$ which is an n-dimensional complex linear space. The complex conjugate of $x + \sqrt{-1}y \in T_\rho^{\mathbb{C}}$ with $x, y \in T_\rho$ is defined by $\overline{x + \sqrt{-1}y} = x - \sqrt{-1}y$ and we have $T_\rho = \{u \in T_\rho^{\mathbb{C}} \,|\, \bar{u} = u\}$.

We denote the cotangent space at ρ and its complexification by $T_\rho^* = \mathrm{span}_{\mathbb{R}}\{(\mathrm{d}\theta^i)_\rho\}_{i=1}^n$ and $T_\rho^{*\mathbb{C}} = T_\rho^* \otimes \mathbb{C} = T_\rho^* \oplus \sqrt{-1}T_\rho^* = \mathrm{span}_{\mathbb{C}}\{(\mathrm{d}\theta^i)_\rho\}_{i=1}^n$, respectively. T_ρ^* is the set of real linear functionals : $T_\rho \to \mathbb{R}$, while $T_\rho^{*\mathbb{C}}$ is the set of complex linear functionals : $T_\rho^{\mathbb{C}} \to \mathbb{C}$. The complex conjugate $\bar{\xi}$ of $\xi \in T_\rho^{*\mathbb{C}}$ is similarly defined.

We sometimes write as $T_\rho = T_\rho^{\mathbb{R}}$ and $T_\rho^* = T_\rho^{*\mathbb{R}}$ to distinguish them from their complexifications.

3 (Pre-)inner products on the cotangent space

Define $\langle A\rangle_\rho := \mathrm{tr}[\rho A]$ for a state $\rho \in \bar{\mathcal{S}}$ and an operator $A \in \mathcal{L}(\mathcal{H})$. Considering $\langle A\rangle : \rho \mapsto \langle A\rangle_\rho$ as a function on $\mathcal{S}_r$, its differential $(\mathrm{d}\langle A\rangle)_\rho$ at a point $\rho \in \mathcal{S}_r$ is defined as an element of $T_\rho^{*\mathbb{C}}$, which we denote by $\Delta_\rho(A) := (\mathrm{d}\langle A\rangle)_\rho$. Then we can represent $T_\rho^{*\mathbb{C}}$ and $T_\rho^{*\mathbb{R}}$ by the image of Δ_ρ as

$$T_\rho^{*\mathbb{C}} = \Delta_\rho(\mathcal{L}) = \Delta_\rho(\mathcal{L}_{0,\rho}), \tag{5}$$

$$T_\rho^{*\mathbb{R}} = \Delta_\rho(\mathcal{L}_{\mathrm{h}}) = \Delta_\rho(\mathcal{L}_{\mathrm{h}0,\rho}), \tag{6}$$

where $\mathcal{L}_{0,\rho} := \{A \in \mathcal{L} \,|\, \mathrm{tr}[\rho A] = 0\}$ and $\mathcal{L}_{\mathrm{h}0,\rho} := \mathcal{L}_{0,\rho} \cap \mathcal{L}_{\mathrm{h}}$.

Following [2], let us define

$$\langle A, B\rangle_\rho^+ := \mathrm{tr}[\rho B A^*], \quad \langle A, B\rangle_\rho^- := \mathrm{tr}[\rho A^* B]$$
$$\text{and} \quad \langle A, B\rangle_\rho := \frac{1}{2}\left(\langle A, B\rangle_\rho^+ + \langle A, B\rangle_\rho^-\right)$$

for $A, B \in \mathcal{L}$. When $r = d$, or equivalently when ρ is faithful, these are complex inner products on $\mathcal{L}$. Otherwise, they are pre-inner products with non-trivial kernels $\mathcal{K}_\rho^\pm := \{A \in \mathcal{L} \,|\, \langle A, A\rangle_\rho^\pm = 0\}$ and $\mathcal{K}_\rho := \{A \in \mathcal{L} \,|\, \langle A, A\rangle_\rho = 0\}$. Note that $\langle\cdot,\cdot\rangle_\rho$ is real on $\mathcal{L}_{\mathrm{h}}$. Since we have

$$\Delta_\rho(A) = 0 \text{ iff } \langle A, A\rangle_\rho = 0,$$

for $A \in \mathcal{L}_{0,\rho}$, it follows from (5) and (6) that

$$T_\rho^{*\mathbb{C}} \simeq \mathcal{L}_{0,\rho}/\mathcal{K}_{0,\rho}, \quad T_\rho^{*\mathbb{R}} \simeq \mathcal{L}_{\mathrm{h}0,\rho}/\mathcal{K}_{\mathrm{h}0,\rho}, \tag{7}$$

where $\mathcal{K}_{0,\rho} = \mathcal{K}_\rho \cap \mathcal{L}_{0,\rho}$ and $\mathcal{K}_{\mathrm{h}0,\rho} = \mathcal{K}_\rho \cap \mathcal{L}_{\mathrm{h}0,\rho}$.

From (7), $\langle\cdot,\cdot\rangle_\rho$ defines a complex inner product on $T_\rho^{*\mathbb{C}}$ and a real inner product on $T_\rho^{*\mathbb{R}}$, the former of which is the complexification of the latter. We denote both of them by g. On the other hand, since $\mathcal{K}_\rho \subset \mathcal{K}_\rho^\pm$, a pair of complex pre-inner products are defined on $T_\rho^{*\mathbb{C}}$ by $\langle\cdot,\cdot\rangle_\rho^\pm$, which we denote by $\tilde{g}^\pm$. Note that $\tilde{g}^+(\xi_1, \xi_2) = \tilde{g}^-(\bar{\xi}_2, \bar{\xi}_1)$ holds for $\xi_1, \xi_2 \in T_\rho^{*\mathbb{C}}$.

4 The SLD and RLD metrics

The inner product g establishes an $\mathbb{R}$-linear isomorphism between T_ρ^* and T_ρ and induces an inner product on T_ρ, which is shown to coincide with the SLD metric g^{S}. This means that g and g^{S} are essentially the same thing, and we can express this fact as $g = g^{\mathrm{S}}$. Similarly, when $r = d$, $\tilde{g}^+$ establishes an anti-linear isomorphism between $T_\rho^{*\mathbb{C}}$ and $T_\rho^{\mathbb{C}}$ and induces an inner product on $T_\rho^{\mathbb{C}}$. This is shown to be the complex RLD metric $\tilde{g}^{\mathrm{R}}$, which we express by $\tilde{g}^+ = \tilde{g}^{\mathrm{R}}$.

5 Generalized RLD metric and its monotonicity

In the general case when the equality $r = d$ does not necessarily hold, we regard $\tilde{g}^+$ as a generalization of the complex RLD metric, although this is a singular pre-inner product on $T_\rho^{*\mathbb{C}}$ and does not induce a (pre-)inner product on $T_\rho^{\mathbb{C}}$.

The monotonicity of a Riemannian metric g' is usually expressed in terms of tangent vectors as [4]

$$g'_\rho(u, u) \geq g'_{\varphi(\rho)}(\varphi_*(u), \varphi_*(u)), \tag{8}$$

where φ is the state change defined by a quantum operation (TPCP map), $\varphi_* = (\mathrm{d}\varphi)_\rho$ is its differential at ρ and u is a tangent vector at ρ. It is equivalently rewritten in terms of cotangent vectors as

$$g'_\rho(\varphi^*(\xi), \varphi^*(\xi)) \leq g'_{\varphi(\rho)}(\xi, \xi), \tag{9}$$

where φ^* is the transposed map of φ_* and ξ is a cotangent vector at $\varphi(\rho)$. The generalized RLD metric $\tilde{g}^+$ and its conjugate $\tilde{g}^-$ satisfy the latter form of monotonicity[3]:

$$\tilde{g}^\pm(\varphi^*(\xi), \varphi^*(\xi)) \leq \tilde{g}^\pm_{\varphi(\rho)}(\xi, \xi), \quad \forall \xi \in T^{*\mathbb{C}}_{\varphi(\rho)}, \tag{10}$$

while the monotonicity corresponding to (8) is meaningless in general.

6 Generalized RLD for tangent vectors

Let

$$\check{T}_\rho^+ := \left\{u \in T_\rho^{\mathbb{C}} \,\middle|\, \exists A \in \mathcal{L},\ \partial_u\rho = \rho A\right\},$$
$$\check{T}_\rho^- := \left\{u \in T_\rho^{\mathbb{C}} \,\middle|\, \exists A \in \mathcal{L},\ \partial_u\rho = A\rho\right\},$$

where $\partial_u\rho := \sum_i c^i \frac{\partial\rho_\theta}{\partial\theta^i} \in \mathcal{L}$ for $u = \sum_i c^i(\frac{\partial}{\partial\theta^i})_\rho$ ($\{c^i\} \subset \mathbb{C}$). $\check{T}_\rho^+$ ($\check{T}_\rho^-$, resp.) is a $\mathbb{C}$-linear subspace of $T_\rho^{\mathbb{C}}$ consisting of complexified tangent vectors for which RLD (LLD, left logarithmic derivative, resp.) operators exists. Note that $\check{T}_\rho^\pm \subsetneq T_\rho^{\mathbb{C}}$ unless $r = d$.

Introducing the commutation operator $D : T_\rho^{\mathbb{C}} \to T_\rho^{\mathbb{C}}$ defined by [2]

$$g(u, Dv) = \frac{1}{2\sqrt{-1}}\left(\tilde{g}^+(u, v) - \tilde{g}^-(u, v)\right)$$
$$= \frac{1}{2\sqrt{-1}}\mathrm{tr}[\rho(L_v L_u^* - L_u^* L_v)],$$

where L_u and L_v are the SLDs for $u, v \in T_\rho^{\mathbb{C}}$, and letting $E^\pm := I \pm \sqrt{-1}D$, we have

$$\check{T}_\rho^+ = E^+(T_\rho^{\mathbb{C}}) \text{ and } \check{T}_\rho^- = E^-(T_\rho^{\mathbb{C}}). \tag{11}$$

[3]This is a direct consequence of the Schwarz inequality $\mathcal{F}(AA^*) \leq \mathcal{F}(A)\mathcal{F}(A)^*$ which holds for any unital CP map $\mathcal{F}$.

From this we see that a complex inner product $\dot{g}^+$ on T_ρ^+ and a complex inner product $\dot{g}^-$ on T_ρ^- are defined by

$$\dot{g}^\pm(E^\pm u, E^\pm v) = g(u, E^\pm v).$$

Since $\dot{g}^\pm$ are nonsingular on $T_\rho^\pm$, they define inner products on the dual spaces $(\dot{T}_\rho^+)^*$ and $(\dot{T}_\rho^-)^*$ which are complex linear subspaces of $T_\rho^{*\mathbb{C}}$. We can show that for $\xi_1, \xi_2 \in T_\rho^{*\mathbb{C}}$,

$$g^\pm(\xi_1, \xi_2) = \dot{g}^\pm(\xi_1|_{\dot{T}_\rho^\pm}, \xi_2|_{\dot{T}_\rho^\pm}). \tag{12}$$

where $\xi_j|_{\dot{T}_\rho^\pm} \in (\dot{T}_\rho^\pm)^*$ is the restriction of $\xi_j : T_\rho^{\mathbb{C}} \to \mathbb{C}$ on the subspaces $\dot{T}_\rho^\pm$.

For $u, v \in \dot{T}_\rho^+$, there exit RLDs A, B such that $\partial_u \rho = \rho A$ and $\partial_v \rho = \rho B$, and we have

$$\dot{g}^+(u, v) = \langle A, B \rangle_\rho^+. \tag{13}$$

The original construction of the RLD metric is thus extended.

7 The case of pure states

Let $r = 1$. Then we have $T_\rho^+ \cap T_\rho^{\mathbb{R}} = \{0\}$ and hence there is no RLD for real tangent vectors. In this case, D satisfies $D^2 = -1$, which is nothing but the almost complex structure of the complex manifold $\mathcal{S}_1 = P^{d-1}(\mathbb{C})$, often denoted by J in the references of complex geometry (e.g. [6]). We have

$$\dot{T}_\rho^\pm = \left\{ u \in T_\rho^{\mathbb{C}} \,\middle|\, Du = \mp\sqrt{-1}\, u \right\}, \tag{14}$$

$$T_\rho^{\mathbb{C}} = \dot{T}_\rho^+ \oplus \dot{T}_\rho^-. \tag{15}$$

$\dot{T}_\rho^+$ and $\dot{T}_\rho^-$ are often called the anti-holomorphic and holomorphic subspaces, respectively, and are denoted as $\dot{T}_\rho^+ = T_\rho^{0,1}$ and $\dot{T}_\rho^- = T_\rho^{1,0}$. It is something interesting to see that $\dot{g}^\pm = \frac{1}{2} g|_{\dot{T}_\rho^\pm}$ as

$$\begin{aligned} \dot{g}^\pm(E^\pm u, E^\pm v) = g(u, E^\pm v) &= g\left(\frac{1}{2}(E^+ + E^-)u, E^\pm v\right) \\ &= \frac{1}{2} g(E^\pm u, E^\pm v), \end{aligned}$$

where the last equality follows from the orthogonality of $\dot{T}_\rho^+$ and $\dot{T}_\rho^-$.

References

[1] C.W. Helstrom, *Quantum Detection and Estimation Theory*, Academic Press, New York, 1976.

[2] A.S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, North-Holland, 1982.

[3] H.P. Yuen and M. Lax, "Multiple-parameter quantum estimation and measurement of nonselfadjoint observables," *IEEE Trans. Inform. Theory*, vol.19, 740-750, 1973.

[4] D. Petz, "Monotone metrics on matrix spaces," *Linear Algebra Appl.*, vol.244, 81-96, 1996.

[5] A. Fujiwara, H. Nagaoka, "An estimation theoretical characterization of coherent states," *J. Math. Phys.*, vol.40, No.9, 4227-4239, 1999.

[6] S. Kobayashi, K. Nomizu, *Foundations of Differential Geometry II*, Wiley, 1969.

Ancilla-Assisted Enhancement of Channel Estimation for Low-Noise Parameters

Masahiro Hotta, Tokishiro Karasawa, and Masanao Ozawa
Graduate School of Information Sciences, Tohoku University
Aoba-ku, Sendai, 980-8579, Japan
jidai@ims.is.tohoku.ac.jp

Abstract

In order to make a unified treatment for estimation problems of a very small noise or a very weak signal in a quantum process, we introduce the notion of a low-noise quantum channel with one noise parameter. It is known in several examples that prior entanglement together with nonlocal output measurement improves the performance of the channel estimation. In this paper, we study this "ancilla-assisted enhancement" for estimation of the noise parameter in a general low-noise channel. For channels on two level systems we prove that the enhancement factor, the ratio of the Fisher information of the ancilla-assisted estimation to that of the original one, is always upper bounded by 3/2.

key words : quantum estimation

1 Introduction

One of the formidable obstacles for the realization of quantum computers is decoherence caused by the coupling between computational qubits and the environment. Recent study of quantum error correction has shown that fault-tolerant quantum computing is in principle possible, but it requires that the noise caused by the decoherence should be lower than the very stringent threshold. Obviously, such a statement has a physical meaning only if we have an efficient method for quantitatively estimating very small noise in quantum devices in real experiments. However, if the noise is very small, so is our success probability of observing the disturbance caused by that noise. This difficulty makes evident the demand for the study of optimal quantum estimation of very small nose in general quantum channels based on well-established quantum estimation theory[1,2,3].

A typical problem of quantum estimation is to ask what is the best observable, possibly in an extended system with ancilla, to measure in order to estimate the true value of θ provided that the system is known to be in one of the state in a given family $\{\rho_\theta\}$. A well-established solution for this problem is given as follows. We call an observable A a (locally) unbiased estimator at $\theta = \theta_0$ if the expectation value $E_\theta[A]$ of A in the state ρ_θ satisfies

$$E_{\theta_0}[A] = \theta_0, \tag{1}$$

$$\partial_\theta E_\theta[A]|_{\theta=\theta_0} = 1. \tag{2}$$

In general there are many unbiased estimators. In order to select a good one, we consider the variance $V_\theta[A]$ of an arbitrary unbiased estimator A in the state ρ_θ. Then, the quantum Cramér-Rao inequality

$$V_\theta[A] \geq \frac{1}{J(\rho_\theta)} \tag{3}$$

holds for any unbiased estimator A at θ, where

$$J(\rho_\theta) = \mathrm{Tr}[\rho_\theta L_\theta^2] \tag{4}$$

is the (quantum) Fisher information defined through the symmetric logarithmic derivative (SLD) L_θ that is characterized by the relations

$$\partial_\theta \rho_\theta = \frac{1}{2}(L_\theta \rho_\theta + \rho_\theta L_\theta), \tag{5}$$

$$L_\theta^\dagger = L_\theta. \tag{6}$$

The SLD is determined uniquely on the range of ρ_θ, i.e., $L_\theta \rho_\theta = L'_\theta \rho_\theta$ holds for any two SLDs L_θ and L'_θ. The Cramér-Rao inequality (3) follows from a simple application of the Schwarz inequality for the Hilbert-Schmidt inner product. From the equality condition for that the lower bound J_θ^{-1} in Eq. (3) is always achieved by any observable A satisfying

$$A\rho_\theta = (J_\theta^{-1} L_\theta + \theta)\rho_\theta, \tag{7}$$

see Refs.[1,2] and for a straightforward derivation see Appendix of Ref.[4].

From the quantum estimation theory for state parameters mentioned above, we can construct an estimation theory for unknown parameters of physical processes, such as coupling constants of the interaction. Suppose that we prepare a quantum system in

an initial state ρ_{in} and leave it in an evolution process characterized by an unknown parameter θ. Then, the final state $\rho_{out}(\theta)$ of this process depends on the parameter θ. The problem of finding the optimal estimation of the parameter θ is solved by maximizing the Fisher information J_θ over all the possible initial states ρ_{in} and all the possible observable A in the final state [5,6]. The above physical process can be represented by a mapping Γ_θ that transform the initial state ρ_{in} to the final state ρ_{out} as

$$\rho_{out} = \Gamma_\theta[\rho_{in}]. \tag{8}$$

It is now fairly well-known that every general state change, called a quantum operation or a quantum channel, such as Γ_θ, physically realizable with probability one should be a trace-preserving completely positive (TPCP) mapping, and conversely that every TPCP map can be realized as a unitary process of the system augmented by an ancilla prepared in a fixed state as shown by Kraus [7,8] see also Ref.[9,10] for the generalization of the above statement to generalized measurements and see Ref.[11] for the latest elaboration.

As pointed out in Ref.[12], one can improve the parameter estimation if a correlation, or in particular an entanglement, is allowed between the input system S and an ancilla A. It should be stressed that in doing so one needs no physical process to occur on the ancilla system A while the system S passes through the channel Γ_θ. In this case, the extended channel is represented as $\Gamma_\theta \otimes id_A$, where id_A stands for the identity channel for A. Then, the improvement can be achieved by the initial preparation of the composite system in an entangled state together with the measurement of the composite system after the process.

In this paper, we are devoted to the ancilla-assisted enhancement of Fisher information derived by the quantum Cramér-Rao bound. This enhancement effect not only projects a theoretical profundity of quantum mechanics, but also suggests many physical applications including the low-noise estimation in quantum computing, where the enhanced noise estimation is expected to contribute to devoting the noise reduction technology. Hence it is very significant to estimate the intensity of the low noise. In the estimation, the above ancilla-assisted enhancement may effectively reduce the trial number of the experiment.

For these reasons, we study the estimation theory of the parameter characterizing a small noise in a general quantum channel on a system with finite dimensional state space. We can formulate natural mathematical requirements for the behavior of the low-noise parameter. It is an interesting problem to figure out how much ancilla-assisted enhancement can be achievable in the estimation of the low-noise parameter. In this paper we shall discuss this problem and obtain several upper bounds for this ancilla-assisted enhancement factor in the low-noise parameter estimation.

In a two level system S_2, we obtain a universal upper bound for the enhancement factor η defined by

$$\eta = \frac{\mathcal{L}[\max[J_{S_2+A}]_{\rho_{S_2+A}}]}{\mathcal{L}[\max[J_{S_2}]_{\rho_{S_2}}]} \tag{9}$$

for any finite level ancilla A. Here, ρ_{S_2} is the input in the system S, J_{S_2} is the Fisher information of $\Gamma_\epsilon[\rho_{S_2}]$, ρ_{S_2+A} is the channel input in the composite system $S_2 + A$, J_{S_2+A} is the Fisher information of the output states $(\Gamma_\epsilon \otimes id_A)[\rho_{S_2+A}]$, and $\max[\cdot]_\rho$ stands for the maximum over all the state ρ. The universal upper bound of the enhancement factor η for all the two level systems is given by

$$\eta \leq \frac{3}{2}. \tag{10}$$

This upper bound is attainable by various channels Γ_ϵ, and the corresponding optimal input state is a maximal entangled state, and holds for any low-noise channels on two level systems.

2 Low-Noise Channels

In this section, we introduce the notion of a low-noise channel Γ_ϵ with unknown parameter ϵ, which takes only small values $\epsilon \sim 0$, by requiring a physically natural assumption of the channel Γ_ϵ for the parameter values near $\epsilon = 0$.

As mentioned in the introduction, we will focus on the ancilla extension of the low-noise channel defined by $\Gamma_\epsilon \otimes id_A$. The ancilla-assisted enhancement factor η is also defined as the ratio of the Fisher information of the ancilla-assisted estimation to that of the original one and is analyzed in detail.

The concept of the noise in a quantum process to implement a target unitary process can be understood under the following consideration. Suppose that we would like to implement a unitary channel $\Lambda^{(U)}$ for a system S, so that the output state corresponding to an input state ρ_{in} of S is designed to be

$$\rho_{out} = \Lambda^{(U)}[\rho_{in}] = U\rho_{in}U^\dagger. \tag{11}$$

In real life, however, the system S is coupled weakly with the environment E which causes the decoherence, and the noise is brought from the environment. Assume that the noise is controlled by one unknown positive parameter ν. Then the actual output state ρ'_{out} deviates from the intended output state ρ_{out} due to the noise It is natural to represent the noisy process by a TPCP map Λ_ν such that

$$\rho'_{out} = \Lambda_\nu[\rho_{in}], \tag{12}$$

where the relation $\Lambda_0 = \Lambda^{(U)}$ holds as the noiseless case. In quantum theory, the channel Λ_ν can be equivalently described by a sequence of two channels (the

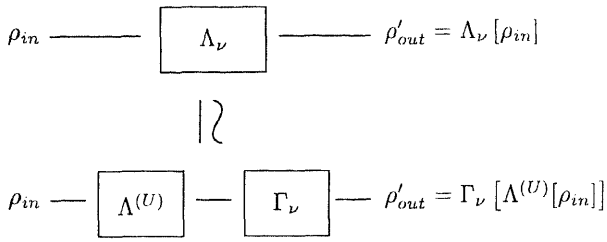

Figure 1: A controlled unitary process usually suffers from noise. The noisy process is described as a TPCP map, a channel Λ_ν, parametrized by one noise parameter ν (the second line). In quantum theory, the disturbed process is equivalently described by a sequence of two channels. The first channel is the originally intended unitary channel $\Lambda^{(U)}$. The second channel Γ_ν describes the genuine noise effect. We call Γ_ν the noise channel (the second line).

second line of Fig. 1). The first one is the target unitary channel $\Lambda^{(U)}$ and the second represents the genuine noise part. This means that the general noisy process is equivalent to the noiseless unitary process followed by an instantaneous noise process. The second channel is called the noise channel Γ_ν and defined by

$$\Gamma_\nu[\rho] := \Lambda_\nu[U^\dagger \rho U] = \Lambda_\nu[(\Lambda^{(U)})^{-1}[\rho]]. \tag{13}$$

When the noise vanishes, the channel reduces to the identity channel:

$$\Gamma_0 = id_S. \tag{14}$$

It is stressed that despite that the noise channel Γ_ν is conceptual constituent, it can be simulated in a real experiment by use of the actual channel Λ_ν. In practice, by adopting a known state $\rho'_{in} = (\Lambda^{(U)})^{-1}[\rho_{in}]$, which is independent of ν, as the input state of the actual channel Λ_ν, instead of the orginal input ρ_{in}, we experimentally obtain the output state of the noise channel $\Gamma_\nu(\rho_{in})$. Then in later analysis, we will concentrate on estimation of the noise parameters for Γ_ν which satisfies relation (14).

Next let us define mathematically the low-noise channel Γ_ϵ. This is a kind of the noise channel and its noise parameter ν takes small positive values, which is denoted by ϵ. We call ϵ the low-noise parameter. Physically, Γ_ϵ is expected to have an analytic ϵ dependence near $\epsilon = 0$. A rigorous mathematical formulation of this requirement is given as follows.

Since the low-noise channel Γ_ϵ is a TPCP map, it has a Kraus representations determined by a family of Kraus operators. We shall define low-noise channels in terms of their Kraus operators. A family of TPCP maps Γ_ϵ with one parameter $\epsilon > 0$ is called a *low-noise channel with low-noise parameter* ϵ if each Γ_ϵ has a Kraus representation

$$\Gamma_\epsilon[\rho] = \sum_a B_a(\epsilon)\rho\, B_a^\dagger(\epsilon) + \epsilon \sum_\alpha C_\alpha(\epsilon)\rho\, C_\alpha^\dagger(\epsilon) \tag{15}$$

with two classes of Kraus operators $\{B_a(\epsilon)\}$ and $\{\sqrt{\epsilon}C_\alpha(\epsilon)\}$ satisfying the following conditions:

(i) $B_a(\epsilon)$ is analytic at $\epsilon = 0$, so that we have the power series expansion

$$B_a(\epsilon) = \kappa_a \mathbf{1}_S - \sum_{n=1}^{\infty} N_a^{(n)} \epsilon^n, \tag{16}$$

in a neighborhood of $\epsilon = 0$, where κ_a and $N_a^{(n)}$ are constant coefficients and operators, respectively, independent of ϵ. The noise channel condition in Eq. (14) requires

$$\sum |\kappa_a|^2 = 1. \tag{17}$$

(ii) $C_\alpha(\epsilon)$ is analytic at $\epsilon = 0$, so that we have the power series expansion

$$C_\alpha(\epsilon) = M_\alpha + \sum_{n=1}^{\infty} M_\alpha^{(n)} \epsilon^n, \tag{18}$$

in a neighborhood of $\epsilon = 0$, where M_α and $M_\alpha^{(n)}$ are constant operators independent of ϵ.

Needless to say, the Kraus operators satisfies the trace-preserving condition

$$\mathbf{1}_S = \sum_a B_a^\dagger(\epsilon) B_a(\epsilon) + \epsilon \sum_\alpha C_\alpha^\dagger(\epsilon) C_\alpha(\epsilon), \tag{19}$$

where $\mathbf{1}_S$ is the identity operator. By definition, the relation

$$\lim_{\epsilon \to +0} \Gamma_\epsilon = id_S \tag{20}$$

is automatically satisfied.

It should be emphasized that our definition of the low-noise channel is general from the physical point of view. Except that Γ_ϵ satisfies Eq. (20) and has analytic dependence of ϵ near the origin, the channel Γ_ϵ can be said to be a general quantum operation acting on the input state. The first-order relation in the ϵ expansion of Eq. (19) is given by

$$\sum_\alpha M_\alpha^\dagger M_\alpha = \sum_a (\kappa_a N_a^{(1)\dagger} + \kappa_a^* N_a^{(1)}). \tag{21}$$

One of our fundamental interests is to ask a question: which input state for the low-noise channel does maximize the Fisher information of its output state ρ_ϵ? It has been shown in [12] that the Fisher information is attained in a pure initial state, so that we can always assume that the input of the channel is a pure state.

Denote the input state by $|\phi\rangle\langle\phi|$. Then, from Eq. (16) and Eq. (18), ρ_ϵ can be expanded as

$$\rho_\epsilon := \Gamma_\epsilon[|\phi\rangle\langle\phi|] = |\phi\rangle\langle\phi| + \epsilon\rho_1 + O(\epsilon^2). \tag{22}$$

Here ρ_1 is given by

$$\rho_1 = \sum_\alpha [K_\alpha|\phi\rangle\langle\phi|N_\alpha^{(1)\dagger} + N_\alpha^{(1)}|\phi\rangle\langle\phi|K_\alpha^\dagger] - \sum_\alpha M_\alpha|\phi\rangle\langle\phi|M_\alpha^\dagger. \tag{23}$$

For this output state ρ_ϵ, we perturbatively solve the equation,

$$\partial_\epsilon\rho_\epsilon = \frac{1}{2}(L_\epsilon\rho_\epsilon + \rho_\epsilon L_\epsilon), \tag{24}$$

in order to get the SLD operator L_ϵ. It is possible to check that the following solution actually satisfies Eq. (24) by substitution.

$$L_\epsilon = \frac{1}{\epsilon}[\mathbf{1} - |\phi\rangle\langle\phi|] - \rho_1 + O(\epsilon). \tag{25}$$

By substituting Eq. (25) into the definition of the Fisher information, we get the value of the information as

$$J_S[\rho_\epsilon] = \frac{1}{\epsilon}\sum_\alpha \left[\langle\phi|M_\alpha^\dagger M_\alpha|\phi\rangle - |\langle\phi|M_\alpha|\phi\rangle|^2\right] + O(\epsilon^0). \tag{26}$$

We can obtain the Fisher information of the low-noise channel in the ancilla-extended system $S+A$ in the same way as the original channel S,

$$J_{S+A}[\tilde{\rho}_\epsilon] = \frac{1}{\epsilon}\sum_\alpha \left[\mathrm{Tr}[\tilde{\rho}M_\alpha^\dagger M_\alpha] - |\mathrm{Tr}[\tilde{\rho}M_\alpha]|^2\right] + O(\epsilon^0), \tag{27}$$

where $\tilde{\rho}$ is S defined by $\tilde{\rho} = \mathrm{Tr}_A[|\Psi\rangle\langle\Psi|]$, and $|\Psi\rangle$ is the input pure state of the extended system. Hence, $\tilde{\rho}$ is able to describe any possible state of S (we assume here the dimension of A is not less than that of S).

By combining both results of J_S and J_{S+A}, we have the ancilla-assisted enhancement factor η such that

$$\eta = \frac{\max\left[\sum_\alpha \left[\mathrm{Tr}\left[\rho_S M_\alpha^\dagger M_\alpha\right] - |\mathrm{Tr}\left[\rho_S M_\alpha\right]|^2\right]\right]_{\rho_S}}{\max\left[\sum_\alpha \left[\langle\phi_S|M_\alpha^\dagger M_\alpha|\phi_S\rangle - |\langle\phi_S|M_\alpha|\phi_S\rangle|^2\right]\right]_{|\phi_S\rangle}}. \tag{28}$$

Here $\max[\,]_{\rho_S}$ means the maximum value over all possible states of S and $\max[\,]_{|\phi_S\rangle}$ the maximum value over all possible pure states of S. Because the set of pure states of S is a subset of the set of states of S, the following inequality trivially holds:

$$\eta \geq 1. \tag{29}$$

3 Channels on Two-Level System

We derive a universal bound on the ancilla-assisted enhancement factor η in a two level system S_2 [13] such that

$$\eta \leq \frac{3}{2}. \tag{30}$$

The bound must hold for all low-noise channels of S_2. The equality $\eta = 3/2$ can be attained by some channels [13] with the maximally-entangled input pure states of $S+A$.

References

[1] C. W. Helstrom, *Quantum Detection and Estimation Theory*, (Academic, New York, 1976).

[2] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory*, (North-Holland, Amsterdam, 1982).

[3] M. Hayashi, ed., *Asymptotic theory of quantum statistical inference: Selected papers* (World Scientific, Singapore, 2005).

[4] M. Hotta and M. Ozawa, Phys. Rev. A, **70**, 022327, (2004).

[5] I. L. Chuang and M. A. Nielsen, J. Mod. Opt. **44**, 2455, (1997).

[6] J. F. Poyatos, J. I. Cirac, and P. Zoller, Phys. Rev. Lett. **78**, 390, (1997).

[7] K. Kraus, Ann. Phys. (N.Y.), **64**, 311, 1971.

[8] K. Kraus, *States, Effects, and Operations: Fundamental Notions of Quantum Theory*, Lecture Notes in Phys. **190**, (Springer, Berlin, 1983)

[9] M. Ozawa, in *Probability Theory and Mathematical Statistics*, Lecture Notes in Math. **1021**, edited by K. Itô and J. V. Prohorov, (Springer, Berlin, 1983), pp. pages(518–525).

[10] M. Ozawa, J. Math. Phys. **25**, 79, (1984).

[11] M. Ozawa, Ann. Phys. (N.Y.), **311**, 350, (2004).

[12] A. Fujiwara, Phys. Rev. A, **63**, 042304, (2001).

[13] M. Hotta, T. Karasawa, M. Ozawa, quant-ph/0507055, to appear in Phys. Rev. A

Characterization of entangled photon pairs generated by spontaneous parametric down conversion

Akihisa Tomita[1 2 3]

[1] *ERATO Quantum Computation and Information Project, JST.*
[2] *Fundamental and Environmental Research Laboratories, NEC Corporation*
Miyukigaoka 34, Tsukuba, Ibaraki 305-8501, Japan.
[3] *Department of Material Science and Engineering, Tokyo Institute of Technology*
4259 Nagatsuda-chou, Midori-ku, Yokohama, Kanagawa 226-0026, Japan.

Abstract. Generation and characterization of entangled photon pairs by the spontaneous parametric down conversion process. Quantum tomography shows that errors of density matrix from the maximally entangled states are not equally probable. For example, two crystal geometry generates likely the mixture of $|\Phi^{(+)}\rangle$ and $|\Phi^{(-)}\rangle$. The hypothesis testing scheme for the entangled photon pairs has been proposed. The test can be improved by optimizing the measurement time on the basis according to the un-isotropic errors on the entangled states.

Keywords: entanglement, fidelity, hypothesis testing

1 Introduction

The concept of entanglement has been thought to be the heart of quantum mechanics. The seminal experiment by Aspect et al [1] has proved the 'spooky' non-local action of quantum mechanics by observing violation of Bell inequality [2] with entangled photon pairs. Recently, entanglement has been also recognized as an important resource for information processing. Quantum information technology has opened the way to novel information processing devices and protocols, such as unconditional security in cryptographic communication and exponential speed-up in some computational tasks. It has been revealed that entanglement plays an essential role, explicitly or implicitly, in the quantum information processing. Entangled states are indispensable in quantum telepotation, a key protocol in quantum repeater. Even in BB84 quantum cryptographic protocol, a hidden entanglement between the legitimate parties guarantee the security. In particulars, maximally entangled states are important resources. The maximally entangled states provides high fidelity on quantum teleportation. It has been shown that universal quantum computation can be done with the maximally entangled states initially prepared and measurement on single qubits. Practical realization of entangled states is therefore one of the most important issues in the quantum information technology.

In the practical implementation, a problem arises how to guarantee the maximal entanglement of the generated (or stored) states. In other words, we need to determine whether the states in hand are good enough to do some quantum information tasks. Imperfections in the generation process, which are unavoidable in practice, reduce the entanglement. Moreover, decoherence and dissipation due to the coupling with the environment degrade the entanglement during the processing. Visibility of two photon interference has been widely used to characterize the entangled states since Aspect's experiment [1]. Quantum tomography [3] has recently applied to obtain full information of the density matrix. However, for practical applications, the characterization is not the goal of the experiment, but only the part of preparation. Therefore, it is favorable to reduce the time and number of the consumed particles as possible. In most application, we don't need to know the full information on the states; we only need to know whether the entanglement is enough or not. The test should be simpler the full characterization. Barbieri et al [4] introduced an entanglement witness to test the entanglement of polarized entangled photon pairs. However, their entanglement witness method considers the statistical fluctuations insufficiently. The statistical hypothesis testing provides appropriate framework for error analysis. Recently, the optimal measurement has been obtained with rigorous statistical treatment [5]. The test can be further improved, if we utilize the knowledge on the tendency of the entanglement degradation. As we discussed in the following section, the error from the maximally entangled states is not isotropic. We can improved the sensitivity for entanglement degradation by focusing the measurement on the expected error directions [6].

In this article, we consider two-photon polarization entangled states generated from spontaneous parametric down conversion (SPDC) process. In section 2, we specify the generation process and show the results of quantum tomography. In section 3, we introduce the statistical hypothesis test scheme for entanglement. In section 4, we show the experimental setting and results.

2 Generation of entangled photon pairs by spontaneous parametric down conversion

SPDC is now widely used to generate entangled photon pairs. This method provide highly entangled states with a simple experimental setting. In particular, Kwait et al [7] have obtained a high flux of the photon pairs from a stack of two type-I phase matched nonlinear crystals. As shown in Fig. 1, the nonlinear crystals (BBO), the optical axis of which are set to orthogonal to on another, are pumped by a pulsed UV light polarized in 45 deg. direction to the optical axis of the crystals. One

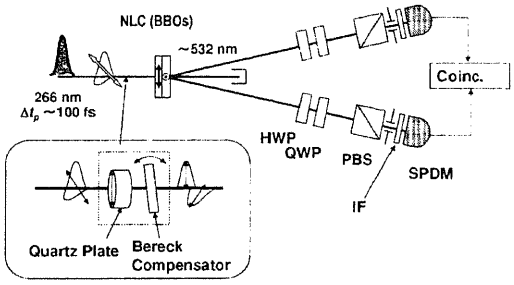

Figure 1: Schematic of the entangled photon pair generation by spontaneous parametric down conversion. Cascade of the nonlinear crystals (NLC) generate the photon pairs. Group velocity dispersion and birefringence in the NLCs are pre-compensated with quartz plates and a Bereck compensator. Two-photon states are analyzed with half wave plates (HWP), quarter wave plates (QWP), and polarization beam splitters (PBS). Interference filters (IF) are placed before the single photon counting modules (SPCM).

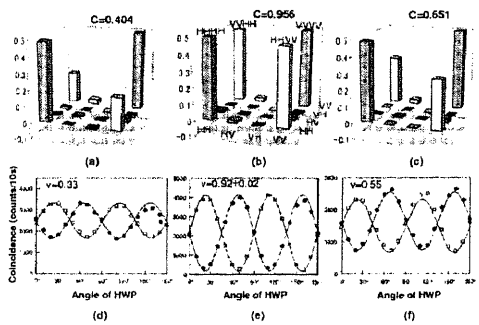

Figure 2: Density matrices estimated by quantum tomography (a)-(c), and the interference fringes (d)-(f) of the two photon states, without compensation (a), (d), optimal compensation (b), (e), and over compensation (c), (f).

nonlinear crystal generates two photons polarized in the horizontal direction ($|HH\rangle$) from the vertical component of the pump light, and the other generates ones polarized in the vertical direction ($|VV\rangle$) from the horizontal component of the pump. If we use very thin (0.13 mm in our experiment) crystals, the directions of the photon waves are almost the same, so that one cannnot distinguish which crystal generates photon from the direction of the photons. Therefore, the two-photon state is given by a super position:

$$\Phi(a,\phi) = a|HH\rangle + \sqrt{1-a^2}e^{i\phi}|VV\rangle. \qquad (1)$$

The amplitude a and phase ϕ of the superposition are determined by the polarization state of the pump light. The 45 deg. polarized pump light will provide $a = 1/\sqrt{2}$ and $\phi = 0$. The two-photon state 1 then refers to the maximally entangled state $|\Phi^{(+)}\rangle = (1/\sqrt{2})(|HH\rangle + |VV\rangle)$

A crucial condition to obtain a highly entangled state in the above scheme is to keep indistinguishability between the two SPDC processes. The group velocity dispersion and birefringence in the crystal may differ the space-time position of the generated photons and make the two processes to be distinguished [8]. For example, in the case of 266 nm pump light wavelength and 532 nm SPDC light wavelength, the horizontally polarized SPDC light travels through the first crystal earlier than the horizontally polarized pump light by 135 fs due to the group velocity dispersion and birefringence. The vertically polarized SPDC light generated in the second crystal takes 33 fs more than the horizontally polarized light to travel through the crystal. Therefore, the horizontally polarized SPDC light arrives at the detector 168fs earlier than the vertically polarized light. This time delay is comparable to the inaccuracy of the SPDC generation equal to the pump pulse duration of 150 fs. The two SPDC processes can be distinguished. Fortunately, this timing information can be erased by compensation; the horizontal component of the pump pulse should arrive at the nonlinear crystals earlier than the vertical component. The compensation can be done by putting a set of birefringence plates (quartz) and a variable wave-plate before the crystals. The two-photon states were analyzed by quantum state tomography and visibility of two-photon interference. The quantum state tomography provides $4times4$ density matrix from the coincidence counts of the 16 combinations, $\{|H\rangle, |V\rangle, |D\rangle, |L\rangle\}_1 \otimes \{|H\rangle, |V\rangle, |D\rangle, |L\rangle\}_2$, where $|D\rangle$ and $|L\rangle$ stand for the linear polarized state in the 45 deg., and the circular polarized state in the anti-clockwise direction, respectively. When the pre-compensation is optimal, the density matrix is close to that of the maximally entangled state, and the visibility is close to unity, as shown in Fig. 2 (b) and (e). It should be noted that only $HHHH$, $VVVV$, $VVHH$, $HHVV$ elements are dominant even in the density matrices for inadequate compensation[8], as seen in Fig. 2 (a) and (c), which implies that the density matrix can be approximately given by the classical mixture of the $|\Phi^{(+)}\rangle\langle\Phi^{(+)}|$ and $|\Phi^{(-)}\rangle\langle\Phi^{(-)}|$. We can improve the hypothesis testing based on this property of the present photon pairs, as given in the following sections.

3 Hypothesis test scheme for entanglement

This section introduces the hypothesis test for entanglement. We consider the two-photon states generated by SPDC described in the previous section. The SPDC generates a state given by a density matrix σ. We assume each two-photon generation process to be identical but individual. The target state is the maximally entangled $|\Phi^{(+)}\rangle$ state. Here we measure the entanglement by the fidelity between the generated state and the target state:

$$\theta = \langle\Phi^{(+)}|\sigma|\Phi^{(+)}\rangle \qquad (2)$$

We introduce the hypothesis H_0 that the entanglement is not enough and H_1 that the entanglement is enough, that is,

$$H_0 : \theta \leq \theta_0 \text{ versus } H_1 : \theta > \theta_0, \tag{3}$$

with a constant θ_0. In the hypothesis testing, two types of the errors may occur. Type I error is rejection of H_0, though H_0 is true, and type II error is acceptance of H_0, though H_1 is true. The type I error is more serious in applications, because insufficient entanglement increases error rate, whereas the type II error only reduces the efficiency. Therefore, we first determine α, the upper limit of the type I error probability, then optimize the test to reduce the type II error probability. The test is called level-α, if the type I error is kept at most α.

The fidelity is estimated from the coincidence counts measured with a basis set of

$$B = \{|VH\rangle, |HV\rangle, |XD\rangle, |DX\rangle, |RR\rangle, |LL\rangle\}, \tag{4}$$

where $|X\rangle$ and $|R\rangle$ stand for the linear polarized state in the 135 deg., and the circular polarized state in the anti-clockwise direction, respectively. The target state $|\Phi^{(+)}\rangle$ yields no coincidence counts in this basis, thus the coincidence counts will detect the error from the target state with high sensitivity. The coincidence counts n_{xy} with the basis $|xy\rangle$ for measurement time t_{xy} will be a random variable according to the Poisson distribution $p(n_{xy})$ of mean $\lambda\mu_{xy}t_{xy}$:

$$p(n_{xy}) = \exp(-\lambda\mu_{xy}t_{xy})\frac{(\lambda\mu_{xy}t_{xy})^{n_{xy}}}{n_{xy}!}, \tag{5}$$

where λ is a normalization constant related to the photon flux and detection efficiency, and μ_{xy} is a density matrix element defined by $\mu_{xy} = \langle xy|\sigma|xy\rangle$.

Inversely, the experimental data of the coincidence counts yield the estimation of the fidelity

$$\hat{\theta} = 1 - \frac{1}{2}\sum_{(xy)\in B}\hat{\mu}_{xy}, \tag{6}$$

through the estimated value $\hat{\mu}_{xy} = \lambda^{-1}\bar{n}_{xy}$, where $\bar{n}_{xy}$ is the average of coincidence counts. We then test the hypothesis based on the rule

$$T = \begin{cases} 0 & (=\text{accept } H_0) \text{ if } \hat{\theta} \leq \theta_0 \\ 1 & (=\text{reject } H_0) \text{ if } \hat{\theta} > \theta_0 \end{cases}, \tag{7}$$

where the level of the test T is determined by the variance $V(\hat{\theta})$ as

$$\alpha = 1 - \int_{-\infty}^{w_\alpha}\exp[-z^2/2]dz \tag{8}$$

$$w_\alpha = \frac{\theta_0 - \hat{\theta}}{\sqrt{V(\hat{\theta})}}, \tag{9}$$

For example, if we obtain $w_\alpha = 1.65$, the test is level-$\alpha = 0.05$. We assume Gaussian distribution of θ, which is a good approximation for a large number of samples. In this case, the probability of type II error equals to α. The variance of $\hat{\theta}$ is calculated by the sum of the variance of μ_{xy}, since the coincidence counting is an independent process. The variance refers to

$$\begin{aligned} V(\hat{\theta}) &= \frac{1}{4}\sum_{(xy)\in B}V(\mu_{xy}) \\ &= \frac{1}{4\lambda^2}\sum_{(xy)\in B}\sigma_{xy}^2, \end{aligned} \tag{10}$$

where σ_{xy}^2 is the unbiased variance of the coincidence counts defined by

$$\sigma_{xy}^2 = \frac{m_{xy}}{m_{xy}-1}\sum_{i=1}^{m_{xy}}\left(\frac{n_{xy}^2}{m_{xy}} - \bar{n}_{xy}^2\right). \tag{11}$$

The limit m_{xy} of Eq. (11) is given by $t_{xy} = m_{xy}\Delta t$, where Δt is unit of measurement time. Performance of the test can be characterized by the level; the smaller α implies the better test.

In the following, we consider a situation often encountered in a actual testing, where the total measurement time is fixed to $t_{tot} = \sum t_{xy}$. The problem is to optimize the measurement time to obtain the smallest value of α, which is equivalent to minimize the variance $V(\hat{\theta})$. If the two-photon state deviates from the maximally entangled state isotropically, the uniform division of time should be optimal, *i.e.*, $t_{VH} = t_{HV} = t_{XD} = t_{DX} = t_{RR} = t_{LL} = t_{tot}/6$. However, if the error is no longer isotopic, the weighted measurement time by Neyman allocation

$$t_{xy} = \frac{\sqrt{\mu_{xy}}}{\sum_{(xy)\in B}\sqrt{\mu_{xy}}}t_{tot}, \tag{12}$$

will yield a better test. Because we don't know the exact values of μ_{xy}, we use the estimated values $\hat{\mu}_{xy}$ to determine the measurement time according to Eq. (12). We employ a two stage measurement strategy: first we estimate μ_{xy} by a uniform measurement $t_{VH} = t_{HV} = t_{XD} = t_{DX} = t_{RR} = t_{LL} = t_1$, and then measure the coincident counts with the weighted time Eq. (12), the total time of which now equals to $t_{tot} - 6t_1$.

4 Experiment

The experimental set-up for the hypothesis testing was almost same as the one shown in Fig. 1. The second harmonic of the mode-locked Ti:S laser light of about 100 fs duration and 150 mW average power was used to pump the nonlinear crystal. The wavelength of SPDC photons is thus 800 nm. Figure 3 shows the coincidence counts measured for 40 second each on the basis given in Eq. (4), when the visibility of the two-photon states was estimated to be 0.92. The distribution of the coincidence events obeys the Poisson distribution. Only small numbers of coincidence were observed on $|HV\rangle$ and $|VH\rangle$ basis. Those observations agree with the prediction, therefore, we expect that the hypothesis testing in the previous section can be applied. In the following, we show the weighted measurement time improves the hypothesis testing. The optimal time for t_1 is derived in [5, 6], however, it requires the information on the two-photon state.

We took $t_1 = \Delta t = 1s$ in the present experiment. We compared the result with the optimal weighted time.

The setting and results are summarized in the table, where (a): uniform time measurement, (b): weighted time measurement with $t_1 = 1$, (c):optimized weighted time measurement; the optimal allocation estimated from the average coincidence counts. Those values yield the estimated values of $\hat{\theta}$ and $\hat{\sigma} = V(\hat{\theta})$. We obtained $(\hat{\theta}, \hat{\sigma}) = (0.960, 4.64 \times 10^{-3}), (0.960, 4.54 \times 10^{-3}), (0.960, 4.54 \times 10^{-3})$, for the uniform time measurement, the weighted time measurement with $t_1 = 1$, the optimal weighted time measurement, respectively. The results of the optimal allocation were same as those using the estimated distribution with the measurement for $t_1 = 1$ in the present experiment. The weighted time measurements provided better testing than the uniform time measurement. For example, if we set the criteria $\theta_0 = 0.952$, the level of the test $\alpha = 0.0432$ in the weighted time measurement, while the uniform time measurement yielded $\alpha = 0.0557$. This implies that we can conclude 'entanglement is enough' with the accuracy of 5 % from the weighted time measurement, whereas we cannot conclude it from the uniform time measurement.

