PHYSICAL REVIEW A **93**, 022104 (2016)

Information, fidelity, and reversibility in general quantum measurements

Hiroaki Terashima
Department of Physics, Faculty of Education, Gunma University, Maebashi, Gunma 371-8510, Japan
(Received 18 November 2015; published 3 February 2016)

We present the amount of information, fidelity, and reversibility obtained by arbitrary quantum measurements on completely unknown states. These quantities are expressed as functions of the singular values of a measurement operator corresponding to the obtained outcome. As an example, we consider a class of quantum measurements with highly degenerate singular values to discuss trade-offs among information, fidelity, and reversibility. The trade-offs are at the level of a single outcome, in the sense that the quantities pertain to each single outcome rather than the average over all possible outcomes.

DOI: 10.1103/PhysRevA.93.022104

I. INTRODUCTION

In quantum theory, information about a physical system cannot be obtained without affecting it because quantum measurement inevitably changes the state of the system via nonunitary state reduction. This property of quantum measurement is profoundly interesting for the foundations of quantum mechanics and is of practical importance in quantum information processing and communication [1], such as in quantum cryptography [2–5]. Therefore, the subject of a trade-off between information gain and state change has been discussed by many authors [6–18] over several years using various formulations. For example, Banaszek [7] showed an inequality between two fidelities quantifying information gain and state change, and Ozawa [12] generalized Heisenberg's uncertainty relation for noise and disturbance in quantum measurements.

On the other hand, state change due to quantum measurement has been shown not to be necessarily irreversible [19–21] if the measurement preserves all the information about the system, although it was once widely believed to be irreversible such that one could not recover the premeasurement state from the postmeasurement state [22]. In fact, in a physically reversible measurement [20,21], the premeasurement state can be recovered from the postmeasurement state with a nonzero probability of success via a second measurement, called a reversing measurement. Reversible measurements have been proposed for various physical systems [23–29] and have been experimentally demonstrated by using a superconducting phase qubit [30] and a photonic qubit [31].

Thus, it is natural to discuss not only the size of the state change but also its reversibility while considering the costs of information gain. Intuitively, as measurements provide more information about a system, one would expect that more information would result in more change of a system's state along with reduced reversibility. Moreover, whenever the reversing measurement recovers the premeasurement state of the first measurement, it erases all the information obtained by the first measurement (see the Erratum of Ref. [24]). In a different type of reversible measurement, known as unitarily reversible measurement [32,33], the premeasurement state can be recovered from the postmeasurement one with unit probability via a unitary operation although the measurement provides no information about the system. Therefore, there are some trade-offs among information gain, state change, and physical reversibility in quantum measurement.

Such trade-offs have been studied in photodetection processes [34] and in single-qubit measurements [35]. These trade-offs are at the level of a single outcome, in contrast to conventional ones [6,7,9,10,14,16]; that is to say that the quantities affected are those pertaining to each single outcome, rather than those averaged over all possible outcomes. This characteristic is desirable for studying state recovery with information erasure in a physically reversible measurement, because it occurs not on average but only when the reversing measurement yields a preferred single outcome. On the other hand, using quantities averaged over outcomes, Cheong and Lee [36] demonstrated that a trade-off exists between information gain and physical reversibility, which has been experimentally verified [37,38] using single photons.

In this paper, we present the general formulas for information gain, state change, and physical reversibility for an arbitrary quantum measurement on a d-level system in a completely unknown state. These formulas are more general versions of those for an arbitrary quantum measurement on a two-level system [35] and those for a projective measurement on a d-level system [39]. We present the evaluation of the amount of information gain by the decrease in Shannon entropy [11,40], the degree of state change by the fidelity [41], and the degree of physical reversibility by the maximum successful probability of the reversing measurement [42]. The formulas are written by using the singular values of a measurement operator corresponding to the outcome of the measurement. Unfortunately, when some singular values are degenerate, the formula for information gain is not useful for numerical calculations due to apparent divergences. Therefore, for the information gain, we show another formula that is free from apparent divergences, even when the singular values are degenerate.

The rest of this paper is organized as follows: Section II explains the procedure for quantifying information gain, state change, and physical reversibility and shows their explicit formulas. Section III deals with the degeneracy of singular values. Section IV considers a class of quantum measurements with highly degenerate singular values and discusses the trade-offs among information gain, state change, and physical reversibility. Section V summarizes our results.

II. FORMULATION

A. Information gain

We first consider the amount of information provided by a quantum measurement. To evaluate this amount, it is first assumed that the premeasurement state of a system to be measured is known to be one of a set of predefined pure states $\{|\psi(a)\rangle\}, a = 1,\ldots,N$, each of which has an equal probability of $p(a) = 1/N$, although the index a of the premeasurement state is unknown. The lack of information about the state is then given by

$$H_0 = -\sum_a p(a)\log_2 p(a) = \log_2 N \tag{1}$$

prior to measurement, where the Shannon entropy has been used as a measure of uncertainty rather than the von Neumann entropy of the mixed state $\hat{\rho} = \sum_a p(a)|\psi(a)\rangle\langle\psi(a)|$ because the uncertain information is the classical variable a rather than the predefined quantum state $|\psi(a)\rangle$. Each state $|\psi(a)\rangle$ can be expanded in an orthonormal basis $\{|i\rangle\}$ as

$$|\psi(a)\rangle = \sum_i c_i(a)|i\rangle, \tag{2}$$

where $i = 1,2,\ldots,d$, and d is the dimension of the Hilbert space associated with the system. For the state to be normalized, the coefficients $\{c_i(a)\}$ must satisfy the normalization condition

$$\sum_i |c_i(a)|^2 = 1. \tag{3}$$

Since, in quantum measurements, the system to be measured is usually in a completely unknown state, the predefined states $\{|\psi(a)\rangle\}$ are assumed to be all of the possible pure states of the system with $N \to \infty$.

A quantum measurement of the system can then be made to obtain information about the state. In general, a quantum measurement is described by a set of measurement operators $\{\hat{M}_m\}$ [1,43] that satisfy

$$\sum_m \hat{M}_m^\dagger \hat{M}_m = \hat{I}, \tag{4}$$

where m denotes the outcome of the measurement and $\hat{I}$ is the identity operator. When the system is in a state $|\psi\rangle$, the measurement $\{\hat{M}_m\}$ yields an outcome m with probability

$$p_m = \langle\psi|\hat{M}_m^\dagger \hat{M}_m|\psi\rangle, \tag{5}$$

changing the state into

$$|\psi_m\rangle = \frac{1}{\sqrt{p_m}}\hat{M}_m|\psi\rangle. \tag{6}$$

Here it has been assumed that the quantum measurement is efficient [8] or ideal [33] in the sense that the postmeasurement state is pure if the premeasurement state is pure, in order to focus on the quantum nature of measurement by ignoring classical noise. Each measurement operator $\hat{M}_m$ can be decomposed by singular-value decomposition as

$$\hat{M}_m = \hat{U}_m \hat{D}_m \hat{V}_m, \tag{7}$$

where $\hat{U}_m$ and $\hat{V}_m$ are unitary operators, and $\hat{D}_m$ is a diagonal operator in the orthonormal basis $\{|i\rangle\}$:

$$\hat{D}_m = \sum_i \lambda_{mi}|i\rangle\langle i|. \tag{8}$$

The diagonal elements $\{\lambda_{mi}\}$, called the singular values of $\hat{M}_m$, are not less than 0 by definition and are not greater than 1 on the basis of Eq. (4); that is,

$$0 \leqslant \lambda_{mi} \leqslant 1 \tag{9}$$

for $i = 1,2,\ldots,d$. In this situation, where the measurement is performed on one of all possible pure states $\{|\psi(a)\rangle\}$, the unitary operator $\hat{V}_m$ can be removed from the measurement operator given in Eq. (7) as

$$\hat{M}_m = \hat{U}_m \hat{D}_m \tag{10}$$

by relabeling the index a as $|\psi'(a)\rangle = \hat{V}_m|\psi(a)\rangle$ without loss of generality. Furthermore, the unitary operator $\hat{U}_m$ is irrelevant to information gain, since the probability given by Eq. (5) is unaffected by $\hat{U}_m$. Although it changes the state of the system as in Eq. (6), the state change caused by $\hat{U}_m$ can be recovered with unit probability and no information loss after the measurement by applying $\hat{U}_m^\dagger$ to the system. Thus, to see the inevitable state change and irreversibility caused by the extraction of information, it suffices to set the measurement operator of Eq. (7) equal to

$$\hat{M}_m = \hat{D}_m. \tag{11}$$

By substituting Eqs. (2) and (11) into Eq. (5), it is evident that the measurement yields outcome m with probability

$$p(m|a) = \sum_i \lambda_{mi}^2 |c_i(a)|^2 \equiv q_m(a) \tag{12}$$

when the premeasurement state of the system is $|\psi(a)\rangle$. Since the probability of $|\psi(a)\rangle$ is $p(a) = 1/N$, the total probability of the outcome m is given by

$$p(m) = \sum_a p(m|a)p(a) = \frac{1}{N}\sum_a q_m(a) = \overline{q_m}, \tag{13}$$

where the overline denotes the average over a:

$$\overline{f} \equiv \frac{1}{N}\sum_a f(a). \tag{14}$$

On the contrary, given the outcome m, the probability of the premeasurement state $|\psi(a)\rangle$ can be calculated to be

