Tight concentration inequalities for quantum adversarial setups exploiting permutation symmetry

Takaya Matsuura[1,2], Shinichiro Yamano[3], Yui Kuramochi[4], Toshihiko Sasaki[3,5], and Masato Koashi[3,5]

[1]Centre for Quantum Computation & Communication Technology, School of Science, RMIT University, Melbourne VIC 3000, Australia

[2]RIKEN Center for Quantum Computing (RQC), Hirosawa 2-1, Wako, Saitama 351-0198, Japan

[3]Department of Applied Physics, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

[4]Department of Physics, Faculty of Science, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, Japan

[5]Photon Science Center, Graduate School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

We developed new concentration inequalities for a quantum state on an N-qudit system or measurement outcomes on it that apply to an adversarial setup, where an adversary prepares the quantum state. Our one-sided concentration inequalities for a quantum state require the N-qudit system to be permutation invariant and are thus de-Finetti type, but they are tighter than the one previously obtained. We show that the bound can further be tightened if each qudit system has an additional symmetry. Furthermore, our concentration inequality for the outcomes of independent and identical measurements on an N-qudit quantum system has no assumption on the adversarial quantum state and is much tighter than the conventional one obtained through Azuma's inequality. We numerically demonstrate the tightness of our bounds in simple quantum information processing tasks.

1 Introduction and the main results

1.1 General introduction

A concentration inequality is a bound on the probability of obtaining atypical events, i.e., events that are far from the expectation. It is essential in lots of information-theoretic primitives such as cryptography and error correction since it gives a failure probability in anticipating that the observed values are close to the expectation values. The same applies to quantum information theory; we need concentration inequalities to obtain an upper bound on the probability of atypical events in quantum information protocols. Evaluation of the probability of atypical events may be easy for independent and identically distributed (i.i.d.) quantum states. There, one can apply Chernoff-Hoeffding theorem [1] or Sanov's theorem [2, 3] to obtain an upper bound on the deviation from the expectation value for an additive observable allowing a small failure probability. For certain applications, however, one even needs a concentration inequality that applies to an adversarial setup. This includes quantum key distribution (QKD), verifiable blind quantum computation [4], and verification and learning of a quantum state prepared by an adversary [5], where the underlying statistics may be governed by an adversary.

In the field of QKD, for example, proving the security against collective attacks in a finite number of communication rounds is relatively easy, and many QKD protocols have been proved secure at least against collective attacks even in the finite-size regime. The situation gets involved when considering the security against general attacks. There are two mainstreams in the way of proving the security against general attacks in finite-size regimes. One of them is called the phase-error correction method [6, 7, 8, 9], which is the main focus of this paper, and evaluates the

Takaya Matsuura: takaya.matsuura@riken.jp

number of phase error patterns or a phase error rate while allowing an exponentially small failure probability. Correcting the phase error with a failure probability ϵ ensures $\sim \epsilon^{1/2}$-secrecy of the protocol against the adversary. In Ref. [10], it was shown that in the phase-error correction method, imposing a permutation invariance to the protocol either actively or passively reduces the problem of general attacks to the collective attacks at the cost of increasing the failure probability ϵ by a polynomial factor. This implies that the error exponent of the failure probability of the phase error correction is the same against collective attacks and against general attacks. After that, a distinct approach using Azuma's inequality, which is based on sequentiality of measurements instead of permutation invariance, was proposed [11]. Since Azuma's inequality is easier to evaluate, it prevails in the QKD community. At the cost of its simplicity, the error exponent of the failure probability when using Azuma's inequality is always less than that of the collective attack case. Intensive research has been made on tightening Azuma's inequality with additional information [12, 13, 14] including the recently found one with unconfirmed knowledge [14], but it is still looser than the Sanov-type inequalities [2, 3] that apply to the i.i.d. random variables. Note that in another approach to the security proofs based on the leftover hash lemma against a quantum adversary [15], a reduction of the security against general attacks to that against collective attacks by exploiting the permutation symmetry has also been studied under the name of de Finetti reduction and the post-selection technique [16, 17, 18, 19, 15, 20, 21]. The problems with these approaches were that the reduction was not very straightforward or that they led to worse performance. Then, the community gradually shifted to an alternative method that uses more recently developed entropy accumulation theorem [22, 23, 24]. Interestingly, both Azuma's inequality and the entropy accumulation utilize sequential conditionings by previous measurement outcomes, yet they may be looser than the bound that applies to the i.i.d. case.