Theory predicts that the improvement should increased as the fidelity. However, the experiment showed almost no gain when the visibility was larger than 0.95. In such high visibility, errors from the maximally entangled state are covered by the noise of the detectors. Since the dark counts are independent of the setting of the measurement apparatus, the allocation shows only a small difference form the uniformly divided one. The levels of the tests are thus similar. The weighted time measurement improves the test in a range of the entanglement. The range would depends mainly on the dark count rate of the detector.

	VH	HV	DX	XD	RR	LL
(a)						
t_{xy}	40	40	40	40	40	40
$\bar{n}_{xy}$	3.625	3.25	16.575	17.1	13.975	15.675
σ^2_{xy}	3.676	2.910	14.866	18.656	9.204	15.917
(b)						
t_{xy}	28	20	42	51	38	55
$\bar{n}_{xy}$	3.536	3.3	16.738	16.922	13.974	15.509
σ^2_{xy}	3.888	3.484	15.954	16.718	9.216	13.069
(c)						
t_{xy}	23	21	49	49	45	47
$\bar{n}_{xy}$	3.565	3.3	16.857	16.959	14.222	15.213
σ^2_{xy}	3.439	3.566	14.417	17.373	10.131	13.389

5 Conclusion

We have discussed the generation of entangled photon pairs by spontaneous parametric down conversion, and state estimation by the quantum tomography. We also consider the hypothesis testing scheme that tells whether the entanglement is enough (above a certain level) or not with a level. The test can be improved by optimizing the measurement time to each coincidence basis. The improvement results from un-isotropic errors in the entangled photon pair generation.

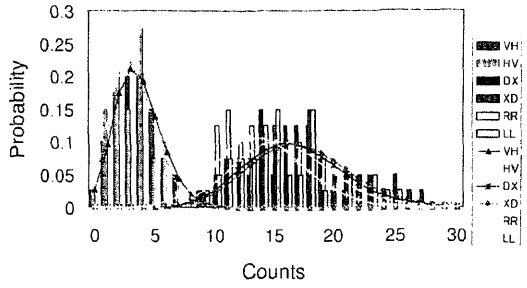

Figure 3: Coincidence counts measured for 40 second each on the basis $|VH\rangle$, $|HV\rangle$, $|XD\rangle$, $|DX\rangle$, $|RR\rangle$, and $|LL\rangle$. Bars present the measured numbers, and lines show the Poisson distribution with the mean values estimated from the experiment.

References

[1] A. Aspect, P. Grangier, and G. Roger. Experimental realization of Einstein-Podolshy-Rosen-Bohm Gedankenexperiment- A new violation of Bell inequalities, *Phys. Rev. Lett.*, 49: 91-94, 1982.

[2] J. S. Bell. *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy.* Cambridge University Press, Cambridge, 1993.

[3] A. G. White, D. F. V. James, P. H. Eberhard, and P. G. Kwiat. Nonmaximally entangled states: Production, characterization, and utilization, *Phys. Rev. Lett.*, 83: 3103-3107, 1999.

[4] M. Barbieri, F. De Martini, G. Di Nepi, P. Mataloni, G. M. D'Ariano, and C. Macchiavello. Experimental detection of entanglement with polarized photons, *Phys. Rev. Lett.*, 91: 227901, 2003.

[5] Y. Tsuda, K. Matsumoto, and M. Hatashi. Hypothesis testing for a maximally entangled state, *quant-ph/0504203.*

[6] Y. Tsuda, B.-S. Shi, A. Tomita, M. Hatashi, K. Matsumoto, and Y.-K. Jiang. Hypothesis testing for an entangled state produced by spontaneous parametric down conversion, *Proceedings of ERATO Conference on Quantum Information Science 2005 (EQIS05).*

[7] P. G. Kwiat, E. Waks, A. G. White, I. Appelbaum, and P.H. Eberhard. Ultra bright source of polarization-entangled photons, *Phys. Rev. A*, 60: 773-776, 1999.

[8] Y. Nambu, K. Usami, Y. Tsuda, K. Matsumoto, and K. Nakamura. Generation of polarization-entangled photon pairs in a cascade of two type-I crystals pumped by femtosecond pulses, *Phys. Rev. A*, 66: 033816, 2002.

Entanglement detection of four-qubit cluster states with local measurements

Yuuki Tokunaga[1 2 3 *] Takashi Yamamoto[2 3] Masato Koashi[2 3] Nobuyuki Imoto[2 3]

[1] *NTT Information Sharing Platform Laboratories, NTT Corporation, 1-1 Hikari-no-oka, Yokosuka, Kanagawa 239-0847, Japan.*
[2] *Division of Materials Physics, Graduate school of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan.*
[3] *CREST Photonic Quantum Information Project, 4-1-8 Honmachi, Kawaguchi, Saitama 331-0012, Japan.*

Abstract. We investigate entanglement witness operators of four-qubit cluster states. We describe a local decomposition of a projector-based entanglement witness and show the number of local measurement settings is optimal. We also show stabilizer-based witnesses with fewer local measurements which have high noise tolerance.

Keywords: entanglement witness, four-qubit entanglement, cluster state, local measurement

1 Introduction

Entanglement is known to be a resource of quantum computation and communication. It was shown that special entangled states called *cluster states* can be utilized for quantum computation with one-qubit measurements [1]. The cluster state $|\Phi\rangle_C$ is defined as

$$\sigma_x^{(a)} \bigotimes_{a' \in ngbh(a)} \sigma_z^{(a')} |\Phi\rangle_C = \pm |\Phi\rangle_C \tag{1}$$

where $ngbh(a)$ specifies the sites of all qubits that interact with the qubit at site $a \in C$. The linearly connected four-qubit cluster states are known to be equivalent to

$$|C_4\rangle = \frac{1}{2}(|0000\rangle + |0011\rangle + |1100\rangle - |1111\rangle) \tag{2}$$

under local unitary transformation [2]. Furthermore,

$$|\chi\rangle = \frac{1}{2}[(|00\rangle + |11\rangle)|00\rangle + (|01\rangle + |10\rangle)|11\rangle] \tag{3}$$

is also equivalent to $|C_4\rangle$ under local unitary transformation. $|\chi\rangle$ has been shown as a resource of teleportation-based controlled-NOT gate [3].

Up to now, some schemes of experimentally producing a four-qubit cluster state have been proposed [4, 5, 6, 7]. In such experiments, we need to verify whether the produced state is a desired genuine multipartite entangled state or not. One of the ways of verification is to use *entanglement witness* operators. An entanglement witness of the four-qubit cluster state which discriminates all biseparable state from $|C_4\rangle$ [8, 9] is known to be

$$\mathcal{W}_{C_4} = \frac{1}{2}\mathbb{1} - |C_4\rangle\langle C_4|. \tag{4}$$

This guarantees that $\mathrm{Tr}(\mathcal{W}_{C_4}\rho_B) \geq 0$ for all biseparable states ρ_B, and that a negative expectation value of the observable $\mathcal{W}_{C_4}$ signifies that the observed state is a genuine four-partite entangled state which is close to $|C_4\rangle$.

*tokunaga.yuuki@lab.ntt.co.jp

If such entanglement witness operators are decomposed into local projection operators, then the measurements can easily be implemented in experiments [10, 11, 12]. For example, we can easily make a projection measurement for a photonic qubit using a polarizing beam splitter and a detector. Moreover, the smallest number of local measurement setting should be selected to decrease experimental effort [11, 13]. However, the necessary number of local measurement setting of the projector-based witnesses seems to grow exponentially with the number of qubits. To overcome the problem, different kinds of witnesses, which are called *stabilizer witnesses*, have been proposed and witnesses with only two measurement settings have been shown in [8, 14].

In this paper, we describe a local decomposition of the projector-based entanglement witness of a four-qubit cluster state with nine measurement settings and show the number of local measurement setting is optimal. We also show stabilizer witnesses of a four-qubit cluster state with more than two local measurement settings which tolerate higher noise compared to the two measurement case. Especially, the noise tolerance of the witness with four local measurement settings is close to that of the projector-based witness which needs nine measurement settings.

2 Local decomposition of entanglement witness

First, we describe a local decomposition of $\mathcal{W}_{C_4}$ and show that it can be measured with nine settings. $|C_4\rangle\langle C_4|$ is locally decomposed as

$$\begin{aligned} |C_4\rangle\langle C_4| = \frac{1}{16}(&IIII + IZXX + ZIXX + XXIZ \\ &+XXZI + ZZII + XYXY + XYYX \\ &+YXXY + YXYX + IIZZ - YYIZ \\ &-YYZI - IZYY - ZIYY + ZZZZ) \end{aligned} \tag{5}$$

using Pauli operators. We use the notation $X = \sigma_x$, $Y = \sigma_y$, $Z = \sigma_z$ for simplicity, and sometimes also use the notation $\sigma_0 = I, \sigma_1 = \sigma_x$, $\sigma_2 = \sigma_y$, $\sigma_3 = \sigma_z$. This

calculation can be done using the following relations

$$|0\rangle\langle 0| = \frac{1}{2}(I+Z), \quad |1\rangle\langle 1| = \frac{1}{2}(I-Z),$$
$$|0\rangle\langle 1| = \frac{1}{2}(X+iY), \quad |1\rangle\langle 0| = \frac{1}{2}(X-iY). \tag{6}$$

Thus, an entanglement witness with local measurements is described as

$$\begin{aligned}\mathcal{W}_{C_4} =& \frac{7}{16}\mathbb{1} - \frac{1}{16}(IZXX + ZIXX + XXIZ \\ &+XXZI + ZZII + XYXY + XYYX \\ &+YXXY + YXYX + IIZZ - YYIZ \\ &-YYZI - IZYY - ZIYY + ZZZZ).\end{aligned} \tag{7}$$

Here, for example, $IZXX$ and $ZIXX$ can be measured by one measurement setting $ZZXX$. Therefore, the witness can be measured with the following nine settings

$$\begin{aligned}&ZZXX, XXZZ, XYXY, XYYX, YXXY, \\ &YXYX, YYZZ, ZZYY, ZZZZ.\end{aligned} \tag{8}$$

Next, we show that $\mathcal{W}_{C_4}$ cannot be decomposed into less than nine local measurement settings, i.e., the decomposition is optimal. The proofs of two-qubit (Bell state) and three-qubit (GHZ and W state) cases were shown in [11, 13]. Here, we extend the proofs to four-qubit cluster states. First, consider a decomposition with eight local measurement settings:

$$\begin{aligned}\sum_{m=1}^{8}\sum_{r,s,t,u=0}^{1} c^m_{rstu}&|A^m_r\rangle\langle A^m_r| \otimes |B^m_s\rangle\langle B^m_s| \\ &\otimes |C^m_t\rangle\langle C^m_t| \otimes |D^m_u\rangle\langle D^m_u|\end{aligned} \tag{9}$$

where $\{|A^m_r\rangle\}$, $\{|B^m_s\rangle\}$, $\{|C^m_t\rangle\}$, and $\{|D^m_u\rangle\}$ are orthonormal bases for $\mathcal{H}_A$, $\mathcal{H}_B$, $\mathcal{H}_C$, and $\mathcal{H}_D$, respectively. We can write any projector of (9) as a vector in the Bloch sphere; $|A^m_0\rangle\langle A^m_0| = \sum_{i=0}^{3} s_i^{A^m_0}\sigma_i$ is represented by the vector $s^{A^m_0} = (1/2, s_1^{A^m_0}, s_2^{A^m_0}, s_3^{A^m_0})$ and $|A^m_1\rangle\langle A^m_1|$ by $s^{A^m_1} = (1/2, -s_1^{A^m_0}, -s_2^{A^m_0}, -s_3^{A^m_0})$. The other projectors can be written similarly. Then, we can expand (9) in the $(\sigma_i \otimes \sigma_j \otimes \sigma_k \otimes \sigma_l)$ basis such as

$$\sum_{m=1}^{8}\sum_{i,j,k,l=0}^{3} \mu^m_{ijkl}\sigma_i \otimes \sigma_j \otimes \sigma_k \otimes \sigma_l. \tag{10}$$

In order to consider measurement settings, we focus on the reduced 3×3 matrices $(\mu^{m,red}_{ijkl})_{j,k=1,\ldots,3}$ for all i, l. They can be written as

$$\alpha^m_{il}(s_1^{B^m}, s_2^{B^m}, s_3^{B^m})^T(s_1^{C^m}, s_2^{C^m}, s_3^{C^m}) \tag{11}$$

where α^m_{il} are constants. Next we write the witness (7) in the $(\sigma_i \otimes \sigma_j \otimes \sigma_k \otimes \sigma_l)$ basis: $\mathcal{W}_{C_4} = \frac{1}{16}\sum_{i,j,k,l=0}^{3}\lambda_{ijkl}\sigma_i \otimes \sigma_j \otimes \sigma_k \otimes \sigma_l$, then the matrices $(\lambda_{ijkl})_{j,k=0,\ldots,3} = (\lambda_{ijkl})_{j,k}$ are written as

$$(\lambda_{0jk0})_{j,k} = \begin{pmatrix} 7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, (\lambda_{0jk1})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix},$$

$$(\lambda_{0jk2})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, (\lambda_{0jk3})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$(\lambda_{1jk0})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, (\lambda_{1jk1})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$(\lambda_{1jk2})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, (\lambda_{1jk3})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$(\lambda_{2jk0})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, (\lambda_{2jk1})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$(\lambda_{2jk2})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, (\lambda_{2jk3})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$(\lambda_{3jk0})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, (\lambda_{3jk1})_{j,k} = \begin{pmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$(\lambda_{3jk2})_{j,k} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, (\lambda_{3jk3})_{j,k} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

The reduced 3×3 matrices that appears when the first rows and the first columns are dropped from the following nine matrices $(\lambda_{0jk1})_{j,k}$, $(\lambda_{0jk2})_{j,k}$, $(\lambda_{1jk0})_{j,k}$, $(\lambda_{1jk1})_{j,k}$, $(\lambda_{1jk2})_{j,k}$, $(\lambda_{2jk0})_{j,k}$, $(\lambda_{2jk1})_{j,k}$, $(\lambda_{2jk2})_{j,k}$, and $(\lambda_{3jk3})_{j,k}$ are linearly independent. However, any linear combination of eight matrices (11) cannot express a nine linearly independent matrices. Therefore, the setting of nine local measurements (8) is optimal.

3 Entanglement witness with fewer local measurements

It has been shown that we can construct stabilizer-based entanglement witnesses with fewer local measurement settings than that of projector-based witnesses [8, 14]. A stabilizer witness for cluster states with two measurement settings which discriminates any biseparable state from a cluster state is described in [14]. In this case, the noise tolerance for a four-qubit cluster state is 33%. The noise considered here is white noise defined in Eq. (16). If there exists witnesses with higher noise

tolerance, we can detect entanglement more sensitive in various situations. Here, we show stabilizer-based witnesses for $|C_4\rangle$ with three or four measurement settings which give 40% or 50% noise tolerance, respectively.

The local decomposition of the projector (5) also derives the stabilizer group of the state. The stabilizing operators of $|C_4\rangle$ are

$$\begin{aligned} S_1 &= IZXX, \quad S_2 = ZIXX, \\ S_3 &= XXIZ, \quad S_4 = XXZI \end{aligned} \tag{12}$$

and their products

$$\begin{aligned} S_5 &= ZZII, & S_6 &= XYXY, \\ S_7 &= XYYX, & S_8 &= YXXY, \\ S_9 &= YXYX, & S_{10} &= IIZZ, \\ S_{11} &= -YYIZ, & S_{12} &= -YYZI, \\ S_{13} &= -IZYY, & S_{14} &= -ZIYY, \\ S_{15} &= ZZZZ, & S_{16} &= IIII. \end{aligned} \tag{13}$$

For all S_k,

$$S_k|C_4\rangle = |C_4\rangle \tag{14}$$

is satisfied. The eigenvectors of the generators of the stabilizer (12) form a complete orthogonal basis. We call it the four-qubit cluster state basis. All the elements of the stabilizing operators are diagonal in this basis.

As shown in [8, 14], a stabilizer witness for $|C_4\rangle$ with two local measurement is described as

$$\begin{aligned} \mathcal{W}_2 = 2\mathbb{1} - \frac{1}{2}(&IZXX + ZIXX + ZZII \\ &+ XXIZ + XXZI + IIZZ). \end{aligned} \tag{15}$$

The two measurement settings are $ZZXX$ and $XXZZ$. Here, $\mathcal{W}_2 - 2\mathcal{W}_{C_4} \geq 0$ holds. Thus, for any state ρ detected by $\mathcal{W}_2$ we have $\mathrm{Tr}(\rho\mathcal{W}_2) \geq 2\mathrm{Tr}(\rho\mathcal{W}_{C_4})$, then the state is also detected by $\mathcal{W}_{C_4}$. Therefore, $\mathcal{W}_2$ also works as a witness. We can simply verify $\mathcal{W}_2 - 2\mathcal{W}_{C_4} \geq 0$ using the expression in the four cluster basis state. Since both $\mathcal{W}_2$ and $\mathcal{W}_{C_4}$ are diagonal in this basis, we can directly check diagonal elements are all non-negative. We consider the following type of noise

$$\rho(p_{noise}) = p_{noise}\mathbb{1}/2^4 + (1 - p_{noise})|C_4\rangle\langle C_4| \tag{16}$$

and from $\mathrm{Tr}W_2\rho(p_{noise}) < 0$, we obtain $p_{noise} < 1/3$ thus it tolerates noise up to 33%.

Here, we construct a witness with four local measurement setting. From the symmetrical structure of the stabilizers, we put a witness

$$\begin{aligned} \mathcal{W}_4 = \beta\mathbb{1} - \gamma(&IZXX + ZIXX + XXIZ + XXZI \\ &- YYIZ - YYZI - IZYY - ZIYY) \\ &- \delta(IIZZ + ZZII) \end{aligned} \tag{17}$$

with four measurement settings $ZZXX$, $XXZZ$, $YYZZ$, and $ZZYY$. Similarly as $\mathcal{W}_2$, $\mathcal{W}_4$ must satisfy the inequality $\mathcal{W}_4 - \alpha\mathcal{W}_{C_4} \geq 0$ for some positive constant α. From $\mathcal{W}_4 - 2\mathcal{W}_{C_4} \geq 0$, we can derive sixteen inequalities by calculating the diagonal elements in the four cluster state basis and the inequalities are put together as

$$\beta - 8\gamma - 2\delta + 1 \geq 0, \tag{18}$$

$$\beta + 8\gamma - 2\delta - 1 \geq 0, \tag{19}$$

$$\beta - 2|\delta| - 1 \geq 0. \tag{20}$$

From $\mathrm{Tr}(W_4\rho(p_{limit})) = 0$, the limit of noise tolerance is calculated [8] as

$$p_{limit} = \frac{-\langle\mathcal{W}\rangle_{|C_4\rangle}}{\langle\mathcal{W}\rangle_{\mathbb{1}/2^4} - \langle\mathcal{W}\rangle_{|C_4\rangle}}. \tag{21}$$

Then β, γ, and δ are chosen such that p_{limit} is maximized on the condition that the inequalities (18) (20) are satisfied. Then, we obtain $\beta = 1$, $\gamma = 1/4$, and $\delta = 0$. Therefore the witness with four measurement settings

$$\begin{aligned} \mathcal{W}_4 = \mathbb{1} - \frac{1}{4}(&IZXX + ZIXX + XXIZ + XXZI \\ &- YYIZ - YYZI - IZYY - ZIYY) \end{aligned} \tag{22}$$

is obtained. This witness tolerates noise up to 50%. The projector-based witness $\mathcal{W}_{C_4}$ tolerates noise up to 53.3%. So the stabilizer-based witness with four measurement settings is comparable in noise tolerance to the projector-based witness with nine measurement settings.

We also show a witness with three local measurement settings. We put a witness from the symmetrical structure

$$\begin{aligned} \mathcal{W}_3 = \beta\mathbb{1} - \gamma(&IZXX + ZIXX + XXIZ + XXZI) \\ &- \delta(IIZZ + ZZII) - \eta(ZZZZ) \end{aligned} \tag{23}$$

with three measurement settings $ZZXX$, $XXZZ$, $ZZZZ$. From $\mathcal{W}_4 - 2\mathcal{W}_{C_4} \geq 0$ and by maximizing the noise tolerance, we obtain $\beta = 3/2$, $\gamma = 1/2$, $\delta = 0$, and $\eta = 1/2$ with similar calculation as in the four local measurement settings. Therefore the witness with three measurement settings

$$\begin{aligned} \mathcal{W}_3 = \frac{3}{2}\mathbb{1} - \frac{1}{2}(&IZXX + ZIXX + XXIZ \\ &+ XXZI + ZZZZ) \end{aligned} \tag{24}$$

is obtained. This witness tolerates noise up to 40%.

4 Summary

We have described a local decomposition of the projector-based entanglement witness of a four-qubit cluster state. It can be measured with nine measurement settings. We have shown that the witness needs at least nine local measurement settings, thus the decomposition is optimal considering the number of measurement setting. We have also described stabilizer-based entanglement witnesses for a four-qubit cluster state with fewer local measurement settings. We have shown a witness with three measurement settings which tolerates noise up to 40%, and a witness with four measurement settings which tolerates noise up to 50%. The projector-based witness with nine measurement settings tolerates noise up to 53.3%. So, the noise tolerance of the stabilizer witness with four local measurement settings is close to that

of the projector-based witness with nine measurement settings for a four-qubit cluster state. These witnesses would be useful for experiments such as verifications of preparing four-photon cluster states.

References

[1] R. Raussendorf and H. J. Briegel, Phys. Rev. Lett. **86**, 5188 (2001).

[2] H. J. Briegel and R. Raussendorf, Phys. Rev. Lett. **86**, 910 (2001).

[3] D. Gottesman and I. L. Chuang, Nature **402**, 390 (1999).

[4] M. Koashi, T. Yamamoto, and N. Imoto, Phys. Rev. A **63**, 030301(R) (2001).

[5] Y. Tokunaga, T. Yamamoto, M. Koashi, and N. Imoto, Phys. Rev. A **71**, 030301(R) (2005).

[6] P. Walther et al., Nature **434**, 169 (2005).

[7] X. Zou and W. Mathis, Phys. Rev. A **71**, 032308 (2005).

[8] G. Tóth and O. Gühne, Phys. Rev. Lett. **94**, 060501 (2005).

[9] In [8], the entanglement witness is described not for $|C_4\rangle$, but for the state originally defined by (1).

[10] B. M. Terhal, Theoretical Computer Science **287**, 313 (2002).

[11] O. Gühne et al., Phys. Rev. A **66**, 062305 (2002).

[12] O. Gühne et al., J. Mod. Opt. **50**, 1079 (2003).

[13] O. Gühne and P. Hyllus, Int. J. Theor. Phys. **42**, 1001 (2002).

[14] G. Tóth and O. Gühne, quant-ph/0501020.

Entanglement spin pairs geometric phase under time independent magnetic field

Xiang-Yu Ge[1] Miki Wadati [1]

[1]*Department of Physics, Graduate School of Science, University of Tokyo*
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

Abstract. The geometric phase for the case of two interacting spins and nonzero magnetic field is calculated. The noncyclic and non-adiabatic entanglement dependence on the geometric phase is explicitly derived. The corresponding unification scheme for that geometric phase is formulated. This unification formulae covers the results for precessing and for isotropically interacting spins.

Keywords: quantum entanglement, geometric phase, interacting spins, magnetic field

1 Introduction

The important ingredient in geometric phases [1, 2, 3] has played a key role in recent advances in the study of quantum mechanics, quantum computation and many other disciplines. Many observable quantum phenomena are related to the geometric phase shift. The Berry's geometric phase has been extensively studied and generalized in various directions, for example, nonadiabatic cyclic evolution [4], noncyclic evolution [5], a kinematic approach [6], and the conditions of adiabaticity, unitarity, and the cyclic nature of the evolution [7]. The geometric phase for mixed states was given within the mathematical context of purification [8], starting from an interferometry setup and a kinematic description [9], and nonunitary evolution for CPMs [10, 11].

Quantum entanglement has been one of the most worthy aspects of quantum theory. Recently, the geometric phases of entangled states (See [12]) have been found useful in quantum information processing, quantum teleportation as well as the geometric quantum computation [13, 14, 15, 16, 17, 18, 19]. i.e., making the quantum gate through the geometric phase shift. In particular, several of them are based on the idea of using the Berry phase [13, 16, 17] shift to the degenerate states. Recently, it was reported [14, 18, 19] that the conditional Berry phase (adiabatic cyclic geometric phase) shift gate can be used in quantum computation. Also a scheme to realize the NMR C-NOT gate through the nonadiabatic cyclic phase shift on the dynamic phase free, and to detect the geometric phase for a Josephson-junction system was proposed [20, 21]. The experimental testing of geometric phase is an interesting and important topic. In Ref. [14], an experiment was done with NMR technique [22, 23, 24, 25] under the adiabatic condition. Recently, the mixed state geometric phase has been confirmed experimentally using NMR [26].

In a recent paper [12], Sjöqvist calculated the geometric phases for both a pair of entangled spins with two spin precessing and with a spin-spin interaction in a time-independent uniform magnetic field. As the study of spin systems effectively permits us to expect design a solid state quantum computer, the entanglement-dependence geometric phase becomes an interesting development in holonomic quantum computer. We first calculate in this paper the geometric phase for a two-interacting spin and nonzero magnetic field model. We then propose an explicit formula and obtain the corresponding unification scheme for entanglement-dependence geometric phase of two interacting spin pairs under a time independent magnetic field. Naturally, one can recover Sjöqvist's results for precessing spins and for isotropically interacting spins models.

2 Geometric phase for two interacting spins

As an example, we determine the geometric phases for two interacting spins. We first present the Hamiltonian operator H in the matrix form,

$$\begin{aligned} H &= \omega(\sigma_z \otimes I + I \otimes \sigma_z) \\ &\quad + \frac{\lambda}{2}(\sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z), \\ &= \begin{pmatrix} 2\omega + \frac{\lambda}{2} & 0 & 0 & 0 \\ 0 & -\frac{\lambda}{2} & \lambda & 0 \\ 0 & \lambda & -\frac{\lambda}{2} & 0 \\ 0 & 0 & 0 & -2\omega + \frac{\lambda}{2} \end{pmatrix}. \end{aligned} \quad (1)$$

where ω is the Larmor frequencies, λ is the strengths of the interaction, and σ_x, σ_y, σ_z are the Pauli matrices. The Hamiltonian operator H describes the isotropic interaction and nonzero magnetic field. In this case, the unitary evolution $U(t)$ is

$$U(t) = e^{-\frac{it\lambda}{2}} e^{-iHt} \begin{pmatrix} e^{-it(\lambda+2\omega)} & 0 & 0 & 0 \\ 0 & \frac{e^{-it\lambda}+e^{it\lambda}}{2} & \frac{e^{-it\lambda}-e^{it\lambda}}{2} & 0 \\ 0 & \frac{e^{-it\lambda}-e^{it\lambda}}{2} & \frac{e^{-it\lambda}+e^{it\lambda}}{2} & 0 \\ 0 & 0 & 0 & e^{-it(\lambda-2\omega)} \end{pmatrix}, \quad (2)$$

if we choose $|\Psi(0)\rangle$ as below

$$|\Psi(0)\rangle = \begin{pmatrix} e^{-i\phi}(\cos\frac{\alpha}{2}\cos\frac{\theta_1}{2}\cos\frac{\theta_2}{2} + \sin\frac{\alpha}{2}\sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2}) \\ \cos\frac{\alpha}{2}\cos\frac{\theta_1}{2}\sin\frac{\theta_2}{2} - \sin\frac{\alpha}{2}\sin\frac{\theta_1}{2}\cos\frac{\theta_2}{2} \\ \cos\frac{\alpha}{2}\sin\frac{\theta_1}{2}\cos\frac{\theta_2}{2} - \sin\frac{\alpha}{2}\cos\frac{\theta_1}{2}\sin\frac{\theta_2}{2} \\ e^{i\phi}(\cos\frac{\alpha}{2}\sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2} + \sin\frac{\alpha}{2}\cos\frac{\theta_1}{2}\cos\frac{\theta_2}{2}) \end{pmatrix}, \quad (3)$$

where α is the Schmidt parameters, ϕ and $\theta_i (i = 1, 2)$ are the initial angle measured from the x-axis and z-axis, respectively.

It is well known that the geometric phase for a pure state $|\Psi(0)\rangle$ and the entangled state $|\Psi(t)\rangle = U(t)|\Psi(0)\rangle$, can be obtained by removing the dynamical phase from the total phase. Let Φ_T, Φ_D, and Φ_G are used to mark total, dynamical, and geometric phases, respectively; and we use t and τ to represent instantaneous time and finite time, respectively. The non-cyclic non-adiabatic geometric phase can be obtained as

$$\Phi_G(\tau) = \Phi_T(\tau) - \Phi_D(\tau) \tag{4}$$

with total phase and dynamical phase

$$\Phi_T(\tau) = \arg\langle\Psi(0)|U(\tau)|\Psi(0)\rangle, \tag{5}$$

$$\Phi_D(\tau) = -i\int_0^\tau \langle\Psi(0)|U(t)^\dagger \dot{U}(t)|\Psi(0)\rangle dt. \tag{6}$$

To obtain the total phase, we substitute that expressions $|\Psi(0)\rangle$ in (3) and $|\Psi(t)\rangle = U(t)|\Psi(0)\rangle$ with $U(t)$ in Eq. (2) into (5) and after some lengthy calculations we get

$$\Phi_T(\tau) = -\arctan\frac{K_1 + K_2\cos\alpha + K_3\sin\alpha}{K_4 + K_5\cos\alpha + K_6\sin\alpha}, \tag{7}$$

with

$$\begin{aligned}
K_1 &= 1 - \tan^2\omega\tau\cos\theta_1\cos\theta_2, \\
K_2 &= -\tan\lambda\tau\tan\omega\tau(\cos\theta_1 + \cos\theta_2), \\
K_3 &= -\tan^2\omega\tau\sin\theta_1\sin\theta_2, \\
K_4 &= \tan\lambda\tau[1 + \frac{1}{2}(1 + \tan^2\omega\tau) \\
&\quad (\sin^2\lambda\tau(\sin\theta_1\sin\theta_2 - 1) + \cos\theta_1\cos\theta_2)], \\
K_5 &= -\tan\omega\tau(\cos\theta_1 + \cos\theta_2), \\
K_6 &= \tan\lambda\tau(1 + \tan^2\omega\tau)\sin^2\frac{\theta_1 + \theta_2}{2}.
\end{aligned}$$

Substituting the $|\Psi(0)\rangle$ in (3) and the entangled state $|\Psi(t)\rangle = U(t)|\Psi(0)\rangle$ with $U(t)$ in Eq. (2) into (6), we obtain the dynamical phase $\Phi_D(\tau)$ as

$$\Phi_D(\tau) = K_7 + K_8\cos\alpha + K_9\sin\alpha \tag{8}$$

with

$$\begin{aligned}
K_7 &= -\frac{\lambda\tau}{2}(1 + \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2), \\
K_8 &= -(\cos\theta_1 + \cos\theta_2)\omega\tau, \\
K_9 &= \frac{\lambda\tau}{2}(1 - \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2).
\end{aligned}$$

From the total phase (7) and dynamical phase (8), the geometric phase Φ_G is readily given by equation (4). The expression in the first line of Eq. (7) is valid only when the resulting angle remains in the $[-\frac{\pi}{2}, \frac{\pi}{2}]$ interval. We emphasize that the above result is new. The physical situation corresponds to experiments using magnetic materials under field. This result should be useful in further study of geometric phases and their applications in quantum information processing.

3 More general result for geometric phase

Consider a quantum system consisting of two spin-$\frac{1}{2}$ particles in a time-independent uniform magnetic field which is applied in the z-direction. Hamiltonian for the systems is

$$\begin{aligned}
H &= \omega_1 S_{1,z} + \omega_2 S_{2,z} \\
&\quad + (8\lambda/\hbar)(a_x S_x^1 S_x^2 + a_y S_y^1 S_y^2 + a_z S_z^1 S_z^2),
\end{aligned} \tag{9}$$

where ω_1 and ω_2 are the Larmor frequencies, and λ and (a_x, a_y, a_z) are the strengths of the interaction. The $S_{1,z}$ and $S_{2,z}$ are the corresponding z components of the spin operators associated with the two particles, and the $\mathbf{S}_1 = (S_x^1, S_y^1, S_z^1)$ and $\mathbf{S}_2 = (S_x^2, S_y^2, S_z^2)$ are the spin operators pertaining to the spin pair. The state $|\Psi(t)\rangle$ of the system obeys Schrödinger equation

$$i\frac{d}{dt}|\Psi(t)\rangle = H|\Psi(t)\rangle \tag{10}$$

where H is the faithful matrix representation of the above Hamiltonian (9),

$$\begin{aligned}
H &= \omega_1\sigma_z \otimes I + \omega_2 I \otimes \sigma_z \\
&\quad + (2\lambda/\hbar)(a_x\sigma_x \otimes \sigma_x + a_y\sigma_y \otimes \sigma_y + a_z\sigma_z \otimes \sigma_z) \\
&= \begin{pmatrix} H_{11} & 0 & 0 & H_{14} \\ 0 & H_{22} & H_{23} & 0 \\ 0 & H_{32} & H_{33} & 0 \\ H_{41} & 0 & 0 & H_{44} \end{pmatrix}.
\end{aligned} \tag{11}$$

Here, $H_{11} = \omega_1 + \omega_2 + a_3$, $H_{22} = \omega_1 - \omega_2 - a_3$, $H_{33} = -\omega_1 + \omega_2 - a_3$, $H_{44} = -\omega_1 - \omega_2 + a_3$, $H_{14} = H_{41} = a_1 - a_2$, $H_{23} = H_{32} = a_1 + a_2$, and $a_1 = (2\lambda/\hbar)a_x$, $a_2 = (2\lambda/\hbar)a_y$, $a_3 = (2\lambda/\hbar)a_z$. The state $|\Psi(t)\rangle$ is a 4×1 matrix, which is written as $|\Psi(t)\rangle = U(t)|\Psi(0)\rangle$. For convenience we parametrize the initial state $|\Psi(0)\rangle$ of the system by

$$\begin{aligned}
|\Psi(0)\rangle &= e^{-i\frac{\beta}{2}}\cos\frac{\alpha}{2}|\mathbf{n}(0), \mathbf{m}(0)\rangle \\
&\quad + e^{i\frac{\beta}{2}}\sin\frac{\alpha}{2}|-\mathbf{n}(0), -\mathbf{m}(0)\rangle.
\end{aligned} \tag{12}$$

where α and β are the Schmidt parameters, $\mathbf{n}$ and $\mathbf{m}$ are two points on the Poincaré sphere and the subscripts denote spin 1 and 2, respectively. We assume that the initial spin states $|\pm\mathbf{n}(0)\rangle$ and $|\pm\mathbf{m}(0)\rangle$ are given by

$$\begin{aligned}
|\mathbf{n}(0)\rangle &= e^{-i\frac{\varphi_1}{2}}\cos\frac{\theta_1}{2}|\uparrow\rangle + e^{i\frac{\varphi_1}{2}}\sin\frac{\theta_1}{2}|\downarrow\rangle, \\
|\mathbf{m}(0)\rangle &= e^{-i\frac{\varphi_2}{2}}\cos\frac{\theta_2}{2}|\uparrow\rangle + e^{i\frac{\varphi_2}{2}}\sin\frac{\theta_2}{2}|\downarrow\rangle,
\end{aligned}$$

and

$$\begin{aligned}
|-\mathbf{n}(0)\rangle &= e^{-i\frac{\varphi_1}{2}}\sin\frac{\theta_1}{2}|\uparrow\rangle - e^{i\frac{\varphi_1}{2}}\cos\frac{\theta_1}{2}|\downarrow\rangle, \\
|-\mathbf{m}(0)\rangle &= e^{-i\frac{\varphi_2}{2}}\sin\frac{\theta_2}{2}|\uparrow\rangle - e^{i\frac{\varphi_2}{2}}\cos\frac{\theta_2}{2}|\downarrow\rangle.
\end{aligned}$$

Assuming that the initial spin state $|\mathbf{n}(0)\rangle$ makes an angle θ from the z-axis, the spin state at any later time t is

$$|\mathbf{n}(t)\rangle = e^{-i\frac{\phi(t)}{2}}\cos\frac{\theta}{2}|\uparrow\rangle + e^{i\frac{\phi(t)}{2}}\sin\frac{\theta}{2}|\downarrow\rangle. \tag{13}$$

Here $\phi(t) = \phi + \omega t$, and $|\uparrow\rangle$ and $|\downarrow\rangle$ denote $(1,0)^T$ and $(0,1)^T$, respectively. Then initial state $|\Psi(0)\rangle$ in Eq. (12) is given by

$$\begin{pmatrix} e^{-i\frac{\phi_1+\phi_2}{2}}(e^{-i\frac{\beta}{2}}\cos\frac{\alpha}{2}\cos\frac{\theta_1}{2}\cos\frac{\theta_2}{2} + e^{i\frac{\beta}{2}}\sin\frac{\alpha}{2}\sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2}) \\ e^{-i\frac{\phi_1-\phi_2}{2}}(e^{-i\frac{\beta}{2}}\cos\frac{\alpha}{2}\cos\frac{\theta_1}{2}\sin\frac{\theta_2}{2} - e^{i\frac{\beta}{2}}\sin\frac{\alpha}{2}\sin\frac{\theta_1}{2}\cos\frac{\theta_2}{2}) \\ e^{i\frac{\phi_1-\phi_2}{2}}(e^{-i\frac{\beta}{2}}\cos\frac{\alpha}{2}\sin\frac{\theta_1}{2}\cos\frac{\theta_2}{2} - e^{i\frac{\beta}{2}}\sin\frac{\alpha}{2}\cos\frac{\theta_1}{2}\sin\frac{\theta_2}{2}) \\ e^{i\frac{\phi_1+\phi_2}{2}}(e^{-i\frac{\beta}{2}}\cos\frac{\alpha}{2}\sin\frac{\theta_1}{2}\sin\frac{\theta_2}{2} + e^{i\frac{\beta}{2}}\sin\frac{\alpha}{2}\cos\frac{\theta_1}{2}\cos\frac{\theta_2}{2}) \end{pmatrix}. \tag{14}$$

The unitary evolution of the state vector is given by

$$\begin{aligned} U(t) &= e^{-itH} = e^{-itSH'S^{-1}} = Se^{-itH'}S^{-1} \\ &= S\mathrm{diag}(e^{-it\lambda_1}, e^{-it\lambda_2}, e^{-it\lambda_3}, e^{-it\lambda_4})S^{-1} \\ &= \begin{pmatrix} U_{11} & 0 & 0 & U_{14} \\ 0 & U_{22} & U_{23} & 0 \\ 0 & U_{32} & U_{33} & 0 \\ U_{41} & 0 & 0 & U_{44} \end{pmatrix}, \end{aligned} \tag{15}$$

where the matrix $H' = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, the $\lambda_i (i = 1, ..., 4)$ are the eigenvalues of the matrix H in (11) which are found to be $\lambda_1 = a_3 + p_+$, $\lambda_2 = -a_3 + p_-$, $\lambda_3 = -a_3 - p_-$, $\lambda_4 = a_3 - p_+$, where $p_\pm = \sqrt{(\omega_1 \pm \omega_2)^2 + (a_1 \mp a_2)^2}$. The normalised eigenvectors of the unitary matrix $S = (x^1, x^2, x^3, x^4)$, which such that $H = SH'S^{-1} = SH'S^\dagger$, are $x^1 = (S_{11}, 0, 0, S_{41})^T$, $x^2 = (0, S_{22}, S_{32}, 0)^T$, $x^3 = (0, S_{23}, S_{33}, 0)^T$ and $x^4 = (S_{14}, 0, 0, S_{44})^T$, the matrix elements S_{ij} are

$$\begin{aligned} S_{11} &= S_{44} = \frac{\omega_1 + \omega_2 + p_+}{\sqrt{2}\sqrt{p_+(p_+ + \omega_1 + \omega_2)}}, \\ S_{22} &= S_{33} = \frac{\omega_1 - \omega_2 + p_-}{\sqrt{2}\sqrt{p_-(p_- + \omega_1 - \omega_2)}}, \\ S_{14} &= -S_{41} = \frac{a_2 - a_1}{\sqrt{2}\sqrt{p_+(p_+ + \omega_1 + \omega_2)}}, \\ S_{32} &= -S_{23} = \frac{a_1 + a_2}{\sqrt{2}\sqrt{p_-(p_- + \omega_1 - \omega_2)}}. \end{aligned}$$

Also the matrix elements U_{ij} are

$$\begin{aligned} U_{11} &= \frac{(a_1 - a_2)^2 e^{-i\lambda_4 t} + (\omega_1 + \omega_2 + p_+)^2 e^{-i\lambda_1 t}}{2p_+(p_+ + \omega_1 + \omega_2)}, \\ U_{22} &= \frac{(a_1 + a_2)^2 e^{-i\lambda_3 t} + (\omega_1 - \omega_2 + p_-)^2 e^{-i\lambda_2 t}}{2p_-(p_- + \omega_1 - \omega_2)}, \\ U_{33} &= \frac{(a_1 + a_2)^2 e^{-i\lambda_2 t} + (\omega_1 - \omega_2 + p_-)^2 e^{-i\lambda_3 t}}{2p_-(p_- + \omega_1 - \omega_2)}, \\ U_{44} &= \frac{(a_1 - a_2)^2 e^{-i\lambda_1 t} + (\omega_1 + \omega_2 + p_+)^2 e^{-i\lambda_4 t}}{2p_+(p_+ + \omega_1 + \omega_2)}, \\ U_{14} &= U_{41} = \frac{(a_1 - a_2)(\omega_1 + \omega_2 + p_+)(e^{-i\lambda_1 t} - e^{-i\lambda_4 t})}{2p_+(p_+ + \omega_1 + \omega_2)}, \\ U_{23} &= U_{32} = \frac{(a_1 + a_2)(\omega_1 - \omega_2 + p_-)(e^{-i\lambda_2 t} - e^{-i\lambda_3 t})}{2p_-(p_- + \omega_1 - \omega_2)}. \end{aligned}$$

The non-cyclic non-adiabatic geometric phase can be obtained as $\Phi_G(\tau) = \Phi_T(\tau) - \Phi_D(\tau)$ with total phase $\Phi_T(\tau) = \arg\langle\Psi(0)|\Psi(\tau)\rangle$ and dynamical phase $\Phi_G(\tau) = -i\int_0^\tau \langle\Psi(t)|\dot{\Psi}(t)\rangle dt$. With $|\Psi(0)\rangle$ in Eq. (12) and $|\Psi(t)\rangle = U(t)|\Psi(0)\rangle$, we arrive at the formulae for the total, dynamical, and geometric phase (5), (6), (4) of the entangled state $|\Psi(t)\rangle$ under a unitary evolution $U(t)$ in (15).

4 Recovering Sjöqvist's result

We use the general formulae (4),(5) and (6) with (12) and (15) to discuss some special cases.

Let $a_i = 0, (i = 1, 2, 3)$. Then, the Hamiltonian (9) becomes

$$H = \omega_1 S_{1,z} + \omega_2 S_{2,z}, \tag{16}$$

which is the Hamiltonian operator in case of the spin precession in [12]. In this case, the unitary evolution $U(t)$ is expressed as

$$U(t) = \mathrm{diag}(e^{-it\frac{\omega_1+\omega_2}{2}}, e^{-it\frac{\omega_1-\omega_2}{2}}, e^{it\frac{\omega_1-\omega_2}{2}}, e^{it\frac{\omega_1+\omega_2}{2}}). \tag{17}$$

If we choose $\beta = 0$ of $|\Psi(0)\rangle$ in (12), the geometric phase of the entangled state $|\Psi(t)\rangle = U(t)|\Psi(0)\rangle$ is given as

$$\begin{aligned} \Phi_G(\tau) &= -\arctan\frac{\cos\alpha[\cos\theta_1\tan\frac{\omega_1\tau}{2} + \cos\theta_2\tan\frac{\omega_2\tau}{2}]}{1 - K_{10}\tan\frac{\omega_1\tau}{2}\tan\frac{\omega_2\tau}{2}} \\ &\quad + \cos\alpha(\frac{\omega_1\tau}{2}\cos\theta_1 + \frac{\omega_2\tau}{2}\cos\theta_2) \end{aligned} \tag{18}$$

with

$$K_{10} = \cos\theta_1\cos\theta_2 + \sin\alpha\sin\theta_1\sin\theta_2.$$

Note that the expression in terms of arctan is valid only when the resulting angle remains in the $[-\frac{\pi}{2}, \frac{\pi}{2}]$ interval. The geometric phase $\Phi_G(\tau)$ in (18) of the entangled state $|\Psi(t)\rangle = U(t)|\Psi(0)\rangle$ for $U(t)$ in (17) and $|\Psi(0)\rangle$ in (12) agrees with Eq. (7) in Ref. [12].

Let $\omega_1 = \omega_2 = 0$ and $a_1 = a_2 = a_3 = a$. Then, the Hamiltonian operator H in Eq. (9) becomes

$$\begin{aligned} H &= (8\lambda/\hbar)(a_x S_x^1 S_x^2 + a_y S_y^1 S_y^2 + a_z S_z^1 S_z^2) \\ &= (2\lambda a/\hbar)(\sigma_x \otimes \sigma_x + \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z), \end{aligned} \tag{19}$$

which is the Hamiltonian operator in case of the isotropic spin-spin interaction in [12]. In this case, if we assume $a = \frac{\hbar}{8}$, the unitary evolution $U(t)$ is

$$U(t) = \begin{pmatrix} e^{-it\lambda} & 0 & 0 & 0 \\ 0 & \frac{e^{-it\lambda}+e^{it\lambda}}{2} & \frac{e^{-it\lambda}-e^{it\lambda}}{2} & 0 \\ 0 & \frac{e^{-it\lambda}-e^{it\lambda}}{2} & \frac{e^{-it\lambda}+e^{it\lambda}}{2} & 0 \\ 0 & 0 & 0 & e^{-it\lambda} \end{pmatrix}. \tag{20}$$

To investigate further the quantum system of two spin-$\frac{1}{2}$ particles with a spin-spin interaction on the Schmidt sphere in the context of the geometric phase, it is convenient to consider a superposition of the type

$$|\Psi(0)\rangle = \cos\frac{\alpha}{2}|\downarrow\rangle_1|\uparrow\rangle_2 + \sin\frac{\alpha}{2}|\uparrow\rangle_1|\downarrow\rangle_2 \tag{21}$$

or

$$\begin{aligned} |\Psi(0)\rangle &= \frac{1}{\sqrt{2}}(\cos\frac{\alpha}{2} + \sin\frac{\alpha}{2})|+\rangle_1|-\rangle_2 \\ &\quad + \frac{1}{\sqrt{2}}(\cos\frac{\alpha}{2} - \sin\frac{\alpha}{2})|-\rangle_1|+\rangle_2. \end{aligned} \tag{22}$$

where $|+\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$ and $|-\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle - |\uparrow\rangle)$.

When $U(t)$ in (20) acts on $|\Psi(0)\rangle$ in (22), we have

$$\begin{aligned} |\Psi(t)\rangle &= \frac{1}{\sqrt{2}}[(e^{-i\lambda t}\cos\frac{\alpha}{2} + e^{i\lambda t}\sin\frac{\alpha}{2})|+\rangle_1|-\rangle_2 \\ &\quad + (e^{-i\lambda t}\cos\frac{\alpha}{2} - e^{i\lambda t}\sin\frac{\alpha}{2})|-\rangle_1|+\rangle_2]. \end{aligned} \tag{23}$$

The time-dependent Schmidt parameters $\alpha(t)$ and $\beta(t)$ is introduced by rewriting Eq. (23) on the symmetric form

$$|\Psi(t)\rangle = e^{-i\frac{\beta_1(t)}{2}}\cos\frac{\alpha(t)}{2}|+\rangle_1|-\rangle_2 + e^{i\frac{\beta_2(t)}{2}}\sin\frac{\alpha(t)}{2})|-\rangle_1|+\rangle_2. \quad (24)$$

Inserting Eq. (24) into Eq. (4), we obtain the noncyclic two-particle geometric phases as

$$\Phi_G(\tau) = -\arctan(\frac{\cos\frac{\alpha(\tau)+\alpha(0)}{2}}{\cos\frac{\alpha(\tau)-\alpha(0)}{2}}\tan\frac{\beta(\tau)}{2}) + \int_0^\tau \frac{\dot{\beta}(\tau)}{2}\cos\alpha(t)dt, \quad (25)$$

where $\alpha(0) = \frac{\pi}{2} - \alpha$, $\cos\alpha(t) = \sin\alpha\cos(2\lambda t)$ and $\tan\beta(t) = \tan\frac{\beta_1(t)+\beta_2(t)}{2} = -\tan\alpha\sin(2\lambda t)$ with $\tan\frac{\beta_1(t)}{2} = \tan(\lambda t)\frac{\cos\alpha}{1+\sin\alpha}$ and $\tan\frac{\beta_2(t)}{2} = -\tan(\lambda t)\frac{1+\sin\alpha}{\cos\alpha}$. In this way, the geometric phase $\Phi_G(\tau)$ in (25) of the entangled state $|\Psi(t)\rangle = U(t)|\Psi(0)\rangle$ for $U(t)$ in (20) and $|\Psi(0)\rangle$ in (22) agrees with Eq. (24) in Ref. [12].