$$p(a|m) = \frac{p(m|a)p(a)}{p(m)} = \frac{q_m(a)}{N\overline{q_m}} \tag{15}$$

according to Bayes' rule. Therefore, after the measurement yields the outcome m, the lack of information about the premeasurement state decreases to the Shannon entropy

$$H(m) = -\sum_a p(a|m)\log_2 p(a|m). \tag{16}$$

Using this decrease in Shannon entropy [11,40], the information provided by the measurement with the outcome m can be expressed as

$$I(m) \equiv H_0 - H(m) = \frac{\overline{q_m \log_2 q_m} - \overline{q_m} \log_2 \overline{q_m}}{\overline{q_m}}, \tag{17}$$

which is always positive and evidently free from the divergent term $\log_2 N \to \infty$ in Eq. (1), due to the assumption that $p(a)$ is uniform. This quantity can be viewed as the relative entropy (or the Kullback–Leibler divergence) [1] of $p(a|m)$ to the uniform distribution $p(a) = 1/N$,

$$I(m) = \sum_a p(a|m) \log_2 \frac{p(a|m)}{p(a)}. \tag{18}$$

To explicitly calculate the information in Eq. (17), it is necessary to average $q_m(a)$ and $q_m(a)\log_2 q_m(a)$ over all possible pure states of the system, $\{|\psi(a)\rangle\}$. As shown in Appendix A, a straightforward calculation gives

$$\overline{q_m} = \frac{1}{d}\sigma_m^2, \tag{19}$$

where σ_m is the Hilbert–Schmidt norm of $\hat{M}_m$,

$$\sigma_m = \sqrt{\mathrm{Tr}(\hat{M}_m^\dagger \hat{M}_m)} = \sqrt{\sum_i \lambda_{mi}^2}. \tag{20}$$

On the other hand, it would be difficult to directly calculate the average of $q_m(a)\log_2 q_m(a)$ by using the method described in Appendix A. However, in different contexts, similar calculations have been performed in various ways [44–46]. By applying the integral formula derived in Ref. [45] to this case, the following expression can be obtained:

$$\overline{q_m \log_2 q_m} = \frac{1}{d}\sum_i \frac{\lambda_{mi}^{2d} \log_2 \lambda_{mi}^2}{\prod_{k\neq i}\left(\lambda_{mi}^2 - \lambda_{mk}^2\right)} - \frac{1}{d \ln 2}[\eta(d) - 1]\sigma_m^2, \tag{21}$$

where $\eta(n)$ is defined by

$$\eta(n) \equiv \sum_{k=1}^{n} \frac{1}{k} = 1 + \frac{1}{2} + \cdots + \frac{1}{n}. \tag{22}$$

Note that in order to obtain the form of Eq. (21) from the integral formula, it is necessary to use the identity

$$\sum_i \frac{\lambda_{mi}^{2d}}{\prod_{k\neq i}\left(\lambda_{mi}^2 - \lambda_{mk}^2\right)} = \sigma_m^2 \tag{23}$$

and the recurrence formula of the digamma function $\psi(z)$, $\psi(z+1) = \psi(z) + 1/z$. By substituting Eqs. (19) and (21) into Eq. (17), the information can finally be expressed as

$$I(m) = \log_2 d - \frac{1}{\ln 2}[\eta(d) - 1] - \log_2 \sigma_m^2 + \frac{1}{\sigma_m^2}\sum_i \frac{\lambda_{mi}^{2d} \log_2 \lambda_{mi}^2}{\prod_{k\neq i}\left(\lambda_{mi}^2 - \lambda_{mk}^2\right)}. \tag{24}$$

This function is invariant under the interchange of any pair of singular values,

$$\lambda_{mi} \longleftrightarrow \lambda_{mj} \quad \text{for any } (i,j), \tag{25}$$

as well as under the rescaling of all singular values by a constant factor c,

$$(\lambda_{m1}, \lambda_{m2}, \ldots, \lambda_{md}) \to (c\lambda_{m1}, c\lambda_{m2}, \ldots, c\lambda_{md}), \tag{26}$$

because of Eq. (23). If the singular values are normalized by the rescaling factor of Eq. (26) to $\sigma_m^2 = 1$, the $\{\lambda_{mi}\}$-dependent part of Eq. (24),

$$Q = \log_2 \sigma_m^2 - \frac{1}{\sigma_m^2}\sum_i \frac{\lambda_{mi}^{2d} \log_2 \lambda_{mi}^2}{\prod_{k\neq i}\left(\lambda_{mi}^2 - \lambda_{mk}^2\right)}, \tag{27}$$

resembles the subentropy discussed in Ref. [46]. However, these quantities have different meanings, since the subentropy is a function of the eigenvalues of the premeasurement density operator $\hat{\rho} = \sum_a p(a)|\psi(a)\rangle\langle\psi(a)|$, rather than a function of the singular values of the measurement operator $\hat{M}_m$. For fixed d, Eq. (27) satisfies the inequality [46]

$$0 \leqslant Q \leqslant \log_2 d - \frac{1}{\ln 2}[\eta(d) - 1]. \tag{28}$$

The lower bound is achieved when only one singular value is nonzero, as in the projective measurement of rank 1, whereas the upper bound is achieved when all singular values are equal, as in the identity operation.

The information in Eq. (17) is at the level of a single outcome in the sense that it has its value when a single outcome m has been obtained. If $I(m)$ is averaged over all outcomes with probabilities given by Eq. (13), the mutual information [1] of the random variables $\{a\}$ and $\{m\}$ is obtained:

$$I \equiv \sum_m p(m) I(m) = \sum_{m,a} p(m,a) \log_2 \frac{p(m,a)}{p(m)p(a)}, \tag{29}$$

where $p(m,a) = p(m|a)p(a)$. However, this is the amount of information that is expected to be obtained on average before the measurement, rather than the actual information $I(m)$. While the average information expressed by Eq. (29) is not discussed further in this paper, the explicit form of I is presented herein, since it cannot be found in the literature. It becomes

$$I = \log_2 d - \frac{1}{\ln 2}[\eta(d) - 1] - \frac{1}{d}\sum_m \left[\sigma_m^2 \log_2 \sigma_m^2 - \sum_i \frac{\lambda_{mi}^{2d} \log_2 \lambda_{mi}^2}{\prod_{k\neq i}\left(\lambda_{mi}^2 - \lambda_{mk}^2\right)}\right] \tag{30}$$

from Eqs. (13), (19), and (24), with an identity resulting from the trace of Eq. (4),

$$\sum_m \sigma_m^2 = d. \tag{31}$$

B. State change

Now the degree of state change caused by the measurement as a cost of the information gain is considered. When the premeasurement state of the system is $|\psi(a)\rangle$, a measurement with outcome m changes it to

$$|\psi(m,a)\rangle = \frac{1}{\sqrt{q_m(a)}}\hat{D}_m|\psi(a)\rangle \tag{32}$$

according to Eq. (6) with Eqs. (11) and (12). This state change can be evaluated using the fidelity [1,41] as

$$F(m,a) = |\langle\psi(a)|\psi(m,a)\rangle| = \frac{1}{\sqrt{q_m(a)}}\sum_i \lambda_{mi}|c_i(a)|^2 \equiv \frac{f_m(a)}{\sqrt{q_m(a)}}, \quad (33)$$

which decreases as the measurement changes the state of the system by a greater extent. By averaging over the premeasurement states $\{|\psi(a)\rangle\}$ with probabilities given by Eq. (15), the fidelity after the measurement with the outcome m can be expressed as

$$F(m) = \sum_a p(a|m)[F(m,a)]^2 = \frac{\overline{f_m^2}}{\overline{q_m}}, \quad (34)$$

where the squared fidelity, rather than the fidelity, has been averaged for simplicity.

To explicitly calculate the fidelity in Eq. (34), it is necessary to average $[f_m(a)]^2$ over all possible pure states of the system, $\{|\psi(a)\rangle\}$. As shown in Appendix A, the average is given by

$$\overline{f_m^2} = \frac{1}{d(d+1)}\left(\sigma_m^2 + \tau_m^2\right), \quad (35)$$

where τ_m is the trace norm of $\hat{M}_m$:

$$\tau_m = \mathrm{Tr}\sqrt{\hat{M}_m^\dagger \hat{M}_m} = \sum_i \lambda_{mi}. \quad (36)$$

By substituting Eqs. (19) and (35) into Eq. (34), the fidelity can be obtained as follows:

$$F(m) = \frac{1}{d+1}\left(\frac{\sigma_m^2 + \tau_m^2}{\sigma_m^2}\right). \quad (37)$$

This function is also invariant under the interchange of Eq. (25) and the rescaling of Eq. (26).

The fidelity in Eq. (34) is also at the level of a single outcome, in the sense that it has its value when a single outcome m has been obtained. If $F(m)$ is averaged over all outcomes with probabilities given by Eq. (13), the mean operation fidelity [7] is obtained:

$$F \equiv \sum_m p(m)F(m) = \sum_m \overline{|\langle\psi|\hat{M}_m|\psi\rangle|^2}, \quad (38)$$

whose explicit form is given by [7]

$$F = \frac{1}{d(d+1)}\left(d + \sum_m \tau_m^2\right) \quad (39)$$

from Eqs. (13), (19), and (31), although the average fidelity of Eq. (38) is not discussed further in this paper.