In this paper, we revisit the idea based on permutation invariance and derive concentration inequalities through the reduction to the i.i.d. scenario. The first case we consider is when a quantum state on a multi-qudit system is invariant under the permutation of the subsystems possibly with additional symmetry on each subsystem. We show that whenever an outcome of the quantum measurement has a small appearance probability for any i.i.d. quantum state, it also has a small appearance probability for these permutation-invariant states up to an explicit factor polynomial in the number of subsystems. This can be proved by observing that any permutation-invariant quantum state is bounded from above by the mixture of i.i.d. quantum states up to a polynomial factor, i.e., we have the following.

Theorem 1 (Informal)**.** *Let ρ_{A^n} be a permutation-invariant state on $\mathcal{H}_A^{\otimes n}$ with $\dim \mathcal{H}_A = d$. Then, we have*

$$\rho_{A^n} \leq f_q(n,d)\ [\textit{Mixture of i.i.d. states}], \quad (1)$$

where $f_q(n,d) = \mathcal{O}(n^{(d^2-1)/2})$ for a large n.

Compared to the known results [15, 20], we tightly evaluate the polynomial prefactor and improve the bound for a general permutation-invariant state. We also obtain a bound that explicitly depends on the irreducible representation space of the permutation group. In particular, when restricted to the totally symmetric subspace, our bound reproduces the one developed for the post-selection technique [15, 20]. The second case we consider is when we perform independent and identical quantum measurements on an arbitrary quantum state prepared by an adversary. In this case, we have an inequality between the probability mass function for an empirical probability of the measurement outcomes and a mixture of multinomial distribution up to a polynomial factor. We then obtain a Sanov-type concentration inequality for the probability of obtaining an empirical probability that lies in a set $\mathcal{A}$ of atypical events.

Theorem 3 (Informal)**.** *Let $\mathcal{M}$ be a k-outcome measurement channel and $\mathcal{P}$ be the probability $(k-1)$-simplex. For a state ρ, $\mathcal{M}(\rho)$ corresponds to a point in $\mathcal{P}$. Let $\mathcal{R} \subset \mathcal{P}$ be a region that is bounded by a hyperplane and contains $\{\mathcal{M}(\rho) : \forall \rho\}$ and all the vertices of $\mathcal{P}$ except one. Expand $\mathcal{R}$ by moving the boundary between $\mathcal{R}$ and $\mathcal{P} \setminus \mathcal{R}$ by $\mathcal{O}(n^{-1})$ to define the region $\mathcal{R}'$. Let $P_{\boldsymbol{x}}$ denotes the empirical probability for a sequence*

$\boldsymbol{x}$. *Then, for any closed convex subset* $\mathcal{A} \subseteq \mathcal{P}$ *and for any* n*-partite quantum state* ρ^n*, the sequence* $\boldsymbol{x}$ *of outcomes of the measurement* $\mathcal{M}^{\otimes n}$ *performed on* ρ^n *satisfies*

$$\mathcal{M}^{\otimes n}(\rho^n)[P_{\boldsymbol{x}} \in \mathcal{A}] \leq \max_{\boldsymbol{p}\in\mathcal{A}} \max_{\boldsymbol{q}\in\mathcal{R}'} f_c(n,k) e^{-nD(\boldsymbol{p}\|\boldsymbol{q})}, \tag{2}$$

where $f_c(n,k) = \mathcal{O}(n^{(k-1)/2})$ *for a large* n *and* $D(\cdot\|\cdot)$ *denotes the Kullback-Leibler divergence.*

As opposed to the first case, the polynomial factor in this case does not depend on the dimension of the underlying quantum system but depends on the number of measurement outcomes.

Both of our bounds are thus Sanov-type and thus expected to be tighter than those obtained by the Azuma-type inequalities. To numerically demonstrate this, we consider two concrete scenarios. The first one is a simple estimation task to obtain an upper bound on a random variable defined through the quantum measurements. The second one is a finite-size key rate of a quantum key distribution protocol. In both cases, we compare the performance obtained through conventional concentration inequalities and our new concentration inequalities.