5 Conclusions

As a special example, the entanglement dependence on the noncyclic two particle geometric phase has been determined for an isotropic interaction and nonzero magnetic field model. We then have extended the noncyclic and non-adiabatic phase to the general case, proposed a generalization of geometric phase and formulated the corresponding unification scheme for that geometric phase. From the present formulae, the geometric phases for both a pair of spin-$\frac{1}{2}$ particles with two spin processing and an isotropic spin-spin interaction are reproduced.

Acknowledgments

One of the authors XYGe acknowledges the support of JSPS. He would like to thank Xiang-Bin Wang for useful discussions.

References

[1] S. Pancharatnam, Proc. Indian Acad. Sci., Sect. A **44**, 247 (1956).

[2] M.V. Berry, Proc. R. Soc. London Ser. A **392**, 45 (1984).

[3] B. Simon, Phys. Rev. Lett. **51**, 2167 (1983).

[4] Y. Aharonov and J. Anandan, Phys. Rev. Lett. **58**, 1593 (1987).

[5] J. Samuel and R. Bhandari, Phys. Rev. Lett. **60**, 2339 (1988).

[6] N. Mukunda and R. Simon, Ann. Phys. (N.Y.) **228**, 205 (1993).

[7] A.K. Pati, Phys. Rev. A **52**, 2576 (1995).

[8] A. Uhlmann, Rep. Math. Phys. **24**, 229 (1986); Lett. Math. Phys.**21**, 229 (1991).

[9] E. Sjöqvist, A.K. Pati, A. Ekert, J.S. Anandan, M. Ericsson, D.K.L. Oi, and V. Vedral, Phys. Rev. Lett. **85**, 2845 (2000).

[10] M. Ericsson, E. Sjöqvist, J. Brännlund, D.K.L. Oi, and A.K. Pati, Phys. Rev. A **67**, 020101(R) (2003).

[11] J.G. Peixoto de Faria *et al.*, Europhys. Lett. **62**, 782 (2003).

[12] E. Sjöqvist, Phys. Rev. A **62**, 022109 (2000).

[13] J. Pachos, P. Zanardi and M. Rasetti, Phys. Rev. A **61**, 010305(R) (2000).

[14] J.A. Jones, V. Vedral, A. Ekert, and G. Castagnoli, Nature (London) **403**, 869 (2000).

[15] G. Falci, R. Fazio, G.M. Palma, J. Siewert and V.Verdal, Nature (London) **407**, 355 (2000).

[16] P. Zanardi and M. Rasetti, Phys. Lett. A **264**, 94 (1999).

[17] L-M. Guan, J.I. Cirac and P. Zoller, Science **292**, 1695 (2001).

[18] A. Ekert, M. Ericsson, P. Hayden, H. Inamori, J.A. Jones, D.K.L. Oi, and V. Vedral, J. Mod. Opt. **47**, 2501 (2000).

[19] X.B. Wang and K. Matsumoto, Phys. Rev. Lett. **87**, 097901 (2001).

[20] X.B. Wang and K. Matsumoto, J. Phys. A: Math. Gen. **34**, L631 (2001).

[21] X.B. Wang and K. Matsumoto, Phys. Rev. B **65**, 172508 (2002).

[22] D.G. Cory, A.F. Fahmy and T.F. Havel, Proc. Natl. Acad. Sci. U.S.A **94**, 1634 (1997).

[23] N.A. Gershenfeld and I.L. Chung, Science **275**, 350 (1997).

[24] J.A. Jones and M. Mosca, J. Chem. Phys **109**, 1648 (1998).

[25] M.A. Nielsen and I.L. Chuang, *Quantum Computation and Quantum Information*(Cambridge University, Cambridge, 2000).

[26] J.F. Du, P. Zou, M. Shi, L.C. Kwek, J.-W. Pan, C.H. Oh, A. Ekert, D.K.L. Oi, and M. Ericsson, Phys. Rev. Lett. **91**, 100403 (2003).

Local Discrimination and Multipartite Entanglement Measures

Damian Markham[1] * Shashank Virmani[2] † Masaki Owari[1] ‡ Mio Murao [1 3]
Masahito Hayashi[4]

[1] *Department of Physics, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033,Japan.*
[2] *Optics Section, Blackett Laboratory & Institute for Mathematical Sciences, Imperial College, London SW7 2AZ, United Kingdom.*
[3] *PRESTO, JST, Kawaguchi, Saitama 332-0012, Japan.*
[4] *Imai Quantum Computation and Information Project, ERATO, JST, Tokyo 113-0033, Japan.*
& Superrobust Computation Project (21st Century COE by MEXT), University of Tokyo, Japan.

Abstract. We find a necessary condition for the local, perfect discrimination of general multipartite states in terms of three entanglement quantities, the global robustness of entanglement, the relative entropy of entanglement and the geometric measure. These results lead to an upper bound on the number of pure, multipartite states that can be perfectly, discriminated locally. The bound is explicitly found for pure bipartite states and some known results are proved in a unified way. This bound is shown to be tight for a set of m-party GHZ states. Further we extend the initial condition to the probabilistic case, leading to a bound on the locally accessible mutual information for a completely random message.

Keywords: LOCC discrimination, multiparty, entanglement, local accessible information

1 Introduction

We consider the connection between distance like measures of general multiparticle entanglement and the problem of LOCC state discrimination [1].

Whenever we want to retrieve classical information from a quantum system, we are essentially making a discrimination of states (or subspaces). In quantum information we often consider situations where separated parties are restricted to using Local Operations and Classical Communication (LOCC). As the final outcomes of any quantum experiment are obtained from measurements, it is hence important for us to understand the effects of the LOCC restriction on the measurement of quantum states. Indeed, the LOCC measurement of quantum states is important for the study of cryptographic protocols [2], channel capacities [3], and distributed quantum information processing [4].

Defining entanglement for more than two parties becomes very complicated. There are two approaches to quantifying entanglement, the first is to define entanglement in terms of units of 'useful' entanglement. In the bipartite case the standard unit is the singlet, which is, for example, essential for faithful teleportation [5] or entanglement based secure communication [2]. We then ask questions like how many singlets is it possible to distill from a state (the entanglement of distillation) or how many singlets are needed to create a state (the entanglement of creation). In the multiparty case however, we have no clear idea of what the units of usefulness are and we have inequivalent types of entanglement [6]. The other approach is to define measures of entanglement in an abstract way, such that they obey certain axioms, and can be called entanglement monotones (the main axiom being they must not increase under LOCC). Typical examples of these are the distance like quantities to the closest separable state [7, 8, 9]. Since the only reference to the multiparty nature is in the definition of separable, these quantities are well defined in the multiparty case. However, there is often no physical interpretation in terms of the meaning for a state as a resource, or the usefulness for tasks.

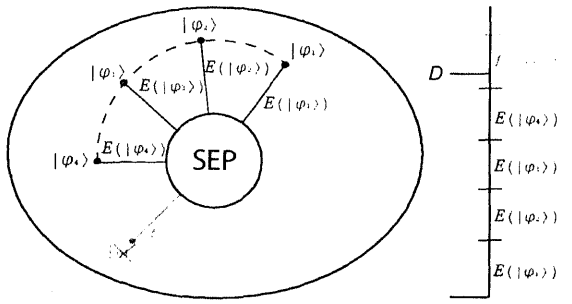

Figure 1: To discriminate the states $\{|\varphi_i\rangle\}$ perfectly under LOCC, the sum of the entanglement distances $E(|\varphi_i\rangle)$ must be less than the total dimension D (Theorem 1 and 2).

The connection between entanglement and LOCC state discrimination is thusfar unclear. One of the earliest results on LOCC discrimination was the observation of 'non-locality without entanglement', where sets of orthogonal product states were presented that cannot be perfectly discriminated locally [10]. This was followed by a proof that *any* two multipartite pure states may be optimally ambiguously, and unambiguously discriminated using LOCC [11]. Subsequently there have been several further interesting results on what can and cannot be done using LOCC in variety of situations [12]. However, due the lack of a simple characterisation of LOCC operations, further general results seem to be difficult to obtain and many results are specific to the bipartite case

*markham@phys.s.u-tokyo.ac.jp
†s.virmani@imperial.ac.uk
‡owari@eve.phys.s.u-tokyo.ac.jp

only, or valid for specific sets of states or scenarios only.

We give a necessary condition for the perfect discrimination of states by LOCC which hold for all states, irrespective of the number of parties, or wether pure or mixed [1]. In the pure state case this condition implies that three distance like measures of entanglement can be interpreted as upper bounds to the number of states that can be discriminated perfectly by LOCC. By using known entanglement results we will give examples of existing and new LOCC discrimination bounds in a unified manner.

We further extend these conditions to the probabilistic case which allows us to give a bound on the LOCC accessible mutual information.

2 The condition and entanglement

Any measurement in physics is described mathematically by a POVM. Our condition stems solely from two facts. First, that there must be a POVM which is deterministic in the outcomes for our states. Second, this POVM must be LOCC. This is a very very hard problem, and we actually use weaker versions of both these conditions. We use the result that any LOCC measurement all POVM elements must be separable [13]. The main mathematical step is to then take an optimization of the trace of the *individual elements* of the POVM, which vastly simplifies the problem. A similar technique was used in [14]. This allows us to translate these conditions to a statement about the entanglement of the states. The existence of a POVM satisfying these conditions, can be rewritten to give the following necessary condition.

- **Necessary condition for deterministic LOCC discrimination of set $\{\rho_i | i = 1..N\}$:**

$$\sum_i t(\rho_i) \leq D \tag{1}$$

where, D is the total dimension of the system, and

$$t(\rho_i) := |P_i| + \min |\Delta| \quad \text{such that} \quad I \geq \Delta \geq 0; P_i + \Delta \in SEP; \mathrm{tr}\{\Delta P_i\} = 0. \tag{2}$$

where P_i is the projector onto the span of each ρ_i and $|A|$ is the trace of an operator A.

Now, the expression for $t(\rho_i)$ is what gives us the link to entanglement. Although it may look obscure at first, it turns out, that it is almost exactly the same as the definition for the *global robustness of entanglement* $R_g(\rho)$ [15]. In fact it can be shown

$$t(\rho_i) \geq s(\rho_i) := |P_i| \left[1 + R_g\left(\frac{P_i}{|P_i|}\right)\right]. \tag{3}$$

More, we can further relate it to the *relative entropy of entanglement* E_R [8] and the *geometric measure* E_G, through the following bound

$$t(\rho_i) \geq s(\rho_i) \geq 2^{E_R(\rho_i)+S(\rho_i)} \geq 2^{E_G(\rho_i)}, \tag{4}$$

where $S(\rho_i)$ is the von Neumann entropy.

This gives, for example, a hierarchy of bounds;

- **Bounds on the number of states N that can be discriminated perfectly by LOCC**

$$N \leq D/\overline{t(\rho_i)} \leq D/\overline{s(\rho_i)} \leq D/\overline{2^{E_R(\rho_i)+S(\rho_i)}} \leq D/\overline{2^{E_G(\rho_i)}}, \tag{5}$$

where for a quantity x_i, we denote $\overline{x_i} := 1/N \sum_{i=1}^{N} x_i$, the 'average'.

Hence, in the pure state case, where the bounding quantities reduce to the geometric measure of entanglement, the relative entropy of entanglement and the robustness of entanglement (from right to left), we can interpret these three distance like entanglement measures as bounds to the number of states we can discriminate perfectly by LOCC.

3 Examples

Given this hierarchy we can apply known results on entanglement to the bounds. Given sets of states with the same entanglement, for the bipartite case ($\mathcal{S}_{bipartite}$), for GHZ states, and for W states, we have the set of bounds, provided by known entanglement results [15, 16]

$$\begin{aligned} N(\mathcal{S}_{bipartite}) &\leq d_1 d_2 / (\sum_i \alpha_i)^2 \\ N(GHZ) &\leq 2^{m-1} \\ N(W) &\leq 2^m (m - 1/m)^{(m-1)} \end{aligned} \tag{6}$$

where α_i are the Schmidt coefficients for any one of the bipartite states in the set $\mathcal{S}_{bipartite}$. In fact, we show that the bound is tight for GHZ by explicit construction. In addition, in the paper we show that these examples prove some known results in a unified way, [17].

4 Probabilistic Case and Locally Accessible Information

It can easily be shown in an analogous way to the deterministic case, that a necessary condition for the ambiguous probabilistic local discrimination of states $\{\rho_i\}$, where for a state ρ_i the probability of correct inference from the measurement is $p(i|i)$, is that the following inequality holds:

$$\sum_i t'(\rho_i) \leq D, \tag{7}$$

where

$$t'(\rho_i) := \min \mathrm{tr}\{M\} \quad \text{such that} \quad I \geq M_i' \geq 0, M_i' \in SEP, \mathrm{tr}\{M_i' \rho_i\} = p(i|i). \tag{8}$$

Another, very much related problem to local state discrimination, is the problem of sending and receiving/decoding messages under LOCC. If we now imagine that the states to be discriminated are actually those chosen to send a message, each state associated to a classical signal, then decoding the message is a similar task to the discrimination of the states. If we can give a bound on the best probability of success p_s for the discrimination,

intuitively we might expect that we can get a bound on the accessible information. This is indeed the case as shown by the known bound [18],

$$H(I:O) \leq H(I) + \log_2(p_s). \tag{9}$$

Given this, for a completely random message encoded on the states $\{\rho_i|i..N\}$, the locally accessible information is bounded from above as

$$H(I:O) \leq \log_2 D - \min_i E_G(\rho_i). \tag{10}$$

We note that in the bipartite pure case, this looks like a special case of the bound given in [19]. There, they give a general bound in terms of the entanglement of formation (which for pure states leads to a generally tighter bound than above). The bound found here is only equivalent for the case where all the encoding states have equal entanglement, are all equally likely and for where geometric entanglement is equal to relative entropy of entanglement (which is equal to entanglement of formation for pure states [8]). However the bound (10) applies to the more general multiparty mixed state case, and in certain mixed state cases can lead to a better bound (since it also takes into account the mixedness of the state).

5 Conclusions

The conditions and bounds found here are apply to the discrimination of general mixed, multipartite states. Further, the simplicity of the bounds and their derivation, leads to sets of weaker conditions, so that for pure states, we can always find at least some calculable condition.

For example, our crucial condition of separability can be changed to a weaker condition, that is possibly easier to compute but still necessary for LOCC, for example PPT, or bi-separability [20]. It can easily be seen that these conditions would follow through to give bounds of entanglement type quantities, defined by the respective sets [20]. In the case of bi-separability, the example for bipartite states above shows it always gives an easily computable bound.

We have given an interpretation of the global robustness of entanglement, the relative entropy of entanglement and the geometric measure of entanglement as a bound to the number of pure states that can be discriminated perfectly by LOCC. Our general mixed state results imply that entanglement guarantees a certain difficulty in the LOCC discrimination of states. It is known that this problem is fundamental to various quantum information tasks, (such as quantum data hiding[13] e.t.c.), which may indicate that all entangled states are useful. This is the topic of ongoing investigations. In this direction, we have seen it is also possible to extend theorem 1 to the case of imperfect discrimination. This leads to bounds on the LOCC accessible information.

References

[1] M. Hayashi, D. Markham, M. Murao, M. Owari and S. Virmani. LOCC State Discrimination and Multiparty Entanglement Measures. quant-ph/0506170 2005. .

[2] C. H. Bennett and G. Brassard, in Proc. IEEE International Conference on Computers, Systems and Signal Processing (IEEE Press, New York,1984); A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).

[3] P. Hayden and C. King, quant-ph/0409026; J. Watrous, quant-ph/0503092.

[4] J. I. Cirac, A. K. Ekert, S. F. Huelga and C. Macchiavello, Phys. Rev. A **59**, 4249 (1999).

[5] M. B. Plenio and V. Vedral, Contemp. Phys. **39**, 431 (1998)

[6] W. Dür, G. Vidal and J. I. Cirac, Phys. Rev. A **62**, 062314 (2000); S. Ishizaka and M. B. Plenio, quant-ph/0503025.

[7] G. Vidal and R. Tarrach, Phys. Rev. A **59**, 141 (1999).

[8] V. Vedral and M. B. Plenio, Phys. Rev. A **57**, 1619 (1998).

[9] A. Shimony, Ann. NY. Acad. Sci **755**, 675 (1995). ; H. Barnum and N. Linden, J. Phys. A: Math. Gen. **34**, 6787 (2001); T-C. Wei and P. M. Goldbart, Phys. Rev. A **68**, 042307 (2003).

[10] C.H. Bennett, D.P. DiVincenzo, C.A. Fuchs, T. Mor, E. Rains, P.W. Shor, J.A. Smolin and W.K. Wootters, Phys. Rev. A. **59**, 1070 (1999).

[11] J. Walgate, A. J. Short, L. Hardy and V. Vedral, Phys. Rev. Lett. **85**, 4972 (2000); S. Virmani, M. F. Sacchi, M. B. Plenio, D. Markham, Phys. Lett. A **288**, 62 (2001); Z. Ji, H. Cao and M. Ying, Phys. Rev. A **71**, 032323 (2005).

[12] Y-X. Chen and D. Yang, Phys. Rev. A. **65**, 022320 (2002); F. Anselmi, A. Chefles and M. B. Plenio, New J. Phys. **6**, 164 (2004); A. Chefles, Phys. Rev. A **69**, 050307(R) (2004); H. Fan, Phys. Rev. Lett. **92**, 177905 (2004); M. Hillery and J. Mimih, quant-ph/0210179; P.X. Chen and C.Z. Li, Phys. Rev. A **68**,062107 (2003); S. de Rinaldis, quant-ph/0304027.

[13] B. M. Terhal, D. P. DiVincenzo and D. W. Leung, Phys. Rev. Lett. **86**, 5807 (2001).

[14] Y. Tsuda, K. Matsumoto and M. Hayashi, quant-ph/0504203.

[15] A. W. Harrow and M. A. Nielsen, Phys. Rev. A **68**, 012308 (2003); M. Steiner, Phys. Rev. A **67**, 054305 (2003); G. Vidal and R. Tarrach, Phys. Rev. A **59**, 141 (1999).

[16] T-C. Wei, M. Ericsson, P. M. Goldbart and W. J. Munro, Quant. Inform. Compu. **4**, 252 (2004).

[17] S. Ghosh, G. Kar, A. Roy, A. Sen (De) and U. Sen,Phys. Rev. Lett **87**, 277902 (2001); S. Ghosh, G. Kar, A. Roy and D. Sarkar, Phys. Rev. A **70**, 022304 (2004); M. Nathanson, quant-ph/041110; M. Horodecki, A. Sen (De), U. Sen and K. Horodecki, Phys. Rev. Lett. **90**, 047902 (2003).

[18] D.P. DiVincenzo, D. W. Leung and B. M. Terhal, IEEE Trans. Inform. Theory **48**, 580 (2002).

[19] P. Badziąg, M. Horodecki, A. Sen, U. Sen, Phys. Rev. Lett. **91**, 117901 (2003).

[20] E. M. Rains, Phys. Rev. A **60**, 179 (1999); Phys. Rev. A **63**, 019902(E) (2000).

State Discrimination without Classical Knowledge

Akihisa Hayashi[1] * Minoru Horibe[1] Takaaki Hashimoto[1]

[1] *Department of Applied Physics, University of Fukui.*
3-9-1 Bunkyo, Fukui, 910-8507, Japan.

Abstract. We address a problem of identifying a given pure state with one of two reference pure states, when no classical knowledge on the reference states is given, but a certain number of copies of them are available. We consider two versions of this problem; one is without no-error conditions ("state identification") and the other is with no-error conditions ("unambiguous state identification"). In both versions we give a complete solution for the optimal mean success probability for an arbitrary number of copies of the reference states in general dimension.

Keywords: state discrimination, unambiguous discrimination

1 Introduction

Suppose we are presented with an unknown quantum pure state ρ on a d dimensional vector space C^d. We know that the input state ρ is either one of two reference states ρ_1 and ρ_2, each being also a pure state on C^d. What is the best strategy to identify the input state with one of the two reference states?

We can consider two cases depending on what kind of information on the reference states is available. In the first case, we are given complete classical knowledge on the reference states ρ_1 and ρ_2. This is the problem of quantum state discrimination, which was solved by Helstrom [1].

On the other hand, we can also consider the case where only a certain number (N) of copies of ρ_1 and ρ_2 are presented, with no classical knowledge on them available. In this case, we could obtain only limited classical information on the reference states, due to the no-cloning theorem. In this report, this problem is called "state identification". If the number of copies N is infinite, the problem is reduced to quantum state discrimination.

In the case of qubit ($d = 2$), similar problems but in different setups have been studied [6]. Sasaki et al. studied quantum matching problem , where a certain number of copies of input and reference qubit states are independently distributed on a great circle of the Bloch sphere.

In the problem of unambiguous discrimination, we are not allowed to make an error but our measurement can produce an inconclusive result [2, 3, 4]. Similarly we can consider the unambiguous state identification problem with no-error conditions. Bergou and Hillery [7] recently discussed this problem when the number of copies $N = 1$ and the dimension $d = 2$ (qubits). They called the optimal strategy a programmable state discriminator since the strategy is not "hard wired" but supplied by the reference states stored in registers in the machine.

In this report we assume that the input state ρ is guaranteed to be prepared in one of the two reference states ρ_1 and ρ_2, and the two reference states are independently distributed on the whole d-dimensional pure state space C^d in a unitary invariant way. No classical knowledge on the reference states are available, but only a certain number (N) of their copies are presented. Our task is to successfully identify the input state with one of the reference states.

We will study two versions of state identification problems; with and without the no-error conditions [8, 9]. In both versions we will give a complete solution of optimal strategies and the mean success probability as a function of the number of copies N of the reference states and the dimension d of the state space. The large N limit of the optical mean success probability is also studied.

2 Mean identification probability

We assume that the input state is given in system 0 and N copies of each reference state ρ_a are prepared in systems $a_1, a_2, \ldots, a_N$ ($a = 1, 2$). We denote the subsystem $a_1 \otimes a_2 \otimes \cdots \otimes a_N$ simply by a ($a = 1, 2$). We specify the system which an operator acts by the system number in the parenthesis; namely, $\rho(0)$ means that this is an operator acting on system 0 for example.

Our task is then distinguish between the states $\rho_1(0)\rho_1(1)^{\otimes N}\rho_2(2)^{\otimes N}$ and $\rho_2(0)\rho_1(1)^{\otimes N}\rho_2(2)^{\otimes N}$. Clearly the corresponding set of POVM $\{E_1, E_2\}$ can be assumed to act on the subsystem $V_{\rm sym}$ where each of systems 1 and 2 is totally symmetric.

The mean success probability is given by

$$p^{(N)}(d) = \frac{1}{2}\sum_{a=1,2}\left\langle \mathrm{tr}\left[E_a\rho_a(0)\rho_1(1)^N\rho_2(2)^N\right]\right\rangle. \quad (1)$$

Here $< \cdots >$ represents the average with respect to the two reference states ρ_1 and ρ_2, each of which is independently distributed over C^d in a unitary invariant way (see [10] for the precise definition).

It is very helpful to use the following formula for the unitary average of the tensor product of n identically prepared pure states [10]:

$$< \rho^{\otimes n} >= \frac{\mathcal{S}_n}{d_n}, \quad (2)$$

where $\mathcal{S}_n$ is the projector onto the totally symmetric subspace of $\{C^d\}^{\otimes n}$ and the dimension of the subspace is given by $d_n = \mathrm{tr}\,[\mathcal{S}_n] = {}_{n+d-1}C_{d-1}$.

Then the optimal mean success probability can be cal-

*hayashi@soliton.fukui-u.ac.jp

culated as follows:

$$\begin{aligned} p^{(N)}(d) &= \frac{1}{2} + \frac{1}{2d_{N+1}d_N}\mathrm{tr}\,[E_1 D] \\ &\leq \frac{1}{2} + \frac{1}{4d_{N+1}d_N}\mathrm{tr}|D| \\ &\equiv p^{(N)}_{\max}(d). \end{aligned} \tag{3}$$

The operator D is defined to be

$$D \equiv S_{N+1}(01) - S_{N+1}(02), \tag{4}$$

where $\mathcal{S}_{N+1}(01)$ is the projector onto the totally symmetric subspace on systems $(0,1) = (0, 1_1, 1_2, \ldots, 1_N)$ and other projectors $\mathcal{S}$ are defined similarly. The optimal POVM measurement can be taken to be a projective measurement, since the equality in Eq. (3) holds if E_1 is the projector onto the positive-eigenvalue invariant space of D. Note that $\mathcal{S}_N(1) = 1$ and $\mathcal{S}_N(2) = 1$ in V_{sym}.

Thus we need to determine non-zero eigenvalues of the operator D in V_{sym} in order to evaluate the optimal mean identification probability $p^{(N)}_{\max}(d)$.

2.1 Case of qubits ($d = 2$)

In this case, the algebra of angular momentum is very useful. In the subspace V_{sym}, the total angular momentum of systems 1 and 2 is $N/2$. Furthermore, the projector $S_{N+1}(0a)$ can be written as

$$S_{N+1}(0a) = \frac{1}{N+1}\left(2\boldsymbol{j}(a)\cdot\boldsymbol{s}(0) + \frac{N}{2} + 1\right), \tag{5}$$

where $\boldsymbol{s}(0)$ is the spin operator of system 0 and $\boldsymbol{j}(a)$ is the angular momentum operator of system $a = 1, 2$. Let us calculate D^2 by the use the properties of the Pauli matrices and the commutation relations of $\boldsymbol{j}(a)$. The result is given by

$$D^2 = \frac{1}{(N+1)^2}\left((N+\frac{1}{2})(N+\frac{3}{2}) - \boldsymbol{J}^2\right), \tag{6}$$

where $\boldsymbol{J} = \boldsymbol{j}(1) + \boldsymbol{j}(2) + \boldsymbol{s}(0)$ is the total angular momentum operator.

Thus we find the optimal mean identification probability for the qubit case to be

$$\begin{aligned} p^{(N)}_{\max}(d=2) &= \frac{1}{2} + \frac{1}{2(N+1)(N+2)} \times \\ &\sum_{J=\frac{1}{2}}^{N-\frac{1}{2}} (2J+1)\sqrt{1 - \left(\frac{J+\frac{1}{2}}{N+1}\right)^2}. \end{aligned} \tag{7}$$

We list explicit values of $p^{(N)}_{\max}(d=2)$ for some small N's.

$$\begin{aligned} p^{(1)}_{\max}(2) &= \frac{1}{2} + \frac{1}{12}\sqrt{3} \simeq 0.644, \\ p^{(2)}_{\max}(2) &= \frac{1}{2} + \frac{1}{18}(\sqrt{2} + \sqrt{5}) \simeq 0.703, \\ p^{(3)}_{\max}(2) &= \frac{1}{2} + \frac{1}{80}(\sqrt{15} + 4\sqrt{3} + 3\sqrt{7}) \simeq 0.734\,. \end{aligned}$$

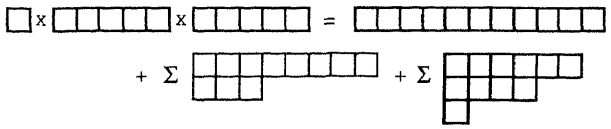

Figure 1: Decomposition of the product of three $U(d)$ irreducible representations $[1] \otimes [N] \otimes [N]$. The decomposition leads to the three orthogonal subspaces V_n $(n = 1, 2, 3)$ according to the number of rows n of Young diagram.

2.2 Case of arbitrary dimension d

Let us introduce the orthonormal base of the total space $(C^d)^{\otimes(2N+1)}$ according to irreducible representations of the symmetric group S_{2N+1} and the unitary group $U(d)$. We write states in this base as

$$|\,\lambda, a, b\,\rangle. \tag{8}$$

Here λ represents an irreducible representation of S_{2N+1}, which is specified by a Young diagram. By the expression $\lambda = [\lambda_1, \lambda_2, \ldots]$, we denote a Young diagram consisting of a set of rows with their lengths given by $\lambda_1, \lambda_2, \ldots$. The label a indexes orthogonal vectors in a particular S_{2N+1} representation space and it runs from 1 to the dimension of the S_{2N+1} representation. It is known that the λ also specifies irreducible representations of the unitary group $U(d)$, and its vectors are indexed by b, which runs from 1 to $m_\lambda(d)$, the multiplicity of representation λ of S_{2N+1} on $(C^d)^{\otimes(2N+1)}$.

Possible Young diagrams λ appearing in V_{sym} and the range of the index a associated with a particular λ can be determined by decomposing the product of three $U(d)$ irreducible representations $[1] \otimes [N] \otimes [N]$. We decompose the space V_{sym} into three orthogonal subspaces V_n $(n = 1, 2, 3)$ according to the number of rows n of Young diagram (see Fig. 1).

It turns out that the all eigenvalues of the operator D are zero in spaces V_1 and V_3. And the eigenvalues of D in V_2 are the same as those in the qubit case $(d = 2)$, since the operator D involves only permutations. Using the results in the preceding subsection, we find the mean optimal success probability to be

$$\begin{aligned} p^{(N)}_{\max}(d) &= \frac{1}{2} + \frac{1}{2d_{N+1}d_N} \times \\ &\sum_{\lambda_1=N+1}^{2N} m_{[\lambda_1,\lambda_2]}(d)\sqrt{1 - \left(\frac{\lambda_1 - N}{N+1}\right)^2}, \end{aligned} \tag{9}$$

where $m_{[\lambda_1,\lambda_2]}(d), (\lambda_2 = 2N + 1 - \lambda_1)$ is the multiplicity of the S_{2N+1} irreducible representation $[\lambda_1, \lambda_2]$, which is realized in V_2.

$$m_{[\lambda_1,\lambda_2]}(d) = \frac{(\lambda_1 + d - 1)!(\lambda_2 + d - 2)!(\lambda_1 - \lambda_2 + 1)}{(d-1)!(d-2)!(\lambda_1 + 1)!\lambda_2!}. \tag{10}$$

2.3 Large N limit

In order to compare the obtained mean identification probability with the discrimination probability [1], we average the optimal discrimination probability for given

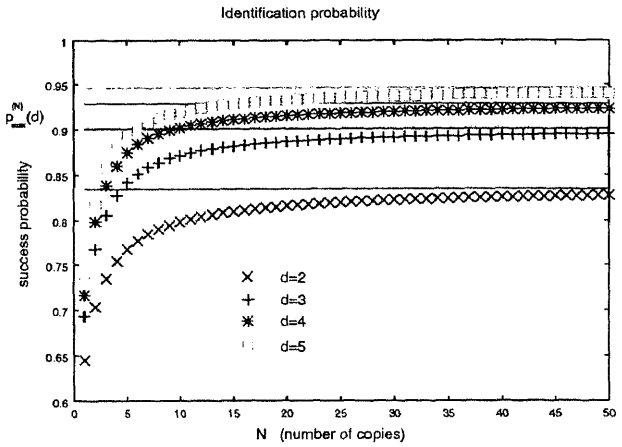

Figure 2: The optimal mean identification probability $p_{\max}^{(N)}(d)$ as a function of the number of copies (N) of the reference states. As N increases, $p_{\max}^{(N)}(d)$ approaches the mean optimal discrimination probability shown by the solid lines.

reference states ρ_1 and ρ_2

$$p_{\max}(d) = \left\langle \frac{1}{2}\left(1 + \frac{1}{2}\mathrm{tr}|\rho_1 - \rho_2|\right)\right\rangle. \tag{11}$$

We find

$$p_{\max}(d) = \frac{1}{2} + \frac{d-1}{2d-1}. \tag{12}$$

Now what is the limit value of the mean identification probability $p_{\max}^{(N)}(d)$ when the number of copies N goes to infinity? In the large N limit, we can replace the sum in $p_{\max}^{(N)}(d)$ of Eq. (9) by an continuous integral. The result is

$$\begin{aligned} p_{\max}^{(N)}(d) &\to \frac{1}{2} + (d-1)\int_0^1 dx\, x(1-x^2)^{d-\frac{3}{2}} \\ &= \frac{1}{2} + \frac{d-1}{2d-1} \\ &= p_{\max}(d),\ (N \to \infty). \end{aligned} \tag{13}$$

This is an expected result, since when the number of copies of the reference states are infinite, we can acquire complete knowledge of the reference states; the problem is reduced to the standard discrimination problem.

Figure 2 displays how the identification probability approaches the discrimination probability as the number of copies increases.

3 Mean unambiguous identification probability

Here we study how to unambiguously distinguish between the states $\rho_1(0)\rho_1(1)^{\otimes N}\rho_2(2)^{\otimes N}$ and $\rho_2(0)\rho_1(1)^{\otimes N}\rho_2(2)^{\otimes N}$, which we call the unambiguous identification problem. The POVM acting on V_{sym} now consists of three elements $\{E_0, E_1, E_2\}$. When the outcome of the POVM is $a(=1,2)$, we identify the input ρ with ρ_a with certainty. Outcome 0 of the POVM means we have an inconclusive result.

The mean success probability is given by the same form as Eq. (1) and we can perform the average by the formula (2) as before.

$$\begin{aligned} q^{(N)}(d) &= \frac{1}{2}\sum_{a=1,2}\left\langle \mathrm{tr}\left[E_a\rho_a(0)\rho_1(1)^N\rho_2(2)^N\right]\right\rangle, \\ &= \frac{1}{2d_{N+1}d_N}\Big(\mathrm{tr}\left[E_1\mathcal{S}_{N+1}(01)\right] + \mathrm{tr}\left[E_2\mathcal{S}_{N+1}(02)\right]\Big). \end{aligned} \tag{14}$$

The big difference is that we are not allowed to make an error in the unambiguous identification setting. This is formulated by the following no-error conditions: for any ρ_1 and ρ_2

$$\begin{aligned} \mathrm{tr}\left[E_1\rho_2(0)\rho_1^{\otimes N}(1)\rho_2^{\otimes N}(2)\right] &= 0, \\ \mathrm{tr}\left[E_2\rho_1(0)\rho_1^{\otimes N}(1)\rho_2^{\otimes N}(2)\right] &= 0, \end{aligned} \tag{15}$$

which are evidently equivalent to

$$\begin{aligned} E_1\mathcal{S}_{N+1}(02) = \mathcal{S}_{N+1}(02)E_1 = 0, \\ E_2\mathcal{S}_{N+1}(01) = \mathcal{S}_{N+1}(01)E_2 = 0. \end{aligned} \tag{16}$$

We observe that the set of POVM's satisfying the no-error conditions Eq. (16) is convex. By this convexity of the legitimate POVM's and the symmetries intrinsic to the problem, we can impose two properties on the optimal POVM without loss of generality (see [9]). One is the exchange symmetry between systems 1 and 2

$$E_2 = TE_1T,\ E_0 = TE_0T, \tag{17}$$

where T is the operator that exchanges systems 1 and 2. The other is that the optimal POVM element is a $U(d)$ scalar; namely, E_a commutes with $U^{\otimes(2N+1)}$ for any unitary U. These two properties are very helpful in determining the optimal POVM.

3.1 Optimal mean unambiguous identification probability

We just sketch the outline of the derivation (see [9] for details). It is convenient to use the base $|\lambda, a, b\rangle$ and the decomposition $V_{\mathrm{sym}} = V_1 \oplus V_2 \oplus V_3$ introduced in Sec. 2.2.

We find that the no-error conditions Eq. (16) and the symmetries of the POVM lead to the following form for the optimal POVM elements:

$$\begin{aligned} E_1 &= e\Big(1 - \mathcal{S}_{N+1}(02)\Big), \\ E_2 &= e\Big(1 - \mathcal{S}_{N+1}(01)\Big), \\ e &= \sum_\lambda e_\lambda \Gamma_\lambda, \end{aligned} \tag{18}$$

where Γ_λ is the projection operator onto the $U(d)$ representation space specified by λ. The positivity of $E_0 = 1 - E_1 - E_2$ requires the operator e should satisfy

$$\frac{1}{1+|A|} \geq e, \tag{19}$$

in subspaces V_2 and V_3, where we introduced the operator A in the subspace $V_{\rm sym}$ as

$$A \equiv \mathcal{S}_{N+1}(01) + \mathcal{S}_{N+1}(02) - 1. \tag{20}$$

The mean success probability Eq.(14) is then expressed as

$$q^{(N)}(d) = \frac{1}{2d_{N+1}d_N} \mathrm{tr}\left[e(1-A^2)\right]. \tag{21}$$

In the above equation, we find that the subspaces V_1 and V_2 have no contribution to the trace sum. And $|A|$ in the upper bound of e in Eq.(19) commutes with $1-A^2$ in the trace. Therefore, we immediately obtain the optimal mean success probability as follows:

$$\begin{aligned} q^{(N)}(d) &\leq \frac{1}{2d_{N+1}d_N} \mathrm{tr}\left[\frac{1}{1+|A|}(1-A^2)\right] \\ &= \frac{1}{2d_{N+1}d_N} \mathrm{tr}\left[1-|A|\right] \\ &\equiv q^{(N)}_{\max}(d). \end{aligned} \tag{22}$$

The optimal success probability is thus attained by the POVM elements given by Eq. (18) with $e = \frac{1}{1+|A|}$ and $E_0 = \frac{A+|A|}{1+|A|}$.

We can compute the non-zero eigenvalues of A in V_2, assuming $d = 2$ in the same way as the computation of the eigenvalues of D. When $d = 2$, A^2 can be calculated by Eq. (5) and the algebra of angular momentum operators.

$$A^2 = \frac{1}{(N+1)^2}\left(\mathbf{J}^2 + \frac{1}{4}\right), \tag{23}$$

from which we can easily read off the eigenvalues of $|A|$.

Finally we obtain the formula for the optimal success probability to be

$$\begin{aligned} q^{(N)}_{\max}(d) &= \frac{1}{d_{N+1}d_N} \times \\ &\sum_{\lambda_1=N+1}^{2N} m_{[\lambda_1\lambda_2]}(d)\left(1 - \frac{\lambda_1 - N}{N+1}\right), \end{aligned} \tag{24}$$

where $m_{[\lambda_1,\lambda_2]}(d)$ ($\lambda_2 = 2N+1-\lambda_1$) is the multiplicity given in Eq. (10).

3.2 Large N limit

Let us study the asymptotic value of $q^{(N)}_{\max}(d)$ when the number of copies N is very large and compare it with the unambiguous discrimination probability [2, 3, 4]. First we average the optimal unambiguous discrimination probability for given reference states $\rho_1 = |\phi_1\rangle\langle\phi_1|$ and $\rho_2 = |\phi_2\rangle\langle\phi_2|$

$$q_{\max}(d) = \left\langle 1 - |\langle\phi_1|\phi_2\rangle| \right\rangle. \tag{25}$$

We obtain

$$q_{\max}(d) = 1 - \frac{2^{d-1}(d-1)!}{(2d-1)!!}, \tag{26}$$

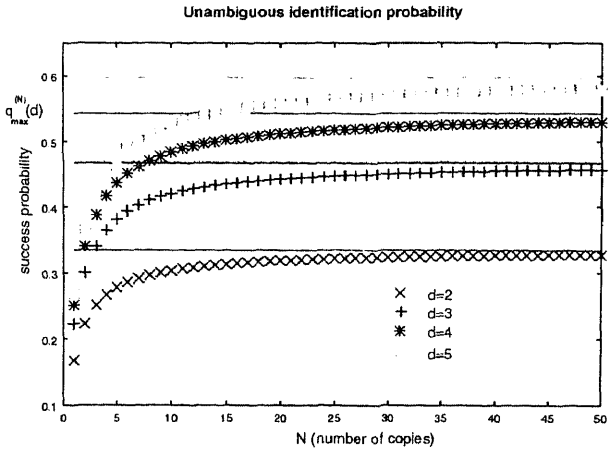

Figure 3: The optimal mean unambiguous identification probability $q^{(N)}_{\max}(d)$ as a function of the number of copies (N) of the reference states. As N increases, $q^{(N)}_{\max}(d)$ approaches the mean optimal unambiguous discrimination probability shown by the solid lines.

which is certainly less than the mean discrimination probability Eq. (12).

On the other hand, the large N limit of $q^{(N)}_{\max}(d)$ can be calculated by replacing the sum in Eq. (24) by an continuous integral. The result is again an expected one.

$$\begin{aligned} q^{(N)}_{\max}(d) &\to 2(d-1)\int_0^1 dx\,(1+x)^{d-2}(1-x)^{d-1} \\ &= 1 - \frac{2^{d-1}(d-1)!}{(2d-1)!!} \quad (N\to\infty), \end{aligned} \tag{27}$$

which is equal to $q_{\max}(d)$ given by Eq. (26).

References

[1] C. W. Helstrom. *Quantum detection and estimation theory* (Academic press, New York, 1976).

[2] I. D. Ivanovic. Phys. Lett. A **123**, 257 (1987).

[3] D. Dieks. Phys. Lett. A **126**, 303 (1988).

[4] A. Peres. Phys. Lett. A **128**, 19 (1988).

[5] Janos A. Bergou and Mark Hillery. Phys. Rev. Lett. **94**, 160501 (2005).

[6] Masahide Sasaki and Alberto Carlini. Phys. Rev. **A66**, 022303 (2002).

[7] Janos A. Bergou and Mark Hillery. Phys. Rev. Lett. **94**, 160501 (2005).

[8] A. Hayashi, M. Horibe, and T. Hashimoto. quant-ph/0507237, to be published in Phys. Rev. A **72**, (2005).

[9] A. Hayashi, M. Horibe, and T. Hashimoto. to be published.

[10] A. Hayashi, T. Hashimoto, and M. Horibe. Phys. Rev. A **72**, 032325 (2005).

Implementation of binary projection measurement with linear optics and photon counting

Masahiro Takeoka[1 2 *] Masahide Sasaki[1 2 †] Norbert Lütkenhaus[3 ‡]

[1] *Quantum Information Technology Group, National Institute of Information and Communications Technology (NICT) 4-2-1 Nukui-kitamachi, Konganei, Tokyo 184-8795, Japan.*
[2] *CREST, Japan Science and Technology Agency, 1-9-9 Yaesu, Chuoh-ku, Tokyo 103-0028, Japan.*
[3] *Quantum Information Theory Group, Zentrum für Moderne Optik, Universität Erlangen-Nürnberg, 91058 Erlangen, Germany.*

Abstract. We discuss the implementation of binary projective measurement with linear optics, photon counting, and classical ancillary states. The problem can be regarded as a discrimination of two orthogonal pure quantum states. We show that any sets of two orthogonal states can be perfectly discriminated via only linear optics, photon counting, coherent ancillary states, and feedforward, at least, in the asymptotic limit of large number of these resources.

Keywords: state discrimination, linear optics, photon counting, feedforward

In various applications of optical quantum information processing, one is often required to make a measurement which projects the state onto complicated superposition states or entangled states. It is, however, a nontrivial problem how to implement such a device physically. One of attractive approach is to use linear optics and classical feedforward associated with photon counting which effectively generate optical nonlinearity in measurement process [1].

Recently, the criteria to decide whether one can implement a given projective measurement with *unit success probability* has been derived under the condition that one can use linear optics, photon counting, arbitrary ancillary states and classical feedforward but the amount of these resources are always finite [2]. In this talk, we extend this scenario to the asymptotic limit of large number of resources. We discuss the implementation of a given binary projection measurement $\{|\Psi\rangle, |\Phi\rangle\}$ via linear optics and photon counting. This problem is equivalent to that of discriminating two orthogonal quantum signals $\{|\Psi\rangle, |\Phi\rangle\}$ unambiguously [3]. We show that, in principle, any set of $\{|\Psi\rangle, |\Phi\rangle\}$, including the sets in which the above criteria state *no-go*, can be perfectly discriminated by using only linear optics, photon counting, and *coherent ancillary states*, at least, in the asymptotic limit of large number of these resources.

Before starting a discussion of linear optics implementation, it would be worth mentioning the distinguishability of two orthogonal states via local operations and classical communication (LOCC). The necessary condition for exact local distinguishability is that, after doing a measurement at some local site, every possible remaining states must be orthogonal to each other. Walgate et al. [4] showed that there always exists a local projective measurement satisfying this orthogonality condition for two orthogonal muti-mode states and thus one can perfectly discriminate them via a series of local projective measurements where the choice of the measurement basis at each local site is conditioned on the previous measurement outcomes. This result means that if one can show a physical scheme that can exactly discriminate any two orthogonal *single-mode* states, its sequential application can achieve an exact discrimination of any two orthogonal *multi-mode* states. In the following, therefore, we concentrate on a discrimination of two single-mode sates.

An arbitrary set of two single-mode orthogonal states are described by

$$|\Psi\rangle = \sum_{m=0}^{\infty} c_m |m\rangle_0, \qquad |\Phi\rangle = \sum_{m=0}^{\infty} d_m |m\rangle_0, \tag{1}$$

where $|m\rangle$ is a m-photon number state and $\langle\Psi|\Phi\rangle = \sum_{m=0}^{\infty} c_m^* d_m = 0$. Figure 1 is the schematic of the measurement apparatus. The states are equally split into N modes by asymmetric $N-1$ beamsplitters [5],

$$\begin{aligned}&\hat{B}_{N-1,0}(\theta_{N-1}) \cdots \hat{B}_{1,0}(\theta_1) |0\rangle^{\otimes N-1} |\Psi\rangle_0 \\ &= e^{-\hat{a}_{N-1}^\dagger \hat{a}_0} \cdots e^{-\hat{a}_1^\dagger \hat{a}_0} e^{\hat{a}_0^\dagger \hat{a}_0 \ln(1/\sqrt{N})} |0\rangle^{\otimes N-1} |\Psi\rangle_0.\end{aligned} \tag{2}$$

where $\hat{B}_{n,0}(\theta_n) = \exp[\theta_n(\hat{a}_n^\dagger \hat{a}_0 - \hat{a}_n \hat{a}_0^\dagger)]$ and

$$\tan\theta_n = \frac{1}{\sqrt{N-n}}. \tag{3}$$

The outputs at N ports are completely symmetric and the effective power reflectance for each output is given by $\sin^2\theta = 1/N$. At each port, one makes some measurement in sequence by using linear optics and photon counters, where the information about the measurement outcome at each port is feedforwarded to design the next measurement.

First, we briefly sketch how two states are discriminated by such a scheme in the limit of $N \to \infty$ and then provide a rigorous proof afterward. Suppose to put $|\Psi\rangle$ and $|\Phi\rangle$ into the first beamsplitter. For sufficiently small $1/N$, the reflectance of more than two photons can be

*takeoka@nict.go.jp
†psasaki@nict.go.jp
‡norbert.luetkenhaus@physik.uni-erlangen.de

negligible. The states after beamsplitting are given by

$$\hat{B}_{1,0}(\theta_1)|0\rangle_1|\Psi\rangle_0 \approx |0\rangle_1|\eta_0\rangle_0 + \frac{1}{\sqrt{N}}|1\rangle_1|\eta_1\rangle_0, \quad (4)$$

$$\hat{B}_{1,0}(\theta_1)|0\rangle_1|\Phi\rangle_0 \approx |0\rangle_1|\nu_0\rangle_0 + \frac{1}{\sqrt{N}}|1\rangle_1|\nu_1\rangle_0. \quad (5)$$

where, $\langle\eta_0|\nu_0\rangle + \langle\eta_1|\nu_1\rangle/N \approx 0$, since a beamsplitting operation is unitary. Then a partial state on the mode 1 is measured. The measurement here is required to keep the orthogonality of any conditional outputs of $|\Psi\rangle$ and $|\Phi\rangle$. The local measurement satisfying this condition is given by a two-dimensional projective measurement [4],

$$|\pi_0\rangle = M_c|0\rangle - \frac{M_s e^{i\varphi}}{\sqrt{N}}|1\rangle, \quad (6)$$

$$|\pi_1\rangle = \frac{M_s e^{-i\varphi}}{\sqrt{N}}|0\rangle + M_c|1\rangle, \quad (7)$$

where, M_s, M_c and φ are the functions of $c_i^* d_j$.

The projective measurement of Eqs. (6) and (7) can be implemented by the displacement operation $\hat{D}(\beta_1/\sqrt{N})$ and photon counting as shown in Fig. 1(b). Since both the signal and displacement are weak enough, we have

$$\hat{D}^\dagger\left(\frac{\beta_1}{\sqrt{N}}\right)|0\rangle \approx e^{-|\beta_1|^2/2N}\left(|0\rangle - \frac{\beta_1}{\sqrt{N}}|1\rangle\right), \quad (8)$$

$$\hat{D}^\dagger\left(\frac{\beta_1}{\sqrt{N}}\right)|1\rangle \approx -e^{-|\beta_1|^2/2N}\left(\frac{\beta_1^*}{\sqrt{N}}|0\rangle + |1\rangle\right), \quad (9)$$

as the measurement vectors of such apparatus. Apparently, they have similar structures to that of Eqs. (6) and (7) and, therefore, by choosing appropriate β_1, one can make a projection measurement that satisfies the orthogonality condition $\langle\Psi'|\Phi'\rangle = 0$ for any conditional states $|\Psi'\rangle$, $|\Phi'\rangle$ after the projection measurement.