C. Physical reversibility

Next, the degree of reversibility of the measurement is considered. A quantum measurement is said to be physically reversible [20,21] if the premeasurement state can be recovered from the postmeasurement state with a nonzero probability of success via a reversing measurement. The necessary and sufficient condition for physical reversibility is that the measurement operator $\hat{M}_m$ has a bounded left inverse $\hat{M}_m^{-1}$. If this condition is satisfied, then the reversing measurement can be constructed by another set of measurement operators $\{\hat{R}_\mu^{(m)}\}$ that satisfy

$$\sum_\mu \hat{R}_\mu^{(m)\dagger}\hat{R}_\mu^{(m)} = \hat{I} \quad (40)$$

and, in addition, for a particular $\mu = \mu_0$,

$$\hat{R}_{\mu_0}^{(m)} = \kappa_m \hat{M}_m^{-1}, \quad (41)$$

where μ denotes the outcome of the reversing measurement and κ_m is a complex constant. When the reversing measurement $\{\hat{R}_\mu^{(m)}\}$ is performed on the postmeasurement state given in Eq. (6) and the preferred outcome μ_0 is obtained, the state of the system successfully reverts to the premeasurement state $|\psi\rangle$, except for an overall phase factor via the second state reduction,

$$|\psi_{m\mu_0}\rangle = \frac{1}{\sqrt{p_{m\mu_0}}}\hat{R}_{\mu_0}^{(m)}|\psi_m\rangle \propto |\psi\rangle, \quad (42)$$

where

$$p_{m\mu_0} = \langle\psi_m|\hat{R}_{\mu_0}^{(m)\dagger}\hat{R}_{\mu_0}^{(m)}|\psi_m\rangle = \frac{|\kappa_m|^2}{p_m} \quad (43)$$

is the probability for the second outcome μ_0 given the first outcome m and thus is the successful probability of the reversing measurement. Then, the physical reversibility can be evaluated by using the maximum successful probability of the reversing measurement [26,36,42,47]. Since the completeness condition given in Eq. (40) requires $\langle\psi|\hat{R}_{\mu_0}^{(m)\dagger}\hat{R}_{\mu_0}^{(m)}|\psi\rangle \leqslant 1$ for any $|\psi\rangle$, the upper bound for $|\kappa_m|^2$ is given by [42]

$$|\kappa_m|^2 \leqslant \inf_{|\psi\rangle}\langle\psi|\hat{M}_m^\dagger\hat{M}_m|\psi\rangle = \lambda_{m,\,\min}^2, \quad (44)$$

where $\lambda_{m,\,\min}$ is the minimum singular value of $\hat{M}_m$:

$$\lambda_{m,\,\min} \equiv \min_j \lambda_{mj}. \quad (45)$$

Therefore, the maximum successful probability of the reversing measurement is

$$\max_{\kappa_m} p_{m\mu_0} = \frac{\lambda_{m,\,\min}^2}{p_m}, \quad (46)$$

which is regarded as a measure of the physical reversibility of measurement.

In this situation, when the measurement on the premeasurement state $|\psi(a)\rangle$ yields an outcome m, the reversibility of Eq. (46) is given by

$$R(m,a) = \frac{\lambda_{m,\,\min}^2}{p(m|a)} = \frac{\lambda_{m,\,\min}^2}{q_m(a)} \quad (47)$$

on the basis of Eq. (12). By averaging over the premeasurement states $\{|\psi(a)\rangle\}$ with probabilities given by Eq. (15), the reversibility of the measurement with the outcome m can be expressed as

$$R(m) = \sum_a p(a|m)R(m,a) = d\left(\frac{\lambda_{m,\,\min}^2}{\sigma_m^2}\right) \quad (48)$$

by using Eq. (19). This function is also invariant under the interchange of Eq. (25) and the rescaling of Eq. (26). Again, this reversibility is at the level of a single outcome in the sense that it has its value when a single outcome m has been obtained. If $R(m)$ is averaged over all outcomes with probabilities given by Eq. (13), the degree of physical reversibility of a measurement that was discussed in Ref. [42] is obtained:

$$R \equiv \sum_m p(m)R(m) = \sum_m \inf_{|\psi\rangle} \langle\psi|\hat{M}_m^\dagger \hat{M}_m|\psi\rangle, \tag{49}$$

whose explicit form is given by [36]

$$R = \sum_m \lambda_{m,\,\mathrm{min}}^2 \tag{50}$$

from Eqs. (13) and (19), although the average reversibility of Eq. (49) is not discussed further in this paper.

D. State estimation

Finally, another measure of information gain called estimation fidelity is introduced to show its general formula at the level of a single outcome, although this paper will mainly use $I(m)$ given in Eq. (17). Suppose that, when the measurement yields an outcome m, the premeasurement state is estimated by a state $|\varphi(m)\rangle$. If the actual premeasurement state is $|\psi(a)\rangle$, the quality of the estimation can be evaluated by the overlap $|\langle\varphi(m)|\psi(a)\rangle|^2$. By averaging over the premeasurement states $\{|\psi(a)\rangle\}$ with probabilities given by Eq. (15), the estimation fidelity after the measurement with the outcome m can be expressed as

$$G(m) = \sum_a p(a|m) |\langle\varphi(m)|\psi(a)\rangle|^2, \tag{51}$$

which depends on the strategy of selecting $|\varphi(m)\rangle$. In the optimal case [7], the estimation $|\varphi(m)\rangle$ is assigned to the eigenvector of $\hat{M}_m^\dagger \hat{M}_m$ corresponding to its maximum eigenvalue. Since $\hat{M}_m^\dagger \hat{M}_m = \hat{D}_m^2$ from Eq. (11), $|\varphi(m)\rangle$ is one of the states in the basis $\{|i\rangle\}$; namely, $|\varphi(m)\rangle = |l\rangle$, with l being one of $1,2,\ldots,d$ that satisfies

$$\lambda_{ml} = \max_j \lambda_{mj} \equiv \lambda_{m,\,\mathrm{max}}. \tag{52}$$

Using this strategy, the estimation fidelity can be written as

$$G(m) = \frac{1}{q_m} \sum_i \lambda_{mi}^2 \frac{1}{N} \sum_a |c_i(a)|^2 |c_l(a)|^2, \tag{53}$$

which is explicitly calculated to be

$$G(m) = \frac{1}{d+1}\left(\frac{\sigma_m^2 + \lambda_{m,\,\mathrm{max}}^2}{\sigma_m^2}\right) \tag{54}$$

by using the calculations in Appendix A. This function is also invariant under the interchange of Eq. (25) and the rescaling of Eq. (26).

This estimation fidelity is at the level of a single outcome. If $G(m)$ is averaged over all outcomes with probabilities given by Eq. (13), the mean estimation fidelity [7] is obtained:

$$G \equiv \sum_m p(m)G(m) = \sum_m \overline{\langle\psi|\hat{M}_m^\dagger \hat{M}_m|\psi\rangle |\langle\varphi(m)|\psi\rangle|^2}, \tag{55}$$

whose explicit form is given by [7]

$$G = \frac{1}{d(d+1)}\left(d + \sum_m \lambda_{m,\,\mathrm{max}}^2\right) \tag{56}$$

from Eqs. (13), (19), and (31).

III. DEGENERACY

When some singular values are degenerate, Eq. (24) for information gain is not useful for numerical calculations due to the apparent divergences of

$$J \equiv \sum_i \frac{\lambda_{mi}^{2d} \log_2 \lambda_{mi}^2}{\prod_{k\neq i}\left(\lambda_{mi}^2 - \lambda_{mk}^2\right)}. \tag{57}$$

Of course, J is finite, because it arises from the integral of a bounded function over a bounded region as in Eq. (21). Even if $\lambda_{mi} = \lambda_{mk}$, a finite result can be obtained by taking the limit as $\lambda_{mi} \to \lambda_{mk}$. However, this limit operation is quite complicated if singular values are highly degenerate. Therefore, another formula will be presented for the information gain that requires no limit operations even when singular values are degenerate.