Applicability of our results is not restricted to a finite-size analysis of QKD. In the verifiable blind quantum computation [4] and learning of a non-i.i.d. quantum state [5] as mentioned earlier, the failure probability of a randomly-permuted property test on an adversarial quantum state is estimated. Our results may fit to these cases as well, where one cares about exponentially rare outcomes of measurements performed on an adversarial state.

1.2 Summary of our main results

We obtained three theorems (Theorems 1–3) for the concentration inequalities in the adversarial setups. Theorems 1 and 2 consider the case where the quantum state is symmetric under the permutation of subsystems, possibly with an additional symmetry restriction on each subsystem (Theorem 2). These two theorems are given in terms of operator inequalities to show that a normalized density operator of any permutation-symmetric states is dominated, up to a polynomial factor in the number of subsystems, by a density operator for a mixture of i.i.d. quantum states. Such a theorem can immediately be combined with any large-deviation theorem for quantum measurements on i.i.d. states to derive a bound for non-i.i.d. but permutation-symmetric states. The polynomial factor can be viewed as a price in extending the applicability to the non-i.i.d. case. Note that this scenario works even in the case where non-i.i.d. quantum measurements are performed. To the best of the authors' knowledge, this approach was first used in the context of quantum information theory for proving the security of the Bennett 1992 (B92) QKD protocol against general attacks [10]. A more general form was derived in relation to quantum extension [19, 15, 20, 21] of the de Finetti theorem, which is given by [20]

$$\rho_{A^n} \leq \binom{n+d^2-1}{n} \int \sigma_A^{\otimes n} \mu(\sigma), \tag{3}$$

where ρ_{A^n} is invariant under the permutation of A^n, d is the dimension of the system A, and μ denotes a probability measure on the set of density operators on the system A. Compared to this known inequality, Theorem 1 provides an improvement in the polynomial factor as well as a further refinement of the factor by allowing its dependence on the i.i.d. state, and applicability to a subsystem which makes it possible to compose multiple uses of the theorem when two or more subsystems separately adhere to permutation symmetry. Theorem 2 is an adaptation of Theorem 1 to the more restricted cases where each subsystem adheres to a general symmetry in addition to the permutation symmetry among n systems. This is in contrast to the previous works [25, 26, 27] in which more general symmetries than the permutation among n systems were considered. Finally, our third result, Theorem 3, treats a somewhat different setup: a concentration inequality for the measurement outcomes when independent and identical measurements are performed on an adversarial quantum state. Unlike the previous two theorems, this theorem does not require any symmetry on the quantum system as long as the concentration phenomenon for the empirical probability of measurement outcomes is considered. Furthermore, the resulting statement does not depend on the dimension of the quantum system but depends on the number of measurement outcomes.

Our first result considers an operator inequality to bound a permutation-invariant state from

above by a mixture of i.i.d. quantum states, which is applicable under the same prerequisite as the conventional de-Finetti-type inequality (3). Let $\mathcal{H}_A$ be a d-dimensional Hilbert space. Let Y_n^d be the set of Young diagrams with n boxes and at most d rows, i.e.,

$$Y_n^d := \{\boldsymbol{n} = (n_1, \dots, n_d) \in Z_n^d : n_1 \geq n_2 \geq \cdots \geq n_d\} \tag{4}$$

with

$$Z_n^d := \Big\{\boldsymbol{n} = (n_1, \dots, n_d) \in \{0, \dots, n\}^{\times d} : \sum_{i=1}^{d} n_i = n\Big\}, \tag{5}$$

where i-th element of $\boldsymbol{n} \in Y_n^d$ denotes the number of boxes in i-th row. From the Schur-Weyl duality, $\mathcal{H}_A^{\otimes n}$ can be decomposed as

$$\mathcal{H}_A^{\otimes n} = \bigoplus_{\boldsymbol{n} \in Y_n^d} \mathcal{U}_{\boldsymbol{n}} \otimes \mathcal{V}_{\boldsymbol{n}}, \tag{6}$$