The conditional states after the first measurement can be rewritten as

$$|\Psi'\rangle = \sum_{m=0}^{\infty} c'_m|m\rangle, \qquad |\Phi'\rangle = \sum_{m=0}^{\infty} d'_m|m\rangle. \quad (10)$$

Since an N-splitter splits a state symmetrically, one can repeat the same procedure for the remaining state with the second beamsplitter, the displacement operation $\hat{D}(\beta_2/\sqrt{N})$, where β_2 is conditioned on the previous measurement outcome, and a photon counter. After repeating the same procedure to the modes 1 to $N-1$ with appropriate β_i's, the final states at the mode 0 include maximally one photon and are still orthogonal to each other. As a consequence, applying the final (N-th) displacement and photon counting, one can exactly discriminate $|\Psi\rangle$ and $|\Phi\rangle$ with unit success probability.

It should be noted that this is a generalized version of the scheme so-called "Dolinar receiver" [6, 7, 8, 9] which was originally proposed as a physical model attaining the minimum error discrimination of the binary coherent signals $\{|\alpha\rangle, |-\alpha\rangle\}$.

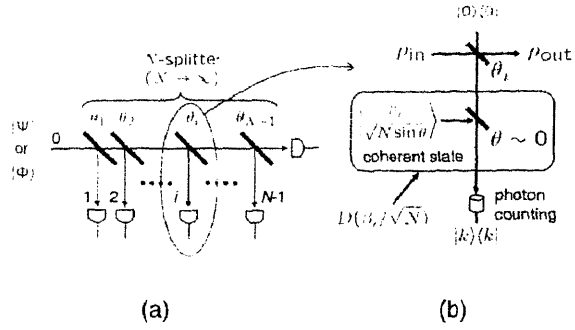

Figure 1: (a) N-splitter, and (b) a measurement apparatus at each step. A displacement operation $\hat{D}(\beta_n/\sqrt{N})$ is realized by combining the signal with a coherent state local oscillator $|\beta_n\rangle$ via a beamsplitter with sufficiently small power reflectance $1/N$.

Now, we discuss the scheme again more rigorously. In the above description, we neglected reflections of more than two photons at each beamsplitter. When it is included, the state space at each port does not remain two-dimensional, and thus the two conditional states after each measurement step are not orthogonal any more. We show that the discrimination error due to such non-orthogonality converges to zero in the limit of lagre N.

Let us denote the operation of the i-th photon counting as

$${}_i\langle k|\hat{D}(\beta_i/\sqrt{N})\hat{B}_{i,0}(\theta_i)|0\rangle_i|\Psi\rangle \equiv \hat{E}_k^{(i)}|\Psi\rangle, \quad (11)$$

where k is the number of the detected photons. Here, we treat only physical states as inputs, that is, the average power of $|\Psi\rangle$ and $|\Phi\rangle$ are always finite. Moreover, we assume that all local oscillators satisfy

$$|\beta_i|^2 \le C_\beta \quad \forall\, i, N, \quad (12)$$

where C_β is some constant independent of N. This also implies that the average power of each local oscillator is finite. At the i-th measurement step, a probability of detecting k photons is given by

$$\begin{aligned} P_k^{(i)} &= \mathrm{Tr}[\hat{E}_k^{(i)}|\Psi\rangle\langle\Psi|\hat{E}_k^{(i)\dagger}] \\ &= \left|\exp\left(-\frac{|\beta_i|^2}{2N}\right)\left(\frac{1}{N}\right)^{k/2}\left(\beta_i - \sqrt{1-\frac{1}{N}}\hat{a}_0\right)^k \right. \\ &\quad \left. \times \exp\left(\hat{a}_0^\dagger\hat{a}_0 \ln\sqrt{1-\frac{1}{N}}\right)|\Psi^{(i-1)}\rangle_0\right|^2 \\ &= \frac{C_k^{(i)}(\beta_i, |\Psi^{(i-1)}\rangle)}{N^k} + O\left(\frac{1}{N^{k+1}}\right) \\ &\le \frac{C_k^{UB}}{N^k} \quad \forall\, i, k \end{aligned} \quad (13)$$

where $|\Psi^{(i-1)}\rangle$ is the conditional output for the input of $|\Psi\rangle$ after the $i-1$-th measurement, and $C_k^{(i)}(\beta_i, |\Psi^{(i-1)}\rangle)$ is a constant depending on β_i and $|\Psi^{(i-1)}\rangle$ but independent of N. Due to the power constraints of the inputs and local oscillators, one can obtain

the inequality at the final line in which $P_k^{(i)}$ is bounded by some constant C_k^{UB} that is independent of N. Apparently, Eq. (13) is also held for $|\Phi\rangle$.

After finishing a whole process of N measurement steps, one can classify the results according to the sequential patterns of detected photon numbers. Let us denote the events in which all the photon counters detected less than one photon by 'success' events and the others by 'failure' events. The probability of resulting the success event P_{succ} is given by a sum of 2^N success events as

$$\begin{aligned} P_{succ} &= \sum^{2^N} P(\sharp^{(N)}) \\ &= \sum^{2^{N-1}} P(\sharp^{(N-1)}) \left(P(0|\sharp^{(N-1)}) + P(1|\sharp^{(N-1)}) \right) \\ &= \sum^{2^{N-1}} P(\sharp^{(N-1)}) \left(1 - P(2|\sharp^{(N-1)}) - \cdots \right) \\ &\geq \sum^{2^{N-1}} P(\sharp^{(N-1)}) \left(1 - \frac{C_2^{UB}}{N^2} + O\left(\frac{1}{N^3}\right) \right) \\ &\geq \left(1 - \frac{C_2^{UB}}{N^2} + O\left(\frac{1}{N^3}\right) \right)^N \\ &= 1 - \frac{C_2^{UB}}{N} + O\left(\frac{1}{N^2}\right), \end{aligned} \tag{14}$$

where $\sharp^{(N)}$ denotes a pattern of the length N sequence of 0- and 1-photon detection results and $P(k|\sharp^{(N-1)})$ is the probability of detecting k photons at the N-th measurement conditioned on the detection pattern of the first $N-1$ measurements. Then, we can easily show that the probability of resulting the failure event P_{fail} is bounded by

$$P_{fail} = 1 - P_{succ} \leq \frac{C_2^{UB}}{N} + O\left(\frac{1}{N^2}\right), \tag{15}$$

which implies that the failure probability approaches to zero in the limit of large N, at least with the order of $1/N$.

To see the residual non-orthogonality, let us revisit to the first beamsplitter $\hat{B}_{1,0}(\theta_1)$. The states after the beamsplitting can be described by

$$\begin{aligned} \hat{B}_{1,0}(\theta_1)|0\rangle|\Psi\rangle &= |0\rangle|\eta_0\rangle + \frac{1}{N^{1/2}}|1\rangle|\eta_1'\rangle + \frac{1}{N}|2\rangle|\eta_2\rangle + \cdots \\ &= |0\rangle|\eta_0\rangle + \frac{1}{N^{1/2}}|1\rangle|\eta_1\rangle \\ &\quad + \frac{1}{N}|1\rangle|\eta_r\rangle + O\left(\frac{1}{N}\right), \end{aligned} \tag{16}$$

where $|\eta_0\rangle = \sum_{m=0}^{\infty} c_m(1-1/N)^{m/2}|m\rangle$, $|\eta_1'\rangle = \sum_{m=1}^{\infty} c_m(m/N)^{1/2}(1-1/N)^{(m-1)/2}|m-1\rangle$, $|\eta_1\rangle = \sum_{m=1}^{\infty} c_m(1-(1-1/N)^m)^{1/2}|m-1\rangle$, and $|\eta_r\rangle = |\eta_1'\rangle - |\eta_1\rangle$. Similarly,

$$\begin{aligned} \hat{B}_{1,0}(\theta_1)|0\rangle|\Phi\rangle &= |0\rangle_1|\nu_0\rangle + \frac{1}{N^{1/2}}|1\rangle|\nu_1\rangle \\ &\quad + \frac{1}{N}|1\rangle|\nu_r\rangle + O\left(\frac{1}{N}\right), \end{aligned} \tag{17}$$

where $\langle\eta_0|\nu_0\rangle + \langle\eta_1|\nu_1\rangle/N = 0$ is exactly satisfied and $O(1/N)$ includes the terms of more than two photon reflections. The terms $|\eta_r\rangle$ and $|\nu_r\rangle$, that have been neglected in the previous discussion, cause the residual non-orthogonality. Note that the leading terms of these vectors are independent of N.

We can design β_1 such that, after the photon counting, the orthogonality between Eqs. (16) and (17) is satisfied up to the order of $1/N^{1/2}$. Then, the conditional outputs for the 0- and 1-photon detection events are given by

$$\begin{aligned} \frac{\hat{E}_0^{(1)}|\Psi\rangle}{\sqrt{P_0^{(1)}}} &= \frac{1}{\sqrt{P_0^{(1)}}} \exp\left(-\frac{|\beta_1|^2}{2N}\right) \\ &\quad \times \left(|\eta_0\rangle - \frac{\beta_1^*}{N}|\eta_1\rangle - \frac{\beta_1^*}{N^{3/2}}|\eta_r\rangle + O\left(\frac{1}{N^2}\right) \right), \end{aligned} \tag{18}$$

$$\begin{aligned} \frac{\hat{E}_1^{(1)}|\Psi\rangle}{\sqrt{P_1^{(1)}}} &= \frac{1}{\sqrt{NP_1^{(1)}}} \exp\left(-\frac{|\beta_1|^2}{2N}\right) \\ &\quad \times \left(\beta_1|\eta_0\rangle + |\eta_1\rangle + \frac{1}{N^{1/2}}|\eta_r\rangle + O\left(\frac{1}{N}\right) \right), \end{aligned} \tag{19}$$

respectively, where the third and fourth terms cause the residual non-orthogonality.

Now, suppose that the same strategy is applied to the choices of β_i for further measurement steps. Here, the leading order of $|\eta_r\rangle$, which is about $1/N$, does not change during these measurement processes because, in Eqs. (16), (18) and (19), the order of the zero photon reflection term $|0\rangle_i|\eta_0\rangle_0$, which is also about $1/N$, does not change during the whole operations and, similarly, after beamsplitting $|\eta_r\rangle$, the state always includes the term $|0\rangle_i|\eta_r'\rangle_0$ which keeps the leading order of $1/N$ the same. After the $N-1$-th measurement, when all the photon counters detected less than one photon, the conditional output is given by

$$\begin{aligned} |\Psi^{(N-1)}\rangle &= \frac{\hat{E}^{(N-1)} \cdots \hat{E}^{(1)}|\Psi\rangle}{\sqrt{P^{(N-1)} \cdots P^{(1)}}} \\ &= c_0^{(N-1)}|0\rangle + \frac{c_1^{(N-1)}}{N^{1/2}}|1\rangle + \cdots \\ &\quad + \frac{1}{N^{3/2}} \left(|H_0^{(i_1)}\rangle + |H_0^{(i_2)}\rangle + \cdots \right) \\ &\quad + \frac{1}{N^{1/2}} \left(|H_1^{(j_1)}\rangle + |H_1^{(j_2)}\rangle + \cdots \right), \end{aligned} \tag{20}$$

where $|H_k^{(l)}\rangle$ is the residual non-orthogonal vector caused by the k-photon detection at the l-th measurement.

Denote the final N-th measurement by $|D_k\rangle \equiv \hat{D}^\dagger(\beta_N/\sqrt{N})|k\rangle$ $(k=0,1)$. Suppose that β_N is designed such that $|\Psi^{(N-1)}\rangle$ and $|\Phi^{(N-1)}\rangle$ are mostly projected onto $|D_0\rangle$ and $|D_1\rangle$, respectively. The error probability

that $|\Psi^{(N-1)}\rangle$ is projected onto $|D_1\rangle$ is given by

$$\begin{aligned} P_{err} &= |\langle D_1|\Psi^{(N-1)}\rangle|^2 \\ &= \left| \frac{1}{N^{3/2}} \sum_{x=1}^{I} \langle D_1|H_0^{(i_x)}\rangle \right. \\ &\quad \left. + \frac{1}{N^{1/2}} \sum_{x=1}^{J} \langle D_1|H_1^{(j_x)}\rangle \right|^2 . \end{aligned} \tag{21}$$

Here, I and J are the numbers of the events of detecting zero and one photons, respectively. The leading order of $\langle D_1|H_k^{(j)}\rangle$ is independent of N for every j and k.

Now, let us consider the limit of large N. One can show that, as N increases, $J^{(N)}$ exponentially approaches to $\sum_{x=1}^{N} P_1^{(x)}$ with the help of Azuma's inequality [10, 11]. Let us define the random variables $Z_i \equiv J^{(i)} - \sum_{x=1}^{i} P_1^{(x)}$ $(i = 0, ..., N)$ where $J^{(i)}$ is the number of the 1-photon detection event up to the i-th measurement. In the sequence $\{Z_0, Z_1, ...\}$, Z_i's are not independent to each other since $P_1^{(i)}$ always depends on the previous $i-1$ measurements. However, it is able to show that Z_i's satisfy the conditions $E[Z_j|Z_0, Z_1, \cdots Z_{j-1}] = Z_{j-1}$ and $|X_j - X_{j-1}| \leq 1$ for $j \geq 1$, where $E[Y|X]$ denotes the conditional expectation. To the random variables satisfying these conditions, one can apply Azuma's inequality and it implies that

$$\mathrm{Prob}[|Z_N - Z_0| \geq N\epsilon] \leq 2e^{-N\epsilon^2/2}, \tag{22}$$

for all $N \geq 0$ and any $\epsilon \geq 0$, which means that, in the limit of large N, $J^{(N)}$ approaches to $\sum_{x=1}^{N} P_1^{(x)}$ exponentially, and also $I^{(N)} = N - J^{(N)}$. Therefore, P_{err} satisfies

$$\begin{aligned} P_{err} &< \frac{1}{N} \left| \left(1 - \frac{C_1^{\max}}{N} + O\left(\frac{1}{N^2}\right) + \epsilon\right) \langle D_1|H_0\rangle_{\mathrm{av}} \right. \\ &\quad \left. + \left(C_1^{\max} + O\left(\frac{1}{N}\right) + N\epsilon\right) \langle D_1|H_1\rangle_{\mathrm{av}} \right|^2 \\ &\leq \frac{C_E}{N} + O\left(\frac{1}{N^2}\right) + \epsilon O\left(\frac{1}{N^0}\right), \end{aligned} \tag{23}$$

where $\langle D_1|H_k\rangle_{\mathrm{av}} = \sum_i \langle D_1|H_k^{(i)}\rangle/N$ $(k = 0, 1)$ and C_E is some constant independent of N. Also, we can choose that C_E satisfies the same relation for $|\langle D_0|\Phi^{(N-1)}\rangle|^2$. Then, summing over all detection patterns, the average error probability is bounded as

$$\begin{aligned} P_{err}^{\mathrm{tot}} &= \sum^{\mathrm{success}} P(\sharp) P_{err}^{succ}(\sharp) + \sum^{\mathrm{failure}} P(\sharp) P_{err}^{fail}(\sharp) \\ &\leq \left(1 - \frac{C_2^{\max}}{N}\right) \left(\frac{C_E}{N} + \tilde{\epsilon}\right) + \frac{C_2^{\max}}{N} + O\left(\frac{1}{N^2}\right) \\ &\leq \frac{C}{N} + O\left(\frac{1}{N^2}\right) + \tilde{\epsilon}, \end{aligned} \tag{24}$$

where C is some constant and $\tilde{\epsilon} = \epsilon O(1/N^0)$. As a consequence, in the limit of $N \to \infty$, one can discriminate $|\Psi\rangle$ and $|\Phi\rangle$ with the success probability arbitrarily close to unit.

Finally, summing over all detection patterns from Eqs. (14), (15), and (23), we can find that the average error probability is bounded by some constant C as

$$\begin{aligned} P_{err}^{\mathrm{tot}} &= \sum^{\mathrm{success}} P(\sharp) P_{err}^{succ}(\sharp) + \sum^{\mathrm{failure}} P(\sharp) P_{err}^{fail}(\sharp) \\ &\leq \left(1 - \frac{C_2^{\mathrm{UB}}}{N}\right) \frac{C_E}{N} + \frac{C_2^{\mathrm{UB}}}{N} + O\left(\frac{1}{N^2}\right) \\ &\leq \frac{C}{N}. \end{aligned} \tag{25}$$

As a consequence, when N is large enough, the scaling of P_{err}^{tot} is at least given by $1/N$ and in the limit of $N \to \infty$, $P_{err} \to 0$, i.e. one can discriminate $|\Psi\rangle$ and $|\Phi\rangle$ with unit success probability.

In summary, we have proved that one can implement arbitrary projection measurement in any two-dimensional signal space by linear optics tools without using any non-classical ancillary states in the asymptotic limit of large number of detections and feedforwards. The resources discussed here are mostly available with current technology. These results are valuable for various quantum information protocols that require binary projection measurements. The remaining interesting questions are the derivation of the optimal convergence rate of the error probability and the possibility to apply the same approach to the problems of more than three states discrimination.

References

[1] E. Knill, R. Laflamme, and G. J. Milburn, Nature **409**, 46 (2001).

[2] P. van Loock and N. Lütkenhaus, Phys. Rev. A **69**, 012302 (2004).

[3] This is true only for the case when all physical operations during a whole measurement can be described by rank 1 operators. As will be shown in the text, our scheme corresponds to this case.

[4] J. Walgate, A. J. Short, L. Hardy, and V. Vedral, Phys. Rev. Lett. **85**, 4972 (2000).

[5] P. van Loock and S. L. Braunstein, Phys. Rev. Lett. **84**, 3482 (2000).

[6] S. J. Dolinar, Research Laboratory of Electronics, MIT, Quarterly Progress Report No. 111, 1973, p. 115.

[7] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).

[8] JM Geremia, LANL arXive quant-ph/0407205 (2004).

[9] M. Takeoka, M. Sasaki, P. van Loock, and N. Lütkenhaus, Phys. Rev. A **71**, 022318 (2005).

[10] K. Azuma, Tohoku Math. J. **19**, 357 (1967).

[11] J.-C. Boileau, K. Tamaki, J. Batuwantudawe, R. Laflamme, and J. M. Renes, Phys. Rev. Lett. **94**, 040503 (2005).

Programmable quantum channels and measurements

Giacomo Mauro D'Ariano[1] * Paolo Perinotti[1 2] †

[1] *QUIT group, Dipartimento di Fisica "A. Volta", via Bassi 6, I-27100, Pavia, Italy*
[2] *CNR-INFM, Unità di Pavia*

Abstract. We review some partial results for two strictly related problems. The first problem consists in finding the optimal joint unitary transformation on system and ancilla which is the most efficient in programming any desired channel on the system by changing the state of the ancilla. In this respect we present a solution for $\dim(\mathscr{H}) = 2$ for both system and ancilla. The second problem consists in finding the optimal universal programmable detector, namely a device that can be tuned to perform any desired measurement on a given quantum system, by changing the state of an ancilla. With a finite dimension d for the ancilla only approximate universal programmability is possible, with $d = d(\varepsilon^{-1})$ increasing function versus ε^{-1}. We show that one can achieve $d(\varepsilon^{-1})$ polynomial, and even linear in specific cases.

Keywords: Quantum information theory; channels; quantum computing; entanglement

1 Introduction

A fundamental problem in quantum computing and, more generally, in quantum information processing [?] is to experimentally achieve any theoretically designed quantum channel or detector using a fixed device, through a suitable program encoded in the state of an ancillary system. While a large branch of theoretical research in quantum information addressed the design of algorithms and of circuits to solve precise problems, a parallel research line is that of designing devices that can be programmed to achieve different tasks, just like classical computers do. Moreover, designing a programmable quantum gate or detector is a problem of relevance for example in proving the equivalence of cryptographic protocols, e. g. proving the equivalence between a multi-round and a single-round quantum bit commitment [?], or when trying to eavesdrop quantum-encrypted information. What makes the problem of gate programmability non trivial is that exact universal programmability of channels is impossible, as a consequence of a no-go theorem for programmability of unitary transformations by Nielsen and Chuang [?]. A similar situation occurs for universal programmability of POVM's [?, ?]. In both cases, it is still possible to achieve programmability probabilistically [?, ?, ?], or even deterministically [?], though within some accuracy. In establishing the theoretical limits to state-programmability of channels or POVM's the starting problem is to find the joint system-ancilla unitary or observable, respectively, which achieves the best accuracy for fixed dimension of the ancilla: this is exactly the problem that is addressed in the present paper. This problem turned out to be hard, even for low dimension. Here we will give a solution for the optimal device for programming unitary channels for dimension two for both system and ancilla. On the other hand, as regards programming observables, we will give an upper bound for the optimal ancilla dimension $d(\varepsilon^{-1})$ versus the accuracy $\varepsilon - 1$ for programmable detectors. As we will see, it turns out [?] that a dimension $d(\varepsilon^{-1})$ increasing polynomially with precision ε^{-1} is possible, and even a linear dependence is achievable for specific cases. This should be compared with the preliminary indications of an exponential growth of Ref. [?]. However, even the linear dependence $d(\varepsilon^{-1})$ is still suboptimal.

*dariano@unipv.it
†perinotti@fisicavolta.unipv.it

2 Statement of the problems

Programmable unitaries We want to program unitary channels by a fixed device as follows

$$\mathscr{P}_{V,\sigma}(\rho) \doteq \mathrm{Tr}_2[V(\rho \otimes \sigma)V^\dagger], \tag{1}$$

with the system in the state ρ interacting with an ancilla in the state σ via the unitary operator V of the programmable device (the state of the ancilla is the *program*). For fixed V the above map can be regarded as a linear map from the convex set of the ancilla states $\mathcal{A}$ to the convex set of channels for the system $\mathcal{C}$. We will denote by $\mathcal{P}_{V,\mathcal{A}}$ the image of the ancilla states $\mathcal{A}$ under such linear map: these are the programmable channels. According to the well known no-go theorem by Nielsen and Chuang it is impossible to program all unitary channels on the system with a single V and a finite-dimensional ancilla, namely the image convex $\mathcal{P}_{V,\mathcal{A}} \subset \mathcal{C}$ is a proper subset of the whole convex $\mathcal{U}$ of unitary channels and their convex combinations. This opens the following problem:

> **Problem:** *For given dimension of the ancilla, find the unitary operators V that are the most efficient in programming unitary channels, namely which minimize the largest distance $\varepsilon(V)$ of each channel $\mathscr{U} \in \mathcal{U}$ from the programmable set $\mathcal{P}_{V,\mathcal{A}}$:*
>
> $$\varepsilon(V) \doteq \max_{\mathscr{U}\in\mathcal{U}} \min_{\mathscr{P}\in\mathcal{P}_{V,\mathcal{A}}} \delta(\mathscr{C},\mathscr{P}) \equiv \max_{\mathscr{U}\in\mathcal{U}} \min_{\sigma\in\mathcal{A}} \delta(\mathscr{C},\mathscr{P}_{V,\sigma}). \tag{2}$$

As a definition of distance it would be most appropriate to use the CB-norm distance $\|\mathscr{C} - \mathscr{P}\|_{CB}$. However, this leads to a very hard problem. We will use instead the following distance

$$\delta(\mathscr{C},\mathscr{P}) \doteq \sqrt{1 - F(\mathscr{C},\mathscr{P})}, \tag{3}$$

where $F(\mathscr{C},\mathscr{P})$ denotes the Raginsky fidelity [?], which for unitary map $\mathscr{C} \equiv \mathscr{U} = U \cdot U^\dagger$ is equivalent to the

channel fidelity [?]

$$F(\mathscr{U},\mathscr{P}) = \frac{1}{d^2}\sum_i |\operatorname{Tr}[C_i^\dagger U]|^2, \tag{4}$$

where $\mathscr{C} = \sum_i C_i \cdot C_i^\dagger$. Such fidelity is also related to the input-output fidelity averaged over all pure states $\overline{F}_{io}(\mathscr{U},\mathscr{P})$, by the formula $\overline{F}_{io}(\mathscr{U},\mathscr{P}) = [1 + dF(\mathscr{U},\mathscr{P})]/(d+1)$. Therefore, our optimal unitary V will maximize the fidelity

$$F(V) \doteq \min_{U\in\mathscr{U}(H)} F(U,V), \quad F(U,V) \doteq \max_{\sigma\in\mathcal{A}} F(\mathscr{U},\mathscr{P}_{V,\sigma}) \tag{5}$$

Programmable detectors The POVM of a measuring apparatus is a set of positive operators $P_i \geqslant 0$, $i = 1,\dots n$, $n < \infty$ normalized to the identity $\sum_{i=1}^n P_i = I$, which gives the probability distribution of the outcomes for each input state ρ via the Born rule

$$p(i|\rho) \doteq \operatorname{Tr}[\rho P_i]. \tag{6}$$

Clearly, the most general programmable detector is described by an observable $\mathbf{F} \doteq \{F_i\}$ jointly measured on system and ancilla. The probability distribution of the outcomes is given by

$$p_\sigma(i|\rho) = \operatorname{Tr}[(\rho\otimes\sigma)F_i],\ \forall i,\forall\rho. \tag{7}$$

By taking the partial trace in Eq. (7) over the ancilla and using the polarization identity (Eq. (7) holds for all states) one obtains the POVM

$$P_{\sigma,i} = \operatorname{Tr}_2[(I\otimes\sigma)F_i]. \tag{8}$$

From Eq. (8) it follows that the convex set of states $\mathscr{A}$ of the ancilla is in correspondence via the map $\mathbf{P}_{\mathbf{F},\sigma} \doteq \operatorname{Tr}_2[(I\otimes\sigma)\mathbf{F}]$ with a convex subset $\mathscr{P}_{\mathbf{F},\mathscr{A}} \subseteq \mathscr{P}_n$ of the convex set $\mathscr{P}_n$ of the system POVM's with n outcomes. The symbol $\mathscr{P}_{\mathbf{F},\mathscr{A}}$ denotes the convex set of programmable POVM's that can be achieved with fixed observable $\mathbf{F}$ and varying state $\sigma\in\mathscr{A}$. The no-go theorem proved in [?, ?] states that for any fixed observable $\mathbf{F}$ the programmable set $\mathbf{P}_{\mathbf{F},\mathscr{A}}$ is strictly contained in $\mathscr{P}_n$, since even just the observables cannot be programmed with a fixed observable $\mathbf{F}$ and a finite dimensional ancilla. We now restrict attention to programmability of observables only, whence $n \equiv \dim(\mathscr{H})$, the case of nonorthogonal POVM's simply resorting to program observables on a larger Hilbert space. In the following we will denote by $\mathcal{O}_n$ the set of observables. The problem of measurement programmability can then be stated in mathematical terms as follows

> **Problem:** *For given dimension of the ancilla, find the joint observables* $\mathbf{F}$ *that are the most efficient in programming system observables, namely which minimize the largest distance* $\varepsilon(\mathbf{F})$ *of each observable* $\mathbf{P}\in\mathcal{O}_n$ *from the programmable set* $\mathbf{P}_{\mathbf{F},\sigma}$*:*

$$\varepsilon(\mathbf{F}) \doteq \max_{\mathbf{P}\in\mathcal{O}_n}\min_{\mathbf{Q}\in\mathbf{P}_{\mathbf{F},\sigma}} \delta(\mathbf{P},\mathbf{Q}) \equiv \max_{\mathbf{P}\in\mathcal{O}_n}\min_{\sigma\in\mathcal{A}} \delta(\mathbf{P},\mathbf{P}_{\mathbf{F},\sigma}). \tag{9}$$

We define the distance between two POVM's as the distance between their respective probabilities, maximized over all possible states, namely

$$\delta(\mathbf{P},\mathbf{Q}) = \max_\rho \sum_i |\operatorname{Tr}[\rho(P_i - Q_i)]|. \tag{10}$$

The distance defined in Eq. (10) is hard to handle analytically, whence we bound it as follows

$$\delta(\mathbf{P},\mathbf{Q}) \leqslant \sum_i \|P_i - Q_i\| \leqslant \sum_i \|P_i - Q_i\|_2, \tag{11}$$

where $\|A\|$ is the usual operator norm of A, and $\|A\|_2 \doteq \sqrt{\operatorname{Tr}[A^\dagger A]}$ is the Frobenius norm.

3 Programming qubit unitaries

By some lengthy calculation we can obtain the Kraus operators for the map $\mathscr{P}_{V,\sigma}(\rho)$

$$\begin{aligned}\mathscr{P}_{V,\sigma}(\rho) &= \sum_{nm} C_{nm}\rho C_{nm}^\dagger,\\ C_{nm} &= \sum_k e^{i\theta_k}\Psi_k|v_n^*\rangle\langle v_m^*|\Psi_k^\dagger\sqrt{\lambda_m}\end{aligned} \tag{12}$$

where $|v_n\rangle$ denotes the eigenvector of σ corresponding to the eigenvalue λ_n and * denotes complex conjugation in the same fixed basis for which the operator Ψ_k have the same matrix elements as the matrix of coefficients of the eigenvector of V corresponding to eigenvalue $e^{i\theta_k}$. We then obtain

$$\begin{aligned}\sum_{nm}|\operatorname{Tr}[C_{nm}^\dagger U]|^2 &= \sum_{kh} e^{i(\theta_k-\theta_h)}\operatorname{Tr}[\Psi_k^\dagger U^\dagger\Psi_k\sigma^\intercal\Psi_h^\dagger U\Psi_h]\\ &= \operatorname{Tr}[\sigma^\intercal S(U,V)^\dagger S(U,V)]\end{aligned} \tag{13}$$

where

$$S(U,V) = \sum_k e^{-i\theta_k}\Psi_k^\dagger U\Psi_k. \tag{14}$$

and $^\intercal$ denotes transposition in the canonical basis. The fidelity (5) can then be rewritten as follows

$$F(U,V) = \frac{1}{d^2}\|S(U,V)\|^2. \tag{15}$$

The operator $S(U,V)$ in Eq. (14) can be written as follows

$$S(U,V) = \operatorname{Tr}_1[(U^\intercal\otimes I)V^*]. \tag{16}$$

Changing V by local unitary operators transforms $S(U,V)$ in the following fashion

$$S(U,(W_1\otimes W_2)V(W_3\otimes W_4)) = W_2^* S(W_1^\dagger U W_3^\dagger, V)W_4^*, \tag{17}$$

namely the local unitaries do not change the minimum fidelity, since the unitaries on the ancilla just imply a different program state, whereas the unitaries on the system just imply that the minimum fidelity is achieved for a different unitary—say $W_1^\dagger U W_3^\dagger$ instead of U.

For system and ancilla both two-dimensional, one can parameterize all possible joint unitary operators as follows [?, ?] $(W_1\otimes W_2)\tilde{V}(W_3\otimes W_4)$, where

$$\tilde{V} = \exp[i(\alpha_1\sigma_1\otimes\sigma_1{}^\intercal + \alpha_2\sigma_2\otimes\sigma_2{}^\intercal + \alpha_3\sigma_3\otimes\sigma_3{}^\intercal)]. \tag{18}$$

The problem is now reduced to study only joint unitary operators of the form of Eq. (18). It can be proved that the coefficients of its eigenvectors are the matrix elements of Pauli matrices σ_j, $j = 0, 1, 2, 3$ where $\sigma_0 = I$, $\sigma_1 = \sigma_x$, $\sigma_2 = \sigma_y$, $\sigma_3 = \sigma_z$. This means that we can rewrite $S(U, V)$ in Eq. (14) as follows

$$S(U,V) = \frac{1}{2}\sum_{j=0}^{3} e^{-i\theta_j}\sigma_j U \sigma_j\,, \tag{19}$$

with

$$\theta_0 = \alpha_1 + \alpha_2 + \alpha_3\,, \quad \theta_i = 2\alpha_i - \theta_0\,. \tag{20}$$

Through the derivation described in Appendix ?? we obtain that the fidelity minimized over all unitaries is given by

$$F(V) = \frac{1}{d^2}\min_j |t_j|^2. \tag{21}$$

where

$$\begin{aligned} t_0 &= \frac{1}{2}\sum_{j=0}^{3} e^{-i\theta_j}, \\ t_j &= e^{-i\theta_0} + e^{-i\theta_j} - t_0,\ 1 \le j \le 3. \end{aligned} \tag{22}$$

The optimal unitary V is now obtained by maximizing $F(V)$. We need then to consider the decomposition Eq. (18), and then to maximize the minimum among the four eigenvalues of $S(U,V)^\dagger S(U,V)$. Notice that $t_j = \sum_\mu H_{j\mu} e^{i\theta_\mu}$, where H is the Hadamard matrix

$$H = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}, \tag{23}$$

which is unitary, and consequently $\sum_j |t_j|^2 = \sum_j |e^{i\theta_j}|^2 = 4$. This implies that $\min_j |t_j| \le 1$. We now provide a choice of phases θ_j such that $|t_j| = 1$ for all j, achieving the maximum fidelity allowed. For instance, we can take $\theta_0 = 0, \theta_1 = \pi/2, \theta_2 = \pi, \theta_3 = \pi/2$, corresponding to the eigenvalues $1, i, -1, i$ for V. Another solution is $\theta_0 = 0, \theta_1 = -\pi/2, \theta_2 = \pi, \theta_3 = -\pi/2$. Also one can set $\theta_i \to -\theta_i$. The eigenvalues of $S(U,V)^\dagger S(U,V)$ are then $1, 1, 1, 1$, while for the fidelity we have

$$F \doteq \max_{V \in \mathscr{U}(H^{\otimes 2})} F(V) = \frac{1}{d^2} = \frac{1}{4}, \tag{24}$$

and the corresponding optimal V has the form

$$V = \exp\left[\pm i\frac{\pi}{4}(\sigma_x \otimes \sigma_x \pm \sigma_z \otimes \sigma_z)\right]. \tag{25}$$

A possible circuit scheme for the optimal V is given in Fig. 1.

Such fidelity cannot be achieved by any V of the controlled-unitary form

$$V = \sum_{k=1}^{2} V_k \otimes |\psi_k\rangle\langle\psi_k|, \qquad \langle\psi_1|\psi_2\rangle = 0, \tag{26}$$

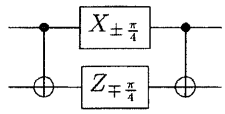

Figure 1: Quantum circuit scheme for the optimal unitary operator V in Eq. (24). W_α denotes $e^{i\frac{\alpha}{2}\sigma_W}$. For the derivation of the circuit consider that $\sigma_x \otimes \sigma_x = C(\sigma_x \otimes I)C$ and $\sigma_z \otimes \sigma_z = C(I \otimes \sigma_z)C$, where C denotes the controlled-not.

where V_1, V_2 are unitaries on $\mathscr{H} \simeq \mathbb{C}^2$. In fact, it is easy to see that in this case the fidelity is given by

$$F(U,V) = \frac{1}{4}|\operatorname{Tr}[V_h^\dagger U]|^2, \qquad h = \arg\max_k |\operatorname{Tr}[V_k^\dagger U]|, \tag{27}$$

and for any couple of unitaries $\{V_k\}$ there always exists a unitary U orthogonal to both $\{V_k\}$, whence $F(V) \doteq \min_{U \in \mathscr{U}(H)} F(U,V) = 0$.

4 Upper bound on optimal size for programmable detectors

We will now derive an upper bound for the function $d = d(\varepsilon^{-1})$, where $\epsilon = \min_{\mathbf{F}} \varepsilon(\mathbf{F})$, that gives the minimal needed dimension of the ancilla to achieve accuracy ε^{-1} in programming observables for a finite-dimensional quantum system. Clearly the function $d = d(\varepsilon^{-1})$ must be increasing, since the higher is the accuracy ε^{-1}, the larger the minimal dimension d needed for the ancilla, namely the "size" of the programmable detector.

Consider now a d-dimensional ancilla and a system-ancilla interaction U of the following *controlled-unitary* form

$$U = \sum_{k=1}^{d} W_k \otimes |\phi_k\rangle\langle\phi_k|, \tag{28}$$

where $\{\phi_k\}$ is an orthonormal complete set of vectors for the ancilla and W_k are generic unitary operators on $\mathscr{H}$. Consider then a POVM $\mathbf{E} = U\mathbf{F}U^\dagger$ of the form

$$E_i = |\psi_i\rangle\langle\psi_i| \otimes I_A\,, \tag{29}$$

where I_A denotes the identity operator on the ancilla space, and $\{\psi_k\}$ is a complete orthonormal set for the system. The observable to be approximated will then be written as follows

$$P_i = W^\dagger|\psi_i\rangle\langle\psi_i|W, \tag{30}$$

W being a unitary operator on $\mathscr{H}$, and we will scan all possible observables by varying W. For the program state of the ancilla we use one of the states ϕ_k, which give the POVM's

$$Q_i = W_k^\dagger|\psi_i\rangle\langle\psi_i|W_k\,. \tag{31}$$

This special form simplifies the calculation of the bound

in Eq. (11), which becomes

$$\begin{aligned}\delta(\mathbf{P},\mathbf{Q}) &\leqslant \sum_i \sqrt{2(1-|\langle\psi_i|W^\dagger W_k|\psi_i\rangle|^2)} \\ &\leqslant \sqrt{2}\sum_i \sqrt{2-\langle\psi_i|(W^\dagger W_k - W_k^\dagger W)|\psi_i\rangle},\end{aligned} \tag{32}$$

and using the Jensen's inequality for the square root function we have

$$\delta(\mathbf{P},\mathbf{Q}) \leqslant \sqrt{2n}\,\|W-W_k\|_2\,. \tag{33}$$

Now we can always take d sufficiently large such that we can choose the d operators $\{W_k\}$ in the unitary transformation U in Eq. (28) in such a way that for each given W there is always a unitary operator W_k in the set for which $\sqrt{2n}\,\|W-W_k\|_2$ is bounded by ε. This will guarantee that for the given observable $\mathbf{P}$ corresponding to W there is a program state for the ancilla such that the POVM $\mathbf{Q}$ achieved by the programmable detector is close to the desired $\mathbf{P}$ less than ε. The set of all possible unitary operators W is a compact manifold of dimension $h = n^2 - n$. We now consider a covering of the manifold with balls of radius $r = \frac{\varepsilon}{\sqrt{2n}}$ centered at the operators W_k. This guarantees that any W would be within a distance $\frac{\varepsilon}{\sqrt{2n}}$ from an operator W_k, which in turns implies that the accuracy of the programmable device is bounded by ε via Eq. (33). Using the volume $V = \frac{\pi^{\frac{h}{2}} r^h}{\Gamma(\frac{1}{2}h+1)}$ of the h-dimensional sphere of radius r, we obtain the number of balls needed for the covering (for sufficiently small ε, corresponding to the upper bound for the minimal dimension of the ancilla

$$d \leqslant \kappa(n)\left(\frac{1}{\varepsilon}\right)^{n(n-1)}\,, \tag{34}$$

where $\kappa(n)$ is a constant that depends on n. Eq. (34) gives an upper bound for the dimension d which is polynomial versus the accuracy ε^{-1}.

For qubits, the observable has only two elements, $P_0 = |\psi\rangle\langle\psi|$ and $P_1 = |\psi_\perp\rangle\langle\psi_\perp| = I - P_0$, and the distance in Eq. (10) can be evaluated analytically as follows

$$\delta(\mathbf{P},\mathbf{Q}) = \max_\rho 2|\,\mathrm{Tr}[\rho(P_0-Q_0)]|\,. \tag{35}$$

As regards now the programmability of all POVM's (i. e. including the nonorthogonal ones), just notice that one just needs to be able to program only the extremal POVM's in $\mathscr{P}_n$, since their convex combinations will corresponds to mixing the program state or to randomly choosing among different detectors. Then, since their maximum number of outcomes is n^2, the extremal POVM's have Naimark's extension to observables in dimension n^2, whence we are reduced to the case of programmability of observables in dimension n^2.

We will now give a programmable detector for qubits that achieves an accuracy that is linear in d. For the ancilla we use a generic d-dimensional quantum system, and relabel the dimension in the angular momentum fashion $d \doteq 2j+1$. The idea is now to design a programmable detector in which the unitary transformation corresponding to the observable $\{P_i\}$ in Eq. (30) is programmed by covariantly changing the program state of the ancilla. By labeling unitary transformations by a group element $g \in \mathbb{SU}(2)$, we write the observable to be programmed as $P_0 \doteq V_g|\frac{1}{2}\rangle\langle\frac{1}{2}|V_g^\dagger$ where $\{V_g\} \equiv (\frac{1}{2})$ is a unitary irreducible representation of $\mathbb{SU}(2)$ with angular momentum $\frac{1}{2}$, whereas the program state will be written as $W_g\sigma W_g^\dagger$, with $\{W_g\} \equiv (j)$ a unitary irreducible representation of $\mathbb{SU}(2)$ on the ancilla space with angular momentum j. As already noticed, without loss of generality we can always choose the state σ as pure. We will now show that a good choice for the program state is $\sigma = |j,j\rangle\langle j,j|$, $\{|j,m\rangle\}$ denoting an orthonormal basis of eigenstates of J_z in the irreducible representation with angular momentum j. The tensor representation $\{V_g \otimes W_g\} \equiv \frac{1}{2}\otimes j$ can be decomposed into the direct sum of two irreducible representations $\frac{1}{2}\otimes j = j_+ \oplus j_-$, where $j_\pm = j \pm \frac{1}{2}$. For the POVM $\mathbf{F}$ of the programmable detector we will use $F_0 = Z_+$ and $F_1 = Z_-$, $Z_\pm$ denoting the orthogonal projector on the invariant space for angular momentum $j_\pm$

$$F_0 = \sum_{m=-j_+}^{j_+} |j_+,m\rangle\langle j_+,m|\,. \tag{36}$$

Using the invariance $(V_g\otimes W_g)F_0(V_g^\dagger\otimes W_g^\dagger) = F_0$, we can write the programmed POVM as follows

$$\begin{aligned}Q_0 &= \mathrm{Tr}_A[(I\otimes W_g^\dagger|j,j\rangle\langle j,j|W_g)F_0] \\ &= V_g^\dagger\,\mathrm{Tr}_A[(I\otimes|j,j\rangle\langle j,j|)F_0]V_g \\ &= V_g\left(|\tfrac{1}{2},\tfrac{1}{2}\rangle\langle\tfrac{1}{2},\tfrac{1}{2}| + \tfrac{1}{2j+1}|\tfrac{1}{2},-\tfrac{1}{2}\rangle\langle\tfrac{1}{2},-\tfrac{1}{2}|\right)V_g^\dagger,\end{aligned} \tag{37}$$

where we used the only non vanishing Clebsch-Gordan coefficients $|\langle j_+,j_+|\frac{1}{2},\frac{1}{2}\rangle|j,j\rangle|^2 = 1$, and $|\langle j_+,j_-|\frac{1}{2},-\frac{1}{2}\rangle|j,j\rangle|^2 = \frac{1}{2j+1}$. Clearly, $Q_0 - P_0 = \frac{1}{2j+1}V_g|\frac{1}{2},-\frac{1}{2}\rangle\langle\frac{1}{2},-\frac{1}{2}|V_g^\dagger$, whence according to Eq. (35) the accuracy is given by $\delta(\mathbf{P},\mathbf{Q}) = 2/d$. The scaling of the dimension with the accuracy is then linear

$$d = 2\varepsilon^{-1}, \tag{38}$$

whereas the bound (34) would be quadratic $d \propto \varepsilon^{-2}$. Sublinear growth of d versus ε^{-1} is not excluded in general, but is not possible for the present model.

Probabilistic cloning with supplementary information

Koji Azuma[2] * Junichi Shimamura[1 3 4] Masato Koashi[1 3 4] Nobuyuki Imoto[1 3 4]

[1] *Division of Materials Physics, Department of Materials Engineering Science, Graduate School of Engineering Science, Osaka University, Toyonaka, Osaka 560-8531, Japan.*
[2] *Department of Applied Physics, University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan.*
[3] *CREST Photonic Quantum Information Project, 4-1-8 Honmachi, Kawaguchi, Saitama 331-0012, Japan.*
[4] *SORST Research Team for Interacting Carrier Electronics, 4-1-8 Honmachi, Kawaguchi, Saitama 331-0012, Japan.*

Abstract. We consider probabilistic cloning helped by a party holding supplementary information. When the number of states is two, we show that the best efficiency of producing m copies is always achieved by a two-step protocol in which the helping party first attempts to produce $m-1$ copies from the supplementary state, and if it fails, then the original state is used to produce m copies. On the other hand, when the number of states exceeds two, we give examples in which the best efficiency is not achieved even if we allow any amount of one-way classical communication from the helping party.

Keywords: Probabilistic cloning, Stronger no-cloning theorem

1 Introduction

The impossibility of deterministic cloning of nonorthogonal pure states is well known as the no-cloning theorem [1, 2]. The best one can do is to carry out weaker tasks, such as allowing the copies to be inaccurate [3, 4, 5, 6, 7, 8, 9, 10], or allowing a failure to occur with a nonzero probability (probabilistic cloning) [11]. Another way to enable the cloning is to provide some hints in the form of a quantum state. Jozsa has considered [12] how much or what kind of supplementary information $\hat{\rho}_i$ is required to make two copies $|\psi_i\rangle|\psi_i\rangle$ from the original information $|\psi_i\rangle$. He has shown that for any mutually nonorthogonal set of original states $\{|\psi_i\rangle\}$, whenever two copies $|\psi_i\rangle|\psi_i\rangle$ are generated with the help of the supplementary information $\hat{\rho}_i$, the state $|\psi_i\rangle$ can be generated from the supplementary information $\hat{\rho}_i$ alone, independently of the original state, i.e.,

$$|\psi_i\rangle\otimes\hat{\rho}_i \stackrel{\mathrm{CPTP}}{\longrightarrow} |\psi_i\rangle|\psi_i\rangle \Longrightarrow \hat{\rho}_i \stackrel{\mathrm{CPTP}}{\longrightarrow} |\psi_i\rangle, \quad (1)$$

where CPTP stands for a completely-positive trace-preserving map, implying that the transformation can be done deterministically. This result, dubbed the stronger no-cloning theorem, implies that the supplementary information must be provided in the form of the result $|\psi_i\rangle$ itself, rather than a help, thereby obliterating the necessity of the cloning task itself.

An interesting question occurring here is whether we can find a similar property in the case of probabilistic cloning when we ask how much increase in the success probability is obtained with the help of supplementary information. Suppose that the success probability of cloning the i-th state $|\psi_i\rangle$ without any help is γ_i. If we are directly given a right copy of state $|\psi_i\rangle$ with probability q_i, the success probability would increase to $\gamma_i' = q_i + (1-q_i)\gamma_i$. Hence the counterpart of the stronger no-cloning theorem in probabilistic cloning will be the implication

$$|\psi_i\rangle\otimes\hat{\rho}_i \stackrel{\gamma_i'}{\longrightarrow} |\psi_i\rangle|\psi_i\rangle \Longrightarrow \hat{\rho}_i \stackrel{q_i}{\longrightarrow} |\psi_i\rangle,\ |\psi_i\rangle \stackrel{\gamma_i}{\longrightarrow} |\psi_i\rangle|\psi_i\rangle \quad (2)$$

with $\gamma_i' = q_i + (1-q_i)\gamma_i$. In other words, it implies that the best usage of the supplementary information is to probabilistically create a copy $|\psi_i\rangle$ from it, independently of the original state.

If there are cases where the above implication is not true, it follows that the supplementary information can help directly the process of the cloning task in those cases. Then, the next question will be to ask what kind of interaction should occur between the supplementary information and the original information.

In this paper, we consider probabilistic cloning of mutually nonorthogonal pure states when supplementary information is given as a pure state. We prove that when the number of the possible original states is two, the above implication is true, namely, the supplementary information only serves to provide a copy with a nonzero probability and it does not directly help the process of the cloning. On the other hand, when we have more than two states to choose from, the above implication is not always true. To see this, it is convenient to assume two parties, Alice and Bob, respectively holding the original information and the supplementary information. We give examples in which there is a gap between the efficiency when Bob only communicates to Alice with a one-way classical channel and the efficiency when they fully cooperate through a quantum channel.

This paper is organized as follows. In Sec. 2, we provide definitions and basic theorems used in later sections. We discuss the two-state problem in Sec. 3 and prove that the property similar to the stronger no-cloning theorem holds in this case. In Sec. 4, we give examples with three or more states and show that there is a gap between the success probabilities in the scenarios with classical communication and quantum communication. Section 5 concludes the paper.