Since the ordering of singular values is insignificant due to the invariance under the interchange of Eq. (25), they can first be divided into groups on the basis of their values:

$$\{\lambda_{mi}\} \longrightarrow \{(\bar{\lambda}_{ms}, n_s)\}, \tag{58}$$

where group s contains n_s singular values of $\bar{\lambda}_{ms}$, and thus $\sum_s n_s = d$. For example, if the singular values are

$$\lambda_{m1} = \lambda_{m2} = \tfrac{1}{4}, \quad \lambda_{m3} = \lambda_{m4} = \lambda_{m5} = \tfrac{1}{2}, \quad \lambda_{m6} = \tfrac{3}{4}, \tag{59}$$

they are divided into three groups as

$$(\bar{\lambda}_{m1}, n_1) = \left(\tfrac{1}{4}, 2\right), \quad (\bar{\lambda}_{m2}, n_2) = \left(\tfrac{1}{2}, 3\right), \quad (\bar{\lambda}_{m3}, n_3) = \left(\tfrac{3}{4}, 1\right). \tag{60}$$

In accordance with this grouping, the summation over i in Eq. (57) can be expressed as a summation over the groups

$$J = \sum_s J_s, \tag{61}$$

where J_s is the sum within the sth group $(\bar{\lambda}_{ms}, n_s)$ defined as a limit of $\lambda_1, \lambda_2, \ldots, \lambda_{n_s} \to \bar{\lambda}_{ms}$:

$$J_s = \lim_{\substack{\lambda_1,\lambda_2,\ldots,\lambda_{n_s} \\ \to \bar{\lambda}_{ms}}} \sum_{i=1}^{n_s} \left(\prod_{k\neq i}^{n_s} \frac{1}{\lambda_i^2 - \lambda_k^2}\right) \frac{\lambda_i^{2d} \log_2 \lambda_i^2}{\prod_{r\neq s}\left(\lambda_i^2 - \bar{\lambda}_{mr}^2\right)^{n_r}}. \tag{62}$$

This limit can be calculated as follows: First, substitute $\bar{\lambda}_{ms}^2$ for λ_1^2 and $\bar{\lambda}_{ms}^2 + \epsilon$ for λ_2^2, and then take the limit as $\epsilon \to 0$. Next, substitute $\bar{\lambda}_{ms}^2 + \epsilon$ for λ_3^2 and take the limit as $\epsilon \to 0$. Repeat similarly one by one for $\lambda_4^2, \lambda_5^2, \ldots, \lambda_{n_s}^2$. As a consequence of these procedures, one finds that at the last step J_s should be of the form

$$J_s = \lim_{\epsilon\to 0}\left[\frac{1}{\epsilon^{n_s-1}} \frac{\left(\bar{\lambda}_{ms}^2 + \epsilon\right)^d \log_2\left(\bar{\lambda}_{ms}^2 + \epsilon\right)}{\prod_{r\neq s}\left(\bar{\lambda}_{ms}^2 + \epsilon - \bar{\lambda}_{mr}^2\right)^{n_r}} + \sum_{n=1}^{n_s-1} \frac{w_n^{(s)}}{\epsilon^n}\right], \tag{63}$$

where $\{w_n^{(s)}\}$ are finite coefficients. Therefore, by using the coefficients $\{z_n^{(s)}\}$ defined by Taylor series

$$\frac{\left(\bar{\lambda}_{ms}^2+\epsilon\right)^d \log_2\left(\bar{\lambda}_{ms}^2+\epsilon\right)}{\prod_{r\neq s}\left(\bar{\lambda}_{ms}^2+\epsilon-\bar{\lambda}_{mr}^2\right)^{n_r}} \equiv \sum_{n=0}^{\infty} z_n^{(s)}\epsilon^n, \tag{64}$$

J_s can be written with no limit operations as

$$J_s = z_{n_s-1}^{(s)}. \tag{65}$$

Note that, when Eq. (64) is substituted into Eq. (63), the divergent terms containing $\{w_n^{(s)}\}$ with $n = 1,2,\ldots,n_s-1$ should be canceled by the divergent terms containing $\{z_n^{(s)}\}$ with $n = n_s-2,\ n_s-3,\ldots,0$, since J_s is finite.

A more explicit form of J_s can be found by separating the left-hand side of Eq. (64) into two parts that can then be expanded as Taylor series. The first part is

$$(\lambda^2+\epsilon)^d \log_2(\lambda^2+\epsilon) \equiv \sum_{n=0}^{d-1} c_n^{(d)}(\lambda)\epsilon^n + O(\epsilon^d), \tag{66}$$

which corresponds to the numerator of Eq. (64). As shown in Appendix B, the coefficients $\{c_n^{(d)}(\lambda)\}$ for $n = 0,1,\ldots,d-1$ are given by

$$c_n^{(d)}(\lambda) = \lambda^{2(d-n)}\left[\binom{d}{n}\log_2\lambda^2 + a_n^{(d)}\right], \tag{67}$$

where the coefficients $\{a_n^{(d)}\}$ are

$$a_n^{(d)} = \frac{1}{\ln 2}\binom{d}{n}[\eta(d)-\eta(d-n)]. \tag{68}$$

The explicit forms of $\{a_n^{(d)}\}$ are

$$a_0^{(d)} = 0,\quad a_1^{(d)} = \frac{1}{\ln 2},\quad a_2^{(d)} = \frac{1}{\ln 2}\left(d-\frac{1}{2}\right),\quad \ldots,$$
$$a_{d-1}^{(d)} = \frac{d}{\ln 2}\left(\frac{1}{2}+\cdots+\frac{1}{d}\right). \tag{69}$$

It is clear that

$$c_n^{(d)}(0) = 0,\quad c_n^{(d)}(1) = a_n^{(d)}. \tag{70}$$

On the other hand, the second part is

$$\frac{1}{\prod_{r\neq s}\left(\bar{\lambda}_{ms}^2+\epsilon-\bar{\lambda}_{mr}^2\right)^{n_r}} \equiv \frac{1}{\prod_{r\neq s}\left(\bar{\lambda}_{ms}^2-\bar{\lambda}_{mr}^2\right)^{n_r}}\sum_{n=0}^{\infty} b_n^{(s)}\epsilon^n. \tag{71}$$

The coefficients $\{b_n^{(s)}\}$ are complicated in general, but they can be described in a compact form with the help of complete Bell polynomials:

$$B_n(x_1,x_2,\ldots,x_n) = \sum_{\{j_r\}} \frac{n!}{j_1!j_2!\cdots j_n!}\left(\frac{x_1}{1!}\right)^{j_1}\left(\frac{x_2}{2!}\right)^{j_2}\cdots\left(\frac{x_n}{n!}\right)^{j_n}, \tag{72}$$

where the summation is taken over all possible sets of non-negative integers $\{j_r\}$ such that

$$\sum_{r=1}^{n} r j_r = n. \tag{73}$$

The explicit forms for $n = 0$, 1, 2, and 3 are

$$B_0 = 1,$$
$$B_1(x_1) = x_1,$$
$$B_2(x_1,x_2) = x_1^2 + x_2,$$
$$B_3(x_1,x_2,x_3) = x_1^3 + 3x_1x_2 + x_3. \tag{74}$$

With these complete Bell polynomials, the coefficients $\{b_n^{(s)}\}$ are given by

$$b_n^{(s)} = \frac{1}{n!}B_n\left(h_1^{(s)},h_2^{(s)},\ldots,h_n^{(s)}\right), \tag{75}$$

where the coefficients $\{h_n^{(s)}\}$ are

$$h_n^{(s)} = (-1)^n(n-1)!\sum_{r\neq s}\frac{n_r}{\left(\bar{\lambda}_{ms}^2-\bar{\lambda}_{mr}^2\right)^n}, \tag{76}$$

as shown in Appendix B. By substituting the Taylor series of Eqs. (66) and (71) into Eq. (64), Eq. (65) can be expressed as

$$J_s = \frac{1}{\prod_{r\neq s}\left(\bar{\lambda}_{ms}^2-\bar{\lambda}_{mr}^2\right)^{n_r}}\sum_{n=0}^{n_s-1} c_n^{(d)}(\bar{\lambda}_{ms})b_{n_s-1-n}^{(s)}. \tag{77}$$

Performing the summation of Eq. (61) over all groups then yields

$$J = \sum_s \frac{1}{\prod_{r\neq s}\left(\bar{\lambda}_{ms}^2-\bar{\lambda}_{mr}^2\right)^{n_r}}\sum_{n=0}^{n_s-1} c_n^{(d)}(\bar{\lambda}_{ms})b_{n_s-1-n}^{(s)}. \tag{78}$$

Since $\bar{\lambda}_{ms} \neq \bar{\lambda}_{mr}$ if $s \neq r$ due to the grouping of Eq. (58), this expression is clearly free from apparent divergences, thus eliminating the need for limit operations even when the singular values are degenerate. In particular, Eq. (78) is more useful than Eq. (57) for numerical calculations, by which the author has verified the consistency of Eq. (24) with Eq. (17) by using the Monte Carlo method for integration.

To outline the calculation of Eq. (78), a simple case is presented wherein the singular values in $d = 6$ are divided into three groups:

$$(\bar{\lambda}_{m1},n_1) = (\lambda,3),\quad (\bar{\lambda}_{m2},n_2) = (\sqrt{2}\lambda,2),$$
$$(\bar{\lambda}_{m3},n_3) = (\sqrt{3}\lambda,1). \tag{79}$$

The first group $s = 1$ can be used to obtain J_1. Since $n_1 = 3$, it is necessary to calculate $b_0^{(1)}$, $b_1^{(1)}$, and $b_2^{(1)}$ from Eq. (77), which themselves require $h_1^{(1)}$ and $h_2^{(1)}$, as in Eq. (75). According to Eq. (76),

$$h_1^{(1)} = \frac{5}{2\lambda^2},\quad h_2^{(1)} = \frac{9}{4\lambda^4}, \tag{80}$$

which gives

$$b_0^{(1)} = 1,\quad b_1^{(1)} = \frac{5}{2\lambda^2},\quad b_2^{(1)} = \frac{17}{4\lambda^4}. \tag{81}$$

By combining these coefficients with $c_0^{(6)}(\lambda)$, $c_1^{(6)}(\lambda)$, and $c_2^{(6)}(\lambda)$, the following equation can be obtained:

$$J_1 = -\lambda^2\left(\frac{137}{8}\log_2\lambda^2 + \frac{4}{\ln 2}\right). \tag{82}$$

Similar calculations should be done for the second and third groups, $s = 2$ and $s = 3$, to obtain J_2 and J_3. Then, J can be obtained by adding J_1, J_2, and J_3, although the result is omitted here.