where $\mathcal{U}_{\boldsymbol{n}}$ denotes an irreducible representation space of $\mathrm{U}(d)$ and $\mathcal{V}_{\boldsymbol{n}}$ denotes that of the permutation group S_n. Let $P_{\boldsymbol{n}}$ be a projection onto $\mathcal{U}_{\boldsymbol{n}} \otimes \mathcal{V}_{\boldsymbol{n}}$. It is known that a state that is invariant under the permutation of n subsystems of A^n has a direct-sum structure over $\boldsymbol{n}$ in Eq. (6) and thus commutes with $P_{\boldsymbol{n}}$. This fact holds even when the permutation-invariant state on the system A^n entangles with an environmental system E. That is, if we define ρ_{A^nE} as a density operator on $\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_E$ invariant under the permutation of n subsystems of A^n, then the following holds:

$$\rho_{A^nE} = \sum_{\boldsymbol{n} \in Y_n^d} P_{\boldsymbol{n}} \rho_{A^nE} P_{\boldsymbol{n}}, \tag{7}$$

where the abbreviated notation $P_{\boldsymbol{n}}$ acting on $\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_E$ means that it acts trivially on $\mathcal{H}_E$. From now on, if X_{AB} is defined as an operator on $\mathcal{H}_A \otimes \mathcal{H}_B$, then X_A denotes $\mathrm{Tr}_B[X_{AB}]$. A map Φ_A with the index A denotes the completely positive trace-preserving (CPTP) map Φ from the set of operators on the system A to itself. For a positive operator τ, $\mathrm{supp}(\tau)$ denotes the support projection, i.e., the minimal projection Q that satisfies $\mathrm{Tr}[Q\tau] = \mathrm{Tr}[\tau]$. Let $\mathrm{Haar}(d)$ denote the Haar measure on the d-dimensional unitary group $\mathrm{U}(d)$. For a d-tuple of real numbers $\boldsymbol{t}$, let $\mathrm{diag}(\boldsymbol{t})$ denote a density operator represented by a diagonal matrix with the diagonal entries $\boldsymbol{t}$ in a fixed basis of $\mathcal{H}_A$. The following functions frequently appear in the later analyses.

Definition 1. *For $n, d \in \mathbb{N}$ and $\boldsymbol{n} \in Y_n^d$, let $s_{\boldsymbol{n}}(\boldsymbol{t})$ be the Schur function of $\boldsymbol{t} \in \mathbb{R}^{\times d}$ with the shape $\boldsymbol{n}$ defined as*

$$s_{\boldsymbol{n}}(\boldsymbol{t}) := \frac{\det(t_i^{n_j+d-j})_{1\leq i,j\leq d}}{\det(t_i^{d-j})_{1\leq i,j\leq d}}. \tag{8}$$

Then, the function $f_q(n, d; \boldsymbol{n})$ is defined with $s_{\boldsymbol{n}}(\boldsymbol{t})$ as

$$f_q(n, d; \boldsymbol{n}) := \frac{\dim \mathcal{U}_{\boldsymbol{n}}}{s_{\boldsymbol{n}}(\boldsymbol{n}/n) \dim \mathcal{V}_{\boldsymbol{n}}}. \tag{9}$$

Furthermore, let $f_q(n, d)$ be defined as

$$f_q(n, d) := \frac{(n+d-1)^{\frac{(d^2-1)}{2}}}{\sqrt{2\pi(d/e^2)^d}\, G(d+1)}, \tag{10}$$

where $G(m)$ denotes the Barnes G-function given for an integer $m \geq 2$ by

$$G(m) = \prod_{i=0}^{m-2} i!. \tag{11}$$

For $\boldsymbol{t} = (t_1, \dots, t_d) \geq 0$ with $\sum_{i=1}^d t_i = 1$ and $\boldsymbol{m} = (n_1, \dots, n_d) \in Z_n^d$, let $\mathrm{Mult}_{\boldsymbol{t}}(\boldsymbol{m})$ be the probability density function of the multinomial distribution given by

$$\mathrm{Mult}_{\boldsymbol{t}}(\boldsymbol{m}) := \frac{n!}{n_1! \cdots n_d!} \prod_{i=1}^{d} t_i^{n_i}. \tag{12}$$

Then, we define $f_c(n, d; \boldsymbol{m})$ as

$$f_c(n, d; \boldsymbol{m}) := [\mathrm{Mult}_{\boldsymbol{m}/n}(\boldsymbol{m})]^{-1}, \tag{13}$$

and define $f_c(n, d)$ as

$$f_c(n, d) := \frac{n^{\frac{d-1}{2}}}{\sqrt{2\pi(d/e^2)^d}}. \tag{14}$$

The following relations between functions defined above will be proved later.