*azuma@appi.t.u-tokyo.ac.jp

2 Probabilistic transformation theorem

Throughout this paper, we consider a class of machines that conducts probabilistic transformation of input pure states into output pure states. We denote by $\{|\Phi_i\rangle \xrightarrow{\gamma_i} |\Psi_i\rangle\}_{i=1,\ldots,n}$ a machine having the following properties: (i) It receives a quantum state as an input, and returns a quantum state as an output, together with one bit of classical output indicating whether the transformation has been successful or not. (ii) When the input quantum state is $|\Phi_i\rangle$, the transformation succeeds with probability γ_i, and the successful output state is $|\Psi_i\rangle$. Note that if the output states $\{|\Psi_i\rangle\}$ forms an orthonormal set, namely, $(\forall i,j)\,(\langle\Psi_i|\Psi_j\rangle = \delta_{ij})$, the machine carries out unambiguous discrimination of the set $\{|\Phi_i\rangle\}$ with success probabilities $\{\gamma_i\}$.

A necessary and sufficient condition for the existence of a machine $\{|\Phi_i\rangle \xrightarrow{\gamma_i} |\Psi_i\rangle\}_{i=1,\ldots,n}$ is given by the following theorem, which can be proved by a similar way as in the probabilistic cloning theorem by Duan and Guo [11].

Theorem 1: *There exists a machine* $\{|\Phi_i\rangle \xrightarrow{\gamma_i} |\Psi_i\rangle\}_{i=1,\ldots,n}$ *if and only if there are normalized states* $|P^{(i)}\rangle$ $(i=1,\ldots,n)$ *such that the matrix* $X - \sqrt{\Gamma}Y\sqrt{\Gamma}$ *is positive semidefinite, where* $X := [\langle\Phi_i|\Phi_j\rangle]$, $Y := [\langle\Psi_i|\Psi_j\rangle\langle P^{(i)}|P^{(j)}\rangle]$ *and* $\Gamma := \mathrm{diag}(\gamma_1,\gamma_2,\ldots,\gamma_n)$ *are* $n\times n$ *matrices.*

When the initial state is chosen from the set $\{|\Phi_i\rangle\}$ with probability p_i, we may define the overall success probability γ_{tot} of a machine as

$$\gamma_{\mathrm{tot}} := \sum_i p_i\gamma_i. \tag{3}$$

In this case, we can define the maximum success probability γ_{totmax} as

$$\gamma_{\mathrm{totmax}} := \max_{\{\gamma_i\}} \sum_i p_i\gamma_i, \tag{4}$$

where the maximum is taken over all combinations $\{\gamma_i\}$ for which there exists a machine $\{|\Phi_i\rangle \xrightarrow{\gamma_i} |\Psi_i\rangle\}_{i=1,\ldots,n}$.

When the number of possible input states is two, we can explicitly determine the achievable region (γ_1,γ_2) from Theorem 1 (the proof omitted):

Corollary 1: *Let* $\eta_{\mathrm{in}} := |\langle\Phi_1|\Phi_2\rangle|$ *and* $\eta_{\mathrm{out}} := |\langle\Psi_1|\Psi_2\rangle|$. *There exists a machine* $\{|\Phi_i\rangle \xrightarrow{\gamma_i} |\Psi_i\rangle\}_{i=1,2}$ *if and only if* $\gamma_1\geq 0$, $\gamma_2\geq 0$, *and*

$$\sqrt{(1-\gamma_1)(1-\gamma_2)} - \eta_{\mathrm{in}} + \eta_{\mathrm{out}}\sqrt{\gamma_1\gamma_2} \geq 0. \tag{5}$$

When $\eta_{\mathrm{in}} > \eta_{\mathrm{out}}$, the region (γ_1,γ_2) determined by Eq. (5) is convex, and is bounded by the line $\gamma_1 = 0$, the line $\gamma_2 = 0$, and the curve specified by the equality in Eq. (5), which connects the points $(\gamma_1,\gamma_2) = (0, 1-\eta_{\mathrm{in}}^2)$ and $(\gamma_1,\gamma_2) = (1-\eta_{\mathrm{in}}^2, 0)$ through the point $\gamma_1 = \gamma_2 = (1-\eta_{\mathrm{in}})/(1-\eta_{\mathrm{out}})$. When $\eta_{\mathrm{in}} \leq \eta_{\mathrm{out}}$, $\gamma_1 = \gamma_2 = 1$ satisfies Eq. (5), namely, a deterministic machine $\{|\Phi_i\rangle \xrightarrow{1} |\Psi_i\rangle\}_{i=1,2}$ exists. Note that Eq. (5) still forbids regions of (γ_1,γ_2) close to $(1,0)$ and $(0,1)$, reflecting the indistinguishability of the two input states.

3 Probabilistic cloning of two states with supplementary information

In this section, we consider the case where one makes m copies of states $\{|\psi_1\rangle, |\psi_2\rangle\}$ with the help of supplementary information in the form of pure states $\{|\phi_1\rangle, |\phi_2\rangle\}$. We show that it is always better to try first the production of $m-1$ copies of the original information from the supplementary information alone, independently of the original state, which is implied by the following theorem (the proof omitted) :

Theorem 2: *If there exists a machine*

$$\{|\psi_i\rangle|\phi_i\rangle \xrightarrow{\gamma_i} |\psi_i\rangle^{\otimes m}\}_{i=1,2},$$

then there exist a machine

$$\{|\psi_i\rangle \xrightarrow{\gamma_i^A} |\psi_i\rangle^{\otimes m}\}_{i=1,2}$$

and a machine

$$\{|\phi_i\rangle \xrightarrow{\gamma_i^B} |\psi_i\rangle^{\otimes m-1}\}_{i=1,2}$$

with

$$\gamma_i^B + (1-\gamma_i^B)\gamma_i^A \geq \gamma_i \quad (i=1,2). \tag{6}$$

If the original information is held by Alice, and the supplementary information by Bob, Theorem 2 implies that the optimal performance is always achieved just by one-bit classical communication from Bob to Alice as follows: Bob, who possesses the supplementary state $|\phi_i\rangle$, first runs the machine $\{|\phi_i\rangle \xrightarrow{\gamma_i^B} |\psi_i\rangle^{\otimes m-1}\}_{i=1,2}$, and tells Alice whether the trial was successful or not. In the successful case, Alice just leaves her state $|\psi_i\rangle$ as it is, and hence they obtain m copies in total. If Bob's attempt has failed, Alice runs the machine $\{|\psi_i\rangle \xrightarrow{\gamma_i^A} |\psi_i\rangle^{\otimes m}\}_{i=1,2}$. The total success probability for input state $|\psi_i\rangle|\phi_i\rangle$ in this protocol is given by $\gamma_i^B + (1-\gamma_i^B)\gamma_i^A$. Hence, by Theorem 2, we see that the above protocol is as good as any other protocol in which Alice and Bob communicate through quantum channels. Note that when $\{|\psi_i\rangle\}$ includes no pair of identical states, $\lim_{m\to\infty}\langle\psi_i|\psi_j\rangle^m = \delta_{ij}$ holds for any $i\neq j$. Hence in the limit $m\to\infty$ the machine $\{|\psi_i\rangle|\phi_i\rangle \xrightarrow{\gamma_i} |\psi_i\rangle^{\otimes m}\}_{i=1,\ldots,n}$ effectively carries out unambiguous discrimination of the set $\{|\psi_i\rangle|\phi_i\rangle\}$. Therefore, in this limit Theorem 2 reproduces the results in [13], namely, local operations and classical communication achieves the global optimality of unambiguous discrimination of any two pure product states with arbitrary a priori probability p_i.

When the initial state $|\psi_i\rangle|\phi_i\rangle$ is chosen with probability p_i, it follows from Theorem 2 that the maximum overall success probability is achieved by the above two-step protocol. For a special case of $p_1 = p_2 = 1/2$, we can directly confirm this as follows. The maximum overall success probability γ_{totmax} can easily be calculated by optimizing $(\gamma_1+\gamma_2)/2$ over the region in Corollary 1, and it is found to be

$$\gamma_{\mathrm{totmax}} = \frac{1-|\alpha\beta|}{1-|\alpha|^m}, \tag{7}$$

where $\alpha := \langle\psi_1|\psi_2\rangle$ and $\beta := \langle\phi_1|\phi_2\rangle$. Corollary 1 also shows the existence of a machine $\{|\phi_i\rangle \xrightarrow{\gamma_i^B} |\psi_i\rangle^{\otimes m-1}\}_{i=1,2}$ with

$$\gamma_1^B = \gamma_2^B = \frac{1-|\beta|}{1-|\alpha|^{m-1}} \tag{8}$$

and a machine $\{|\psi_i\rangle \xrightarrow{\gamma_i^A} |\psi_i\rangle^{\otimes m}\}_{i=1,2}$ with

$$\gamma_1^A = \gamma_2^A = \frac{1-|\alpha|}{1-|\alpha|^m}. \tag{9}$$

Hence, using these machines in the two-step protocol, we obtain an overall success probability

$$\gamma_1^B + (1-\gamma_1^B)\gamma_1^A = \frac{1-|\alpha\beta|}{1-|\alpha|^m}, \tag{10}$$

which coincides with $\gamma_{\rm totmax}$. For cases with general (p_1, p_2), it is even difficult to represent $\gamma_{\rm totmax}$ in an explicit form, but Theorem 2 states that $\gamma_{\rm totmax}$ is always achieved by the two-step protocol.

4 Probabilistic cloning with supplementary information for three or more states

When the number of the possible states is three or more, Theorem 2 is not always true, and there may exist a better protocol than just running machines $\{|\phi_i\rangle \xrightarrow{\gamma_i^B} |\psi_i\rangle^{\otimes m-1}\}_{i=1,2,\ldots,n}$ and $\{|\psi_i\rangle \xrightarrow{\gamma_i^A} |\psi_i\rangle^{\otimes m}\}_{i=1,2,\ldots,n}$. We will give such an example in this section, and also show that a somewhat stronger statement holds about how the supplementary and the original information should be combined to give the optimal performance. For that purpose, we assume that two separated parties, Alice and Bob, have the original information $|\psi_i\rangle$ and supplementary information $|\phi_i\rangle$, respectively. We do not care which of the parties produces the copies, as long as they produce m copies of $|\psi_i\rangle$ in total, namely, the task is successful when

$$|\psi_i\rangle_A|\phi_i\rangle_B \longrightarrow |\psi_i\rangle_A^{\otimes m-k}|\psi_i\rangle_B^{\otimes k}, (i=1,2,\ldots,n), \tag{11}$$

for any integer k. We consider two scenarios depending on the allowed communication between Alice and Bob:

Scenario (I): *Alice and Bob can use a one-way quantum channel from Bob to Alice.*

Note that this scenario is equivalent to the case where a single party having both the original and the supplementary information runs a machine $\{|\psi_i\rangle|\phi_i\rangle \xrightarrow{\gamma_i} |\psi_i\rangle^{\otimes m}\}_{i=1,\ldots,n}$, and its success probabilities are determined by Theorem 1.

Scenario (II): *Alice and Bob can use only a one-way classical channel from Bob to Alice.*

Note that the two-step protocol in the last section is included in this scenario. In what follows, we construct an example showing a gap between the two scenarios.

Consider an n-dimensional Hilbert space, and choose an orthonormal basis $\{|j\rangle\}_{j=1,\ldots,n}$. Let us define n normalized states $\{|\mu_j\rangle\}_{j=1,\ldots,n}$ as follows:

$$|\mu_j\rangle := \sqrt{1-(n-1)z^2}|j\rangle - z\sum_{i\neq j}|i\rangle, \tag{12}$$

where $z \geq 0$. The inner product between any pair of the states is given by

$$\langle\mu_i|\mu_j\rangle = z\left[(n-2)z - 2\sqrt{1-(n-1)z^2}\right] \tag{13}$$

for $i \neq j$. The right-hand side is zero for $z = 0$, and is $-1/(n-1)$ for $z = 1/\sqrt{n(n-1)}$. By continuity, we see that for any $\alpha \in [-1/(n-1), 0]$, there exists a set of n normalized states $\{|\mu_j\rangle\}_{j=1,\ldots,n}$ satisfying $\langle\mu_i|\mu_j\rangle = \alpha$ for $i \neq j$.

Now we consider a problem of producing m copies of a state chosen randomly ($p_j = 1/n$) from the set $\{|\psi_j\rangle\}_{j=1,\ldots,n}$ satisfying

$$\langle\psi_i|\psi_j\rangle = \alpha = -|\alpha| \ \ (0 < |\alpha| < \frac{1}{n-1}) \tag{14}$$

for any $i \neq j$, each accompanied by supplementary information $\{|\phi_j\rangle\}_{j=1,\ldots,n}$ satisfying

$$\langle\phi_i|\phi_j\rangle = \beta = -\frac{1}{n-1} \tag{15}$$

for any $i \neq j$. Both sets of states, $\{|\psi_j\rangle\}_{j=1,\ldots,n}$ and $\{|\phi_j\rangle\}_{j=1,\ldots,n}$, exist because they are special cases of the set $\{|\mu_j\rangle\}_{j=1,\ldots,n}$ above.

Let $\gamma_{\rm totmax}^{(I)}$ and $\gamma_{\rm totmax}^{(II)}$ be the maximum overall probabilities in Scenario (I) and in Scenario (II), respectively. We show that for any $n \geq 3$ and any $m \geq 2$, there is a gap between the two scenarios ($\gamma_{\rm totmax}^{(I)} > \gamma_{\rm totmax}^{(II)}$) for sufficiently small (but nonzero) $|\alpha|$.

First, we derive a lower bound for $\gamma_{\rm totmax}^{(I)}$, written as

$$\gamma_{\rm totmax}^{(I)} \geq \frac{1-|\beta||\alpha|}{1-|\beta||\alpha|^m} = \frac{n-1-|\alpha|}{n-1-|\alpha|^m}. \tag{16}$$

This relation can be proved via Theorem 1 with $|\Phi_i\rangle = |\psi_i\rangle|\phi_i\rangle$ and $|\Psi_i\rangle = |\psi_i\rangle^{\otimes m}$.

Next, we derive an upper bound on $\gamma_{\rm totmax}^{(II)}$. We use the following lemma (the proof omitted).

Lemma 1: *Consider a linearly independent set of n states $\{|\Psi_1\rangle, |\Psi_2\rangle, \ldots, |\Psi_n\rangle\}$, and another set of n states $\{|\Phi_1\rangle, |\Phi_2\rangle, \ldots, |\Phi_n\rangle\}$ satisfying*

$$\sum_{i=1}^{n} b_i|\Phi_i\rangle = 0. \tag{17}$$

If $b_j \neq 0$, there is no machine $\{|\Phi_i\rangle \xrightarrow{\gamma_i} |\Psi_i\rangle\}_{i=1,\ldots,n}$ with $\gamma_j > 0$.

In the problem at hand, the set of states $\{|\psi_i\rangle^{\otimes k}\}$ is linearly independent for any integer $k \geq 1$ since the eigenvalues of the $n \times n$ matrix $[\langle\psi_i|\psi_j\rangle^k]$ are only $1-\alpha^k > 0$ and $1+(n-1)\alpha^k > 0$. The set $\{|\phi_i\rangle\}$ satisfies $\sum_i |\phi_i\rangle = 0$ since $\sum_{i,j}\langle\phi_i|\phi_j\rangle = n+n(n-1)\beta = 0$. Then, we see from Lemma 1 that any machine $\{|\phi_i\rangle \xrightarrow{\gamma_i^B} |\psi_i\rangle^{\otimes k}\}_{i=1,\ldots,n}$ has zero success probability, $\gamma_1^B = \gamma_2^B = \cdots = \gamma_n^B = 0$. In Scenario (II), this fact implies that all of the m copies must be produced by Alice, and Bob's role is just to provide classical information to help Alice's operation. Hence, we are allowed to limit Bob's action to a POVM measurement $\{\hat{E}_\mu\}$ applied to his initial state $|\phi_i\rangle$, providing outcome μ with probability $p_{\mu|i} := \langle\phi_i|\hat{E}_\mu|\phi_i\rangle$.

Depending on the outcome μ received from Bob, Alice runs a machine $\{|\psi_i\rangle \xrightarrow{\gamma_i^{(\mu)}} |\psi_i\rangle^{\otimes m}\}_{i=1,\ldots,n}$ to produce m copies of state $|\psi_i\rangle$. Since the initial state $|\psi_i\rangle$ is randomly chosen ($p_i = 1/n$), the overall success probability is

$$\gamma_{\rm tot}^{(II)} = \sum_\mu p_\mu \left(\sum_{i=1}^n p_{i|\mu} \gamma_i^{(\mu)} \right). \tag{18}$$

where $p_\mu := \sum_i p_{\mu|i} p_i$ and $p_{i|\mu} := p_{\mu|i} p_i / p_\mu$.

From Theorem 1, $\Gamma^{(\mu)} := {\rm diag}(\gamma_1^{(\mu)}, \gamma_2^{(\mu)}, \ldots, \gamma_n^{(\mu)})$ should satisfy

$$\boldsymbol{b}^\dagger \left(X - \sqrt{\Gamma^{(\mu)}} Y \sqrt{\Gamma^{(\mu)}} \right) \boldsymbol{b} \geq 0 \tag{19}$$

for any $\boldsymbol{b}$. Here the elements of matrices X and Y are given by $X_{ij} = (1-\alpha)\delta_{i,j} + \alpha$ and $Y_{ij} = [(1-\alpha^m)\delta_{i,j} + \alpha^m] \langle P^{(i)} | P^{(j)} \rangle$. If we choose $\boldsymbol{b} = (\sqrt{p_{1|\mu}}, \sqrt{p_{2|\mu}}, \ldots, \sqrt{p_{n|\mu}})$, after a somewhat lengthy computation, we have

$$\gamma_{\rm totmax}^{(II)} \leq \frac{1-|\alpha|}{1-(n-1)|\alpha|^m}. \tag{20}$$

Combining Eqs. (16) and (20), we obtain

$$\gamma_{\rm totmax}^{(I)} - \gamma_{\rm totmax}^{(II)} \geq \frac{(n-2)|\alpha|(1+|\alpha|^m - n|\alpha|^{m-1})}{(n-1-|\alpha|^m)[1-(n-1)|\alpha|^m]}, \tag{21}$$

which shows that $\gamma_{\rm totmax}^{(I)} > \gamma_{\rm totmax}^{(II)}$ when $|\alpha|$ is a small enough positive number.

5 Summary

In this paper, we have discussed probabilistic cloning of a mutually nonorthogonal set of pure states $\{|\psi_1\rangle, \ldots, |\psi_n\rangle\}$, with the help of supplementary information. It has turned out that the situation is quite different for $n = 2$ and for other cases. When $n = 2$, the role of the supplementary information is limited to just produce copies on its own, independently of the original state. This property is quite similar to the property in deterministic cloning, stated in the stronger no-cloning theorem. For $n \geq 3$, such a simple property does not hold any longer. We assumed that the original and the supplementary information are held by separated parties, and asked what kind of communication is required to achieve the optimal performance. We have found examples in which the optimum performance cannot be achieved even if we allow any amount of classical communication from the party with the supplementary information to the other. If we limit to the one-way communication scenarios, this result means that a non-classical interaction between the supplementary and the original information helps to improve the performance. On the other hand, if we allow the flow of information in the other direction, we are not sure the gap still exists. Analysis of such two-way protocols will be an interesting problem. The cases where the set $\{|\psi_i\rangle\}$ includes a mutually orthogonal pair, or the cases where supplementary information is provided as a mixed state are also worth investigating.

Acknowledgements

We thank R. Namiki, S.K. Ozdemir, and T. Yamamoto for helpful discussions. This work was supported by 21st Century COE Program by the Japan Society for the Promotion of Science and by a MEXT Grant-in-Aid for Young Scientists (B) 17740265.

References

[1] W. K. Wootters and W. H. Zurek, Nature (London) **299**, 802 (1982).

[2] H.P. Yuen, Phys. Lett. **113A**, 405 (1986).

[3] V. Buzek and M. Hillery, Phys. Rev. A **54**, 1844 (1996).

[4] D. Mozyrsky, V. Privman, and M. Hillery, Phys. Lett. A **226**, 253 (1997).

[5] N. Gisin and, B. Huttner, Phys. Lett. A **228**, 13 (1997).

[6] V. Buzek, V. Vedral, M.B. Plenio, P.L. Knight, and M. Hillery, Phys. Rev. A **55**, 3327 (1997).

[7] M. Hillery and V. Buzek, Phys. Rev. A **56**, 1212 (1997).

[8] D. Bruß, D.P. DiVincenzo, A. Ekert, C.A. Fuchs, C. Macchiavello, and J.A. Smolin, Phys. Rev. A **57**, 2368 (1998).

[9] N. Gisin and S. Massar, Phys. Rev. Lett. **79**, 2153 (1997).

[10] D. Bruß, A. Ekert, and C. Macchiavello, Phys. Rev. Lett. **81**, 2598 (1998).

[11] L.M. Duan and G.C. Guo, Phys. Rev. Lett. **80**, 4999 (1998).

[12] R. Jozsa, quant-ph/0204153.

[13] Y.X. Chen and D. Yang, Phys. Rev. A **64**, 064303 (2001).

A quantum protocol to win the graph colouring game on all Hadamard graphs

David Avis[1] * Jun Hasegawa[2 3] † Yosuke Kikuchi[3] ‡ Yuuya Sasaki [2 §]

[1]*Department of Computer Science, McGill University, 3480 University, Montreal, Quebec, Canada H3A 2A7*

[2]*Department of Computer Science, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

[3]*ERATO QCI Project, JST, Hongo White Building, Hongo 5-28-3, Bunkyo-ku, Tokyo 113-0033, Japan*

Abstract. This paper deals with graph colouring games, an example of pseudo-telepathy, in which two provers can convince a verifier that a graph G is c-colourable where c is less than the chromatic number of the graph. They win the game if they convince the verifier. It is known that the players cannot win if they share only classical information, but they can win in some cases by sharing entanglement. The smallest known graph where the players win in the quantum setting, but not in the classical setting, was found by Galliard, Tapp and Wolf and has 32,768 vertices. It is a connected component of the Hadamard graph G_N with $N = c = 16$. Their protocol applies only to Hadamard graphs where N is a power of 2. We propose a protocol that applies to all Hadamard graphs. Combined with a result of Frankl, this shows that the players can win on any induced subgraph of G_{12} having 1609 vertices, with $c = 12$. Combined with a result of Frankl and Rodl, our result shows that all sufficiently large Hadamard graphs yield pseudo-telepathy games.

Keywords: graph colouring game, pseudo-telepathy, Hadamard graph, quantum chromatic number

1 Introduction

It is known that quantum entanglement allows for a phenomenon called pseudo-telepathy, that is, two parties pretend to be endowed with telepathic powers, as described in a survey paper by Brassard, Broadbent and Tapp [1]. For two parties A and B, a pair of question $(q_a, q_b) \in Q_A \times Q_B$ are given and then a pair of answer $(a_a, a_b) \in A_A \times A_B$ are returned. In an initial phase, the parties can communicate with each other and share information. If the parties share an entangled state, then this setting is called shared entanglement. In the second phase, the parties are no longer allowed to communication with each other before revealing their answers. The parties *win* this instance, if $a_a = a_b \Leftrightarrow q_a = q_a$. A protocol is successful with probability p if it wins any instance that satisfies the promise with probability at least p. This exchange is called a *pseudo-telepathy game* if there is a protocol that is successful with probability 1 with shared entanglement and does not admit such a protocol that is successful with probability 1 without sharing entanglement.

The graph colouring game is an example of a pseudo-telepathy game (Brassard, Cleve and Tapp [1], Cleve, Høyer, Toner and Watrous [3]). In a *graph colouring game*, there are two provers, called Alice and Bob, and a verifier. A graph $G(V, E)$ and a integer c is given to Alice and Bob. Alice and Bob agree on a protocol to convince the verifier that G is c-colourable. The verifier sends $a \in V$ to Alice and sends $b \in V$ to Bob such that $a = b$ or $(a, b) \in E$. Alice and Bob are not permitted to communicate after receiving a and b. Alice sends the colour c_A of a to the verifier and Bob sends the colour c_B of b to the verifier. Alice and Bob win if $a \neq b$ and $c_A \neq c_B$ or $a = b$ and $c_A = c_B$, and lose otherwise. They win the game if they convince the verifier.

The chromatic number $\chi(G)$ of a graph G is the the smallest number of colours that can be assigned to vertices such that no two adjacent vertices receive the same colour. If $c \geq \chi(G)$ then there exists a protocol to win with probability 1 by using a colouring of G with c colours. Otherwise, Alice and Bob cannot win the game with probability 1 using classical methods [2]. Using shared entanglement however, there are graphs where they can win in this situation. The Hadamard graphs, defined by Ito [9, 10], provide such examples.

Let $N = 4k$ for any positive integer k. The *Hadamard graph* G_N is defined as the graph whose vertex set $V_N = \{0, 1\}^N$ and edge set $E_N = \{(u, v) \in V_N^2 | d_H(u, v) = N/2\}$, where $d_H(u, v)$ means Ham-

*avis@cs.mcgill.ca

†hasepyon@is.s.u-tokyo.ac.jp

‡kikuchi@qci.jst.go.jp

§y_sasaki@is.s.u-tokyo.ac.jp

ming distance of u and v. Hadamard graphs are related to Hadamard matrices, which are an important object of study in combinatorics, especially design theory (e.g., see Stinson [12]). One of the open problems is to decide whether for every k, there exists a Hadamard matrix of order $4k$.

With shared entanglement Brassard , Cleve and Tapp [2] showed that Alice and Bob win the graph colouring game with probability 1 for G_{2^n} and $c = 2^n$ by using the Deutsch-Jozsa protocol and by sharing an entangled state. A result of Frankl and Rodl [5] (Theorem 1.11) implies that for all large enough n, $\chi(G_{2^n}) > 2^n$, and so asymptotically the game is an example of a pseudo-telepathy game. Galliard, Tapp and Wolf [6] showed that this was already the case for $n = 16$ using a rather complicated combinatorial argument. Thus the graph colouring game for G_{16} and $c = 16$ is a pseudo-telepathy game.

We extend these results in this paper to all Hadamard graphs. In the next section we state known results on the chromatic number of these graphs. In Section 3, we design a protocol to win the graph colouring game for all Hadamard graph with probability 1. Combing these results it is shown that the graph colouring game for G_{12} with $c = 12$ is a pseudo-telepathy game. Furthermore this is the smallest value of N for which G_N is a pseudo-telepathy game with $c = N$, and this holds for any induced subgraph with at least 1069 vertices.. The concluding section proposes a definition of quantum chromatic number and gives some open problems.

2 Chromatic number of Hadamard graphs

It is easily seen that $\chi(G_4) = 4$ and Ito [10] proved $\chi(G_8) = 8$. Given a graph $G = (V, E)$, the *independence number* of the graph, denoted $\alpha(G)$, is the cardinality of the largest subset of vertices such that no two of them are joined by an edge. Let $p \geq 3$ be an odd prime, $q \geq 1$, and $k = p^q$. Frankl [4] showed that

$$\alpha(G_{4k}) = 4 \sum_{i=0}^{k-1} \binom{4k-1}{i} < \frac{4^{4k}}{3^{3k}}.$$

When $p = 3$ and $q = 1$ we get $\alpha(G_{12}) = 268$. An elementary result of graph theory is that $\chi(G) \geq |V|/\alpha(G)$. Therefore $\chi(G_{12}) \geq 4096/268 > 12$. In fact G_N consists of two identical connected components, each having independence number half of that of G_N. Let H be any induced subgraph, having 1609 vertices, of one of the connected components of G_{12}. (In an induced subgraph, two vertices are adjacent if and only if they are adjacent in the original graph.) Then $\chi(H) \geq 1609/134 > 12$.

Since $\chi(G_4) = 4$, G_4 is not a pseudo-telepathy game with $c = 4$. Similarly, G_8 is not a pseudo-telepathy game with $c = 8$. However we will see that H is a pseudo-telepathy game with $c = 12$.

3 A protocol with probability 1 using QFT

In this section we extend the protocol by Brassard, Cleve and Tapp [2]. Their protocol employs the quantum Hadamard transform while our protocol employs the quantum Fourier transform (QFT) with any order, which can be exactly done as shown by Mosca and Zalka [11].

We describe a protocol such that Alice and Bob win the graph colouring game of G_N with probability 1 where $2^{n-1} < N \leq 2^n$. In this protocol, we use the following two operations QFT_N and P_{l_i}: QFT_N is a *general* quantum Fourier transform with order N, not necessarily 2^n, defined as

$$\mathrm{QFT}_N : |i\rangle \longrightarrow \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} (\omega)^{i \cdot j} |j\rangle, \qquad (1)$$

where $\omega = \exp\left(\frac{2\pi\sqrt{-1}}{N}\right)$. Mosca and Zalka [11] show that this QFT with any order can be performed exactly. The operation P_{l_i} is a phase shift corresponding to the i-th bit of an input string l

$$P_{l_i} : |i\rangle \mapsto (-1)^{l_i} |i\rangle. \qquad (2)$$

Our protocol has four steps. Alice and Bob can communicate with each other at only step 1.

Step 1: Prepare initial state $|\Psi_{AB}\rangle$
In step I, Alice and Bob prepare $2n$-qubits $|0\rangle^{\otimes n} \otimes |0\rangle^{\otimes n}$. Alice has the first n-qubits and Bob has the second. Alice first applies QFT_N to her N-qubits:

$$|0\rangle^{\otimes n} \otimes |0\rangle^{\otimes n} \overset{QFT_N}{\mapsto} \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \otimes |0\rangle. \qquad (3)$$

Bob then applies controlled-NOT operations to his n-qubits with Alice's qubits as control qubits for sharing the initial entanglement states:

$$\frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \otimes |i\rangle =: |\Psi_{AB}\rangle. \qquad (4)$$

Step 2: Apply phaseshift P_{a_i} and P_{b_i}
Alice and Bob compute

$$(P_{a_i} \otimes P_{b_i})|\Psi_{AB}\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} (-1)^{a_i \oplus b_i} |i\rangle \otimes |i\rangle. \qquad (5)$$

Step 3: Apply the general quantum Fourier transform
Alice and Bob compute

$$\begin{aligned}&\left(\mathrm{QFT}_N \otimes \mathrm{QFT}_N^{-1}\right)\left(P_{a_i} \otimes P_{b_i}\right)|\Psi_{AB}\rangle \\ &=\left(\frac{1}{\sqrt{N}}\right)^3 \sum_{j_A=0}^{3N-1} \sum_{j_B=0}^{N-1} \sum_{i=0}^{N-1} (\omega)' (-1)^{a_i \oplus b_i} |j_A\rangle \otimes |j_B\rangle,\end{aligned} \tag{6}$$

where $(\omega)' = (\omega)^{i \cdot (j_A - j_B)}$.
Step 4: Measure
Alice and Bob measure her(his) qubits using the computational bases and each obtain one of N basis states. They output colours corresponding to their measurement result to verifier.

We have the following theorem for our protocol.

Theorem 1 *Alice and Bob win the game with probability* 1 *by our protocol.*

Proof. Suppose that Alice(Bob) receives $a(b)$ and sends $c_A(c_B)$. The probability that Alice and Bob obtain basis states $|j\rangle \otimes |j\rangle$ after measurement is

$$\begin{aligned}&\left|\langle j| \otimes \langle j| \left(\mathrm{QFT}_N \otimes \mathrm{QFT}_N^{-1}\right)\left(P_{a_i} \otimes P_{b_i}\right)|\Psi_{AB}\rangle\right|^2 \\ &= \left| \left(\frac{1}{\sqrt{N}}\right)^3 \sum_{i=0}^{N-1} (-1)^{a_i \oplus b_i} \right|^2.\end{aligned}$$

In case of $a = b$, it holds that $a_i \oplus b_i = 0$ for any i. Thus the probability of $c_A = c_B$ is

$$\Pr[c_A = c_B] = 1.$$

On the other case of $a \neq b$, it holds that $d_H(a, b) = \frac{N}{2}$, because of the definition of the Hadamard graph. It means that

$$\Pr[c_A = c_B] = 0.$$

□

By combining Frankl's result [4] and Theorem 1, there is a gap between the shared entanglement setting and otherwise for G_{12}, and for the smaller subgraph H mentioned in the previous subsection. We have obtained the following result.

Theorem 2 *The smallest Hadamard graph G_N such that the graph colouring game is a pseudo-telepathy game with $c = N$ is G_{12}. Any of its induced subgraphs with 1609 vertices also has this property.*

Godsil and Newman have proved that a Hadamard graph G_N has chromatic number strictly larger than N whenever $N = 4m > 8$ [8]. Then next result holds.

Theorem 3 *The graph colouring game for Hadamard graph G_{4m} is a pseudo-telepathy game with $c = 4m$ for all $m \geq 3$.*

The final statement of this theorem uses the result of Frankl and Rodl [5] mentioned in Section 1.

4 Concluding remarks

In this paper, we have dealt with two party case for the quantum colouring game. It may be interesting to investigate the multi-party case for quantum colouring game.

The chromatic number $\chi(G)$ of a graph G is equal to the minimum number of colours such that Alice and Bob win the graph colouring game for G with probability 1 without shared entanglement. Patrick Hayden [private communication] suggested we define the quantum chromatic number $\chi_Q(G)$ as the minimum number of colours such that Alice and Bob win the graph colouring game for G, using shared entanglement, with probability 1. It is easy to see that $\chi_Q(G) \leq \chi(G)$, and the pseudo-telepathy graph colouring game is concerned with graphs with $\chi_Q(G) < \chi(G)$. Characterizing such graphs G would be interesting from both the standpoint of quantum communication and of combinatorics. What is the smallest such graph?

For Hadamard graphs G_{4p^q}, there is an exponential gap between the chromatic number and the quantum chromatic number. What is the largest such gap as a function of the number of vertices of G?

Acknowledgments

The authors would like to thank Anne Broadbent, Patrick Hayden, Hiroshi Imai and Francois Le Gall for their helpful comments and discussions.

References

[1] G. Brassard, A. Broadbent and A. Tapp, Quantum pseudo-telepathy, *Foundations of Physics*, to appear, Preprint: arXiv.org e-Print quant-ph/0407221, 2004.

[2] G. Brassard, R. Cleve and A. Tapp, Cost of exactly simulating quantum entanglement with classical communication *Physical Review Letters*, 83(9), 1874 1877, 1999.

[3] R. Cleve, P. Høyer, B. Toner and J. Watrous, Consequences and limits of nonlocal strategies, *Proc. 19th IEEE Annual Conference on Computational Complexity*, 236-249, 2004.

[4] P. Frankl, Orthogonal vectors in the n-dimensional cube and codes with missing distances, *Combinatorica*, 6(3), 279 285, 1986.

[5] P. Frankl and V. Rodl. Forbidden intersections. *Trans. American Mathematical Society*, 300, 259-286, 1987.

[6] V. Galliard A. Tapp and S. Wolf. The impossibility of pseudo-telepathy without quantum entanglement. *Proceedings of ISIT 2003*, 457, 2003, Preprint: arXiv.org e-Print quant-ph/0211011, 2002.

[7] V. Galliard and S. Wolf. Pseudo-telepathy, entanglement, and graph colorings. *Proceedings of ISIT 2002*, 101, 2002.

[8] C. D. Gosil and M. W. Newman. Colouring an Orthogonality Graph. Preprint: arXiv.org e-Print math.CO/0509151, 2005.

[9] N. Ito. Hadamard graphs. I *Graphs and Combinatorics*, 1, 57 64, 1985.

[10] N. Ito, Hadamard graphs. II *Graphs and Combinatorics*, 1, 331 337, 1985.

[11] M. Mosca and C. Zalka. Exact quantum fourier transforms and discrete logarithm algorithms. *International Journal of Quantum Information*, 2, 91 100(2004).

[12] D. Stinson. *Combinatorial Designs: Construction and Analysis*. Springer, 2004.

Network Quantum Shannon Theory

Jon Yard[1]

[1] *Center for the Physics of Information*
California Institute of Technology
Pasadena, California, 91125, USA

Abstract. I will talk about quantum channels which have many senders or many receivers. As such channels can be used to simultaneously distribute classical and quantum information among separated parties in many different ways, I will focus on some special cases which generalize results from the literature of classical information theory on broadcast and multiple access channels. Focus will be placed on channels for which single-letter characterizations of their associated rate regions are obtainable.

Claude Shannon initiated the study of communication theory as an abstract mathematical discipline. By modeling communication channels using conditional probabilities, it is possible to obtain a theory which, in many cases, reasonably models the underlying physics of the communication media, while also leading to a rich mathematical framework with well-posed problems and enlightening solutions. In this area of research, nowadays known as "Shannon theory," the central theme is that of proving whether arbitrarily good codes exist which send information at a given rate. Mathematically, this amounts to proving whether or not there exist sequences of codes *achieving* a particular rate of R bits per channel use, in the sense that the probability of a decoder error can be made arbitrarily small, at a potential cost of requiring many channel uses. The boundary between achievable and non-achievable rates is the capacity C of the channel. The capacity is an efficiently computable function of the transition matrix $p(y|x)$ of the channel, and has a simple characterization as the maximum of the mutual information between the input and output of the channel over all probability distributions on the input symbols.

The extension of Shannon's initial theory to channels with possibly many senders and receivers has been called network Shannon theory. For a general such channel, there are a multitude of ways in which it can be used to distribute information from the senders to the receivers. While no solution to this general problem is known, there is a large literature on simpler versions of the problem. Among these, two of the most central results are those on multiple access channels [1, 2] and broadcast channels [3].

A multiple access channel $p(z|x,y)$ with two senders and a single receiver can be used, for instance, to send independent information from each sender to the receiver. The pairs of nonnegative rates (R_X, R_Y) at which this can be done is called a *capacity region*, whose boundary is analogous to the single-user capacity described above. For a generic two-dimensional capacity region, consider the points in $\mathbb{R}^2$ underneath the curve pictured in Figure 1. This picture reveals the tradeoff between the rates at which each sender can transmit reliably. Namely, each senders rate is maximal when the other's is zero. For this reason, the boundary of the capacity region is referred to as a *tradeoff curve*.

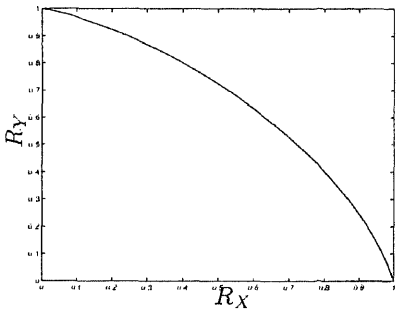

Figure 1: Generic capacity region

A broadcast channel $p(y,z|x)$ with one sender and two receivers can be used to send a common message to both receivers, as well as (say) a personalized or even private one to one of the receivers. It is possible to characterize the capacity region of pairs of rates at which such tasks are possible in a similar manner as with the multiple access channel. In this talk, I will discuss work that I have done in the last year or so with Igor Devetak and Patrick Hayden which extends these early results from the network Shannon theory literature to the quantum domain [4, 5, 6, 7].

When quantum physics dominates the properties of a channel, the replacement of conditional probabilities by completely positive, trace preserving maps yields, on the one hand, more refined channel models, and on the other hand, the ability to prove capacity theorems which essentially include earlier classical results as special cases. As such quantum channels can be used to transmit classical and quantum information at the same time, there is a classical-quantum tradeoff curve associated with each channel [8]. Using n instances of a given channel, sending classical information at a rate of R bits per channel use amounts to enabling the receiver to distinguish a set of of 2^{nR} input preparations arbitrarily well via measurement. In this talk, I will focus on the simplest of the many equivalent notions of quantum communication: namely, that of generating entanglement across the channel. The users of the n channels are said to *generate entanglement at rate* Q if they are able to create, using the channels, a state which is arbitrarily close to a *rate* Q *maximally*

entangled state

$$|\Phi\rangle \equiv \frac{1}{\sqrt{2^{nQ}}} \sum_{m=1}^{2^{nQ}} |m\rangle|m\rangle.$$

This therefore opens up the possibilities for analyzing quantum versions of classical network Shannon-theoretic problems in which classical and or quantum information is transmitted simultaneously. For instance, we will see in this talk that associated to each quantum multiple access channel is a four-dimensional capacity region containing the quadruples of rates at which each of the two senders can send classical and quantum information simultaneously.

One may do the same with a broadcast channel, but there are new possibilities for states to generate. Analogous to transmitting a common classical message from one sender to two receivers is to generate some sort of large tripartite state. We have shown that this can be done when that state is a *rate Q GHZ state*

$$|\Gamma\rangle \equiv \frac{1}{\sqrt{2^{nQ}}} \sum_{m=1}^{2^{nQ}} |m\rangle|m\rangle|m\rangle,$$

where the subsystems are held by the sender and the two receivers.

There is, however, somewhat of a stumbling block within this program. In the culture of classical Shannon theory, a capacity region is not generally considered to be solved unless it is given in terms of a *single-letter characterization*. These are usually only obtainable when the given characterization is *additive*. Quantum information theory is currently plagued with additivity problems [9]. To begin, the best known characterization of the quantum capacity of a quantum channel is known not to be additive for every channel. In addition, it is still not known whether the classical capacity of an arbitrary quantum channel is additive. Worse yet, the collections of channels for which additivity has been proved in each case have a somewhat narrow intersection, which limits the classes of channels whose classical-quantum tradeoff curves can be proven to be additive. Another issue is that certain standard "tricks" involving entropy manipulations for proving converse theorems in classical network Shannon theory fail to carry over to the quantum extension, except for certain special cases which I will discuss in detail.

This insistence on single-letter formulae describing the capacity regions is not merely a cultural artifact. There are generally many different ways to characterize a capacity region in terms of nonadditive quantities. We will see how this arises for the case where each sender of a multiple access channel sends quantum information to the common receiver. I will give two characterizations of this region – for the first, we have only been able to identify trivial channels for which it is additive. In the second, the formulas are directly analogous to those which arise in the original classical solution (which, incidentally, is single-letter for *every* classical channel), although we are able to find a nontrivial channel for which this second characterization is additive, allowing us to explicitly write down its associated single-letter capacity region.

It is thus reasonable to expect that as more becomes known about the additivity properties of single-user channels for transmitting classical and or quantum information, more possibilities will be opened for providing single-letter characterizations in the network theory counterpart. On the other hand, network problems show us that, while regularizations of one-shot capacity formulas can often easily be shown to equal the actual capacity or capacity region for the problem at hand, such characterizations are not necessarily the end of the road, perhaps indicating that there is still more to be known about the quantum capacities of single-user channels.

References

[1] R. Ahlswede, "Multi-way communication channels," Second Intern. Sympos. on Inf. Theory, Thakadsor, 1971, Publ. House of the Hungarian Adad. of Sciences, 23–52, 1973.

[2] Liao, "Multiple access channels", Ph.D. dissertation, Dept. of Electrical Engineering, University of Hawaii, 1972.

[3] T. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. 18, pp. 2–14, January 1972.

[4] J. Yard, I. Devetak, P. Hayden, "Capacity theorems for quantum multiple access channels – classical-quantum and quantum-quantum capacity regions," submitted to *IEEE Trans. Inform. Theory*, `quant-ph/0501045`.

[5] J. Yard, "Simultaneous classical-quantum capacities of quantum multiple access channels," Ph.D. dissertation, Electrical Engineering Dept., Stanford University, March 2005, `quant-ph/0506050`.

[6] J. Yard, I. Devetak, P. Hayden, "Capacity theorems for quantum multiple access channels," *Proc. IEEE Intern. Sympos. Inform. Theory*, Adelaide, Australia, September 2005, `cs.IT/0508031`.

[7] J. Yard, I. Devetak, P. Hayden, "Quantum broadcast channels," in preparation.

[8] I. Devetak, P. Shor, "The capacity of a quantum channel for simultaneous transmission of classical and quantum information," `quant-ph/0311131`.

[9] P. Shor, "Equivalence of additivity questions in quantum information theory," `quant-ph/0305035`.

Dualities in quantum information theory

Igor Devetak[1] *

[1]*Electrical Engineering Department, University of Southern California, Los Angeles, CA 90089, USA*

Abstract.
Two quantum information processing protocols are said to be dual under resource reversal if the resources consumed (generated) in one protocol are generated (consumed) in the other. We define quantum feedback channels, and show that they may be reversibly decomposed into a perfect quantum channel and pure entanglement. The dual protocols responsible for this decomposition are the "feedback father" (FF) protocol and the "fully quantum reverse Shannon" (FQRS) protocol. Moreover, the "fully quantum Slepian-Wolf" protocol (FQSW), a generalization of the recently discovered "quantum state merging", is related to FF by source-channel duality, and to FQRS by time reversal duality, thus forming a triangle of dualities.

Keywords: quantum information theory, time-reversal, source coding, channel coding, resource inequality, duality

The canonical example of an entangled state is an ebit, or EPR pair, $\Phi^{AB} = 1/\sqrt{2}\,(|0\rangle^A|0\rangle^B + |1\rangle^A|1\rangle^B)$ shared between two spatially separated parties Alice and Bob. The systematic study of entanglement was initiated by the realization that a general pure bipartite state $|\phi\rangle^{AB}$ is *asymptotically equivalent* to a real number E of ebits, where $E = H(A)_\phi$, and $H(A)_\phi = -\mathrm{Tr}\,\phi^A \log \phi^A$ is the von Neumann entropy of the restriction $\phi^A = \mathrm{Tr}_B(\phi^{AB})$ of the state $\phi^{AB} = |\phi\rangle\langle\phi|^{AB}$ to Alice's system. For any $\epsilon, \delta > 0$ and sufficiently large number n, there exists a protocol which transforms n copies of $|\phi\rangle^{AB}$ to a state that is ϵ-close (say, in trace distance) to $\lfloor n(E-\delta)\rfloor$ ebits. This protocol is called entanglement concentration [1], and we can symbolically write the statement of its existence as a *resource inequality* [2]

$$\langle\phi\rangle \geq H(A)_\phi\,[q\,q].$$

Here $\langle\phi\rangle$ is the infinite sequence $(\phi^{\otimes n})_{n=1}^\infty$. The notation used for ebits $[q\,q] := \langle\Phi\rangle$ was introduced in [3], along with corresponding notation for qubit channels $[q \to q]$, classical bit channels $[c \to c]$ and bits of common randomness $[c\,c]$. $R\langle\xi\rangle$ is defined as $(\xi^{\otimes\lfloor Rn\rfloor})_{n=1}^\infty$. In general we write an inequality $\geq$ between $(\phi_n)_{n=1}^\infty$ and $(\psi_n)_{n=1}^\infty$ if for any $\epsilon, \delta > 0$ and sufficiently large n there is a protocol transforming ϕ_n into an ϵ-approximation of $\psi_{\lfloor(1-\delta)n\rfloor}$. As it turns out, the reverse is also true. The entanglement dilution [4] resource inequality reads

$$H(A)_\phi\,[q\,q] \geq \langle\phi\rangle.$$

Dilution additionally consumes a sublinear amount classical communication, but this corresponds to an asymptotic rate of 0, and as such does not enter into the resource count. The two may be combined to give a resource *equality*

$$\langle\phi\rangle = H(A)_\phi\,[q\,q]. \tag{1}$$

A single number E thus suffices to characterize the asymptotic properties of the state $|\phi\rangle^{AB}$.

Another known resource equality regards "coherent communication" and follows from coherent versions of teleportation and super-dense coding [5]

$$2[q \to qq] = [q \to q] + [q\,q]. \tag{2}$$

Here $[q \to qq]$ represents the coherent classical bit channel (or *cobit*), an isometry $\Delta : A' \to AB$ defined by

$$\Delta : |i\rangle^{A'} \to |i\rangle^A|i\rangle^B, i = 0, 1,$$

where $\{|0\rangle, |1\rangle\}$ is a preferred orthonormal basis of a qubit system.