IV. EXAMPLE

As an example, a class of quantum measurements with highly degenerate singular values is considered next to discuss trade-offs among the information, fidelity, and reversibility that are given by Eqs. (24), (37), and (48), respectively. The measurement considered here is described by a measurement operator whose singular values are

$$\begin{aligned} \lambda_{m1} &= \lambda_{m2} = \cdots = \lambda_{mk} = 1, \\ \lambda_{m(k+1)} &= \lambda_{m(k+2)} = \cdots = \lambda_{m(k+l)} = \lambda, \\ \lambda_{m(k+l+1)} &= \lambda_{m(k+l+2)} = \cdots = \lambda_{md} = 0, \end{aligned} \tag{83}$$

when it yields an outcome m. The singular values are sorted in descending order by the interchange of Eq. (25), and the maximum singular values are normalized to 1 by the rescaling of Eq. (26). Note that, if $k = 0$, $l = 0$, $\lambda = 0$, or $\lambda = 1$, this measurement becomes a projective measurement, as was discussed in Ref. [39]. Therefore, it is assumed that

$$k = 1,2,\dots,d-1, \quad l = 1,2,\dots,d-k, \quad 0 < \lambda < 1. \tag{84}$$

First, the calculation of the information given by Eq. (24) is presented with dividing the singular values into groups as in Eq. (58) to handle their degeneracies:

$$\begin{aligned} (\bar{\lambda}_{m1}, n_1) &= (1,k), \quad (\bar{\lambda}_{m2}, n_2) = (\lambda, l), \\ (\bar{\lambda}_{m3}, n_3) &= (0, d-k-l). \end{aligned} \tag{85}$$

In this case, Eq. (65) should be used rather than Eq. (77) to calculate the dangerous term of Eq. (57), because it is easy to expand the left-hand side of Eq. (64) as a Taylor series. In fact, for the first group $s = 1$, Eq. (64) becomes

$$\frac{(1+\epsilon)^{k+l}\log_2(1+\epsilon)}{(1+\epsilon-\lambda^2)^l} = \sum_{n=0}^{\infty} z_n^{(1)}\epsilon^n. \tag{86}$$

The numerator is expanded as in Eq. (66), with coefficients $c_n^{(k+l)}(1) = a_n^{(k+l)}$, while the remaining part can be expanded by the generalized binomial theorem as

$$\frac{1}{(1+\epsilon-\lambda^2)^l} = (-1)^l \sum_{n=0}^{\infty} \binom{l+n-1}{l-1} \frac{1}{(\lambda^2-1)^{l+n}}\epsilon^n. \tag{87}$$

Using these Taylor series, for the first group $s = 1$, J_1 can be found to be

$$J_1 = (-1)^l \sum_{n=0}^{k-1} \binom{k+l-n-2}{l-1} \frac{a_n^{(k+l)}}{(\lambda^2-1)^{k+l-n-1}} \tag{88}$$

and, similarly, for the second group $s = 2$,

$$J_2 = (-1)^k \sum_{n=0}^{l-1} \binom{k+l-n-2}{k-1} \frac{c_n^{(k+l)}(\lambda)}{(1-\lambda^2)^{k+l-n-1}}. \tag{89}$$

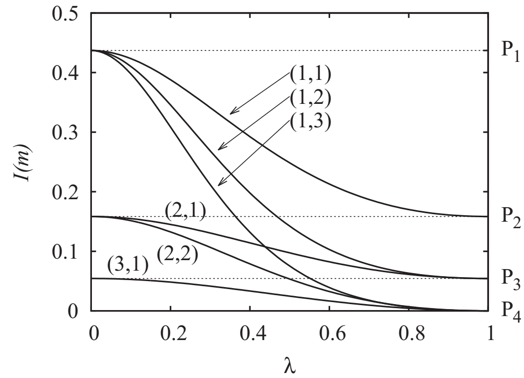

FIG. 1. Information $I(m)$ as a function of singular value λ in $d = 4$ for $(k,l) = (1,1)$, $(1,2)$, $(1,3)$, $(2,1)$, $(2,2)$, and $(3,1)$. The symbols $\{P_r\}(r = 1,2,3,4)$ denote projective measurements of rank r.

On the other hand, for the third group $s = 3$, $J_3 = 0$ can be obtained from Eq. (77) because $c_n^{(d)}(0) = 0$. The dangerous term in Eq. (57) is then given by $J = J_1 + J_2$, as in Eq. (61). From the resultant J, the information of Eq. (24) can be calculated as

$$\begin{aligned} I(m) &= \log_2 d - \frac{1}{\ln 2}[\eta(d) - 1] - \log_2(k + l\lambda^2) \\ &+ \frac{1}{k+l\lambda^2}\Bigg[(-1)^l \sum_{n=0}^{k-1} \binom{k+l-n-2}{l-1} \frac{a_n^{(k+l)}}{(\lambda^2-1)^{k+l-n-1}} \\ &+ (-1)^k \sum_{n=0}^{l-1} \binom{k+l-n-2}{k-1} \frac{c_n^{(k+l)}(\lambda)}{(1-\lambda^2)^{k+l-n-1}}\Bigg], \end{aligned} \tag{90}$$

since the Hilbert–Schmidt norm of Eq. (20) is $\sigma_m^2 = k + l\lambda^2$ in this case. Figure 1 shows this information $I(m)$ as a function of λ in $d = 4$ for various (k,l). In the figure, the symbols $\{P_r\}(r = 1,2,3,4)$ denote projective measurements of rank r, even though P_4 in $d = 4$ is nothing more than the identity operation. The information for P_r is given by [39]

$$I(m) = \log_2 \frac{d}{r} - \frac{1}{\ln 2}[\eta(d) - \eta(r)]. \tag{91}$$

As shown in Fig. 1, the information of Eq. (90) for (k,l) is equal to that for P_k when $\lambda = 0$ and is equal to that for P_{k+l} when $\lambda = 1$, as expected; these facts can be confirmed mathematically from Eq. (90) as shown in Appendix C. The estimation fidelity $G(m)$ given in Eq. (54) also changes in a way similar to $I(m)$ between 1/4 and 2/5.

At the same time, the fidelity of Eq. (37) and reversibility of Eq. (48) can be calculated to be

$$F(m) = \frac{1}{d+1}\left[\frac{k(k+1) + 2kl\lambda + l(l+1)\lambda^2}{k + l\lambda^2}\right] \tag{92}$$

and

$$R(m) = d\left(\frac{\lambda^2}{k + l\lambda^2}\right)\delta_{d,(k+l)}, \tag{93}$$

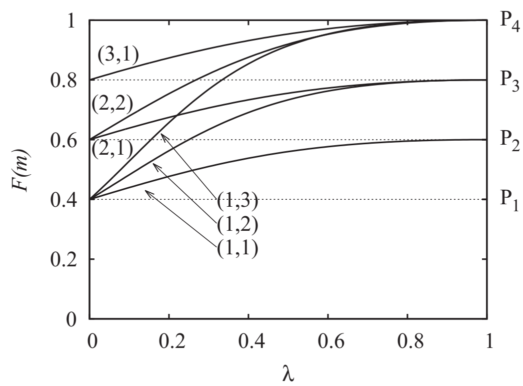

FIG. 2. Fidelity $F(m)$ as a function of singular value λ in $d = 4$ for $(k,l) = (1,1)$, $(1,2)$, $(1,3)$, $(2,1)$, $(2,2)$, and $(3,1)$. The symbols $\{\mathrm{P}_r\}(r = 1,2,3,4)$ denote projective measurements of rank r.

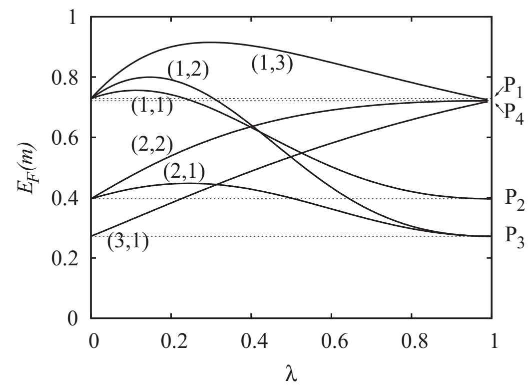

FIG. 4. Efficiency with respect to fidelity, $E_F(m)$, as a function of singular value λ in $d = 4$ for $(k,l) = (1,1)$, $(1,2)$, $(1,3)$, $(2,1)$, $(2,2)$, and $(3,1)$. The symbols $\{\mathrm{P}_r\}(r = 1,2,3,4)$ denote projective measurements of rank r.

respectively, since the trace norm of Eq. (36) is $\tau_m = k + l\lambda$ and the minimum singular value of Eq. (45) is $\lambda_{m,\min} = \lambda\delta_{d,(k+l)}$. Figures 2 and 3 show this fidelity $F(m)$ and reversibility $R(m)$ as functions of λ in $d = 4$ for various (k,l), while those for P_r are given by [39]

$$F(m) = \frac{r+1}{d+1} \tag{94}$$

and

$$R(m) = \delta_{d,r}. \tag{95}$$

The reversibility of Eq. (93) is 0 for each of $(k,l) = (1,1)$, $(1,2)$, and $(2,1)$ since $\lambda_{m,\min} = 0$, as shown in Fig. 3.

Now the trade-offs among information gain, state change, and physical reversibility can be discussed for this class of measurements, since the three quantities have been expressed as functions of the same single parameter λ. As the parameter λ increases, the information of Eq. (90) monotonically decreases, as in Fig. 1, whereas the fidelity of Eq. (92) and reversibility of Eq. (93) monotonically increase, as in Figs. 2 and 3. Thus, as a measurement provides more information about the state of the system, it changes the state less reversibly and to a greater extent. Therefore, loss of fidelity and loss of reversibility are both regarded as costs of information gain.

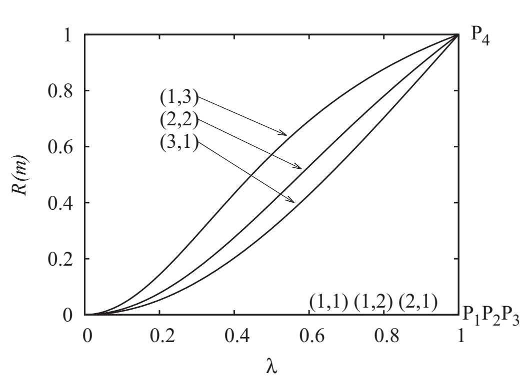

FIG. 3. Reversibility $R(m)$ as a function of singular value λ in $d = 4$ for $(k,l) = (1,1)$, $(1,2)$, $(1,3)$, $(2,1)$, $(2,2)$, and $(3,1)$. The symbols $\{\mathrm{P}_r\}(r = 1,2,3,4)$ denote projective measurements of rank r.

To explore the balance between costs and gains, two kinds of measurement efficiencies can be defined: one is the ratio of information gain to fidelity loss,

$$E_F(m) \equiv \frac{I(m)}{1 - F(m)}, \tag{96}$$

and the other is the ratio of information gain to reversibility loss,

$$E_R(m) \equiv \frac{I(m)}{1 - R(m)}. \tag{97}$$

Figures 4 and 5 show these efficiencies, $E_F(m)$ and $E_R(m)$, as functions of λ in $d = 4$ for various (k,l). As shown in Fig. 4, the efficiency $E_F(m)$ is not always a monotonic function, although it is difficult to analytically find its extreme value. In contrast, as shown in Fig. 5, the efficiency $E_R(m)$ is a monotonic function like the information function $I(m)$. In fact, for $(k,l) = (1,1)$, $(1,2)$, and $(2,1)$, the efficiency $E_R(m)$ is identical to the information $I(m)$ because of the irreversibility, $R(m) = 0$. The efficiencies $E_F(m)$ and $E_R(m)$ for P_r can also be calculated from Eqs. (91), (94), and (95) when $r = 1$, 2, or 3. However, it is not straightforward

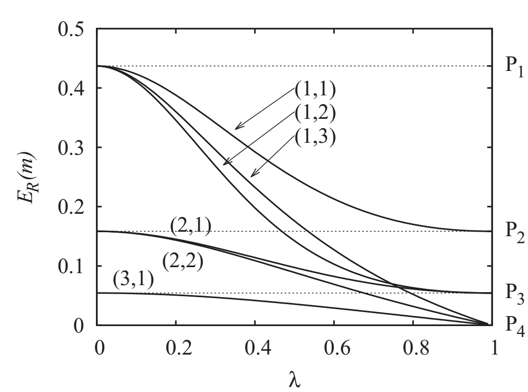

FIG. 5. Efficiency with respect to reversibility, $E_R(m)$, as a function of singular value λ in $d = 4$ for $(k,l) = (1,1)$, $(1,2)$, $(1,3)$, $(2,1)$, $(2,2)$, and $(3,1)$. The symbols $\{\mathrm{P}_r\}(r = 1,2,3,4)$ denote projective measurements of rank r.

to calculate the efficiencies for the identity operation P_4, since $I(m)=0$ and $F(m)=R(m)=1$. The limit values at P_4 can be calculated by considering the measurement of Eq. (83) with $(k,l)=(d-1,1)$ and $\lambda^2=1-\epsilon$. In this case, the information, fidelity, and reversibility given by Eqs. (90), (92), and (93), respectively, can be expanded as

$$I(m)=\frac{1}{2d^2\ln 2}\left(\frac{d-1}{d+1}\right)\epsilon^2+O(\epsilon^3), \tag{98}$$

$$F(m)=1-\frac{1}{4d}\left(\frac{d-1}{d+1}\right)\epsilon^2+O(\epsilon^3), \tag{99}$$

$$R(m)=1-\frac{d-1}{d}\epsilon-\frac{d-1}{d^2}\epsilon^2+O(\epsilon^3). \tag{100}$$

By taking the limit as $\epsilon\to 0$, the limits of the efficiencies $E_F(m)$ and $E_R(m)$ at P_4 are found to be

$$E_F(m)\to\frac{2}{d\ln 2},\quad E_R(m)\to 0. \tag{101}$$

V. CONCLUSION

The information, fidelity, and reversibility of an arbitrary quantum measurement have been shown in a d-level system whose premeasurement state is assumed to be completely unknown. These quantities have been expressed as functions of the singular values $\{\lambda_{mi}\}$ of the measurement operator $\hat{M}_m$ corresponding to the outcome m of the measurement, as shown in Eqs. (24), (37), and (48). Unfortunately, when some singular values are degenerate, Eq. (24) for the information gain is not useful due to the apparent divergence of the dangerous term shown in Eq. (57). Therefore, another expression for the dangerous term was presented in Eq. (78), which is free of any apparent divergence even when singular values are degenerate. As an example, a class of quantum measurements was considered whose singular values, as shown in Eq. (83), are highly degenerate. According to the general formulas, the information, fidelity, and reversibility were calculated as shown in Eqs. (90), (92), and (93), respectively. For $d=4$, these quantities are shown in Figs. 1–3, which indicate the trade-offs among the information, fidelity, and reversibility. That is, as a measurement provides more information about the state of the system, it changes the state by a greater degree and more irreversibly. Two measurement efficiencies were also defined, as shown in Eqs. (96) and (97), to show their different behavior.

The formulas shown in this paper are applicable to any efficient quantum measurement in systems with a finite-dimensional Hilbert space, such as multiple qubits or a qudit in quantum information theory. When an outcome is obtained by measurements, it is possible to calculate how much information is provided and how greatly and reversibly the state of the system is changed directly from the singular values of the measurement operator corresponding to the obtained outcome with no optimization problems [7,9,14]. The three quantities are for each single outcome rather than those averaged over all possible outcomes with probabilities given by Eq. (13), as shown in Eqs. (29), (38), and (49). It is not necessary to know the measurement operators corresponding to other outcomes. Therefore, the trade-offs at the level of a single outcome are more fundamental in quantum measurement. Although the trade-offs were shown only in a specific class of measurements in this paper, a general theory for such trade-offs will be presented in future studies. For general measurements, increasing information does not necessarily result in decreasing fidelity or reversibility. This is because the three quantities are functions of $d-1$ parameters and hence their relations are expressed by regions of finite size rather than lines. However, the boundaries of the regions show trade-offs among information, fidelity, and reversibility.

APPENDIX A: AVERAGES OVER STATES

Herein, the averages of $q_m(a)$ and $[f_m(a)]^2$ over all possible pure states of a d-level system are shown to prove Eqs. (19) and (35). They are given by

$$\overline{q_m}=\frac{1}{N}\sum_a\sum_i\lambda_{mi}^2|c_i(a)|^2, \tag{A1}$$

$$\overline{f_m^2}=\frac{1}{N}\sum_a\sum_{i,j}\lambda_{mi}\lambda_{mj}|c_i(a)|^2\left|c_j(a)\right|^2, \tag{A2}$$

from Eqs. (12) and (33), together with Eq. (14). First, the constants C, D, and E can be defined as

$$\frac{1}{N}\sum_a|c_i(a)|^2\equiv C \tag{A3}$$

and

$$\frac{1}{N}\sum_a|c_i(a)|^2|c_j(a)|^2\equiv\begin{cases}D & (\text{if } i=j)\\ E & (\text{if } i\neq j).\end{cases} \tag{A4}$$

Note that these constants do not depend on i or j, because there is no preferred state $|i\rangle$ when the index a runs over all pure states of the system. Using these constants, Eqs. (A1) and (A2) can be written as

$$\overline{q_m}=C\sum_i\lambda_{mi}^2=C\sigma_m^2, \tag{A5}$$

$$\overline{f_m^2}=D\sum_i\lambda_{mi}^2+E\sum_{i\neq j}\lambda_{mi}\lambda_{mj}=(D-E)\sigma_m^2+E\tau_m^2, \tag{A6}$$

where σ_m and τ_m are defined by Eqs. (20) and (36), respectively.