Fully quantum reverse Shannon inequality. The first result of this paper is a resource equality that generalizes both (1) and (2). We first introduce the concept of a *relative resource*. Usually, when Alice and Bob are connected by a quantum channel $\mathcal{N} : A' \to B$, no restriction is placed on Alice's input density operator, as long as it lives on a Hilbert space of the right dimension. For a fixed blocklength n, possessing a relative resource $\langle\mathcal{N} : \rho^{A'}\rangle$ means that only if Alice inputs a density operator close to $(\rho^{A'})^{\otimes n}$ is she guaranteed that the channel will behave like $\mathcal{N}^{\otimes n}$. Relative resources come about naturally in the context of quantum compression. Using Schumacher compression [6], Alice is able to convey a good approximation to n copies of some state $\rho^{A'}$ to Bob using $\approx nH(A')_\rho$ qubits of communication (for sufficiently large n, as usual). Letting $\varphi^{RA'}$ be a purification of $\rho^{A'}$ (i.e. a pure state such that $\mathrm{Tr}_R\varphi^{RA'} = \rho^{A'}$), this may be written as a resource inequality

$$H(R)_\varphi\,[q \to q] \geq \langle \mathrm{id}^{A'\to B} : \rho^{A'}\rangle, \tag{3}$$

i.e. we are able to simulate the identity channel $\mathrm{id}^{A'\to B}$, *assuming* that the input density operator is close to $(\rho^{A'})^{\otimes n}$. The same simulation could never work for an input density matrix of a higher entropy, by the converse to Schumacher's theorem [6]. What if one wishes to simulate an arbitrary channel $\mathcal{N} : A' \to B$? The quantum reverse Shannon theorem [7] gives us a way to do this:

$$H(B)_\sigma\,[q\,q] + I(R;B)_\sigma\,[c \to c] \geq \langle\mathcal{N}^{A'\to B} : \rho^{A'}\rangle, \tag{4}$$

where

$$\sigma^{RB} = (I^R \otimes \mathcal{N}^{A'\to B})\varphi^{RA'},$$

*devetak@usc.edu

and the *quantum mutual information* is defined as $I(R;B) = H(R) + H(B) - H(RB)$. In fact, the protocol that achieves (4) accomplishes slightly more [7]. A noisy channel $\mathcal{N}$ normally arises from an isometry $\tilde{\mathcal{U}} : A' \to BE$ with a larger target Hilbert space that includes the unobserved environment E, followed by the tracing out of E. In the simulation of $\mathcal{N}$ the environment is also simulated, and ends up in Alice's possession at the end of the protocol. Thus the channel we end up simulating is the *quantum feedback channel* $\mathcal{U} : A' \to AB$, an isometry from a system belonging to Alice to a system shared between Alice and Bob, and the resource inequality becomes

$$H(B)_\sigma [q\,q] + I(R;B)_\sigma [c \to c] \geq \langle \mathcal{U}^{A' \to AB} : \rho^{A'} \rangle.$$

It is shown in [8] that the protocol can be made "coherent" [5, 2], yielding the *fully quantum reverse Shannon* (FQRS) inequality

$$1/2\, I(R;B)_\psi [q \to q] + 1/2\, I(A;B)_\psi [q\,q] \geq \langle \mathcal{U}^{A' \to AB} : \rho^{A'} \rangle, \tag{5}$$

where

$$\psi^{RAB} = (I^R \otimes \mathcal{U}^{A' \to AB}) \varphi^{RA'} \tag{6}$$

is a purification of σ^{RB}. Schumacher compression is a special case of this inequality in which the feedback system A is absent.

Feedback father. The "father" inequality [2] regards entanglement-assisted quantum communication over the noisy quantum channel $\mathcal{N}$:

$$\langle \mathcal{N}^{A' \to B} \rangle + 1/2\, I(R;A)_\psi [q\,q] \geq 1/2\, I(R;B)_\psi [q \to q]. \tag{7}$$

The state ψ^{RAB} is defined by (6), noting that entropic coefficients are independent of the choice of $\mathcal{U} : A' \to AB$ for which $\mathcal{N} = \mathrm{Tr}_A \circ \mathcal{U}$. The first observation is that there is a protocol implementing (7) that merely requires the relative resource $\langle \mathcal{N}^{A' \to B} : \rho^{A'} \rangle$ instead of the full $\langle \mathcal{N}^{A' \to B} \rangle$,

$$\langle \mathcal{N}^{A' \to B} : \rho^{A'} \rangle + 1/2\, I(R;A)_\psi [q\,q] \geq 1/2\, I(R;B)_\psi [q \to q].$$

The second observation is that if the feedback channel $\mathcal{U}$ is given instead of the weaker $\mathcal{N}$, then applying the protocol from [2] achieving (7), Bob is left with a purification of Alice's system A, yielding an additional $H(A)[q\,q]$ after entanglement concentration. Thus

$$\begin{aligned} &\langle \mathcal{U}^{A' \to AB} : \rho^{A'} \rangle + 1/2\, I(R;A)_\psi [q\,q] \\ &\quad \geq \quad 1/2\, I(R;B)_\psi [q \to q] + H(A)_\psi [q\,q]. \end{aligned}$$

Canceling terms and using

$$H(A)_\psi = 1/2\, I(R;A)_\psi + 1/2\, I(A;B)_\psi, \tag{8}$$

gives the *feedback father* (FF) inequality:

$$\langle \mathcal{U}^{A' \to AB} : \rho^{A'} \rangle \geq 1/2\, I(R;B)_\psi [q \to q] + 1/2\, I(A;B)_\psi [q\,q]. \tag{9}$$

A special case of (9) where there is no actual feedback is the reverse of Schumacher compression:

$$\langle \mathrm{id}^{A' \to B} : \rho^{A'} \rangle \geq H(R)_\varphi [q \to q]. \tag{10}$$

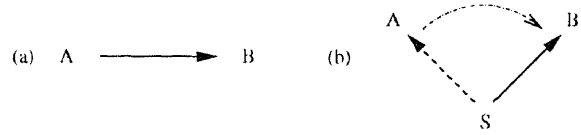

Figure 1: A channel (a) between Alice and Bob may be used in a source coding problem (b) to convert the channel from the Source to Alice, into a channel from the Source to Bob.

Duality #1: FQRS is related to FF by resource reversal. Clearly, (5) and (9) are reverses of each other, and together they give the resource equality

$$\langle \mathcal{U}^{A' \to AB} : \rho^{A'} \rangle = 1/2\, I(R;B)_\psi [q \to q] + 1/2\, I(A;B)_\psi [q\,q]. \tag{11}$$

A special case is (1) in which $\mathcal{U}$ is the *appending* channel $\mathcal{A}_\phi : A' \to A_1 A_2 B$, which relabels A' as A_1, and appends the state $\phi^{A_2 B}$ (i.e., it maps some $\rho^{A'}$ to $\rho^{A_1} \otimes \phi^{A_2 B}$). Another special case is (3) and (10), in which $\mathcal{U}$ is $\mathrm{id}^{A' \to B}$ and A is null. The third special case is (2), where $\mathcal{U}$ is the coherent classical bit channel Δ, and ρ is a maximally mixed qubit state τ (it can be shown that $\langle \Delta : \tau \rangle$ is equivalent to $\langle \Delta \rangle$).

A resource equality of this generality can tell us a lot about optimal transformations between the resources involved, such as the capacity of the quantum feedback channel $\mathcal{U}^{A' \to AB}$ for simultaneous generation of quantum communication and entanglement. The task is to find the set $\mathcal{S}(\mathcal{U})$ of rate pairs (Q, E) for which

$$\langle \mathcal{U}^{A' \to AB} \rangle \geq Q[q \to q] + E[q\,q].$$

The answer is given by $\mathcal{S}(\mathcal{U}) = \bigcup_{n \to \infty} 1/n\, \mathcal{S}^{(1)}(\mathcal{U}^{\otimes n})$, where

$$\mathcal{S}^{(1)} = \bigcup_\psi \{1/2(I(R;B)_\psi, 1/2 I(A;B)_\psi)\}.$$

Fully quantum Slepian-Wolf So far we have been dealing with what is traditionally known as *channel coding*: there are two parties Alice and Bob, and their task is to effect conversions between resources, whether static, dynamic or relative. In *source coding* there is an additional protagonist, the Source. The Source holds a state purified by some reference system R. Alice's and Bob's job is to effect conversions between relative resources originating at the Source. A simple example of a source type resource inequality is

$$\langle \mathrm{id}^{S \to \hat{A}} : \rho^S \rangle + \langle \mathrm{id}^{A' \to B} : \rho^{A'} \rangle \geq \langle \mathrm{id}^{S \to \hat{B}} : \rho^S \rangle,$$

illustrated in figure 1. Combining it with Schumacher compression (3) gives

$$\langle \mathrm{id}^{S \to \hat{A}} : \rho^S \rangle + H(S)_\rho [q \to q] \geq \langle \mathrm{id}^{S \to \hat{B}} : \rho^S \rangle.$$

While the two formulations are equivalent, compression is traditionally thought of in terms of source coding. More generally, the source may be initially distributed between Alice and Bob via a general isometry

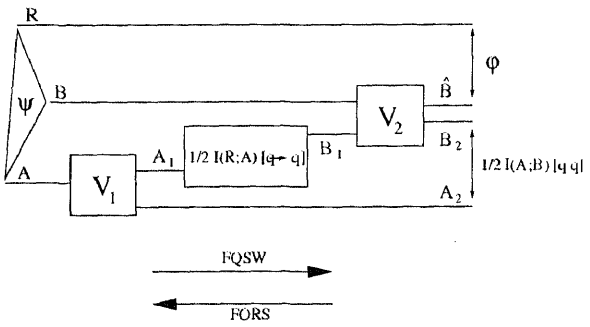

Figure 2: The protocol implementing FQSW consists of an isometry $V_1 : A \to A_1A_2$ performed by Alice, sending A_1, of size $\lfloor n(1/2\,I(R;A)_\psi + \delta)\rfloor$ qubits, through a quantum channel $\mathrm{id}^{A_1\to B_1}$ to Bob, and an isometry $V_2 : BB_1 \to B_2\hat{B}$ performed by Bob; it approximately transforms $(\psi^{RAB})^{\otimes n}$ into $(\varphi^{R\hat{B}})^{\otimes n}$ and $\lfloor n/2\,I(A;B)_\psi\rfloor$ ebits shared between Alice and Bob. The time reversal of this protocol implements the FQRS resource inequality.

$\mathcal{U} : S \to AB$, and the goal is to divert it entirely to Bob. A problem of this kind, the quantum Slepian-Wolf problem, was first solved in [9]. The result of [9] is generalized in [10] to the *fully quantum Slepian-Wolf* (FQSW) inequality, which reads

$$\begin{aligned}\langle \mathcal{U}^{S\to AB} : \rho^S\rangle + 1/2\,I(R;A)_\psi[q\to q] \\ \geq\ 1/2\,I(A;B)_\psi[q\,q] + \langle \mathrm{id}^{S\to\hat{B}} : \rho^S\rangle, \qquad (12)\end{aligned}$$

where

$$\psi^{RAB} = (I^R \otimes \mathcal{U}^{S\to AB})\varphi^{RS}$$

and $\varphi^{RS} = |\varphi\rangle\langle\varphi|^{RS}$ is a purification of ρ^S, cf. (6).

Since neither Alice nor Bob have control over the Source, (12) holds when applied to ρ^S, giving the "mother" inequality [2]

$$\langle \sigma^{AB}\rangle + 1/2\,I(R;A)_\psi[q\to q] \geq 1/2\,I(A;B)_\psi[q\,q],$$

which concerns quantum-communication-assisted distillation of entanglement from $\sigma^{AB} = \mathcal{U}^{S\to AB}\langle\rho^S\rangle$.

Duality #2: FF is related to FQSW via source-channel duality. The FF inequality (9) may be combined with Schumacher compression (3), to give, after cancellation of terms,

$$\begin{aligned}\langle \mathcal{U}^{A'\to AB} : \rho^{A'}\rangle + 1/2\,I(R;A)_\psi[q\to q] \\ \geq\ 1/2\,I(A;B)_\psi[q\,q] + \langle \mathrm{id}^{A'\to\hat{B}} : \rho^{A'}\rangle.\end{aligned}$$

This is a channel version of the FQSW, obtained by formally replacing S with A'! We refer to this phenomenon as *source-channel duality*. In the case where A is null, the inequalities reduce to the two equivalent formulations of Schumacher compression; in general, however, the two are incomparable. This observation sheds new light on the mysterious mother-father duality [2], as the mother and father inequalities stem from FQSW and FF, respectively.

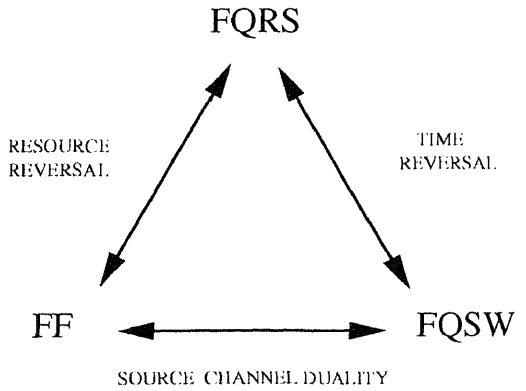

Figure 3: A triangle of dualities.

Duality #3: FQRS is related to FQSW by time reversal. We can make FQRS (5) into a source type inequality by adding $\langle \mathrm{id}^{S\to A'} : \rho\rangle$ to both sides of the equation:

$$\begin{aligned}\langle \mathrm{id}^{S\to A'} : \rho^S\rangle + 1/2\,I(A;B)_\psi[q\,q] + 1/2\,I(R;B)_\psi[q\to q] \\ \geq \langle \mathcal{U}^{S\to AB} : \rho^S\rangle.\end{aligned}$$

Interchanging the roles of A and B gives

$$\begin{aligned}\langle \mathrm{id}^{S\to B'} : \rho\rangle + 1/2\,I(A;B)_\psi[q\,q] + 1/2\,I(R;A)_\psi[q\leftarrow q] \\ \geq \langle \mathcal{U}^{S\to AB} : \rho^S\rangle.\end{aligned}$$

This is precisely the time-reversal of the FQSW inequality (12)! Unlike the previous two dualities, this one has *operational* implications: a protocol achieving (5) may be transformed into a protocol achieving (9), and vice versa (figure 2).

Our three dualities are summarized in figure 3.

The classical counterpart. To what degree do these dualities carry over to classical information theory? Let us define a classical (relative) feedback channel to take a random variable X as Alice's input, and output a related random variable Y to Bob, while feeding XY back to Alice. We shall use the simplified classical notation $\langle X_A \to (XY)_A Y_B\rangle$ for such a resource. The same role played by purification in the quantum world is played by copying in the classical world. Initially the reference system R contains a copy of Alice's input state X; whereas in the quantum case the feedback to A was a purification of the RB system, here it is a classical copy of the RB system. Notice the breaking of "purification symmetry": while in the quantum case each of the R, A and B systems purifies the other two, here only A is left with a copy of both R and B.

The classical analogue of FQRS is the classical reverse Shannon theorem [11]

$$I(X;Y)[c\to c] + H(Y|X)[c\,c] \geq \langle X_A \to (XY)_A Y_B\rangle.$$

The classical analogue of FF is a feedback version of Shannon's channel coding theorem [12]:

$$\langle X_A \to (XY)_A Y_B\rangle \geq I(X;Y)[c\to c] + H(Y|X)[c\,c].$$

We observe immediately that the resource reversal duality #1 holds.

The classical analogue of FQSW is, not surprisingly, the original Slepian-Wolf theorem [13]:

$$\langle (XY)_S \to X_A Y_B \rangle + H(X|Y)[c \to c] \geq \langle (XY)_S \to (XY)_B \rangle.$$

The symmetry is now broken in a different way, with R containing a copy of AB, and there is no basis for dualities #2 and #3 to hold.

Conclusion. In summary, we have investigated three resource inequalities: FQRS, FF and FQSW. These are implemented by variations on protocols exhibited elsewhere, via the adding of feedback or placing restrictions on channel inputs. All three involve only *closed* quantum resources, meaning that there is no mixing with an unobserved environment, but rather non-trivial distribution of quantum information among the protagonists. With this simplification, a beautiful structure emerges (figure 3): FF and FQRS are related by the resource reversal duality #1; FF and FQSW are related by the source-channel duality #2; FQRS and FQSW are related by the time reversal duality #3. Along the way we provide insights into the difference between source and channel coding, the mother-father duality [2], and the breaking of "purification symmetry" in classical information theory.

Acknowledgments. We are grateful to T. Brun, P. Hayden, A. Harrow, A. Winter and J. Yard for useful comments on the manuscript.

References

[1] C. H. Bennett, H. J. Bernstein, S. Popescu, B. Schumacher, Phys. Rev. A **53**, 2046 (1996)

[2] I. Devetak, A. W. Harrow, A. Winter, Phys. Rev. Lett. **93**, 230504 (2004)

[3] I. Devetak, A. Winter, IEEE Trans. Inf. Theory **50**, 3183 (2004)

[4] H.-K. Lo, S. Popescu, Phys. Rev. Lett. **83**, 1459 (1999)

[5] A. W. Harrow, Phys. Rev. Lett. **92**, 097902 (2004)

[6] B. Schumacher, Phys. Rev. A **51**, 2738 (1995); R. Jozsa, B. Schumacher, J. Mod. Optics **41**, 2343 (1994)

[7] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, A. Winter "The quantum reverse Shannon theorem", in preparation.

[8] I. Devetak, P. Hayden, D. W. Leung, P. W. Shor, "Triple trade-offs in quantum Shannon theory", in preparation.

[9] M. Horodecki, J. Oppenheim, A. Winter, "Quantum information can be negative", submitted to Nature.

[10] A. Abeyesinghe, I. Devetak, P. Hayden, A. Winter, "Distributed source compression", in preparation.

[11] C. H. Bennett, P. W. Shor, J. A. Smolin, A. Thapliyal. IEEE Trans. Inf. Theory **48**, 2637 (2002)

[12] C. E. Shannon, Bell Sys. Tech. J. **27**, 379 (1948)

[13] D. Slepian, J. K. Wolf, IEEE Trans. Inf. Theory **19**, 471 (1973)

An Application of Entanglement to Leader Election in Anonymous Networks

Seiichiro Tani[12*]

[1] *NTT Communication Science Laboratories, NTT Corporation*
3-1, Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan
[2] *ERATO Quantum Computation and Information Project, JST*
Hongo White Building, 5-28-3 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Abstract. It is well-known that no classical algorithm can solve exactly (i.e., in bounded time without error) the leader election problem in anonymous networks. Recently, the author proved in collaboration with Kobayashi and Matsumoto that the problem can be exactly solved when the parties are connected by quantum communication links. A certain kind of entanglement plays a central role in the proposed algorithms. This paper focuses on applying entanglement to the leader election problem.

Keywords: leader election, quantum communication, entanglement, anonymous networks

1 Introduction

Quantum entanglement plays a central role in quantum information processing. It is well-known that, with prior entanglement, $n/2$-qubit communication is necessary and sufficient to transfer an n-bit classical message. In quantum distributed computing, which consists of quantum computation and communication, entanglement sometimes makes communication much more efficient than the classical equivalents.

The quantum communication complexity model first proposed by Yao [17] has been extensively studied areas in the field of quantum distributed computing. In the standard setting, two parties, which are connected via a bidirectional communication link, have to follow a communication protocol *or* distributed algorithm to compute values that depend in some predefined way on both of the inputs given to the two parties.

The goal is minimizing the number of communication (qu)bits or messages between parties, even though the local computation time required may be large (excellent surveys are found in [8, 5]). Some of the major results are as follows: there exists an exponential gap in communication cost [4, 11] when inputs are assumed to meet some predefined condition, while there is a case where the gap is quadratic [1, 12] when each of the outputs is the value of a Boolean function of the inputs. The communication complexity model can naturally be extended to the multiparty case, for which there are also several studies [6, 3, 7]. Since the communication complexity model is often used as just a mathematical tool to analyze the amount of resources required by other computation models such as Boolean circuits and the Turing machine, most applications of this model assume that the underlying network topology is a complete graph, i.e., every party is directly connected to each of the other parties. Another traditional research area of distributed computing (we call this research area "traditional distributed computing") involves the mathematical analysis of problems that arise in practical distributed computing environments, whose network topologies are extremely varied.

Recently the author, together with Kobayashi and Matsumoto, showed, by giving polynomial-time exact quantum algorithms for the leader election problem, that quantum distributed computing via entanglement is distinctly superior to the classical equivalent in the setting of traditional distributed computing [13, 14].

This paper reviews the algorithms in [13], and discusses a future research direction. The leader election problem is a core problem in traditional distributed computing, and has been studied for decades (see, e.g., [10]); there are many situations where parties have to decide which party should do some task in order to solve distributed computing problems. The goal of the leader election problem is to elect a unique leader from among distributed parties. Obviously, it is possible to deterministically elect a unique leader if each party has a unique identifier, and many classical deterministic algorithms with this assumption have been proposed. As the number of parties grows, however, it becomes difficult to preserve the uniqueness of the identifiers. Thus, other studies have examined the cases wherein the network is anonymous, i.e., each party has the same identifier [2, 9, 15, 16], as an extreme case. In this setting, no classical exact algorithm (i.e., an algorithm that runs in bounded time and solves the problem with zero error) exists for a broad class of network topologies including regular graphs, even if the network topology (and thus the number of parties) is known to each party prior to algorithm invocation [15]. Moreover, to the best of our knowledge, no zero-error probabilistic algorithm is known that works for *any* topology and runs in time/communication expected polynomial in the number of parties.

The key to solving the leader election problem in an anonymous network is to break symmetry, i.e., to have at least one pair of parties whose classical states are different from those of each other. Initially, all parties are eligible to become a unique leader. We can reduce the number of eligible parties by at least one each time eligible parties break symmetry: if some eligible party are in state 0 and others are in state 1, then only the latter parties are allowed to remain eligible. For the number n of parties, they can elect the unique leader after breaking symmetry $n-1$ times among eligible parties. We call a classical string obtained by concatenating (the binary expressions of) all parties' classical states among which at least two states are different from each other "inconsistent". Our contribution is to develop quantum protocols that prepare a superposition of inconsistent classical strings shared by the eligible parties in bounded time and without error. This implies that eligible parties can break symmetry by measuring such superposition, and that the unique leader can be elected in bounded time and without error by repeating symmetry breaking and

*tani@theory.brl.ntt.co.jp

eligible party reduction.

This paper considers the quantum model (i.e., every party can perform quantum computation and communication and each adjacent pair of parties has a bidirectional quantum communication link between them, but parties do not share any prior entanglement). Both of the two exact algorithms in [13] elect a unique leader from among n parties in polynomial time for *any* topology of synchronous and anonymous networks. We analyze the algorithms in terms of several complexity measures: time complexity (i.e., the maximum number of steps, including steps for the local computation, necessary for each party to execute the protocol, where the maximum is taken over all parties), communication complexity (i.e., the number of (quantum/classical) communication bits), and round complexity (i.e., the number of simultaneous message passing in synchronous networks). The two algorithms have their own characteristics in terms of these complexity measures. More strictly, our first algorithm runs in $O(n^3)$ time. The total communication complexity of this algorithm is $O(n^4)$, but this includes the quantum communication of $O(n^4)$ qubits. To reduce the quantum communication complexity, our second algorithm incurs $O(n^6(\log n)^2)$ time complexity, but demands the quantum communication of just $O(n^2 \log n)$ qubits (plus classical communication of $O(n^6(\log n)^2)$ bits). While our first algorithm needs $\Theta(n^2)$ rounds of quantum communication, our second algorithm needs only one round of quantum communication at the beginning of the protocol to share a sufficient amount of entanglement, and after the first round, the protocol performs only local quantum operations and classical communications (LOCCs) of $O(n \log n)$ rounds. Both algorithms are easily modified to support their use in asynchronous networks. Furthermore, both algorithms can be modified so that they work well even when each party initially knows only the upper bound of the number of parties. This implies that the exact number of parties can be computed when its upper bound is given. No classical zero-error algorithm exists in such cases for any topology that has a cycle as its subgraph [9].

2 Preliminaries

A *distributed system* (or *network*) is composed of multiple parties and bidirectional classical communication links that connect parties. In a quantum distributed system, every party can perform quantum computation and communication and each adjacent pair of parties has a bidirectional quantum communication link between them. Every party has *ports* corresponding one-to-one to communication links incident to the party. Every port of party l has a unique label i, $(1 \leq i \leq d_l)$, where d_l is the number of parties adjacent to l. For ease of explanation, we assume that port i corresponds to the link connected to the ith adjacent party of l. In our model, each party knows the number of its ports and the party can choose the appropriate port whenever it transmits or receives a message.

Initially, every party has local information, such as its internal state, and global information, such as the number of nodes in the system (or its upper bound). Every party runs the same algorithm, which has local and global information as its arguments. If all parties have the same local and global information except for the number of ports the parties have, the system is said to be *anonymous*. This is essentially equivalent to the situation in which every party has the same identifier since we can regard the local/global information of the party as his identifier. If message passing is performed synchronously, such a distributed system is called *synchronous*. The unit interval of synchronization is called a *round* (see [10] for more detailed descriptions).

Next we define the *leader election (LE) problem*. Suppose that there is a distributed system and each party in the system has a variable initialized to 0. The task is to set the variable of exactly one of the parties to 1 and the variables of all the other parties to 0. In the case of anonymous networks, Yamashita and Kameda [15] proved that, if the "symmetricity" (defined in [15]) of the network topology is more than one, LE cannot be solved exactly by any classical algorithm even if all parties know the topology of the network (and thus the number of nodes). In fact, "symmetricity" is more than one for a broad class of graphs such as regular graphs.

When the parties initially know only the upper bound of the number of the parties, the result by Itai and Rodeh [9] implies that LE cannot be solved with zero error by any classical algorithm (including the one that may not always halt).

3 Quantum Leader Election Algorithm I

For simplicity, we assume that the network is synchronous and that each party knows the number of parties, n, prior to the algorithm invocation. It is easy to generalize our algorithm to the asynchronous case and to the case where only the upper bound N of the number of parties is given [13].

First we introduce the concept of *consistent* and *inconsistent* strings. Suppose that each party l has a c-bit string x_l. That is, the n parties share cn-bit string $x = x_1 x_2 \cdots x_n$. For convenience, we may consider that each x_l expresses an integer, and identify string x_l with the integer it expresses. Given a set $E \subseteq \{1, \ldots, n\}$, string x is said to be *consistent* over E if x_l has the same value for all l in E. Otherwise x is said to be *inconsistent* over E. We also say that the cn-qubit pure state $|\psi\rangle = \sum_x \alpha_x |x\rangle$ shared by the n parties is *consistent (inconsistent)* over E if $\alpha_x \neq 0$ only for x's that are consistent (inconsistent) over E. Further, for positive integer m, we denote the state that is of the form of $(|0^m\rangle + |1^m\rangle)/\sqrt{2}$, by the m-cat state.

3.1 The Algorithm

The algorithm repeats one procedure exactly $(n-1)$ times, each of which is called a *phase*. Initially, all parties are eligible to become the unique leader. Formally, every party has a variable $\mathtt{status} \in \{\text{"eligible"}, \text{"ineligible"}\}$, which indicates whether the party is eligible or not. In each phase, the number of eligible parties either decreases or remains the same, but never increases or becomes zero. After $(n-1)$ phases the number of eligible parties becomes one with certainty.

Each phase has a parameter denoted by k, whose value is $(n-i+1)$ in phase i. In each phase i, let $E_i \subseteq \{1, \ldots, n\}$ be the set of all ls such that party l is still eligible. First, each eligible party prepares the state $(|0\rangle + |1\rangle)/\sqrt{2}$ in register $\mathbf{R}_0$, while each ineligible party prepares the state $|0\rangle$ in $\mathbf{R}_0$. Next every party calls Subroutine A, followed by partial measurement. This transforms the system state, i.e., the state in all parties' $\mathbf{R}_0$s into either $(|0^{|E_i|}\rangle + |1^{|E_i|}\rangle) \otimes |0^{n-|E_i|}\rangle/\sqrt{2}$ or a state that is inconsistent over E_i, where the first $|E_i|$ qubits

Table 1: The definition of commute operator "$\circ$"

x	y	$x \circ y$	x	y	$x \circ y$	x	y	$x \circ y$
0	0	0	1	1	1	$*$	$*$	$*$
0	1	$\times$	1	$*$	1	$*$	$\times$	$\times$
0	$*$	0	1	$\times$	$\times$			
0	$\times$	$\times$				$\times$	$\times$	$\times$

represent the qubits in eligible parties' $\mathbf{R}_0$s. In the former case, each eligible party calls Subroutine B, which uses a new ancilla qubit, initialized to $|0\rangle$, in register $\mathbf{R}_1$. If k equals $|E_i|$, Subroutine B always succeeds in transforming the $|E_i|$-cat state in eligible parties' $\mathbf{R}_0$s into a $2|E_i|$-qubit state that is inconsistent over E_i by using the $|E_i|$ ancilla qubits. Next, each eligible party l measures his qubits in $\mathbf{R}_0$ and $\mathbf{R}_1$ in the computational basis to obtain (a binary expression of) some two-bit integer z_l. Parties then compute the maximum value of z_l over all eligible parties l, by calling Subroutine C. Finally, parties with the maximum value remain eligible, while the other parties become ineligible. In what follows, we focus on only Subroutines A and B, which handle entanglement, while we omit the description of Subroutine C, which is just a classical algorithm.

Subroutine A

Subroutine A is essentially for the purpose of checking the consistency over E_i of each string that is superposed to a quantum state shared by parties. We use a commute operator "$\circ$" over a set $\{0, 1, *, \times\}$ whose operations are summarized in Table 1. Intuitively, "0" and "1" represent the possible values all eligible parties will have when the string finally turns out to be consistent; "$*$" represents "don't care," which means that the corresponding party has no information on the values any of the eligible parties have; and "$\times$" represents "inconsistent," which means that the corresponding party already knows that the string is inconsistent. The precise description of Subroutine A is given below. Subroutine A is called with $\mathbf{R}_0$, S, **status**, n, and d, where the content of S is initially "consistent," and d is the number of adjacent parties. Therefore, after every party finishes Subroutine A, the state shared by parties in their $\mathbf{R}_0$s is decomposed into a consistent state for which each party has the content "consistent" in his S, and an inconsistent state for which each party has the content "inconsistent" in his S.

Subroutine A

Input: one-qubit quantum registers $\mathbf{R}_0$ and S, a classical variable **status** $\in$ {"eligible", "ineligible"}, integers n, d

Output: one-qubit quantum registers $\mathbf{R}_0$ and S

1. Prepare two-qubit quantum registers $\mathbf{X}_0^{(1)}, \ldots, \mathbf{X}_d^{(1)}, \ldots, \mathbf{X}_0^{(n-1)}, \ldots, \mathbf{X}_d^{(n-1)}, \mathbf{X}_0^{(n)}$. If **status** = "eligible," copy the content of $\mathbf{R}_0$ to $\mathbf{X}_0^{(1)}$, otherwise set the content of $\mathbf{X}_0^{(1)}$ to "$*$."
2. For $t := 1$ to $n-1$, do the following:
   2.1 Copy the content of $\mathbf{X}_0^{(t)}$ to each of $\mathbf{X}_1^{(t)}, \ldots, \mathbf{X}_d^{(t)}$.
   2.2 Exchange the qubit in $\mathbf{X}_i^{(t)}$ with the party connected via port i for $1 \le i \le d$ (i.e., the original qubit in $\mathbf{X}_i^{(t)}$ is sent via port i, and the qubit received via that port is newly set in $\mathbf{X}_i^{(t)}$).
   2.3 Set the content of $\mathbf{X}_0^{(t+1)}$ to $x_0^{(t)} \circ x_1^{(t)} \circ \cdots \circ x_d^{(t)}$, where $x_i^{(t)}$ denotes the content of $\mathbf{X}_i^{(t)}$ for $0 \le i \le d$.
3. If the content of $\mathbf{X}_0^{(n)}$ is "$\times$," turn the content of S over (i.e., if initially the content of S is "consistent," it is flipped to "inconsistent," and vice versa).
4. Invert every computation and communication in Step 2.
5. Invert every computation in Step 1.
6. Output quantum registers $\mathbf{R}_0$ and S.

Subroutine B

Suppose that, among n parties, k parties are still eligible and share the k-cat state $(|0^k\rangle + |1^k\rangle)/\sqrt{2}$ in their $\mathbf{R}_0$'s. Subroutine B has the purpose of transforming the k-cat state to a superposition of inconsistent strings with certainty by using k fresh ancilla qubits that are initialized to $|0\rangle$, if k is given. The precise description of Subroutine B is given below, where $\{U_k\}$ and $\{V_k\}$ are two families of unitary operators,

$$U_k = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & e^{-i\frac{\pi}{k}} \\ -e^{i\frac{\pi}{k}} & 1 \end{pmatrix},$$

$$V_k = \frac{1}{\sqrt{R_k + 1}} \begin{pmatrix} 1/\sqrt{2} & 0 & \sqrt{R_k} & e^{i\frac{\pi}{k}}/\sqrt{2} \\ 1/\sqrt{2} & 0 & -\sqrt{R_k}e^{-i\frac{\pi}{k}} & e^{-i\frac{\pi}{k}}/\sqrt{2} \\ \sqrt{R_k} & 0 & \frac{e^{-i\frac{\pi}{2k}} I_k}{i\sqrt{2} R_{2k}} & -\sqrt{R_k} \\ 0 & \sqrt{R_k + 1} & 0 & 0 \end{pmatrix},$$

where R_k and I_k are the real and imaginary parts of $e^{i\frac{\pi}{k}}$, respectively.

Subroutine B

Input: one-qubit quantum registers $\mathbf{R}_0, \mathbf{R}_1$, an integer k

Output: one-qubit quantum registers $\mathbf{R}_0, \mathbf{R}_1$

1. If k is even, apply U_k to the qubit in $\mathbf{R}_0$; otherwise perform CNOT controlled by the qubit in $\mathbf{R}_0$ to that in $\mathbf{R}_1$, and then apply V_k to the qubits in $\mathbf{R}_0$ and $\mathbf{R}_1$.
2. Output quantum registers $\mathbf{R}_0$ and $\mathbf{R}_1$.

The point is that the amplitudes of the states $|00\rangle^{\otimes k}$, $|01\rangle^{\otimes k}$, $|10\rangle^{\otimes k}$, and $|11\rangle^{\otimes k}$ shared by k eligible parties in their registers $\mathbf{R}_0$ and $\mathbf{R}_1$ are simultaneously zero after each eligible party applies Subroutine B with parameter k, if the qubits in $\mathbf{R}_0$s of all eligible parties form the k-cat state.

4 Quantum Leader Election Algorithm II

4.1 The Algorithm

As in the previous section, we assume that the network is synchronous and each party knows the number n of parties prior to the algorithm. Again our algorithm is easily generalized to the asynchronous case. It is also possible to modify our algorithm so that it can work well even if only the upper bound N of the number of parties is given. This needs a bit more elaboration, but is not mentioned further here.

The algorithm consists of two stages, which we call Stages 1 and 2 hereafter. Stage 1 aims to have the n parties share a certain type of entanglement, and thus, this stage requires the parties to exchange quantum messages. In Stage 1, each party performs Subroutine Q $s = \lceil \log n \rceil$ times in parallel to share s pure quantum states $|\phi^{(1)}\rangle, \ldots, |\phi^{(s)}\rangle$ of n qubits. Here, each $|\phi^{(i)}\rangle$ is of the form $(|x^{(i)}\rangle + |\overline{x^{(i)}}\rangle)/\sqrt{2}$ for an n-bit string $x^{(i)}$ and its bitwise negation $\overline{x^{(i)}}$, and the lth qubit of each $|\phi^{(i)}\rangle$ is possessed by the lth party. It is stressed that only one round of quantum communication is necessary in Stage 1.

In Stage 2, the algorithm decides a unique leader among the n parties only by local quantum operations and classical communications with the help of the shared entanglement prepared in Stage 1. This stage consists of at most s phases, each of which reduces the number of eligible parties by at least half. In each phase i, let $E_i \subseteq \{1, \ldots, n\}$ be the set of all ls such that party l is still eligible. First every party runs Subroutine $\tilde{\mathrm{A}}$ to decide if state $|\phi^{(i)}\rangle$ is consistent or inconsistent over E_i. If state $|\phi^{(i)}\rangle$ is consistent, every party performs Subroutine $\tilde{\mathrm{B}}$, which first transforms $|\phi^{(i)}\rangle$ into the $|E_i|$-cat state $(|0^{|E_i|}\rangle + |1^{|E_i|}\rangle)/\sqrt{2}$ shared only by eligible parties and then calls Subroutine B described in the previous section to obtain an inconsistent state. Now each party l measures his qubits to obtain a label x_l and performs Subroutine $\tilde{\mathrm{C}}$ that works in the classical way to reduce the number of eligible parties by at least half via minority voting. In what follows, we describe the details of only Subroutine Q, which plays a central role in Stage 1.

Subroutine Q is mainly for the purpose of sharing a cat-like quantum state $|\phi\rangle = (|x\rangle + |\overline{x}\rangle)/\sqrt{2}$. It also outputs a classical string, which is used in Stage 2 for each party to obtain the information on $|\phi\rangle$ via classical communication. This subroutine can be performed in parallel, and thus Stage 1 involves only one round of quantum communication. First, each party prepares the state $\frac{|0\rangle+|1\rangle}{\sqrt{2}}$ in register $\mathbf{R}_0$ and computes the XOR of the contents of its own and each adjacent party's registers. The party then measures the qubits whose contents are the results of the XORs. After these operations, the qubits in $\mathbf{R}_0$ is in the state described as $(|x\rangle + |\overline{x}\rangle)/\sqrt{2}$. The precise description of Subroutine Q is given below. Subroutine Q also outputs classical value y, which is used to check if $|\phi^{(i)}\rangle$ is consistent or not in Subroutine $\tilde{\mathrm{A}}$.

Subroutine Q

Input: a one-qubit quantum register $\mathbf{R}_0$, an integer d

Output: a one-qubit quantum register $\mathbf{R}_0$, a binary string y of length d

1. Prepare $2d$ one-qubit quantum registers $\mathbf{R}'_1, \ldots, \mathbf{R}'_d$ and $\mathbf{S}_1, \ldots, \mathbf{S}_d$, each of which is initialized to the $|0\rangle$ state.
2. Create the $(d+1)$-cat state $(|0^{d+1}\rangle + |1^{d+1}\rangle)/\sqrt{2}$ in registers $\mathbf{R}_0, \mathbf{R}'_1, \ldots, \mathbf{R}'_d$.
3. Exchange the qubit in $\mathbf{R}'_i$ with the party connected via port i for $1 \leq i \leq d$ (i.e., the original qubit in $\mathbf{R}'_i$ is sent via port i, and the qubit received via that port is newly set in $\mathbf{R}'_i$).
4. Set the content of $\mathbf{S}_i$ to $x_0 \oplus x_i$, for $1 \leq i \leq d$, where x_0 and x_i denote the contents of $\mathbf{R}_0$ and $\mathbf{R}'_i$, respectively.
5. Measure the qubit in $\mathbf{S}_i$ in the $\{|0\rangle, |1\rangle\}$ basis to obtain a bit y_i, for $1 \leq i \leq d$.
   Set $y := y_1 \cdots y_d$.
6. Apply CNOT controlled by the content of $\mathbf{R}_0$ and targeting to the content of each $\mathbf{R}'_i$ for $i = 1, 2, \ldots, d$ to disentangle $\mathbf{R}'_i$s.
7. Output $\mathbf{R}_0$ and y.

---

5 Concluding Remarks

Our algorithms require a set of elementary unitary gates whose cardinality is linear in the number n of parties. It is open whether inconsistent states can be produced by using a fixed set of gates in an anonymous network.

References

[1] S. Aaronson and A. Ambainis. Quantum search of spatial regions. In *Proc. 44th IEEE FOCS*, pages 200–209, 2003.

[2] D. Angluin. Local and global properties in networks of processors (extended abstract). In *Proc. 20th ACM STOC*, pages 82–93, 1980.

[3] G. Brassard, A. Broadbent, and A. Tapp. Multi-party pseudo-telepathy. In *Proc. 8th WADS*, volume 2748 of *Lecture Notes in Computer Science*, pages 1–11, 2003.

[4] H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs. classical communication and computation. In *Proc. 30th ACM STOC*, 1998.

[5] H. Buhrman and H. Röhrig. Distributed quantum computing. In *Proc. 28th MFCS*, volume 2747 of *Lecture Notes in Computer Science*, pages 1–20. Springer, 2003.

[6] H. M. Buhrman, W. van Dam, P. Høyer, and A. Tapp. Multiparty quantum communication complexity. *Physical Review A*, 60(4):2737–2741, 1999.

[7] R. E. Cleve and H. M. Buhrman. Substituting quantum entanglement for communication. *Physical Review A*, 56(2):1201–1204, 1997.

[8] R. de Wolf. Quantum communication and complexity. *Theoretical Comput. Sci.*, 287(1):337–353, 2002.

[9] A. Itai and M. Rodeh. Symmetry breaking in distributed networks. *Inf. Comput.*, 88(1):60–87, 1990.

[10] N. A. Lynch. *Distributed Algorithms*. Morgan Kaufman Publishers, 1996.

[11] R. Raz. Exponential separation of quantum and classical communication complexity. In *Proc. 31st ACM STOC*, pages 358–367, 1999.

[12] A. Razborov. Quantum communication complexity of symmetric predicates. Technical Report quant-ph/0204025, http://arxiv.org/, 2002.

[13] S. Tani, H. Kobayashi, and K. Matsumoto. Exact quantum algorithms for the leader election problem. In *Proc. of 22nd STACS*, volume 3404 of *Lecture Notes in Computer Science*, pages 581–592, 2005.

[14] S. Tani, H. Kobayashi, and K. Matsumoto. Quantum leader election via exact amplitude amplification. In *Proceedings of ERATO conference on Quantum Information Science*, pages 11–12, 2005.

[15] M. Yamashita and T. Kameda. Computing on anonymous networks: Part I – characterizing the solvable cases. *IEEE Trans. Parallel Distrib. Syst.*, 7(1):69–89, 1996.

[16] M. Yamashita and T. Kameda. Computing on anonymous networks: Part II – decision and membership problems. *IEEE Trans. Parallel Distrib. Syst.*, 7(1):90–96, 1996.

[17] A. C.-C. Yao. Quantum circuit complexity. In *Proc 34th IEEE FOCS*, pages 352–361, 1993.

Quantum noise, universal uncertainty principle, and modal interpretation of quantum mechanics

Masanao Ozawa[1] *

[1] *Graduate School of Information Sciences, Tohoku University*
Aoba-ku, Sendai, 980- 8579, Japan

Abstract. In order to construct a rigorous and universal formulation of the uncertainty principle on noise and disturbance in quantum measurement, we reexamine the notions of quantum correlation, quantum measurement, quantum noise, and quantum disturbance in the light of the modal interpretation of quantum mechanics.

Keywords: quantum noise, quantum disturbance, uncertainty principle, modal interpretation

1 Introduction

The long standing mathematical formulation of the uncertainty principle, established by Heisenberg [1], Kennard [2], and Robertson [3], has not served to provide a reliable and general precision limit of measurements. In fact, it has been clarified through the controversy [4, 5] on the validity of the standard quantum limit for the gravitational wave detection [6, 7] that the purported reciprocal relation on noise and disturbance were found not generally true [8, 9]. However, the rapid development of the theory in the last two decades has made it possible to establish a universally valid uncertainty principle [10] for the most general class of quantum measurements, which will be useful for precision measurement and quantum information processing.

In Ref. [11] it was shown that the statistical properties of any physically possible quantum measurement is described by a normalized completely positive map valued measure (instrument), and conversely that any instrument arises in this way. Thus, we naturally conclude that measurements are represented by instruments, just as states are represented by density operators and observables by self-adjoint operators. Under this formulation, we have generalized the Heisenberg-type noise-disturbance relation to a relation that holds for any measurements, from which conditions have been obtained for measurements to satisfy the original Heisenberg-type relation [10]. In particular, every measurements with the noise and the disturbance statistically independent from the measured object is proven to satisfy the Heisenberg-type relation [12].

In this paper, we revisit the foudations of the the universal uncertainty principle. In particular, we show that the observable to be measured and the meter observable after the measuring interaction, though they are not necessarily commuting, are perfectly correlated and in the same maximal beable algebra of observables, so that the Copenhagen interpretation warrants that the meter observable after the measuring interaction possesses the same value as one that the observable to be measured has possessed just before the measuring interaction. The formal aspect of this result also clarifies the meaning of the root-mean-square noise and disturbance of meausurements that generally obey the universal uncertainty principle.

*ozawa@math.is.tohoku.ac.jp

2 Beable algebras and modal interpretation

Let $\mathcal{H}$ be a Hilbert space. Denote by $\mathcal{H}_1$ the unit sphere of $\mathcal{H}$ and by $\mathcal{L}(\mathcal{H})$ the algebra of bounded linear operators on $\mathcal{H}$. Let $\mathcal{B}$ be a unital C*-algebra acting on $\mathcal{H}$. A *dispersion-free state* on $\mathcal{B}$ is a state ω satisfying $\omega(A^*A) = |\omega(A)|^2$ for all $A \in \mathcal{B}$. Let ρ be a density operator on $\mathcal{H}$. We say that $\mathcal{B}$ is *beable* for ρ iff there is a probability measure μ on the space S of dispersion-free state of $\mathcal{B}$ such that

$$\mathrm{Tr}[A\rho] = \int_S \omega(A)\, d\mu(\omega) \tag{1}$$

for all $A \in \mathcal{B}$. Any observables in a common beable can be considered to possess simultaneous determinate values in the given state.

Let R be an observable (self-adjoint operator densely defined) on $\mathcal{H}$, and let ρ be a density operator on $\mathcal{H}$. Then, for any C*-algebra $\mathcal{B}$ on $\mathcal{H}$, we say that $\mathcal{B}$ is *R-beable* for ρ just in case:

(i) (*Beable*) $\mathcal{B}$ is beable for ρ.

(ii) (*R-Priv*) $R \in \mathcal{B}$.

(iii) (*Def*) For any unitary $U \in \mathcal{A}$, if $[U, R] = [U, \rho] = 0$, then $U\mathcal{B}U^* = \mathcal{B}$.

We say that $\mathcal{B}$ is *maximal R-beable* for ρ if and only if $\mathcal{B}$ is maximal with respect to the properties (Beable), (R-Priv), and (Def).

The *cyclic subspace* of $\mathcal{H}$ spanned by an observable R and a state vector $\psi \in \mathcal{H}_1$ is the closed subspace $\mathcal{C}(R, \psi)$ defined by

$$\mathcal{C}(R, \psi) = \text{the closure of } \{f(R)\psi \mid f \in B(\mathbf{R})\}, \tag{2}$$

where $B(\mathbf{R})$ stands for the space of bounded Borel functions on the real line $\mathbf{R}$. Denote by $\mathcal{C}_1(R, \psi)$ the unit sphere of $\mathcal{C}(R, \psi)$ and by $P_{R,\psi}$ the projection of $\mathcal{H}$ onto $\mathcal{C}(R, \psi)$.

The following characterization is given by Halvorson and Clifton [13], generalizing the Bub-Clifton uniqueness theorem [14, 15].

Theorem 2.1 *Let ψ be a vector state in $\mathcal{H}$. Then, the maximal R-beable algebra for ψ is uniquely determined as the form $\mathcal{L}(\mathcal{C}(R, \psi)^\perp) \oplus \{R\}'' P_{R,\psi}$.*

3 Quantum correlation

We say that two observables X and Y are *perfectly correlated* [16, 17] in a vector state ψ iff their spectral

measures E^X and E^Y satisfy

$$\langle E^X(\Delta)\psi, E^Y(\Gamma)\psi\rangle = 0 \tag{3}$$

for any disjoint Borel sets Δ and Γ. We say that two observables X and Y are *identically distributed* in a state ψ iff

$$\langle\psi|E^X(\Delta)|\psi\rangle = \langle\psi|E^Y(\Delta)|\psi\rangle \tag{4}$$

for all $\Delta \in \mathcal{B}(\mathbf{R})$. Then we obtain the following theorem.

Theorem 3.1 *For any two observables X and Y and any vector state ψ in $\mathcal{H}$, the following conditions are equivalent.*

(i) X and Y are perfectly correlated in ψ.

(ii) X and Y are perfectly correlated in all $\phi \in \mathcal{C}_1(X,\psi)$.

(iii) X and Y are identically distributed in all $\phi \in \mathcal{C}_1(X,\psi)$.

(iv) The maximal X-beable algebra and the maximal Y-beable algebra are the same, so that $\mathcal{C}(X,\psi) = \mathcal{C}(Y,\psi)$, and X and Y satisfy $X = Y$ on $\mathcal{C}(X,\psi)$.

4 Quantum instruments

Denote by $\tau c(\mathcal{H})$ the space of trace class operators on $\mathcal{H}$, by $\mathcal{L}(\tau c(\mathcal{H}))$ the space of bounded linear transformations on $\tau c(\mathcal{H})$, and by $\mathcal{S}(\mathcal{H})$ the space of density operators. A linear transformation $T \in \mathcal{L}(\tau c(\mathcal{H}))$ is called *completely positive* iff $T \otimes \mathrm{id}_n \in \mathcal{L}(\tau c(\mathcal{H} \otimes \mathbf{C}^n))$ is a positive transformation for any positive integer n. Denote by $\mathcal{CP}(\tau c(\mathcal{H}))$ the space of completely positive maps on $\tau c(\mathcal{H})$. An *instrument* is a countably additive normalized completely positive map valued measure from $\mathcal{B}(\mathbf{R})$ to $\mathcal{L}(\tau c(\mathcal{H}))$, i.e., a mapping $\mathcal{I} : \mathcal{B}(\mathbf{R}) \to \mathcal{CP}(\tau c(\mathcal{H}))$ satisfying that $\mathcal{I}(\mathbf{R})$ is trace-preserving and $\sum_{j=1}^{\infty} \mathcal{I}(\Delta_j) = \mathcal{I}(\mathbf{R})$ in the strong operator topology for any disjoint Borel sets $\Delta_1, \Delta_2, \ldots$ such that $\bigcup_j \Delta_j = \mathbf{R}$ [11].