To calculate the constants C, D, and E, a parametrization of the coefficients $\{c_i(a)\}$ can be introduced. If $\alpha_i(a)$ and $\beta_i(a)$ are the real and imaginary parts of $c_i(a)$, respectively, then the normalization condition of Eq. (3) becomes

$$\sum_i[\alpha_i(a)^2+\beta_i(a)^2]=1, \tag{A7}$$

which is the condition for a point to be on the unit sphere in $2d$ dimensions. Thus, $\{\alpha_i(a)\}$ and $\{\beta_i(a)\}$ should be parametrized by the hyperspherical coordinates $(\theta_1,\theta_2,\ldots,\theta_{2d-2},\phi)$ as

$$\alpha_1(a)=\sin\theta_{2d-2}\sin\theta_{2d-3}\cdots\sin\theta_3\sin\theta_2\sin\theta_1\cos\phi,$$
$$\beta_1(a)=\sin\theta_{2d-2}\sin\theta_{2d-3}\cdots\sin\theta_3\sin\theta_2\sin\theta_1\sin\phi,$$

$$\alpha_2(a) = \sin\theta_{2d-2}\sin\theta_{2d-3}\cdots\sin\theta_3\sin\theta_2\cos\theta_1,$$
$$\beta_2(a) = \sin\theta_{2d-2}\sin\theta_{2d-3}\cdots\sin\theta_3\cos\theta_2,$$
$$\vdots$$
$$\alpha_d(a) = \sin\theta_{2d-2}\cos\theta_{2d-3},$$
$$\beta_d(a) = \cos\theta_{2d-2}, \tag{A8}$$

where $0 \leqslant \phi < 2\pi$ and $0 \leqslant \theta_p \leqslant \pi$ for $p = 1,2,\ldots,2d-2$. The index a can be replaced with the angles $(\theta_1,\theta_2,\ldots,\theta_{2d-2},\phi)$, and the summation over a can be replaced with the integral over the angles:

$$\frac{1}{N}\sum_a \longrightarrow \frac{(d-1)!}{2\pi^d}\int_0^{2\pi} d\phi \prod_{p=1}^{2d-2}\int_0^{\pi} d\theta_p \sin^p\theta_p. \tag{A9}$$

Then, if $i = 1$ and $j = d$,

$$C = \frac{1}{N}\sum_a |c_1(a)|^2 = \frac{(d-1)!}{\pi^{d-1}}\prod_{p=1}^{2d-2}\int_0^{\pi} d\theta_p \sin^{p+2}\theta_p, \tag{A10}$$

$$D = \frac{1}{N}\sum_a |c_1(a)|^4 = \frac{(d-1)!}{\pi^{d-1}}\prod_{p=1}^{2d-2}\int_0^{\pi} d\theta_p \sin^{p+4}\theta_p, \tag{A11}$$

$$E = \frac{1}{N}\sum_a |c_1(a)|^2|c_d(a)|^2 = C - \frac{(d-1)!}{\pi^{d-1}}\prod_{p=2d-3}^{2d-2}\int_0^{\pi} d\theta_p \sin^{p+4}\theta_p \times \prod_{p=1}^{2d-4}\int_0^{\pi} d\theta_p \sin^{p+2}\theta_p. \tag{A12}$$

These integrals can easily be calculated to be

$$C = \frac{1}{d}, \quad D = \frac{2}{d(d+1)}, \quad E = \frac{1}{d(d+1)} \tag{A13}$$

by using the integral formula

$$\int_0^{\pi} d\theta \sin^n\theta = \sqrt{\pi}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n+2}{2}\right)} \tag{A14}$$

for $n > -1$ with the Gamma function $\Gamma(n)$. Therefore, Eqs. (19) and (35) can be proven by substituting Eq. (A13) into Eqs. (A5) and (A6).

APPENDIX B: COEFFICIENTS OF SERIES

Herein, the coefficients of the Taylor series in Eqs. (66) and (71) are presented. To find the coefficients $\{c_n^{(d)}(\lambda)\}$ in Eq. (66), the following Taylor series is first considered:

$$(1+\epsilon)^d \log_2(1+\epsilon) \equiv \sum_{n=0}^{d-1} a_n^{(d)}\epsilon^n + O(\epsilon^d). \tag{B1}$$

By expanding $(1+\epsilon)^d$ and $\log_2(1+\epsilon)$ in the Taylor series, the coefficients $\{a_n^{(d)}\}$ can be determined to be $a_0^{(d)} = 0$ for $n = 0$ and

$$a_n^{(d)} = \frac{1}{\ln 2}\sum_{k=1}^{n}\frac{(-1)^{k+1}}{k}\binom{d}{n-k} \tag{B2}$$

for $n = 1,2,\ldots,d-1$.

Next, a proof of the equivalence between Eqs. (B2) and (68) will be presented by mathematical induction. As the first step, it will be shown that the statement holds for $a_1^{(d)}$ and $a_{d-1}^{(d)}$. It is easy to see that both equations yield $a_1^{(d)} = 1/\ln 2$. At the same time, by using the identity

$$\frac{1}{k}\binom{d}{d-1-k} = d\left[\frac{1}{k} - \frac{1}{k+1}\right]\binom{d-1}{k} \tag{B3}$$

and the summation formulas

$$\sum_{k=1}^{n}\frac{(-1)^{k+1}}{k}\binom{n}{k} = \eta(n), \quad \sum_{k=1}^{n}\frac{(-1)^{k+1}}{k+1}\binom{n}{k} = \frac{n}{n+1}, \tag{B4}$$

$a_{d-1}^{(d)}$ in Eq. (B2) becomes

$$a_{d-1}^{(d)} = \frac{d}{\ln 2}[\eta(d) - 1], \tag{B5}$$

which is equal to that in Eq. (68). As the second step, it will be shown that if the statement holds for $a_n^{(d-1)}$ with $n = 1,2,\ldots,d-2$, then it holds for $a_n^{(d)}$ with $n = 2,3,\ldots,d-2$, on the basis of the recurrence relation

$$a_n^{(d)} = a_n^{(d-1)} + a_{n-1}^{(d-1)}, \tag{B6}$$

which originates from

$$(1+\epsilon)^d\log_2(1+\epsilon) = (1+\epsilon)(1+\epsilon)^{d-1}\log_2(1+\epsilon). \tag{B7}$$

Since this recurrence relation is satisfied by both equations, the second step can be shown. Accordingly, by mathematical induction starting from $d = 2$ and $n = 1$, the statement that Eq. (B2) is equal to Eq. (68) for all d and n has been proven. Note that Eq. (68) can include the case of $n = 0$, since $a_0^{(d)} = 0$.

By using the coefficients $\{a_n^{(d)}\}$, the coefficients $\{c_n^{(d)}(\lambda)\}$ can be found. The left-hand side of Eq. (66) can be written as

$$(\lambda^2+\epsilon)^d\log_2(\lambda^2+\epsilon) = \lambda^{2d}\left(1+\frac{\epsilon}{\lambda^2}\right)^d\left[\log_2\lambda^2 + \log_2\left(1+\frac{\epsilon}{\lambda^2}\right)\right], \tag{B8}$$

while from Eq. (B1),

$$\left(1+\frac{\epsilon}{\lambda^2}\right)^d\log_2\left(1+\frac{\epsilon}{\lambda^2}\right) = \sum_{n=0}^{d-1}a_n^{(d)}\left(\frac{\epsilon}{\lambda^2}\right)^n + O(\epsilon^d). \tag{B9}$$

By substituting Eq. (B9) into Eq. (B8), the coefficients $\{c_n^{(d)}(\lambda)\}$ can be obtained, as in Eq. (67).

Finally, the coefficients $\{b_n^{(s)}\}$ of Eq. (71) will be derived. The coefficients $\{b_n^{(s)}\}$ can be found by defining K_s as

$$K_s \equiv \prod_{r\neq s}\left(1+\frac{\epsilon}{\bar{\lambda}_{ms}^2 - \bar{\lambda}_{mr}^2}\right)^{-n_r} = \sum_{n=0}^{\infty} b_n^{(s)}\epsilon^n \tag{B10}$$

and expanding $\ln K_s$ rather than K_s itself as a Taylor series:

$$\ln K_s = \sum_{n=1}^{\infty} \frac{1}{n!} h_n^{(s)} \epsilon^n, \tag{B11}$$

where the coefficients $\{h_n^{(s)}\}$ are given by Eq. (76). Therefore, K_s can be expressed as the exponential of a Taylor series:

$$K_s = \exp\left(\sum_{n=1}^{\infty} \frac{1}{n!} h_n^{(s)} \epsilon^n\right). \tag{B12}$$

According to Faà di Bruno's formula, the exponential of a Taylor series can be expanded as a Taylor series by the complete Bell polynomials shown in Eq. (72) as

$$\exp\left(\sum_{n=1}^{\infty} \frac{1}{n!} x_n \epsilon^n\right) = \sum_{n=0}^{\infty} \frac{1}{n!} B_n(x_1, x_2, \ldots, x_n) \epsilon^n. \tag{B13}$$

By applying this formula to Eq. (B12), the coefficients $\{b_n^{(s)}\}$ of Eq. (75) can be obtained. Note that the complete Bell polynomials satisfy the following formulas: for a constant c and a positive integer m,

$$B_n(cx_1, c^2 x_2, \ldots, c^n x_n) = c^n B_n(x_1, x_2, \ldots, x_n), \tag{B14}$$

$$B_n(0!m, 1!m, \ldots, (n-1)!m) = n! \binom{m+n-1}{m-1}. \tag{B15}$$

The first formula is valid on the basis of the definition in Eq. (72), and the second formula can be derived from Eq. (B13) because

$$\sum_{n=0}^{\infty} \frac{1}{n!} B_n(0!m, 1!m, \ldots, (n-1)!m) \epsilon^n = \exp\left(m \sum_{n=1}^{\infty} \frac{1}{n} \epsilon^n\right) = e^{-m \ln(1-\epsilon)} = \frac{1}{(1-\epsilon)^m}. \tag{B16}$$

APPENDIX C: LIMITS TO PROJECTIVE MEASUREMENT

Herein, Eq. (90) for (k,l) is shown to be equal to Eq. (91) for $r = k$ when $\lambda = 0$ and to that for $r = k + l$ when $\lambda = 1$, as expected from the definition of Eq. (83). When $\lambda = 0$, the dangerous term $J = J_1 + J_2$ given in Eqs. (88) and (89) becomes

$$\lim_{\lambda \to 0} J = \sum_{n=0}^{k-1} \binom{k+l-n-2}{l-1} (-1)^{k-n-1} a_n^{(k+l)}, \tag{C1}$$

since $c_n^{(d)}(0) = 0$. This expression can be simplified by the identity

$$\sum_{n=0}^{k-1} \binom{k+l-n-2}{l-1} (-1)^{k-n-1} a_n^{(k+l)} = a_{k-1}^{(k)}, \tag{C2}$$

which is derived from

$$\frac{1}{(1+\epsilon)^l} (1+\epsilon)^{k+l} \log_2(1+\epsilon) = (1+\epsilon)^k \log_2(1+\epsilon) \tag{C3}$$

by expanding $1/(1+\epsilon)^l$, $(1+\epsilon)^{k+l} \log_2(1+\epsilon)$, and $(1+\epsilon)^k \log_2(1+\epsilon)$ in Taylor series and comparing terms of order ϵ^{k-1} on both sides. The dangerous term is then found to be

$$\lim_{\lambda \to 0} J = a_{k-1}^{(k)} = \frac{k}{\ln 2}[\eta(k) - 1] \tag{C4}$$

by using Eq. (68). This shows that, when $\lambda = 0$, Eq. (90) becomes

$$I(m) = \log_2 \frac{d}{k} - \frac{1}{\ln 2}[\eta(d) - \eta(k)], \tag{C5}$$

which is equal to Eq. (91) for $r = k$.

On the other hand, when $\lambda = 1$, the dangerous term $J = J_1 + J_2$ has apparent divergences as in Eqs. (88) and (89). However, it can be calculated by substituting $1 + \epsilon$ for λ^2 and taking the limit as $\epsilon \to 0$. Note that the divergent terms in Eqs. (88) and (89) should cancel each other, since J is finite. The dangerous term is thus given by

$$\lim_{\lambda \to 1} J = \sum_{n=0}^{l-1} \binom{k+l-n-2}{k-1} (-1)^{l-n-1} \times \left[\binom{k+l}{n} a_{k+l-n-1}^{(k+l-n)} + (k+l-n) a_n^{(k+l)}\right]. \tag{C6}$$

Moreover, by using

$$\binom{k+l}{n} a_{k+l-n-1}^{(k+l-n)} + (k+l-n) a_n^{(k+l)} = \binom{k+l-1}{n} a_{k+l-1}^{(k+l)} \tag{C7}$$

derived from Eq. (68) and

$$\sum_{n=0}^{l-1} (-1)^{l-n-1} \binom{k+l-n-2}{k-1} \binom{k+l-1}{n} = 1 \tag{C8}$$

derived from

$$\frac{1}{(1+\epsilon)^k} (1+\epsilon)^{k+l-1} = (1+\epsilon)^{l-1} \tag{C9}$$

by comparing terms of order ϵ^{l-1} on both sides, it is found to be

$$\lim_{\lambda \to 1} J = a_{k+l-1}^{(k+l)} = \frac{k+l}{\ln 2}[\eta(k+l) - 1]. \tag{C10}$$

This shows that, when $\lambda = 1$, Eq. (90) becomes

$$I(m) = \log_2 \frac{d}{k+l} - \frac{1}{\ln 2}[\eta(d) - \eta(k+l)], \tag{C11}$$

which is equal to Eq. (91) for $r = k + l$.

[1] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[2] C. H. Bennett and G. Brassard, in *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE, New York, 1984), pp. 175–179.
[3] A. K. Ekert, Phys. Rev. Lett. **67**, 661 (1991).
[4] C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992).
[5] C. H. Bennett, G. Brassard, and N. D. Mermin, Phys. Rev. Lett. **68**, 557 (1992).
[6] C. A. Fuchs and A. Peres, Phys. Rev. A **53**, 2038 (1996).
[7] K. Banaszek, Phys. Rev. Lett. **86**, 1366 (2001).
[8] C. A. Fuchs and K. Jacobs, Phys. Rev. A **63**, 062305 (2001).
[9] K. Banaszek and I. Devetak, Phys. Rev. A **64**, 052307 (2001).
[10] H. Barnum, arXiv:quant-ph/0205155.
[11] G. M. D'Ariano, Fortschr. Phys. **51**, 318 (2003).
[12] M. Ozawa, Ann. Phys. (NY) **311**, 350 (2004).
[13] L. Maccone, Phys. Rev. A **73**, 042307 (2006).
[14] M. F. Sacchi, Phys. Rev. Lett. **96**, 220502 (2006).
[15] F. Buscemi and M. F. Sacchi, Phys. Rev. A **74**, 052320 (2006).
[16] K. Banaszek, Open Sys. Info. Dyn. **13**, 1 (2006).
[17] F. Buscemi, M. Hayashi, and M. Horodecki, Phys. Rev. Lett. **100**, 210504 (2008).
[18] L. Fan, W. Ge, H. Nha, and M. S. Zubairy, Phys. Rev. A **92**, 022114 (2015).
[19] M. Ueda and M. Kitagawa, Phys. Rev. Lett. **68**, 3424 (1992).
[20] M. Ueda, N. Imoto, and H. Nagaoka, Phys. Rev. A **53**, 3808 (1996).
[21] M. Ueda, in *Frontiers in Quantum Physics: Proceedings of the International Conference on Frontiers in Quantum Physics, Kuala Lumpur, Malaysia, 1997*, edited by S. C. Lim, R. Abd-Shukor, and K. H. Kwek (Springer-Verlag, Singapore, 1999), pp. 136–144.
[22] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics (Non-Relativistic Theory)*, 3rd ed. (Butterworth–Heinemann, Oxford, 1977).
[23] A. Imamoḡlu, Phys. Rev. A **47**, R4577 (1993).
[24] A. Royer, Phys. Rev. Lett. **73**, 913 (1994); **74**, 1040(E) (1995).
[25] H. Terashima and M. Ueda, Phys. Rev. A **74**, 012102 (2006).
[26] A. N. Korotkov and A. N. Jordan, Phys. Rev. Lett. **97**, 166805 (2006).
[27] H. Terashima and M. Ueda, Phys. Rev. A **75**, 052323 (2007).
[28] Q. Sun, M. Al-Amri, and M. S. Zubairy, Phys. Rev. A **80**, 033838 (2009).
[29] Y.-Y. Xu and F. Zhou, Commun. Theor. Phys. **53**, 469 (2010).
[30] N. Katz, M. Neeley, M. Ansmann, R. C. Bialczak, M. Hofheinz, E. Lucero, A. O'Connell, H. Wang, A. N. Cleland, J. M. Martinis, and A. N. Korotkov, Phys. Rev. Lett. **101**, 200401 (2008).
[31] Y.-S. Kim, Y.-W. Cho, Y.-S. Ra, and Y.-H. Kim, Opt. Express **17**, 11978 (2009).
[32] H. Mabuchi and P. Zoller, Phys. Rev. Lett. **76**, 3108 (1996).
[33] M. A. Nielsen and C. M. Caves, Phys. Rev. A **55**, 2547 (1997).
[34] H. Terashima, Phys. Rev. A **83**, 032111 (2011).
[35] H. Terashima, Phys. Rev. A **83**, 032114 (2011).
[36] Y. W. Cheong and S.-W. Lee, Phys. Rev. Lett. **109**, 150402 (2012).
[37] G. Chen, Y. Zou, X.-Y. Xu, J.-S. Tang, Y.-L. Li, J.-S. Xu, Y.-J. Han, C.-F. Li, G.-C. Guo, H.-Q. Ni, Y. Yu, M.-F. Li, G.-W. Zha, Z.-C. Niu, and Y. Kedem, Phys. Rev. X **4**, 021043 (2014).
[38] H.-T. Lim, Y.-S. Ra, K.-H. Hong, S.-W. Lee, and Y.-H. Kim, Phys. Rev. Lett. **113**, 020504 (2014).
[39] H. Terashima, Phys. Rev. A **85**, 022124 (2012).
[40] H. Terashima and M. Ueda, Phys. Rev. A **81**, 012110 (2010).
[41] A. Uhlmann, Rep. Math. Phys. **9**, 273 (1976).
[42] M. Koashi and M. Ueda, Phys. Rev. Lett. **82**, 2598 (1999).
[43] E. B. Davies and J. T. Lewis, Commun. Math. Phys. **17**, 239 (1970).
[44] W. K. Wootters, Found. Phys. **20**, 1365 (1990).
[45] K. R. W. Jones, J. Phys. A: Math. Gen. **24**, 1237 (1991). In this reference, the relevant formula shown as Eq. (29) has a typographical error that the first factor should be $1/d$ instead of $\Gamma(d)$.
[46] R. Jozsa, D. Robb, and W. K. Wootters, Phys. Rev. A **49**, 668 (1994).
[47] M. Ban, J. Phys. A: Math. Gen. **34**, 9669 (2001).