The dual map of $\mathcal{I}(\Delta)$ is the linear transformation $\mathcal{I}(\Delta)^*$ on $\mathcal{L}(\mathcal{H})$ defined by

$$\mathrm{Tr}[(\mathcal{I}(\Delta)^* A)\rho] = \mathrm{Tr}[A\mathcal{I}(\Delta)\rho] \tag{5}$$

for all $A \in \mathcal{L}(\mathcal{H})$, $\rho \in \tau c(\mathcal{H})$, and $\Delta \in \mathcal{B}(\mathbf{R})$. The relation

$$\Pi^{\mathcal{I}}(\Delta) = \mathcal{I}(\Delta)^* I \tag{6}$$

where $\Delta \in \mathcal{B}(\mathbf{R})$ defines a POVM, called the *POVM* of $\mathcal{I}$, which satisfies

$$\mu_\rho^{\mathcal{I}}(\Delta) = \mathrm{Tr}[\Pi^{\mathcal{I}}(\Delta)\rho] \tag{7}$$

for all $\Delta \in \mathcal{B}(\mathbf{R})$ and $\rho \in \mathcal{S}(\mathcal{H})$.

5 Measuring processes

A *measuring process* for $\mathcal{H}$ is defined to be a quadruple $(\mathcal{K}, \xi, U, M)$ consisting of a separable Hilbert space $\mathcal{K}$, a state vector ξ in $\mathcal{K}$, a unitary operator U on $\mathcal{H} \otimes \mathcal{K}$, and an observable M on $\mathcal{K}$ [11]. For any measuring process $\mathcal{M} = (\mathcal{K}, \xi, U, M)$, the relation

$$\mathcal{I}_{\mathcal{M}}(\Delta)\rho = \mathrm{Tr}_{\mathcal{K}}[(I \otimes E^M(\Delta))U(\rho \otimes |\xi\rangle\langle\xi|)U^\dagger], \tag{8}$$

where $\rho \in \mathcal{S}(\mathcal{H})$ and $\Delta \in \mathcal{B}(\mathbf{R})$, defines an instrument $\mathcal{I}_{\mathcal{M}}$, called the *instrument* of $\mathcal{M}$. Conversely, it has been proved in Ref. [11] that for any instrument $\mathcal{I}$, there exists a measuring process $\mathcal{M} = (\mathcal{K}, \xi, U, M)$ such that $\mathcal{I} = \mathcal{I}_{\mathcal{M}}$.

6 Measurement operator formalism

A family $\mathbf{M} = \{M_m\}$ of operators with one real parameter m is called a family of *measurement operators* iff $\sum_m M_m^\dagger M_m = I$. The relation

$$\mathcal{I}_{\mathbf{M}}(\Delta)\rho = \sum_{m\in\Delta} M_m \rho M_m^\dagger \tag{9}$$

defines a CP instrument called the *instrument of* $\mathbf{M}$. The POM Π of the instrument $\mathcal{I}_{\mathbf{M}}$ is given by

$$\Pi(\Delta) = \sum_{m\in\Delta} M_m^\dagger M_m, \tag{10}$$

where $\Delta \subseteq \mathbf{R}$. The TPCP map T of the instrument $\mathcal{I}_{\mathbf{M}}$ and its dual map T^* are given by

$$T\rho = \sum_m M_m \rho M_m^\dagger, \tag{11}$$

so that $\{M_m\}$ is a family of Kraus operators of T.

7 Noise of POVMs

Let Π be a POVM on a Hilbert space $\mathcal{H}$. The *first and the second moment operators of* Π, denoted by $O(\Pi)$ and $O^{(2)}(\Pi)$, are defined by

$$O(\Pi) = \int x\, d\Pi(x), \tag{12}$$

$$O^{(2)}(\Pi) = \int_{\mathbf{R}} x^2\, d\Pi(x). \tag{13}$$

Let A and ψ be an observable and a vector state. We define the *root-mean-square (rms) noise* $\varepsilon(\Pi, A, \psi)$ of POVM Π for A in ψ by

$$\varepsilon(\Pi, A, \psi) = \langle\psi|O^{(2)}(\Pi) - O(\Pi)A - AO(\Pi) + A^2|\psi\rangle^{1/2}. \tag{14}$$

Then, we have

Theorem 7.1 *(i) If a POVM Π has a measuring process $(\mathcal{K}, \xi, U, M)$, then we have*

$$\varepsilon(\Pi, A, \psi) = \|U^\dagger(I \otimes M)U(\psi \otimes \xi) - (A \otimes I)(\psi \otimes \xi)\| \tag{15}$$

for all $\psi \in \mathcal{H}_1$.

(ii) If a POVM Π has measurement operators $\{M_m\}$, we have

$$\varepsilon(\Pi, A, \psi) = (\sum_m \|M_m(m - A)\psi\|^2)^{1/2} \tag{16}$$

for all $\psi \in \mathcal{H}_1$.

The rms noise can be statistically estimated from the experimental data. In fact, we have [12]

$$\begin{aligned} d_\psi(\Pi, A)^2 = {} & \langle\psi|A^2|\psi\rangle + \langle\psi|O^{(2)}(\Pi)|\psi\rangle \\ & + \langle\psi|O(\Pi)|\psi\rangle + \langle A\psi|O(\Pi)|A\psi\rangle \\ & - \langle(A+I)\psi|O(\Pi)|(A+I)\psi\rangle. \end{aligned} \tag{17}$$

In the above, $\langle\psi|A^2|\psi\rangle$ is the theoretical mean value of A^2 in state ψ, $\langle\psi|O^{(2)}(\Pi)|\psi\rangle$ is the mean of the squared output $\mathbf{x}^2$ in state ψ, and the other terms are the means of the output $\mathbf{x}$ in the respective input states. Thus, the error $\varepsilon(\Pi, A, \psi)$ can be statistically estimated, in principle, from experimental data of the measurements in states ψ, $A\psi/\|A\psi\|$, and $(A+I)\psi/\|(A+I)\psi\|$.

8 Precise measurements of observables

A POVM Π is said to *perfectly correlated* with an observable iff we have $\langle \Pi(\Delta)\psi, E^X(Ga)\psi\rangle = 0$. An instrument is said to satisfy the *Born statistical formula* (BSF) in a state ψ

$$\mathrm{Tr}[\mathcal{I}(\Delta)|\psi\rangle\langle\psi|] = \langle\psi|E^A(\Delta)|\psi\rangle, \tag{18}$$

where $\Delta \in \mathcal{B}(\mathbf{R})$. An instrument $\mathcal{I}$ with POVM Π is said to *precisely measure* an observable A on input state ψ iff Π and A are perfectly correlated in state The following theorem characterizes, up to statistical equivalence, the precise measurements of an observable in a given state.

Theorem 8.1 *Let $\mathcal{I}$ be an instrument. Let Π be the POVM of $\mathcal{I}$ and $\mathcal{M} = (\mathcal{K}, \xi, U, M)$ be a measuring process for $\mathcal{I}$. For any observable A and any state ψ, the following conditions are all equivalent.*

(i) $\mathcal{I}$ precisely measures A in ψ.

(ii) $A \otimes I$ is perfectly correlated with $U^\dagger(I \otimes M)U$ in ψ.

(iii) $\Pi^{\mathcal{I}}$ is perfectly correlated with A in all $\phi \in \mathcal{C}_1(A, \psi)$

(iv) $\mathcal{I}$ satisfies the BSF for A for all $\phi \in \mathcal{C}_1(A, \psi)$.

(v) $\varepsilon(\Pi, A, \psi) = 0$ for all $\phi \in \mathcal{C}_1(A, \psi)$.

9 Disturbance of TPCP maps

Let T be a TPCP map and let B be an observable. We denote by T^*E^B the POVM defined by

$$T^*E^B(\Delta) = T^*[E^B(\Delta)] \tag{19}$$

for all Borel set $\Delta \subseteq \mathbf{R}$ and by $T^*[f(B)]$ the operator defined by

$$T^*(f(B)) = \int f(\lambda) T^*[dE^B(\lambda)] \tag{20}$$

for any real-valued Borel function f.

The rms disturbance $\eta(B, T, \psi)$ of B caused by T in ψ is defined by

$$\eta(B, T, \psi) = \varepsilon(T^*E^B, B, \psi), \tag{21}$$

or equivalently

$$\begin{aligned}&\eta(B, T, \psi)^2 \\ &\quad = \langle\psi|T^*(B^2) - BT^*(B) - T^*(B)B + B^2|\psi\rangle. \end{aligned} \tag{22}$$

Then, we have

Theorem 9.1 *(i) If a TPCP map T has a measuring process $(\mathcal{K}, \xi, U, M)$, then we have*

$$\eta(T, B, \psi) = \|U^\dagger(B \otimes I)U(\psi \otimes \xi) - (A \otimes I)(\psi \otimes \xi)\| \tag{23}$$

for all $\psi \in \mathcal{H}_1$.

(ii) If a TPCP map T has Kraus operators $\{M_m\}$, we have

$$\eta(T, B, \psi) = (\sum_m \|[M_m, B]\psi\|^2)^{1/2}. \tag{24}$$

for all $\psi \in \mathcal{H}_1$.

In what follows, we abbreviate $\varepsilon(A) = \varepsilon(\Pi, A, \psi)$ and $\eta(B) = \eta(T, B, \psi)$, and $\langle\cdots\rangle = \langle\psi|\cdots|\psi\rangle$, where Π, T, and ψ are clearly identified in the context.

10 Projective measurements do not obey the Heisenberg-type noise-disturbance relation

An instrument $\mathcal{I}$ is said to be of *projective measurement* of a discrete observable A with spectral decomposition $A = \sum_m mE^A_m$ if

$$\mathcal{I}(\{m\})\rho = E^A_m \rho E^A_m. \tag{25}$$

Now we shall show the following

Theorem 10.1 *The disturbance of a bounded operator B caused by any TPCP map T is at most $2\|B\|$, i.e.,*

$$\eta(B) \leq 2\|B\|. \tag{26}$$

Theorem 10.2 *No instruments of projective measurement satisfy the Heisenberg-type noise-disturbance relation for (A, B), i.e.,*

$$\varepsilon(A)\eta(B) \geq \frac{1}{2}|\langle[A, B]\rangle|, \tag{27}$$

if B is bounded and $\langle[A, B]\rangle \neq 0$.

11 Universal uncertainty principle

We have argued that the Heisenberg-type noise-disturbance relation is often unreliable. Recently, the present author [10] proposed a new relation for noise and disturbance with a rigorous proof of the universal validity.

Theorem 11.1 (Universal Uncertainty Principle) *Any instrument with POVM Ψ and TPCP map T satisfies the relation*

$$\varepsilon(A)\eta(B) + \varepsilon(A)\sigma(B) + \sigma(A)\eta(B) \geq \frac{1}{2}|\langle[A, B]\rangle|. \tag{28}$$

for any A, B and state ψ, where $\langle\cdots\rangle = \langle\psi|\cdots|\psi\rangle$, and σ stands for the standard deviation in the state ψ.

12 When the Heisenberg-type noise-disturbance relation holds?

We introduce the *mean noise operator* and the *mean disturbance operator* of the measuring process $\mathcal{M} = (\mathcal{K}, \xi, U, M)$ by

$$n_A = \langle\xi|N_A|\xi\rangle_{\mathcal{K}}, \tag{29}$$

$$d_B = \langle\xi|D_B|\xi\rangle_{\mathcal{K}}, \tag{30}$$

where $\langle\xi|\cdots|\xi\rangle_{\mathcal{K}}$ stands for the partial mean on $\mathcal{K}$. The noise operator N_A is said to be *statistically independent of the object* $\mathbf{S}$ if n_A is scalar, and moreover the disturbance operator D_B is *statistically independent of the object system* $\mathbf{S}$ if d_B is scalar. Then, we have the following characterizations of measurements that obey the Heisenberg-type noise-disturbance relation.

Theorem 12.1 *(i) For any measuring process $\mathcal{M}$ and observables A, B, we have*

$$\varepsilon(A)\eta(B) + \frac{1}{2}|\langle[n_A, B]\rangle - \langle[d_B, A]\rangle| \geq \frac{1}{2}|\langle[A, B]\rangle|. \tag{31}$$

(ii) If the noise and disturbance are statistically independent of the object system, we have the Heisenberg-type noise-disturbance relation.

(iii)An instrument with measurement operators $\{M_m\}$ satisfies the Heisenberg-type noise-disturbance relation if we have

$$[\sum_m mM_m^\dagger M_m - A, B] = [\sum_m M_m^\dagger B M_m - B, A]. \quad (32)$$

13 Typical violations of the Heisenberg-type noise-disturbance relation

If the Heisenberg-type noise-disturbance relation were to hold for bounded observables A, B with $\langle [A,B] \rangle \neq 0$, we would have no precise measurements with $\varepsilon(A) = 0$ nor non-disturbing measurements with $\eta(B) = 0$. From the universal uncertainty principle, we have correct limitations on the noiseless or nondisturbing measurements [10].

The *uncertainty principle for non-disturbing measurements*, i.e., $\eta(B) = 0$, is given by

$$\varepsilon(A)\sigma(B) \geq \frac{1}{2}|\langle [A,B] \rangle|. \quad (33)$$

The *uncertainty principle for noiseless measurements*, i.e., $\varepsilon(A) = 0$, is given by

$$\sigma(A)\eta(B) \geq \frac{1}{2}|\langle [A,B] \rangle|. \quad (34)$$

From the above, we have the following statements.

Theorem 13.1 *For any instrument with measurement operators $\{M_m\}$, the relation Eq. (33) holds if $[M_m, B]\psi = 0$ for all m, and the relation Eq. (34) holds if $mM_m\psi = M_m A\psi$ for all m.*

14 Projective measurements of Pauli operators

In order to figure out the disturbance in projective measurements, let X, Y, Z be the Pauli operators on the 2 dimensional state space $\mathbf{C}^2$, and consider the projective measurement of Z. In this case, the measurement operators are given by $M_{-1} = (I - Z)/2$, $M_1 = (I + Z)/2$, and $M_m = 0$ if $m \neq \pm 1$. Let ψ be an arbitrary state vector. Then, from Eq. (16) we have

$$\varepsilon(Z) = 0. \quad (35)$$

On the other hand, we have

$$\eta(X)^2 = \sum_{m=\pm 1} \|[M_m, X]\psi\|^2 = 2\|Y\psi\|^2,$$

and since $\|Y\psi\| = 1$, we have

$$\eta(X) = \sqrt{2}. \quad (36)$$

We actually have $\eta(X) = \sqrt{2} \leq 2 = 2\|X\|$ as Eq. (26), and we have $\varepsilon(Z)\eta(X) = 0$. Thus, the Heisenberg-type noise-disturbance relation is violated in the state with $\langle [X,Z] \rangle \neq 0$. On the other hand, the universal uncertainty relation holds, as we have

$$\begin{aligned} &\varepsilon(Z)\eta(X) + \varepsilon(Z)\sigma(X) + \sigma(Z)\eta(X) \\ &\quad = \sigma(Z)\eta(X) = \sqrt{2}\sigma(Z) \geq \sigma(X)\sigma(Z) \\ &\quad \geq \frac{1}{2}|\langle [Z,X] \rangle|. \end{aligned}$$

Acknowledgments

This work was supported by the SCOPE Project of the MPHPT of Japan and by the Grant-in-Aid for Scientific Research of the JSPS.

References

[1] W. Heisenberg, Z. Phys. **43**, 172 (1927).

[2] E. H. Kennard, Z. Phys. **44**, 326 (1927).

[3] H. P. Robertson, Phys. Rev. **34**, 163 (1929).

[4] H. P. Yuen, Phys. Rev. Lett. **51**, 719 (1983), [see also *ibid.* p. 1603].

[5] C. M. Caves, Phys. Rev. Lett. **54**, 2465 (1985).

[6] V. B. Braginsky, Y. I. Vorontsov, and K. S. Thorne, Science **209**, 547 (1980).

[7] C. M. Caves *et al.*, Rev. Mod. Phys. **52**, 341 (1980).

[8] M. Ozawa, Phys. Rev. Lett. **60**, 385 (1988).

[9] M. Ozawa, in *Squeezed and Nonclassical Light*, edited by P. Tombesi and E. R. Pike (Plenum, New York, 1989), pp. 263 286.

[10] M. Ozawa, Phys. Rev. A **67**, 042105 (2003).

[11] M. Ozawa, J. Math. Phys. **25**, 79 (1984).

[12] M. Ozawa, Ann. Phys. (N.Y.) **311**, 350 (2004).

[13] H. Halvorson and R. Clifton, Int. J. Theor. Phys. **38**, 2441 (1999).

[14] J. Bub and R. Clifton, Stud. Hist. Phil. Mod. Phys. **27**, 181 (1996).

[15] J. Bub, *Interpreting the Quantum World* (Cambridge University Press, Cambridge, 1997).

[16] M. Ozawa, Phys. Lett. A **335**, 11 (2005).

[17] M. Ozawa, Ann. Phys. (N.Y.) to appear (2005), online preprint: LANL quant-ph/0501081.

Unconditional security of QKD and the uncertainty principle

Masato Koashi[1 2 *]

[1] *Division of Materials Physics, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan.*

[2] *CREST Photonic Quantum Information Project, 4-1-8 Honmachi, Kawaguchi, Saitama 331-0012, Japan.*

Abstract. An approach to the unconditional security of quantum key distribution protocols is presented, which is based on the uncertainty principle. The approach applies to every case that has been treated via the argument by Shor and Preskill, but it is not necessary to find quantum error correcting codes. It can also treat the cases with uncharacterized apparatuses. The proof can be applied to cases where the secret key rate is larger than the distillable entanglement.

Keywords: Quantum key distribution, Uncertainty principle, Unconditional security

1 Introduction

One of the aims of the cryptography is to allow two legitimate parties, Alice and Bob, to exchange messages secretly without leak to a third party, Eve, who tries to eavesdrop. It is well known that once Alice and Bob share a secret key, which is a common random bit sequence unknown to Eve, they can communicate a secret message of the same length as the key. The task of quantum key distribution (QKD) is a way to produce or to amplify the secret key using the properties of quantum mechanics. For any protocol of QKD, it is vital to have a proof of the unconditional security because the robustness against any kind of attack allowed by the law of physics is the main advantage of QKD over classical schemes aiming at the same task. One of the well-known strategies for the security proof is the argument [1] given by Shor and Preskill, in which a reduction to an entanglement distillation protocol (EDP) based on Calderbank-Shor-Steane (CSS) quantum error correcting codes (QECC) [2, 3] is used to show that the information leak on the final key is negligible. This approach has turned out to be quite versatile due to the simplicity of the idea: for example, the original proof for the BB84 protocol [4] has been extended [5, 6] to cover the B92 protocol [7]. On the other hand, invoking the CSS-QECC in the proof requires the actual users to find a quantum code satisfying a certain property, which is not always an easy task. Even the innocent-looking formula [(1) below] for the asymptotic key gain needs a complicated argument [8] for strict derivation. Decoupling of the error correction and the privacy amplification can be made by encrypting the former [9], but only when it satisfies a constraint coming from the CSS-QECC.

If we look back to the first proof [10] of unconditional security by Mayers, we notice that it also has its own merits. One disadvantage, the complexity of the proof, was recently remedied by a simple proof [11] by Koashi and Preskill based on the same spirit, namely, reduction to a two-party protocol by omitting one of the legitimate users by a symmetry argument. In this line of approach, the error correction and the privacy amplification are decoupled from the start, and we can just use any conventional scheme for the error correction. The proof also shows a peculiar and useful property, which allows the use of basis-independent uncharacterized sources or detectors. For example, if we use an ideal detector, the source can be anything as long as it does not reveal which basis is used in the BB84 protocol. We can still use the same formula for the key rate, indicating that any fault in the source can be automatically caught in the form of an increase in the observed bit errors. Unfortunately, the argument of omitting one party relies heavily on the symmetry of the BB84 protocol, and it cannot be applied to the protocols with no such symmetry.

Here we present an approach to the unconditional security based on uncertainty principle. This argument has the same advantages in the Mayers-Koashi-Preskill argument, while retaining the versatility of the Shor-Preskill argument. In fact, in any protocol having a proof that relies on the Shor-Preskill argument, we can decouple the error correction and the privacy amplification just by encrypting the former, thereby relieve it from the constraint of CSS-QECC. The new approach allows us to solve security problems with imperfect devices that were beyond either of the previous arguments. For example, we can derive a key rate formula for the BB84 protocol with an arbitrary source, the properties of which are unknown except for a bound on the fidelity between the averaged states for two bases [12]. Our proof also provides an insight into the recently predicted phenomenon of secure key from bound entanglement [13].

2 Basic ideas in the security proof

Most of the QKD protocols can be equivalently described by an entanglement-based protocol, in which Alice and Bob share a pair of quantum systems $\mathcal{H}_A \otimes \mathcal{H}_B$ after discarding other systems used for random sampling tests. The state ρ_0 of $\mathcal{H}_A \otimes \mathcal{H}_B$ at this point is not fixed and may be highly correlated among subsystems due to Eve's intervention, but the results of the tests may give a set of promises on the possible state. For example, in the case of Shor-Preskill proof, $\mathcal{H}_A \otimes \mathcal{H}_B$ is composed of N pairs of shared qubits, and there is a promise that the following statements hold except for an exponentially small probability: Suppose that each qubit is measured on z or x basis. Then the number $n_{\rm bit}$ of qubits showing the bit error ($\sigma_z \otimes \sigma_z = -1$) satisfies $n_{\rm bit}/N \leq \delta_{\rm bit}$, and the

*koashi@mp.es.osaka-u.ac.jp

number $n_{\rm ph}$ with the phase error ($\sigma_x \otimes \sigma_x = -1$) satisfies $n_{\rm ph}/N \leq \delta_{\rm ph}$. Here $\delta_{\rm bit}$ and $\delta_{\rm ph}$ are determined from the results of the test. Here we consider more general cases, in which the size of $\mathcal{H}_A \otimes \mathcal{H}_B$ is arbitrary. We give a proof for the unconditional security of the protocols having the following form:

Actual Protocol — Alice and Bob make measurements on $\mathcal{H}_A$ and on $\mathcal{H}_B$, respectively. Through an encrypted classical communication consuming r bits of secret key, they agree on an N-bit reconciled key $\boldsymbol{\kappa}_{\rm rec}$, except for a negligible failure probability. In the binary vector space on N bits, one party chooses a linearly-independent set $\{\boldsymbol{V}_k\}_{k=1,\ldots,N-m}$ of N-bit sequences randomly and announce it. The k-th bit of the final key $\boldsymbol{\kappa}_{\rm fin}$ is defined as scalar product $\boldsymbol{\kappa}_{\rm rec} \cdot \boldsymbol{V}_k$.

This protocol newly produces $N-m$ bits of secret key, and the net secret key gain is $G = N - r - m$ bits. We first give an overview of our security proof, taking the Shor-Preskill (SP) case as an example. The core of our approach is to regard $\boldsymbol{\kappa}_{\rm rec}$ as the outcome of z-basis measurements on N virtual qubits $\mathcal{K}^{\otimes N}$. In the SP case, we may just identify $\mathcal{H}_B$ with $\mathcal{K}^{\otimes N}$. Next, we ask how we could have predicted the N-bit outcome $\boldsymbol{X}$ if the N qubits had been measured in the x-basis. In the SP case, we could have measured $\mathcal{H}_A$ on the x-basis to obtain an N-bit outcome μ. The random sampling tests assure that this outcome coincides with $\boldsymbol{X}$ within $\sim N\delta_{\rm ph}$-bit errors, namely, the conditional entropy is bounded as $H(\boldsymbol{X}|\mu) \leq N\xi$ with $\xi \sim h(\delta_{\rm ph})$, where $h(y) \equiv -y\log y - (1-y)\log(1-y)$. Then, the uncertainty of the complementary observable, namely, the z-basis outcome $\boldsymbol{\kappa}_{\rm rec}$, should satisfy $H(\boldsymbol{\kappa}_{\rm rec}) \geq N - N\xi$ according to the entropic uncertainty relation [14]. Hence, it is not surprising that Eve has negligible information on the final key $\boldsymbol{\kappa}_{\rm fin}$ when $m = N[h(\delta_{\rm ph}) + \epsilon]$. Since the error correction consumes $r = N[h(\delta_{\rm bit}) + \epsilon]$ bits of secret key, we arrive at the familiar asymptotic net key gain

$$G = N[1 - h(\delta_{\rm bit}) - h(\delta_{\rm ph})]. \tag{1}$$

3 Main theorem

The rough sketch of the proof in the previous section can be made strict and generalized as follows. First we choose a quantum operation Λ that converts state ρ on $\mathcal{H}_A \otimes \mathcal{H}_B$ to state $\Lambda(\rho)$ on $\mathcal{H}_R \otimes \mathcal{K}^{\otimes N}$, where $\mathcal{H}_R$ stands for an ancillary system R. We further consider a measurement M_R on $\mathcal{H}_R$, and define μ to be its outcome. As we have seen, in the SP case we may choose Λ to be a trivial operation Λ_0 that just changes the definition as $\mathcal{H}_A \cong \mathcal{H}_R$ and $\mathcal{H}_B \cong \mathcal{K}^{\otimes N}$, and take M_R to be the x-basis measurement. But the security proof here allows almost free choices of Λ and M_R, except for the following requirement:

Assumption 1 — The application of Λ followed by the standard z-basis measurements on $\mathcal{K}^{\otimes N}$ is equivalent to the measurement of $\boldsymbol{\kappa}_{\rm rec}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$ in Actual Protocol.

Note that within the constraint of Assumption 1, it is even allowed to take Λ involving collective operations over $\mathcal{H}_A$ and $\mathcal{H}_B$.

Let $\boldsymbol{X}$ be the outcome of x-basis measurements on $\mathcal{K}^{\otimes N}$. The next step is to rephrase the condition $H(\boldsymbol{X}|\mu) \leq N\xi$ in the rough sketch in a more rigorous and flexible form:

Assumption 2 — There exists a set T_μ of N-bit sequences with cardinality $|T_\mu| \leq 2^{N\xi}$ for each μ, such that the pair of measurement outcomes $(\mu, \boldsymbol{X})$ satisfies $\boldsymbol{X} \in T_\mu$ except for an exponentially small probability η.

Now we can state the main theorem about the security:

Theorem 1 *If Assumptions 1 and 2 hold for $m = N(\xi + \epsilon)$ with $\epsilon > 0$, Eve's information on $\boldsymbol{\kappa}_{\rm fin}$ in Actual Protocol is at most $h(\eta') + N\eta'$ with $\eta' = \eta + 2^{-N\epsilon}$.*

This theorem can be used as follows. First we choose Λ and M_R under Assumption 1. Next, combined with the promises obtained from the random sampling tests, we obtain a value of ξ with which Assumption 2 holds. Then, Theorem assures that the unconditionally secure key gain of at least $G = N - r - N(\xi + \epsilon)$ is achievable. For a good key gain, Λ and M_R should be chosen such that ξ is as large as possible.

4 Proof of the main theorem

Thanks to Assumption 1, Eve's knowledge on $\boldsymbol{\kappa}_{\rm fin}$ in Actual Protocol is the same as that on $\boldsymbol{\kappa}_{\rm fin}$ obtained from $\mathcal{H}_A \otimes \mathcal{H}_B$ by the following procedure.

Protocol 1 — Apply Λ and discard $\mathcal{H}_R$. For the N qubits $\mathcal{K}^{\otimes N}$, measure each qubit on z-basis to determine the N-bit key $\boldsymbol{\kappa}_{\rm rec}$. Choose a linearly-independent set $\{\boldsymbol{V}_k\}_{k=1,\ldots N-m}$ randomly, and announce it to Eve. Let $\boldsymbol{\kappa}_{\rm rec} \cdot \boldsymbol{V}_k$ be the k-th bit of the final key $\boldsymbol{\kappa}_{\rm fin}$.

In order to show that Eve has negligible information on $\boldsymbol{\kappa}_{\rm fin}$, we consider yet another protocol, which is later shown to be equivalent to Protocol 1. Define operator $\Sigma_\nu(\boldsymbol{W}) \equiv \sigma_\nu^{b_1}\sigma_\nu^{b_2}\cdots\sigma_\nu^{b_N} (\nu = x, z)$ acting on $\mathcal{K}^{\otimes N}$ for N-bit sequence $\boldsymbol{W} = [b_1 b_2 \cdots b_N]$. The new protocol is defined as follows:

Protocol 2 — (a) Apply Λ and make measurement M_R on $\mathcal{H}_R$ to obtain outcome μ. (b) Choose N-bit sequences $\boldsymbol{W}_j (j = 1, \ldots m)$ randomly, and take an arbitrary linearly-independent set $\{\boldsymbol{V}_k\}_{k=1,\ldots N-m}$ of N-bit sequences satisfying $\boldsymbol{V}_k \cdot \boldsymbol{W}_j = 0$ for any j, k. Announce $\{\boldsymbol{V}_k\}$ to Eve. (c) Measure m observables $\{\Sigma_x(\boldsymbol{W}_j)\}$ to determine an N-bit sequence $\boldsymbol{X}^*$ as we will explain later. (d) Apply unitary operation $\Sigma_z(\boldsymbol{X}^*)$. (e) Measure $\{\Sigma_z(\boldsymbol{V}_k)\}$ to determine the $(N-m)$-bit final key $\boldsymbol{\kappa}_{\rm fin}$.

If we measured $\mathcal{K}^{\otimes N}$ on the x-basis before step (c), the outcome $\boldsymbol{X}$ would be one of $2^{N\xi}$ candidates T_μ except for probability η (Assumption 2). Each measurement of $\Sigma_x(\boldsymbol{W}_j)$ in step (c) gives a random parity bit $\boldsymbol{X} \cdot \boldsymbol{W}_j$, which halves the number of candidates. Hence, as in the

hushing method of EDP [15], by knowing $m = N(\xi + \epsilon)$ random parity bits we can derive an estimate $\boldsymbol{X}^*$ of $\boldsymbol{X}$ with an exponentially small failure probability $Pr(\boldsymbol{X}^* \neq \boldsymbol{X}) \leq \eta' \equiv \eta + 2^{-N\epsilon}$. Then, if we measured $\mathcal{K}^{\otimes N}$ on the x-basis after the phase flip in step (d), the outcome would be $\boldsymbol{X}^* - \boldsymbol{X}$, which is 0 except for probability η'. This implies that the state σ of the qubits after step (d) is a nearly-pure state satisfying $\langle 0_x^{\otimes N}|\sigma|0_x^{\otimes N}\rangle \geq 1 - \eta'$, where $|0_x^{\otimes N}\rangle$ is the x-basis eigenstate for $\boldsymbol{X} = \boldsymbol{0}$. Since the measurement in step (e) is applied on this nearly-pure state, Eve has only negligible (at most $S(\sigma) \leq [h(\eta') + N\eta']$-bit) information about $\boldsymbol{\kappa}_{\rm fin}$.

The equivalence of the two protocols are easy to be seen. In Protocol 2, the operators $\{\Sigma_z(\boldsymbol{V}_k)\}$ commute with $\Sigma_z(\boldsymbol{X}^*)$ and with $\Sigma_x(\boldsymbol{W}_j)$ since $\boldsymbol{V}_k \cdot \boldsymbol{W}_j = 0$. Hence we can omit steps (c) and (d) and still obtain the same final key. We further notice that M_R is now redundant, and the choosing method of $\{\boldsymbol{V}_k\}$ can be simplified to a random selection. Noting that $\{\Sigma_z(\boldsymbol{V}_k)\}$ can be also obtained through a z-basis measurement on each qubit, we are lead to Protocol 1. This completes the proof.

5 Discussion

We have described a method of proving the unconditional security which unifies two major previous approaches and retains the advantages in both of them. The proof relies on the observation that Alice can guess the z-basis outcomes of virtual N qubits with r-bit uncertainty in the actual protocol, and Alice and Bob can guess the x-basis outcomes with m-bit uncertainty in a equivalent protocol. The "excess" over the uncertainty limit, $N - r - m$, amounts to the key gain. Note that if they share a maximally entangled state (MES), Alice alone can guess for both of the bases. The condition for the secrecy is weaker than that since it allows her to collaborate with Bob nonlocally for the x basis, through any operation Λ satisfying Assumption 1. This difference is considered to be a reason for the gap between distillable entanglement and secret key gain [13]. In fact, examples in [13] are constructed by applying a nonlocal "twisting" operation to $\rho_{AB} \otimes \rho_{A'B'}$, where ρ_{AB} is an MES. Their twisting operations do not change the outcome of z-basis measurement on $\mathcal{H}_B$, which can be regarded as $\boldsymbol{\kappa}_{\rm rec}$. Hence, we can define Λ to be the reverse of the twisting followed by Λ_0, which satisfies Assumption 1. This shows that the present method potentially gives a key rate exceeding the amount of distillable entanglement.

Acknowledgment

The author thanks N. Imoto, J. Preskill and H. -K. Lo for helpful discussions. This work was supported by a MEXT Grant-in-Aid for Young Scientists (B) 17740265.

References

[1] P. W. Shor and J. Preskill. Simple proof of security of the BB84 quantum key distribution protocol. *Phys. Rev. Lett.*, 85:441, 2000.

[2] A. R. Calderbank and P. W. Shor. Good quantum error-correcting codes exist. *Phys. Rev. A*, 54:1098, 1996.

[3] A. M. Steane. Multiple-particle interference and quantum error correction. *Proc. R. Soc. Lond. A*, 452:2551, 1996.

[4] C. H. Bennett and G. Brassard. Quantum cryptography: public key distribution and coin tossing. In *Proceeding of the IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India*, pages 175–179. IEEE, New York, 1984.

[5] K. Tamaki, M. Koashi, and N. Imoto. Unconditionally secure key distribution based on two nonorthogonal states. *Phys. Rev. Lett.*, 90:167904, 2003.

[6] M. Koashi. Unconditional security of coherent-state quantum key distribution with a strong phase-reference pulse. *Phys. Rev. Lett.*, 93:120501, 2004.

[7] C. H. Bennett. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett*, 68:3121, 1992.

[8] M. Hamada. Reliability of Calderbank-Shor-Steane codes and security of quantum key distribution, quant-ph/0308039.

[9] H. K. Lo. Method for decoupling error correction from privacy amplification. *New J. Phys.*, 5:36, 2003.

[10] D. Mayers. Quantum key distribution and string oblivious transfer in noisy channels. *Lect. Notes Comput. Sci.*, 1109:343, 1996.

[11] M. Koashi and J. Preskill. Secure quantum key distribution with an uncharacterized source. *Phys. Rev. Lett.*, 90:057902, 2003.

[12] M. Koashi. Simple security proof of quantum key distribution via uncertainty principle, quant-ph/0505108.

[13] K. Horodecki, M. Horodecki, P. Horodecki, and J. Oppenheim. Secure key from bound entanglement. *Phys. Rev. Lett.*, 94:160502, 2005.

[14] H. Maassen and J. B. M. Uffink. Generalized entropic uncertainty relations. *Phys. Rev. Lett.*, 60:1103, 1988.

[15] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters. Mixed-state entanglement and quantum error correction. *Phys. Rev. A*, 54:3824, 1996.

Security proof of the BB84 protocol in practical implementation

Yodai Watanabe[1]

[1]*National Institute of Informatics, Research Organization of Information and Systems*
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 1018430, Japan

Abstract. This paper provides a security proof of the Bennett-Brassard (BB84) quantum key distribution protocol for arbitrary source and detector. The only assumption in the security proof is that the source is characterized; that is, the proof requires no restrictions on the source and detector, and, moreover, the detector can be uncharacterized. The proof is performed by lower-bounding adversary's Rényi entropy about the key before privacy amplification. The bound reveals the leading factor which reduces the key generation rate.

Keywords: Practical quantum key distribution, Unconditional security

1 Introduction

One of the fundamental problems in cryptography is to provide a way of sharing a secret random number between two parties, Alice and Bob, in the presence of an adversary Eve. A quantum key distribution scheme is a solution to this problem[1, 2]; indeed it allows Alice and Bob to generate a shared secret key securely against Eve with unbounded resources of computation. The security of quantum key distribution against general attacks was first proved by Mayers[10]. After that, Shor-Preskill[11] provided a simple security proof based on the observation that a quantum key distribution (BB84 protocol) is closely related to quantum error-correcting codes (CSS codes). Gottesmann et al.[6] showed that the Shor-Preskill proof is still valid as long as the source and detector are perfect enough so that all defects can almost be absorbed into Eve's attack (see also [7, 13] for the achievable rate of quantum codes in the security proof). In contrast to the security proof based on quantum codes, the Mayers proof has a remarkable characteristic. Namely, in the Mayers proof, although the source has to be (almost) perfect, there is no restriction on the detector; in particular, it can be uncharacterized. By exchanging the role of the source and detector in the Mayers proof, Koashi-Preskill[9] provided a security proof which applies to the case where the detector is perfect, but the source can be uncharacterized (except that the averaged states are independent of Alice's basis). The aim of this work is to extend these security proofs in ideal settings to the one in the real setting where both the source and detector can be arbitrary. We provide a security proof of the BB84 protocol in which the only assumption is that the source is characterized. In the same way as Koashi-Preskill[9], this can be transformed into a security proof which is based on the characteristics of the detector. Further we note that the security proof also applies to other quantum key distribution protocols such as the B92 protocol[1] and the DPSQKD (Differential phase shift quantum key distribution) protocol[12].

2 Preliminaries

Let us first recall the BB84 protocol[2]. Let $\mathcal{H}$ be a Hilbert space, and for a Hilbert space $\mathcal{H}$, $I_{\mathcal{H}}$ denotes the identity on $\mathcal{H}$. Let $\mathcal{A} = \{1, \cdots, N\}$, and for $\mathcal{B} \subset \mathcal{A}$, denote the cardinality of $\mathcal{B}$ by $n_{\mathcal{B}}$. The BB84 protocol is described as follows: (i) Alice generates two binary strings $a^{\mathcal{A}} = \{a_i\}_{i\in\mathcal{A}}$ and $x^{\mathcal{A}} = \{x_i\}_{i\in\mathcal{A}}$ according to the probability distribution $p(a^{\mathcal{A}}, x^{\mathcal{A}}) = \prod_i p(a_i)p(x_i)$. (ii) Bob generates a binary string $b^{\mathcal{A}} = \{b_i\}_{i\in\mathcal{A}}$ according to the probability distribution $p(b^{\mathcal{A}})$. (iii) Alice sends the quantum state $\rho^{\mathcal{A}}_{a,x} = \bigotimes_{i\in\mathcal{A}} \rho_{a_i,x_i}$ on $\mathcal{H}^{\otimes N}$ to Bob. (iv) Bob applies the measurement $\{E^{\mathcal{A}}_{b,y}\}_{y^{\mathcal{A}}} = \left\{\bigotimes_{i\in\mathcal{A}} E_{b_i,y_i}\right\}_{y^{\mathcal{A}}\in\{0,1,\phi\}^N}$ on $\mathcal{H}^{\otimes N}$ to the received quantum state, where $E_{0,\phi} = E_{1,\phi}$ is the measurement corresponding to the result that Bob cannot detect a state. (v) Alice and Bob respectively open $a^{\mathcal{A}}$ and $b^{\mathcal{A}}$. Let $\mathcal{D} = \{i \in \mathcal{A} | y_i \neq \phi\}$ and $\mathcal{C} = \{i \in \mathcal{D} | a_i = b_i\}$. Alice and Bob select a random subset $\mathcal{T} \subset \mathcal{C}$ according to a binomial distribution $p_{\mathcal{C}}(\mathcal{T}) = B(|\mathcal{T}|; n_{\mathcal{C}}, p_{\mathcal{T}})$. Let $\mathcal{K} = \mathcal{C} - \mathcal{T}$. (vi) Alice and Bob compare $x^{\mathcal{T}}$ and $y^{\mathcal{T}}$, and count the number of errors, $n^{\mathrm{e}}_{\mathcal{T}} = |\{i \in \mathcal{T} | x_i \neq y_i\}|$. (vii) Bob estimates $x^{\mathcal{K}}$ by exchanging error-correction information with Alice. (viii) Alice and Bob generate a secret key s by applying a compression function to $x^{\mathcal{K}}$.

We next provide basic definitions which will be used later (see e.g., [8] for details). Let p and q be probability distributions. The relative entropy between p and q is given by $D(p\|q) = \sum_\omega p(\omega)(\log p(\omega) - \log q(\omega))$. The variation distance between p and q is given by $d_V(p, q) = \sum_\omega \{p(\omega) - q(\omega) > 0\}^*$, where, for a random variable X, $\{X(\omega) > 0\}^*$ takes the value $X(\omega)$ if $X(\omega) > 0$ and 0 otherwise, i.e. $\{X > 0\}^* = X\{X > 0\}$. The quantum analogue of the variation distance is called the trace distance. Let ρ and σ be quantum states. For an Hermitian operator X with the spectral decomposition $X = \sum_i x_i E_i$, define the projection $\{X > 0\}$ by $\{X > 0\} = \sum_{i:x_i>0} E_i$. The trace distance between ρ and σ is then given by $d_T(\rho, \sigma) = \mathrm{Tr}(\rho - \sigma)\{\rho - \sigma > 0\}$.

3 Security of the BB84 protocol

To prove the security of the BB84 protocol, previous works[6, 9, 10, 11] assume that either Alice's source or Bob's detector is sufficiently perfect in the sense that all defects in the device can almost be absorbed into Eve's attack. Note that the previous security proofs have been based on directly bounding Eve's mutual information about the final key, i.e. the key after privacy amplification. In the present work, we first lower-bound Eve's

Rényi entropy about the key before privacy amplification, and then apply privacy amplification given in the classical information theory, which makes use of a compression function chosen at random from a universal hash family (see [3] for the classical theory of privacy amplification).

We begin with introducing a simpler protocol whose security implies that of the original protocol. Let $\mathcal{X} = \{(0,0),(0,1),(1,0),(1,1)\}$, and $\alpha, \beta \in \mathcal{X}$. We construct a set of pure states, $\{\hat{\rho}_\alpha\}_{\alpha\in\mathcal{X}}$, such that a physical transformation from $\{\hat{\rho}_\alpha\}_{\alpha\in\mathcal{X}}$ to $\{\rho_\alpha\}_{\alpha\in\mathcal{X}}$ exists. Let us first write ρ_α in the form of a decomposition

$$\rho_\alpha = \int \lambda_\alpha(\xi)|\psi_\alpha(\xi)\rangle\langle\psi_\alpha(\xi)|d\xi. \tag{1}$$

Then we can define the Gram matrix G by

$$[G]_{\alpha\beta} = \int (\lambda_\alpha\lambda_\beta)^{\frac{1}{2}}\langle\phi_\alpha|\phi_\beta\rangle\langle\psi_\alpha|\psi_\beta\rangle d\xi,$$

where $|\phi_\alpha\rangle = |\phi_\alpha(\xi)\rangle$ is a state on an ancilla system $\mathcal{H}_\phi$. Here it should be stated that the decomposition (1) and the choice of $|\phi_\alpha(\xi)\rangle$ are not unique; they should be determined so that the length of the final key will be maximized. It is clear from the definition of G that $G \geq 0$, and so a square matrix C exists such that $G = C^\dagger C$. Further, since all the diagonal elements of G are 1, we can define a pure state $\hat{\rho}_\alpha$ on a four-dimensional Hilbert space $\mathcal{H}_4$ by

$$\hat{\rho}_\alpha = |C_\alpha\rangle\langle C_\alpha|,$$

where C_α denotes the α-th column of C. It follows from this construction that there exists a physical transformation from $\{\hat{\rho}_\alpha\}_{\alpha\in\mathcal{X}}$ to $\{\rho_\alpha\}_{\alpha\in\mathcal{X}}$ (see [4]). Consider here a modified protocol in which Alice uses $\hat{\rho}_\alpha$ instead of ρ_α. By using the above transformation, one can convert an arbitrary adversary attacking the original protocol into an adversary attacking the modified protocol. Therefore, if there exists an adversary which can break the original protocol, i.e. if the original protocol is insecure, then the modified protocol is also insecure; or equivalently (contraposition), if the latter is secure, then the former is also secure. Based on this consideration, we will treat the modified protocol instead of the original protocol.

We next introduce a protocol which approximates the modified protocol. Let $\mathcal{H}_2$ be a two-dimensional subspace of $\mathcal{H}_4$, and $\{\sigma_\alpha\}_{\alpha\in\mathcal{X}}$ be a set of states on $\mathcal{H}_2$ such that

$$\bar{\sigma}_0 = \bar{\sigma}_1, \quad \bar{\sigma}_a \equiv \sum_{x\in\{0,1\}} p(x)\sigma_{a,x}.$$

Here the choice of σ_α is not unique; it should be determined so that the distance $d_T(\hat{\rho}^{\mathcal{A}}_{a,x}, \sigma^{\mathcal{A}}_{a,x})$ will be minimized. Now we can introduce an approximate protocol in which Alice sends $\sigma^{\mathcal{A}}_{a,x}$ with probability $q(a^{\mathcal{A}}, x^{\mathcal{A}}) = \prod_i p_U(a_i)p(x_i)$, where p_U denotes the uniform distribution.

Let $n \in \mathbb{N}$, and for $1 \leq i \leq n$, let X_i be a random variable distributed over $\{0,1\}$ with $\Pr[X_i = 1] = p$. Define $X = \sum_{i=1}^n X_i$. Then, for $\delta > 0$,

$$\Pr[X \geq n(p+\delta)] \leq n\exp_2(-nD(B_1(p)||B_1(p+\delta))),$$

where $\exp_2(x) = 2^x$ and B_1 denotes a Bernoulli distribution (see e.g., [5]). This inequality can be used to estimate the error rate $p^e_{\mathcal{K}}$ at $\mathcal{K}$ from $p^e_{\mathcal{T}} = n^e_{\mathcal{T}}/n_{\mathcal{T}}$, the error rate at $\mathcal{T}$. Define for $\delta_p > 0$,

$$p^+_{\mathcal{K}} = p^e_{\mathcal{T}} + \frac{n_{\mathcal{K}}}{n_{\mathcal{C}}}\delta_p,$$
$$\epsilon^e_{\mathcal{T}} = n_{\mathcal{T}}\exp_2\left(-n_{\mathcal{T}}D(B_1(p^e_{\mathcal{T}})||B_1(p^e_{\mathcal{T}}+\delta_p))\right).$$

Then the probability that $p^e_{\mathcal{K}} > p^+_{\mathcal{K}}$ is bounded as

$$\Pr_{\mathcal{T}}[p^e_{\mathcal{K}} > p^+_{\mathcal{K}}] \leq \epsilon^e_{\mathcal{T}}, \tag{2}$$

where probability $\Pr_{\mathcal{T}}$ is taken over the randomness in choosing $\mathcal{T}$.

Let z be the output of the measurement by Eve. Then, without loss of generality, the probability distribution of the random variables can be written as

$$p^{\mathcal{A}}(a,b,x,y,z,\mathcal{T}) = p(a^{\mathcal{A}},x^{\mathcal{A}})p(b^{\mathcal{A}})p_{\mathcal{C}}(\mathcal{T})\mathrm{Tr}\,\hat{\rho}^{\mathcal{A}}_{a,x}M^{\mathcal{A}}_{b,yz},$$

where $M^{\mathcal{A}}_{b,yz}$ denotes the positive operator-valued measurement (POVM) performed by Bob and Eve. In the approximate protocol, the corresponding probability distribution can also be written as

$$q^{\mathcal{A}}(a,b,x,y,z,\mathcal{T}) = q(a^{\mathcal{A}},x^{\mathcal{A}})p(b^{\mathcal{A}})p_{\mathcal{C}}(\mathcal{T})\mathrm{Tr}\,\sigma^{\mathcal{A}}_{a,x}M^{\mathcal{A}}_{b,yz}.$$

It thus follows from the monotonicity of the trace distance that

$$d_V(p^{\mathcal{A}},q^{\mathcal{A}}) \leq \epsilon_{\mathcal{A}}, \tag{3}$$
$$\epsilon_{\mathcal{A}} \equiv d_V(p(a^{\mathcal{A}}),p_U(a^{\mathcal{A}})) + \sum_{a^{\mathcal{A}},x^{\mathcal{A}}} p(a^{\mathcal{A}},x^{\mathcal{A}})d_T(\hat{\rho}^{\mathcal{A}}_{a,x},\sigma^{\mathcal{A}}_{a,x}).$$

Consider now the probability distributions $p^{\mathcal{T}}_{ab}$ and $q^{\mathcal{T}}_{ab}$ defined by

$$p^{\mathcal{T}}_{ab}(x,y,z) = p(x^{\mathcal{T}},y^{\mathcal{T}},z|a^{\mathcal{A}},b^{\mathcal{A}},\mathcal{D},\mathcal{T}),$$
$$q^{\mathcal{T}}_{ab}(x,y,z) = q(x^{\mathcal{T}},y^{\mathcal{T}},z|a^{\mathcal{A}},b^{\mathcal{A}},\mathcal{D},\mathcal{T}),$$

respectively. It will help to remember Markov's inequality for a non-negative random variable X and a constant $c > 0$,

$$\Pr[X \geq cE(X)] \leq c^{-1}.$$

Here the expectation $E(X)$ may be replaced by its upper bound. If we define $X = \{p^{\mathcal{A}} - q^{\mathcal{A}} > 0\}^*/p(a^{\mathcal{A}},b^{\mathcal{A}},\mathcal{D},\mathcal{T})$, then Markov's inequality for X and a constant $c_{\mathcal{T}} > 0$, together with (3), gives

$$\Pr_{p^{\mathcal{A}}}[\{p^{\mathcal{A}} - q^{\mathcal{A}} > 0\}^* \geq c_{\mathcal{T}}\epsilon_{\mathcal{A}}p(a^{\mathcal{A}},b^{\mathcal{A}},\mathcal{D},\mathcal{T})] \leq c^{-1}_{\mathcal{T}}.$$

Similarly, it can be shown that

$$\Pr_{p^{\mathcal{A}}}[\{q^{\mathcal{A}} - p^{\mathcal{A}} > 0\}^* \geq c_{\mathcal{T}}\epsilon_{\mathcal{A}}p(a^{\mathcal{A}},b^{\mathcal{A}},\mathcal{D},\mathcal{T})] \leq c^{-1}_{\mathcal{T}}.$$

Consequently, we obtain

$$\Pr_{p^{\mathcal{A}}}[d_V(p^{\mathcal{T}}_{ab},q^{\mathcal{T}}_{ab}) \geq c_{\mathcal{T}}\epsilon_{\mathcal{A}}] \leq 2c^{-1}_{\mathcal{T}}. \tag{4}$$

The probability distribution $q^{\mathcal{T}}_{ab}$ can be represented as

$$q^{\mathcal{T}}_{ab}(x,y,z) = \frac{p(x^{\mathcal{T}})\mathrm{Tr}\,\bar{\sigma}^{\mathcal{A}}_{a,x^{\mathcal{T}}}M^{\mathcal{A}}_{b,y^{\mathcal{T}}z}}{\mathrm{Tr}\,\bar{\sigma}^{\mathcal{A}}_{a}M^{\mathcal{A}}_{b}}.$$

Here we have used the definitions

$$\bar{\sigma}^{\mathcal{A}}_{a,x^T} = \sum_{\hat{x}^{\mathcal{A}}} p(\hat{x}^{\mathcal{A}}|x^T)\sigma^{\mathcal{A}}_{a,\hat{x}}, \quad \bar{\sigma}^{\mathcal{A}}_{a} = \sum_{x^T} p(x^T)\bar{\sigma}^{\mathcal{A}}_{a,x^T},$$
$$M^{\mathcal{A}}_{b,y^T z} = \sum_{\hat{y}^{\mathcal{A}}:C^T_{\hat{y}}} M^{\mathcal{A}}_{b,\hat{y}z}, \qquad M^{\mathcal{A}}_{b} = \sum_{y^T,z} M^{\mathcal{A}}_{b,y^T z},$$

with $C^T_{\hat{y}} \equiv \{\hat{y}^T = y^T \wedge \hat{y}^{\mathcal{D}-\mathcal{T}} \in \{0,1\}^{\mathcal{D}-\mathcal{T}} \wedge \hat{y}^{\mathcal{A}-\mathcal{D}} = \phi^{\mathcal{A}-\mathcal{D}}\}$. Now, for a basis $a^{\mathcal{A}}$, let $\tilde{a}^{\mathcal{A}}$ denote the basis such that

$$\tilde{a}^{\mathcal{K}} = \bar{a}^{\mathcal{K}}, \quad \tilde{a}^{\mathcal{A}-\mathcal{K}} = a^{\mathcal{A}-\mathcal{K}},$$

where $\bar{a}$ denotes the bit-wise inversion of binary string a. It can then be verified that $\bar{\sigma}^{\mathcal{A}}_{a,x^T} = \bar{\sigma}^{\mathcal{A}}_{\tilde{a},x^T}$ and $M^{\mathcal{A}}_{b,y^T z} = M^{\mathcal{A}}_{\bar{b},y^T z}$, because $\bar{\sigma}_0 = \bar{\sigma}_1$ and $E_{0,0} + E_{0,1} = E_{1,0} + E_{1,1}$, respectively. Therefore, it is deduced that the probability distributions $q^{\mathcal{T}}_{ab}$ and $q^{\mathcal{T}}_{\bar{a}b}$ are identical. It thus follows from (4) that

$$d_V(p^{\mathcal{T}}_{ab}, q^{\mathcal{T}}_{\bar{a}b}) \leq c_{\mathcal{T}}\epsilon_{\mathcal{A}}, \tag{5}$$

which holds with probability at least $1 - 2c_{\mathcal{T}}^{-1}$.

Consider next the probability distributions $p^{\mathcal{K}}_{ab}$ and $q^{\mathcal{K}}_{ab}$ defined by

$$p^{\mathcal{K}}_{ab}(x,y,z) = p(x^{\mathcal{K}}, y^{\mathcal{K}}, z|a^{\mathcal{A}}, b^{\mathcal{A}}, \xi^T, \eta^T, \mathcal{D}, \mathcal{T}),$$
$$q^{\mathcal{K}}_{ab}(x,y,z) = q(x^{\mathcal{K}}, y^{\mathcal{K}}, z|a^{\mathcal{A}}, b^{\mathcal{A}}, \xi^T, \eta^T, \mathcal{D}, \mathcal{T}),$$

respectively. The probability distribution $q^{\mathcal{K}}_{ab}(x,y,z)$ can be represented as

$$q^{\mathcal{K}}_{ab}(x,y,z) = \frac{p(x^{\mathcal{K}})\mathrm{Tr}\,\bar{\sigma}^{\mathcal{A}}_{a,x^{\mathcal{K}}} M^{\mathcal{A}}_{b,y^{\mathcal{K}}z}}{\mathrm{Tr}\,\bar{\sigma}^{\mathcal{A}}_{a;\xi} M^{\mathcal{A}}_{b;\eta}}.$$

Here we have used the definitions

$$\bar{\sigma}^{\mathcal{A}}_{a,x^{\mathcal{K}}} = \sum_{\hat{x}^{\mathcal{A}}} p(\hat{x}^{\mathcal{A}}|x^{\mathcal{K}}, \xi^T)\sigma^{\mathcal{A}}_{a,\hat{x}}, \quad \bar{\sigma}^{\mathcal{A}}_{a;\xi} = \sum_{x^{\mathcal{K}}} p(x^{\mathcal{K}})\bar{\sigma}^{\mathcal{A}}_{a,x^{\mathcal{K}}},$$
$$M^{\mathcal{A}}_{b,y^{\mathcal{K}}z} = \sum_{\hat{y}^{\mathcal{A}}:C^{\mathcal{K}}_{\hat{y}}} M^{\mathcal{A}}_{b,\hat{y}z}, \qquad M^{\mathcal{A}}_{b;\eta} = \sum_{y^{\mathcal{K}},z} M^{\mathcal{A}}_{b,y^{\mathcal{K}}z},$$

with $C^{\mathcal{K}}_{\hat{y}} \equiv \{\hat{y}^{\mathcal{K}} = y^{\mathcal{K}} \wedge \hat{y}^T = \eta^T \wedge \hat{y}^{\mathcal{D}-\mathcal{C}} \in \{0,1\}^{\mathcal{D}-\mathcal{C}} \wedge \hat{y}^{\mathcal{A}-\mathcal{D}} = \phi^{\mathcal{A}-\mathcal{D}}\}$. Note that $\bar{\sigma}^{\mathcal{A}}_{a,x^{\mathcal{K}}}$ can be expressed in the form

$$\bar{\sigma}^{\mathcal{A}}_{a,x^{\mathcal{K}}} = \sigma^{\mathcal{K}}_{a,x} \otimes \bar{\sigma}^{\mathcal{A}-\mathcal{K}}_{a;\xi},$$

where, for $\mathcal{B} \subset \mathcal{A}$ and a state $\rho^{\mathcal{A}}$ on $\mathcal{H}^{\mathcal{A}}$, $\rho^{\mathcal{B}}$ is given by taking the partial trace over $\mathcal{H}^{\mathcal{A}-\mathcal{B}}$, i.e. $\rho^{\mathcal{B}} \equiv \mathrm{Tr}_{\mathcal{H}^{\mathcal{A}-\mathcal{B}}}\rho^{\mathcal{A}}$. Here $\bar{\sigma}^{\mathcal{A}-\mathcal{K}}_{a;\xi}$ is independent of $x^{\mathcal{K}}$, and hence there exists a POVM $\{M^{\mathcal{K}}_{b,yz}\}$ on $\mathcal{H}^{\mathcal{K}}$ such that

$$q^{\mathcal{K}}_{ab}(x,y,z) = (\nu_{\mathcal{K}})^{-1}\mathrm{Tr}\,\sigma^{\mathcal{K}}_{a,x}M^{\mathcal{K}}_{b,yz},$$

where $\nu_{\mathcal{K}} \equiv \mathrm{Tr}\,\bar{\sigma}^{\mathcal{K}}_{a}M^{\mathcal{K}}_{b;\eta} = \mathrm{Tr}\,\bar{\sigma}^{\mathcal{K}}_{\bar{a}}M^{\mathcal{K}}_{b;\eta}$ is the normalization constant.

Now, inequalities (2) and (5) lead to

$$\sum_{x^{\mathcal{K}},y^{\mathcal{K}},z:|x\oplus y|>n_{\mathcal{K}}p^+_{\mathcal{K}}} q^{\mathcal{K}}_{a\bar{b}}(x,y,z) \leq \omega_{\mathcal{K}}, \quad \omega_{\mathcal{K}} \equiv \epsilon^{\epsilon}_{\mathcal{T}} + c_{\mathcal{T}}\epsilon_{\mathcal{A}};$$

or equivalently

$$\sum_{y^{\mathcal{K}},z} \mathrm{Tr}(\bar{\sigma}^{\mathcal{K}}_{\tilde{a}} - \bar{\sigma}^{\mathcal{K}}_{\tilde{a},y})M^{\mathcal{K}}_{b,yz} \leq \nu_{\mathcal{K}}\omega_{\mathcal{K}}, \tag{6}$$

where we have defined

$$\bar{\sigma}^{\mathcal{K}}_{\tilde{a},y} = \sum_{x^{\mathcal{K}}:|x\oplus y|\leq n_{\mathcal{K}}p^+_{\mathcal{K}}} p(x^{\mathcal{K}})\sigma^{\mathcal{K}}_{\tilde{a},x}.$$

It should be stated that $E_{0,0} + E_{0,1} = E_{1,0} + E_{1,1}$, and so $p^{\mathcal{K}}_{ab}(x,z) = p^{\mathcal{K}}_{a\bar{b}}(x,z)$. That is, the joint probability of the random variables x and z is independent of the basis used for the Bob's measurement. Hence, in the sequel, we will consider $p^{\mathcal{K}}_{ab}(x,y,z)$ rather than $p^{\mathcal{K}}_{ab}(x,y,z)$.

Let $R^{\mathcal{K}}_{ab}(X|y,z)$ denote the conditional Rényi entropy defined by

$$R^{\mathcal{K}}_{ab}(X|y,z) = -\log_2 \sum_{x^{\mathcal{K}}} \left(p^{\mathcal{K}}_{ab}(X = x|Y = y, Z = z)\right)^2,$$

where a capital letter (say X) denotes the random variable which samples the corresponding small letter (say x). To lower-bound $R^{\mathcal{K}}_{ab}$, we now upper-bound the conditional probabilities $p^{\mathcal{K}}_{ab}(x|y,z)$ and $q^{\mathcal{K}}_{ab}(x|y,z)$. Let $\bar{\pi}_{\mathcal{K}}$ be the smallest eigenvalue of $\bar{\sigma}^{\mathcal{K}}_{a}$. Then $\bar{\pi}_{\mathcal{K}}$ is given by

$$\bar{\pi}_{\mathcal{K}} = \exp_2(-n^0_{\mathcal{K}}\lambda_{\bar{\sigma}_0} - n^1_{\mathcal{K}}\lambda_{\bar{\sigma}_1}),$$

where $n^b_{\mathcal{K}} = |\{i \in \mathcal{K}|a_i = b\}|$ for $b \in \{0,1\}$, and λ_σ denotes the smaller eigenvalue of a state σ. By definition of $\bar{\pi}_{\mathcal{K}}$, it is clear that

$$q^{\mathcal{K}}_{a\bar{b}}(y,z) \geq \bar{\pi}_{\mathcal{K}}\tau_{\mathcal{K}}, \quad \tau_{\mathcal{K}} \equiv (\nu_{\mathcal{K}})^{-1}\mathrm{Tr}\,M^{\mathcal{K}}_{\bar{b},yz}.$$

To upper-bound $q^{\mathcal{K}}_{a\bar{b}}(x,y,z)$, we expand $q^{\mathcal{K}}_{a\bar{b}}(x,y,z)$ by using the identity $\sigma^{\mathcal{K}}_{a,x} = (V_y + \bar{V}_y)^\dagger\sigma^{\mathcal{K}}_{a,x}(V_y + \bar{V}_y)$, where

$$V_y = \left(\bar{\sigma}^{\mathcal{K}}_{a}\right)^{-\frac{1}{2}}\left(\bar{\sigma}^{\mathcal{K}}_{\tilde{a}} - \bar{\sigma}^{\mathcal{K}}_{\tilde{a},y}\right)^{\frac{1}{2}}, \quad \bar{V}_y = I_{\mathcal{H}^{\mathcal{K}}} - V_y.$$

Consider first the term including $\bar{V}^\dagger_y\sigma^{\mathcal{K}}_{a,x}\bar{V}_y$. On using $\bar{\sigma}^{\mathcal{K}}_{a} = \bar{\sigma}^{\mathcal{K}}_{\tilde{a}}$ and $\bar{\sigma}^{\mathcal{K}}_{\tilde{a}} \geq \bar{\sigma}^{\mathcal{K}}_{\tilde{a}} - \bar{\sigma}^{\mathcal{K}}_{\tilde{a},y}$, it can be shown that

$$\mathrm{Tr}\,\bar{V}^\dagger_y\sigma^{\mathcal{K}}_{a,x}\bar{V}_y \leq \mathrm{Tr}\,\sigma^{\mathcal{K}}_{a,x}\left(\bar{\sigma}^{\mathcal{K}}_{\tilde{a}}\right)^{-\frac{1}{2}}\bar{\sigma}^{\mathcal{K}}_{\tilde{a},y}\left(\bar{\sigma}^{\mathcal{K}}_{\tilde{a}}\right)^{-\frac{1}{2}}.$$

Therefore

$$(\nu_{\mathcal{K}})^{-1}p(x^{\mathcal{K}})\mathrm{Tr}\,\bar{V}^\dagger_y\sigma^{\mathcal{K}}_{a,x}\bar{V}_yM^{\mathcal{K}}_{b,yz} \leq \pi_{\mathcal{K}}\tau_{\mathcal{K}},$$
$$\pi_{\mathcal{K}} \equiv p(x^{\mathcal{K}})\mathrm{Tr}\,\sigma^{\mathcal{K}}_{a,x}\left(\bar{\sigma}^{\mathcal{K}}_{\tilde{a}}\right)^{-\frac{1}{2}}\bar{\sigma}^{\mathcal{K}}_{\tilde{a},y}\left(\bar{\sigma}^{\mathcal{K}}_{\tilde{a}}\right)^{-\frac{1}{2}}.$$

It is now convenient to define

$$\bar{\sigma}^{\otimes\mathcal{K}}_{\tilde{a},y} = \bigotimes_{i\in\mathcal{K}} \exp_2(h(p^+_{\mathcal{K}}) - 1)((1-p^+_{\mathcal{K}})\sigma^{\mathcal{K}}_{\tilde{a},y_i} + p^+_{\mathcal{K}}\sigma^{\mathcal{K}}_{\tilde{a},\bar{y}_i}),$$

where $h(p) = -p\log_2 p - (1-p)\log_2(1-p)$. Since $\bar{\sigma}^{\mathcal{K}}_{\tilde{a},y} \leq \bar{\sigma}^{\otimes\mathcal{K}}_{\tilde{a},y}$, we can take $\pi_{\mathcal{K}}$ in a more convenient form

$$\pi_{\mathcal{K}} = p(x^{\mathcal{K}})\mathrm{Tr}\,\sigma^{\mathcal{K}}_{a,x}\left(\bar{\sigma}^{\mathcal{K}}_{\tilde{a}}\right)^{-\frac{1}{2}}\bar{\sigma}^{\otimes\mathcal{K}}_{\tilde{a},y}\left(\bar{\sigma}^{\mathcal{K}}_{\tilde{a}}\right)^{-\frac{1}{2}}.$$

Consider next the term including $V^\dagger_y\sigma^{\mathcal{K}}_{a,x}V_y$, i.e.

$$q'_{a\bar{b}}(x,y,z) \equiv (\nu_{\mathcal{K}})^{-1}p(x^{\mathcal{K}})\mathrm{Tr}\,V^\dagger_y\sigma^{\mathcal{K}}_{a,x}V_yM^{\mathcal{K}}_{\bar{b},yz}.$$

Inequality (6) now leads to

$$\sum_{x^{\mathcal{K}},y^{\mathcal{K}},z} q^{\mathcal{K}}_{a\bar{b}}(x,y,z)\frac{q'_{a\bar{b}}(x,y,z)}{q^{\mathcal{K}}_{a\bar{b}}(x,y,z)} \leq \omega_{\mathcal{K}}.$$

Hence Markov's inequality for $X = q'_{a\tilde{b}}/q^{\mathcal{K}}_{a\tilde{b}}$ and a constant $c > 0$ becomes

$$\Pr_{q^{\mathcal{K}}_{a\tilde{b}}}[q'_{a\tilde{b}}(x,y,z) \geq c\omega_{\mathcal{K}} q^{\mathcal{K}}_{a\tilde{b}}(x,y,z)] \leq c^{-1}.$$

Finally, the remaining terms can be bounded, by use of Schwarz's inequality, as

$$\mathrm{Tr}\, V_y^\dagger p(x^{\mathcal{K}}) \sigma^{\mathcal{K}}_{a,x} \bar{V}_y M^{\mathcal{K}}_{\tilde{b},yz} \leq (\pi_{\mathcal{K}}\tau_{\mathcal{K}} q'_{a\tilde{b}}(x,y,z))^{\frac{1}{2}},$$
$$\mathrm{Tr}\, \bar{V}_y^\dagger p(x^{\mathcal{K}}) \sigma^{\mathcal{K}}_{a,x} V_y M^{\mathcal{K}}_{\tilde{b},yz} \leq (\pi_{\mathcal{K}}\tau_{\mathcal{K}} q'_{a\tilde{b}}(x,y,z))^{\frac{1}{2}}.$$

It therefore follows that

$$q^{\mathcal{K}}_{a\tilde{b}}(x,y,z) \leq \left((\pi_{\mathcal{K}}\tau_{\mathcal{K}})^{\frac{1}{2}} + (q'_{a\tilde{b}}(x,y,z))^{\frac{1}{2}}\right)^2,$$

and so

$$\Pr_{q^{\mathcal{K}}_{a\tilde{b}}}[q^{\mathcal{K}}_{a\tilde{b}}(x|y,z) \geq \Pi_{\mathcal{K}}] \leq c^{-1}, \ \Pi_{\mathcal{K}} \equiv \frac{\pi_{\mathcal{K}}}{\bar{\pi}_{\mathcal{K}}\left(1 - (c\omega_{\mathcal{K}})^{\frac{1}{2}}\right)^2}.$$

In the same way as before (see (4)), Markov's inequality for a constant $c_{\mathcal{K}} > 0$ can be used to show that

$$\Pr_{p^{\mathcal{A}}}[d_V(p^{\mathcal{K}}_{a\tilde{b}}(x|y,z), q^{\mathcal{K}}_{a\tilde{b}}(x|y,z)) \geq c_{\mathcal{K}}\epsilon_{\mathcal{A}}] \leq 2c_{\mathcal{K}}^{-1},$$

and hence

$$\Pr_{p^{\mathcal{K}}_{a\tilde{b}}}[p^{\mathcal{K}}_{a\tilde{b}}(x|y,z) \geq \Pi_{\mathcal{K}} + c_{\mathcal{K}}\epsilon_{\mathcal{A}}] \leq c^{-1},$$

which holds with probability at least $1 - 2c_{\mathcal{T}}^{-1} - 2c_{\mathcal{K}}^{-1}$ (with respect to probability distribution $p^{\mathcal{A}}$). We note that the constants c, $c_{\mathcal{T}}$ and $c_{\mathcal{K}}$ should be determined so that the length of the final key will be maximized.

Define

$$R^{\mathcal{K}}_E = -\log_2 \max_{x^{\mathcal{K}}, y^{\mathcal{K}}} \{\Pi_{\mathcal{K}} + c_{\mathcal{K}}\epsilon_{\mathcal{A}}\},$$

and let m be an integer such that $l \equiv R^{\mathcal{K}}_E - m > 0$. Choose a function g at random from a universal family of hash functions from $\{0,1\}^n$ to $\{0,1\}^m$. If Alice and Bob choose $s = g(x^{\mathcal{K}})$ as their secret key, then Eve's expected information about S, given Z and G, satisfies $I(S : Z, G) \leq m\epsilon_c + 2^{-l}/\ln 2$, where $\epsilon_c \equiv c^{-1} + 2c_{\mathcal{T}}^{-1} + 2c_{\mathcal{K}}^{-1}$, I denotes the mutual information, and we consider Y as an auxiliary random variable (see [3] for details). Note that $R^{\mathcal{K}}_E$ is not explicitly dependent on the characteristics of the detector, and hence the detector can be uncharacterized. Further, the term $\epsilon_{\mathcal{A}}$ approaches 1 as $n_{\mathcal{K}} \to \infty$ unless $\hat{\rho}_{a,x} = \sigma_{a,x}$. This shows that the leading factor reducing the key generation rate is the asymmetry of the source represented by the term $\epsilon_{\mathcal{A}}$, and hence eliminating the asymmetry is of practical importance in implementing efficient quantum key distribution schemes.

To see that our result is consistent with the previous ones, we suppose that the source is perfect. In this case, we can take $\rho_{a,x} = \hat{\rho}_{a,x} = \sigma_{a,x}$, $\bar{\sigma}_0 = \bar{\sigma}_1 = (1/2)I_{\mathcal{H}_2}$, $\bar{\sigma}^{\mathcal{K}}_{\tilde{a}} = (1/2)^{n_{\mathcal{K}}} I_{\mathcal{H}^{\mathcal{K}}_2}$, $\epsilon_{\mathcal{A}} = 0$, $p(x^{\mathcal{K}}) = \bar{\pi}_{\mathcal{K}} = (1/2)^{n_{\mathcal{K}}}$, and $\mathrm{Tr}\sigma^{\mathcal{K}}_{a,x}\bar{\sigma}^{\otimes\mathcal{K}}_{\tilde{a},y} = \exp_2(-n_{\mathcal{K}}(2 - h(p^+_{\mathcal{K}})))$. Since $\omega_{\mathcal{K}} = \epsilon^e_{\mathcal{T}} \to 0$ as $n_{\mathcal{K}} \to \infty$ for fixed δ_p, $R^{\mathcal{K}}_E/n_{\mathcal{K}}$ approaches $1 - h(p^e_{\mathcal{T}})$ for sufficiently small c^{-1}, $c_{\mathcal{T}}^{-1}$, $c_{\mathcal{K}}^{-1}$ and δ_p. This is consistent with the results in the previous works[6, 9, 10, 11].

4 Future problems

We close this paper with mentioning some extensions of the above security proof. In the same way as Koashi-Preskill[9], we can provide a security proof of the BB84 protocol where the only assumption is that the detector and basis dependence of the averaged states are characterized. It is also of importance to give a security proof of the B92 protocol[1] in practical implementation. Suppose that the source generates ρ_0 with probability p_0 and ρ_1 with probability p_1. Define then $\hat{\rho}_a$ by introducing the Gram matrix as above. Note that $\hat{\rho}_a$ is a pure state on a two-dimensional Hilbert space $\mathcal{H}_2$. Hence, the asymmetry mentioned above automatically vanishes in this case, which could be considered as an advantage of the B92 protocol. More detailed investigation concerning these extensions will be the subject of future work.

Acknowledgements

The author is grateful to Dr. Keiji Matsumoto for useful comments. This work was supported in part by MEXT, Grant-in-Aid for Encouragement of Young Scientists (B) No. 15760289.

References

[1] C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992)

[2] C. H. Bennett and G. Brassard, in *Proceedings of IEEE Conference on Computers, Systems and Signal Processing*, Banglore (India), 175–179 (1984)

[3] C. H. Bennett et al., IEEE Trans. Inform. Theory **41**, 1915 (1995)

[4] A. Chefles, R. Jozsa and A. Winter, quant-ph/0307227.

[5] I. Csizár and J. Körner, Information theory, coding theorems for discrete memoryless systems, Academic (1981)

[6] D. Gottesman et al., *Quantum Inform. Comput.* **4**, 325–360 (2004)

[7] M. Hamada, J. Phys. A: Math. and Gen. **37**, 8303 (2003)

[8] M. Hayashi, An Introduction to Quantum Information Theory, Springer, to be published

[9] M. Koashi and J. Preskill, Phys. Rev. Lett. **90**, 057902 (2003)

[10] D. Mayers, J. Assoc. Comput. Mach. **48**, 351 (2001)

[11] P.W. Shor and J. Preskill, Phys. Rev. Lett. **85**, 441 (2000)

[12] H. Takesue et al., quant-ph/0507110

[13] S. Watanabe et al., quant-ph/0412070

[14] Y. Watanabe, quant-ph/0506246

On Information-Disturbance Theorem

Takayuki Miyadera[1] * Hideki Imai[1] [2]

[1] *Research Center for Information Security (RCIS), National Institute of Industrial Science and Technology (AIST). Akihabara Daibiru 1102, 1-18-13 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan.*

[2] *Institute of Industrial Science, University of Tokyo. 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan.*

Abstract. We derive a novel version of information-disturbance theorems which are crucial in the security proof of BB84 quantum key distribution protocol. We show that the information gain by Eve inevitably makes the outcomes by Bob in the conjugate basis not only erroneous but random.

Keywords: Information-Disturbance Theorem, BB84 QKD protocol.

1 Introduction

In 1984, Bennett and Brassard[1] proposed a quantum key distribution protocol which is now called as BB84 protocol. Its unconditional security was first proved by Mayers[2] in 1996 and after his proof the various proofs[3, 4, 5] have appeared. Among them, a proof by Biham, Boyer, Boykin, Mor, and Roychowdhury[4] is based upon a so-called *information disturbance theorem* and is related with the present paper. According to the theorem, the information gain by Eve inevitably induces *errors* in outcomes obtained by Bob. This disturbance enables Alice and Bob to notice the existence of an eavesdropper. As well as its application to BB84 protocol, since it can be regarded as an information theoretic version of uncertainty relation, the theorem is very interesting by itself. Recently, Boykin and Roychowdhury[6] showed a simple proof of the theorem in an arbitrary dimension by using purification technique and trace norm inequality. We, in this paper, derive a different version of the theorem. Our information-disturbance theorem is an inequality between the information gain by Eve and the *randomness* (rather than error probability) of the outcomes obtained by Bob. We compare our theorem with the former one and discuss its implication.

2 Setting

Let us begin with a setting. Three characters, Alice, Bob and Eve play their roles. Our setting is a simplified version of BB84 quantum key distribution protocol. The following analysis, however, can be easily applied to the full BB84 protocol with proceeding public discussion procedures. Let us consider two pairs of orthogonal states, $b := \{|0\rangle, |1\rangle\}$ and its *conjugate* $\bar{b} := \{|\bar{0}\rangle, |\bar{1}\rangle\}$ in $\mathbf{C}^2$. They are assumed mutually unbiased with each other. That is,

$$\langle i|\bar{k}\rangle = \sqrt{\frac{1}{2}}(-1)^{ik}$$

holds for each pair of $i, k \in \{0, 1\}$. Alice first selects b or $\bar{b}$ which is used to encode a random number. Alice next randomly generates an N-bits sequence $i \in \{0,1\}^N$ with probability $p(i) = \frac{1}{2^N}$. We write A a random variable representing this N-bits sequence. Alice encodes this information on N-qubits and sends them to Bob. For instance, suppose that Alice selects b and generates a sequence $i = i_1 i_2 \cdots i_N$, she sends the corresponding state $|i\rangle = |i_1\rangle \otimes |i_2 \otimes \cdots \otimes |i_N\rangle \in \mathbf{C}^2 \otimes \cdots \otimes \mathbf{C}^2 =: \mathcal{H}_A \simeq \mathcal{H}_B$ to Bob. If the conjugate basis $\bar{b}$ and a sequence $i = i_1 i_2 \cdots i_N$ are chosen, the state sent to Bob is $|\bar{i}\rangle = |\bar{i_1}\rangle \otimes |\bar{i_2}\rangle \otimes \cdots \otimes |\bar{i_N}\rangle \in \mathcal{H}_A$. Alice, after confirming that Bob actually has received N-qubits, informs him of the basis she used. Bob makes a measurement with respect to the basis and obtains an outcome. Let us write B the random variable representing this outcome. If there is no eavesdropper, $A = B$ naturally follows. Eve wants to obtain the information of the random variable A. For the purpose, Eve prepares an apparatus and makes it interact with the N-qubits sent to Bob by Alice. Let us denote $\mathcal{H}_E$ a Hilbert space describing Eve's apparatus. In general, Eve's operation is described by a unitary operator U,

$$\begin{aligned} U : \mathcal{H}_E \otimes \mathcal{H}_A &\to \mathcal{H}_E \otimes \mathcal{H}_B \\ |0\rangle \otimes |i\rangle &\mapsto \sum_j |E_{ij}\rangle \otimes |j\rangle, \end{aligned} \tag{1}$$

where $|0\rangle$ is a normalized vector in $\mathcal{H}_E$ and $\{|E_{ij}\rangle\} \subset \mathcal{H}_E$ satisfies unitarity condition: $\sum_{j\in\{0,1\}^N} \langle E_{ij}|E_{kj}\rangle = \delta_{ik}$. After this interaction, Eve tries to make an optimal measurement on her apparatus to extract the information of A.

3 Information-Disturbance Theorem

3.1 Information v.s. Error

One can show that if Eve's operation yields herself to gain large information, error probability in qubits sent to Bob in the conjugate basis becomes inevitably large. It has been called as *information-disturbance theorem* and was proved by [4, 6].

The representation (1) depends upon the choice of the basis. It is useful to rewrite the same unitary operator in the conjugate basis, $\bar{b}$. Using $|\bar{i}\rangle = \sum_{l\in\{0,1\}^N} |l\rangle\langle l|\bar{i}\rangle$ and $|i\rangle = \sum_{j\in\{0,1\}^N} |\bar{j}\rangle\langle\bar{j}|i\rangle$, we obtain

$$U|0\rangle \otimes |\bar{l}\rangle = \sum_{s\in\{0,1\}^N} |\overline{E}_{ls}\rangle \otimes |\bar{s}\rangle,$$

where $|\overline{E}_{ls}\rangle = \sum_{i,j\in\{0,1\}^N} |E_{ij}\rangle\langle\bar{s}|j\rangle\langle i|\bar{l}\rangle$.

*miyadera-takayuki@aist.go.jp

When Alice chooses basis b and a sequence $i \in \{0,1\}^N$, a state obtained by Eve is computed as

$$\rho^i_{Eve} := \sum_{j\in\{0,1\}^N} |E_{ij}\rangle\langle E_{ij}|.$$

Later we consider how much information Eve can extract from it. When Alice chooses another basis $\overline{b}$ and a sequence i, Bob obtains a state

$$\overline{\rho}^i_{Bob} = \sum_{j,l\in\{0,1\}^N} \langle\overline{E}_{il}|\overline{E}_{ij}\rangle|\overline{j}\rangle\langle\overline{l}| \qquad (2)$$

in the presence of Eve. Later we consider the error it induces to the outcome.

Let us begin with Eve's information gain. Eve performs a measurement (POVM) $X := \{X_\alpha\}$ on her state. (POVM is a family of positive operators satisfying $\sum_\alpha X_\alpha = \mathbf{1}$.) We put $E[X]$ a random variable representing this outcome. Probability to obtain an outcome α is $p(\alpha|i,b) = \mathrm{tr}\left(X_\alpha \rho^i_{Eve}\right)$. Information gain (mutual information) by Eve with respect to a POVM X is calculated as,

$$I(A : E[X]|b) = \frac{1}{2^N}\sum_\alpha\sum_i p(\alpha|i)\left(\log p(\alpha|i) - \log\sum_j p(\alpha|j)\right) + N.$$

What we are interested in is its optimal value with respect to all the possible measurements by Eve:

$$I(A : E|b) := \sup\{I(A : E[X]|b)\,|\,X = \{X_\alpha\}\text{is a POVM in } \mathcal{H}_E\}.$$

Now we consider outcomes obtained by Bob in the conjugate basis. Remind that when Alice chooses basis $\overline{b}$, the state sent to Bob is (2). Bob makes a measurement of an observable $\sum j|\overline{j}\rangle\langle\overline{j}|$ on it. We put B a random variable for this outcome. The probability to obtain each outcome is expressed as $p(j|i,\overline{b}) = \langle\overline{E}_{ij}|\overline{E}_{ij}\rangle$. Thus probability to obtain an outcome whose difference from input is c, is

$$\begin{aligned} p(B = A\oplus c|\overline{b}) &:= \sum_i \frac{1}{2^N} p(i\oplus c|i,\overline{b}) \\ &= \frac{1}{2^N}\sum_i \langle\overline{E}_{i\ i\oplus c}|\overline{E}_{i\ i\oplus c}\rangle, \end{aligned} \qquad (3)$$

where $c \in \{0,1\}^N$ and the symbol "$\oplus$" is a a bit-wise XOR operation. By use of these quantities, the information-disturbance theorem obtained by Boykin and Roychowdhury is expressed as

$$I(A : E|b) \le 4N\sqrt{\sum_{c\neq 0} p(B = A\oplus c|\overline{b})}, \qquad (4)$$

whose right hand side means the square root of the error probability in Bob's outcome. That is, their theorem claims that the information gain by Eve makes Bob's outcome in conjugate basis *erroneous*.

3.2 Information v.s. Randomness

In this subsection, we derive a new information-disturbance theorem which relates information gain by Eve with randomness in Bob's outcome.

To estimate the information gain by Eve, we introduce a symmetrized attack as in [4]. We add N auxiliary qubits to Eve's apparatus and thus the Eve's Hilbert space is dilated to $\mathcal{H}_{E'} := \mathbf{C}^2\otimes\cdots\otimes\mathbf{C}^2\otimes\mathcal{H}_E$. Introduce a set of new vectors $\{|E^s_{ij}\rangle\}$ in this Hilbert space $\mathcal{H}_{E'}$ as

$$|E^s_{ij}\rangle := \sqrt{\frac{1}{2^N}}\sum_{m\in\{0,1\}^N}(-1)^{m\cdot(i\oplus j)}|m\rangle\otimes|E_{i\oplus m\ j\oplus m}\rangle,$$

where "$\oplus$" is again a bit-wise XOR operation and "$\cdot$" represents bit-wise multiplications followed by their summation. Consider a new *symmetrized* attack as

$$\begin{aligned} U^s : \mathcal{H}_{E'}\otimes\mathcal{H}_A &\to \mathcal{H}_{E'}\otimes\mathcal{H}_B \\ (|0\rangle\otimes|0\rangle)\otimes|i\rangle &\mapsto \sum_j |E^s_{ij}\rangle\otimes|j\rangle \end{aligned}$$

which can be extended to satisfy unitarity condition [4]. Although this symmetrized attack is different from the original attack, it is shown below that to treat this new attack is useful.

If we employ the symmetrized attack, Eve has a state described as

$$\rho^i_{Eve,sym} := \sum_{j\in\{0,1\}^N} |E^s_{ij}\rangle\langle E^s_{ij}|.$$

To extract the information from it, she can measure the value of the auxiliary N-qubits and then apply a POVM $X = \{X_\alpha\}$ on the original apparatus $\mathcal{H}_E$. It is shown that this strategy gives same amount of information with the original attack. The values obtained by the first measurement are equally distributed, that is, each value m is obtained with probability $\frac{1}{2^N}$. After obtaining a value m, the reduction of wave packet forces the state into

$$\rho^i_m := \sum_j |E_{i\oplus m\ j\oplus m}\rangle\langle E_{i\oplus m\ j\oplus m}|.$$

The second measurement gives a probability

$$p^s(\alpha|i,m) = \sum_j \langle E_{i\oplus m\ j\oplus m}|X_\alpha|E_{i\oplus m\ j\oplus m}\rangle,$$

from which it is easy to see that

$$p^s(\alpha|i,m) = p(\alpha|i\oplus m)$$

holds. Thus by using conditional probability $p^s(\alpha,m|i) = \frac{1}{2^N}p(\alpha|i\oplus m)$, mutual information can be computed to coincide with $I(A : E[X]|b)$. Taking a supremum over all the possible POVM over the full Hilbert space $\mathcal{H}_{E'}$ can make it larger and therefore the following inequality holds,

$$I(A : E|b) \le I(A : E|b)_{sym}, \qquad (5)$$

where the right hand side is the optimal information gain by the symmetrized attack.

Now we can state our theorem.

Theorem 1 *The following inequality holds:*

$$I(A:E|b) \leq H(A \oplus B|\overline{b}), \tag{6}$$

where $H(\cdot)$ is the Shannon entropy. That is, the information gain by Eve in the basis b makes the outcome of measurement by Bob in the conjugate basis $\overline{b}$ random.

Proof: We can prove the theorem by first symmetrizing the attack as in [4] and next bound Eve's information gain by Holevo's inequality. Thanks to (5), it is sufficient to estimate the quantity $I(A,E|b)_{sym}$ for our purpose. Holevo's theorem[7] bounds it from above as

$$\begin{aligned} & I(A:E|b)_{sym} \\ \leq\ & S\left(\frac{1}{2^N}\sum_i \rho^i_{Eve,sym}\right) - \sum_i \frac{1}{2^N} S\left(\rho^i_{Eve,sym}\right) \\ =:\ & \chi\left(\{\rho^i_{Eve,sym}\}\right), \end{aligned}$$

where $S(\rho)$ is von Neumann entropy of a state ρ. There exists a useful representation of this quantity χ. Consider another additional N-qubits Hilbert space $\mathcal{H}_R$ and a state over $\mathcal{H}_R \otimes \mathcal{H}_{E'}$,

$$\Theta := \sum_i \frac{1}{2^N} |i\rangle\langle i| \otimes \rho^i_{Eve,sym}.$$

Its quantum mutual entropy between $\mathcal{H}_R$ and $\mathcal{H}_E$ is shown to coincide with the quantity $\chi\left(\{\rho^i_{Eve,sym}\}\right)$,

$$I(\Theta) := S\left(\Theta|_{E'}\right) + S\left(\Theta|_R\right) - S\left(\Theta\right) = \chi\left(\{\rho^i_{Eve,sym}\}\right),$$

where $\Theta|_{E'}$ is a restricted state to $\mathcal{H}_{E'}$ of Θ and $\Theta|_R$ is defined in the same manner. To estimate this quantity, we consider a purification of $\rho^i_{Eve,sym}$. Introduce another N-qubits system $\mathcal{H}_P$ and states over $\mathcal{H}_{E'} \otimes \mathcal{H}_P$ [4],

$$|\varphi_i\rangle := \sum_j |E^s_{ij}\rangle \otimes |i \oplus j\rangle.$$

One can easily show that they are normalized. A state $\tilde{\Theta}$ over $\mathcal{H}_R \otimes \mathcal{H}_{E'} \otimes \mathcal{H}_P$ defined by

$$\tilde{\Theta} := \sum_i \frac{1}{2^N} |i\rangle\langle i| \otimes |\varphi_i\rangle\langle\varphi_i|$$

gives Θ if restricted to $\mathcal{H}_R \otimes \mathcal{H}_{E'}$. By using subadditivity for the entropy difference[8], the mutual entropy $I(\tilde{\Theta})$ between $\mathcal{H}_R$ and $\mathcal{H}_{E'} \otimes \mathcal{H}_P$ is shown to be larger than $I(\Theta)$. Therefore we estimate the quantity,

$$I(\tilde{\Theta}) := S\left(\tilde{\Theta}\Big|_{E'P}\right) + S\left(\tilde{\Theta}\Big|_R\right) - S\left(\tilde{\Theta}\right)$$

Now we compute the restricted states over the subsystems,

$$\begin{aligned} \tilde{\Theta}\Big|_R &= \sum_{ij} \frac{1}{2^N} \langle E^s_{ij}|E^s_{ij}\rangle |i\rangle\langle i| = \frac{1}{2^N}\mathbf{1} \\ \tilde{\Theta}\Big|_{E'P} &= \sum_i \frac{1}{2^N} |\varphi_i\rangle\langle\varphi_i|. \end{aligned}$$

The von Neumann entropy of $\tilde{\Theta}\Big|_R$ is N.

To compute the von Neumann entropy of $\tilde{\Theta}$ itself, we purify this by adding an additional N-qubits $\mathcal{H}_T$ and define a state over $\mathcal{H}_T \otimes \mathcal{H}_R \otimes \mathcal{H}_{E'} \otimes \mathcal{H}_P$,

$$|\Psi\rangle := \sum_i \sqrt{\frac{1}{2^N}} |i\rangle \otimes |i\rangle \otimes |\varphi_i\rangle.$$

Taking partial trace over $\mathcal{H}_R \otimes \mathcal{H}_{E'} \otimes \mathcal{H}_P$ leads

$$\sum_i \frac{1}{2^N} \langle\varphi_i|\varphi_i\rangle |i\rangle\langle i|. = \frac{1}{2^N}\mathbf{1}$$

whose entropy also is N. Thus the mutual entropy is determined by $\tilde{\Theta}\Big|_{E'P}$.

$$I(\tilde{\Theta}) = S\left(\sum_i \frac{1}{2^N} |\varphi_i\rangle\langle\varphi_i|\right).$$

Now let us calculate the von Neumann entropy of $\tilde{\Theta}\Big|_{E'P}$. Again a purification using an additional N-qubits $\mathcal{H}_Z$ to $\mathcal{H}_{E'} \otimes \mathcal{H}_P$ gives a state

$$|\Phi\rangle := \sum_i \sqrt{\frac{1}{2^N}} |i\rangle \otimes |\varphi_i\rangle$$

on $\mathcal{H}_Z \otimes \mathcal{H}_{E'} \otimes \mathcal{H}_P$. Its restriction to $\mathcal{H}_Z$ gives

$$\sigma := \frac{1}{2^N} \sum_{ij} \sum_n \langle E^s_{j\ j\oplus u}|E^s_{i\ i\oplus u}\rangle |i\rangle\langle j|$$

whose entropy agrees with $I(\tilde{\Theta})$. Let us consider its components with respect to the basis $\{|i\rangle\}$. Since

$$\sum_u \langle E^s_{j\ j\oplus u}|E^s_{i\ i\oplus u}\rangle = \frac{1}{2^N} \sum_n \sum_u \langle E_{j\oplus u\ j\oplus n\oplus u}|E_{i\oplus u\ i\oplus n\oplus u}\rangle$$

holds, it depends upon only $i \oplus j$. Let us write it as $f(i \oplus j)$ to represent σ as

$$\sigma = \frac{1}{2^N} \sum_{ij} f(i \oplus j) |i\rangle\langle j|.$$

It can be diagonalized by an orthonormalized vectors

$$|\mu_l\rangle := \sqrt{\frac{1}{2^N}} \sum_l (-1)^{i\cdot l} |l\rangle$$

as

$$\sigma = \sum_l \lambda_l |\mu_l\rangle\langle\mu_l|,$$

with $\lambda_l := \frac{1}{2^N} \sum_t f(t)(-1)^{t\cdot l}$. The eigenvalue λ_l is calculated as

$$\begin{aligned} \lambda_l &:= \frac{1}{2^N} \sum_t f(t)(-1)^{t\cdot l} \\ &= \frac{1}{2^N} \sum_{t,n,v} \langle E_{v\ v\oplus n}|E_{t\oplus v\ t\oplus v\oplus n}\rangle (-1)^{t\cdot l} \\ &= \left(\frac{1}{2^N}\right)^2 \sum_{t,n,v} \sum_{ij} \sum_{i'j'} \langle\overline{E}_{ij}|\overline{E}_{i'j'}\rangle \langle\overline{j}|v \oplus n\rangle\langle v|\overline{i}\rangle \\ &\quad \langle t \oplus v \oplus n|\overline{j'}\rangle\langle\overline{i'}|t \oplus v\rangle (-1)^{t\cdot l}. \end{aligned}$$

Since we are treating mutually unbiased case,

$$\langle i|\overline{k}\rangle = \sqrt{\frac{1}{2^N}}(-1)^{i\cdot k}$$

holds, where $i \cdot k := \sum_{n=1}^{N} i_n k_n$. It leads

$$\begin{aligned}\lambda_l &= \left(\frac{1}{2^N}\right)^4 \sum \delta_{i\oplus j\oplus i'\oplus j',0}\delta_{j\oplus j',0}\delta_{j'\oplus i'\oplus l,0}\langle \overline{E}_{ij}|\overline{E}_{i'j'}\rangle \\ &= \frac{1}{2^N}\sum_i \langle \overline{E}_{i\ i\oplus l}|\overline{E}_{i\ i\oplus l}\rangle\end{aligned}$$

which is nothing but $p(B = A \oplus l|\overline{b})$ introduced in (3). Finally we obtain the following inequality,

$$I(A:E|b) \leq H(A \oplus B|\overline{b}).$$

Q.E.D.

4 Discussions

Below we discuss the implication of our theorem by comparing it with the former one. Since the right hand side of our inequality is determined by $\{p(B = A \oplus c|\overline{b})\}$ it can be reduced to include only the term $\sum_{c\neq 0} p(B = A \oplus c|\overline{b})$.

Corollary 2 *The following inequality between the information gain by Eve and the error probability in Bob's outcome holds:*

$$I(A:E|b) \leq -\delta\log\delta - (1-\delta)\log(1-\delta) + N\delta,$$

where $\delta := \sum_{c\neq 0} p(B = A \oplus c|\overline{b})$.

Proof:
Under the constraint $\delta = \sum_{c\neq 0} p(B = A\oplus c|\overline{b})$ for fixed δ, the distribution which makes the Shannon entropy $H(A\oplus B|\overline{b})$ maximum is $p(B = A|\overline{b}) = 1-\delta$ and $p(B = A \oplus c|\overline{b}) = \frac{\delta}{2^N-1}$ for all $c \neq 0$. It gives

$$H(A \oplus B|\overline{b}) = -\delta\log\delta - (1-\delta)\log\delta + \delta\log(2^N - 1)$$

and ends the proof. Q.E.D.

For a fixed error probability $\delta = \sum_{c\neq 0} p(B = A \oplus c|\overline{b})$, for sufficiently large N, the term $N\delta$ becomes dominant in the right hand side of the above equation. Thus our inequality becomes tighter than (4) in such a case.

Finally we present a situation which shows a drastic difference between the two inequalities. Suppose that Eve employs the following "attack": Eve does not make the qubits sent by Alice interact with any apparatus, but she just converts the each value. That is, for each qubit, Eve performs a unitary operation $|i\rangle \mapsto (-1)^i|i\rangle$ $(i = 0, 1)$. One can easily see that also for the conjugate basis this operation works as conversion. In this case the error probability δ becomes 1 and thus if they employ the inequality (4) Alice and Bob cannot rule out the possibility of Bob's information gain. On the other hand, since the error in Bob's outcome is deterministic, the right hand side of (6) vanishes. Thus Alice and Bob can be convinced that there is no information gain by Eve.

In this paper we showed a novel version of information-disturbance theorems. According to our theorem, one can see that the information gain by Eve induces randomness to Bob's outcome in the conjugate basis. The both sides of the inequality are expressed in terms of entropy and thus seems to be natural. For large N case, in which we are usually interested in, our inequality gives tighter bound than the previously proposed ones. Moreover, our theorem can rule out the cases when Eve just turns over the qubits and gains no information. Our theorem, as previous one, also relies upon the assumptions of fair probability of the random variable A and mutually unbiasedness between b and $\overline{b}$. It will be very interesting and crucial to generalize the theorem to more general setting.

References

[1] C. H. Bennett and G. Brassard. Quantum Cryptography: Public Key Distribution and Coin Tossing. In *Proc. of IEEE Int. Conf. on Computers, Systems and Signal Processing*, pages 175–179, 1984.

[2] D. Mayers. Quantum key distribution and string oblivious transfer in noisy channel. In *Advances in cryptology - CRYPTO'96*, LNCS 1109, pages 343–357, 1996.

[3] H-K. Lo and H-F. Chau. Unconditional security of quantum key distribution over arbitrary long distances. *Science*, 283, pages 2050–2056, 1999.

[4] E. Biham, M. Boyer, P. O. Boykin, T. Mor, and V. Roychowdhury. A proof of the security of quantum key distribution. in *Proc. of the 32nd Annual ACM Symposium on Theory of Computing*, pages 715–724, 2000.

[5] P. W. Shor and J. Preskill. Simple proof of security of the BB84 quantum key distribution protocol. *Phys.Rev.Lett.*, 85, pages 441–444, 2000.

[6] P. O. Boykin and V. P. Roychowdhury. Information vs. disturbance in dimension D. quant-ph/0412028.

[7] A.S. Holevo, Problemy Peredachi Informacii, **9** (1973) pp.3-11.

[8] W. Thirring, *Quantum Mathematical Physics*, Springer-Verlag, (1983).