Lemma 1. *Let $f_q(n, d; \boldsymbol{n})$, $f_q(n, d)$, $f_c(n, d; \boldsymbol{m})$, and $f_c(n, d)$ be as defined in Definition 1. Then, we have*

$$f_q(n, d; \boldsymbol{n}) \leq f_q(n, d) \tag{15}$$

and

$$f_c(n, d; \boldsymbol{m}) \leq f_c(n, d). \tag{16}$$

Now, we state our first result.

Theorem 1. *Consider a composite system A^nE associated with a Hilbert space $\mathcal{H}_A^n \otimes \mathcal{H}_E$ with $\dim \mathcal{H}_A = d$. Then, there exists a pair $(\mathcal{S}_{A^n}, \mathcal{R}_{A^n})$ of CPTP maps that satisfies the following conditions (i), (ii), and (iii).*

(i) For any density operator ρ_{A^nE} and for any $\boldsymbol{n} \in Y_n^d$, we have

$$\mathcal{S}_{A^n} \otimes \mathrm{Id}_E(P_{\boldsymbol{n}} \rho_{A^nE} P_{\boldsymbol{n}}) = P_{\boldsymbol{n}} \mathcal{S}_{A^n} \otimes \mathrm{Id}_E(\rho_{A^nE}) P_{\boldsymbol{n}}. \tag{17}$$

(ii) For any ρ_{A^nE} that is invariant under permutation of n systems of A^n, we have

$$\mathcal{R}_{A^n} \circ \mathcal{S}_{A^n} \otimes \mathrm{Id}_E(\rho_{A^nE}) = \rho_{A^nE}. \tag{18}$$

(ii) For any ρ_{A^n} that is invariant under permutation of n systems of A^n and for any $\boldsymbol{n} \in Y_n^d$, we have

$$\begin{aligned} &\mathrm{supp}(\mathcal{S}_{A^n}(P_{\boldsymbol{n}} \rho_{A^n} P_{\boldsymbol{n}})) \\ &\le f_q(n, d; \boldsymbol{n}) \\ &\quad \times \mathcal{S}_{A^n}\Big(\mathbb{E}_{U_A \sim \mathrm{Haar}(d)} \Big[\Big(U_A \,\mathrm{diag}\Big(\frac{\boldsymbol{n}}{n}\Big) U_A^\dagger\Big)^{\otimes n}\Big]\Big), \end{aligned} \tag{19}$$

where $f_q(n, d; \boldsymbol{n})$ is defined in Definition 1.

This theorem can immediately be cast into the conventional form of Eq. (3). Since $\mathcal{S}_{A^n}$ is a CPTP map, we trivially have, for any $\boldsymbol{n} \in Y_n^d$,

$$\mathcal{S}_{A^n}(P_{\boldsymbol{n}} \rho_{A^n} P_{\boldsymbol{n}}) \le \mathrm{Tr}[P_{\boldsymbol{n}} \rho_{A^n}] \,\mathrm{supp}(\mathcal{S}_{A^n}(P_{\boldsymbol{n}} \rho_{A^n} P_{\boldsymbol{n}})), \tag{20}$$

and thus, for any permutation-invariant state ρ_{A^n}, we have

$$\mathcal{S}_{A^n}(\rho_{A^n}) = \sum_{\boldsymbol{n} \in Y_n^d} \mathcal{S}_{A^n}(P_{\boldsymbol{n}} \rho_{A^n} P_{\boldsymbol{n}}) \tag{21}$$

$$\le \sum_{\boldsymbol{n} \in Y_n^d} \mathrm{Tr}[P_{\boldsymbol{n}} \rho_{A^n}] \,\mathrm{supp}(\mathcal{S}_{A^n}(P_{\boldsymbol{n}} \rho_{A^n} P_{\boldsymbol{n}})). \tag{22}$$

Combining the above with Eqs. (18) and (19), we have

$$\begin{aligned} \rho_{A^n} \le \sum_{\boldsymbol{n} \in Y_n^d} & f_q(n, d; \boldsymbol{n}) \mathrm{Tr}[P_{\boldsymbol{n}} \rho_{A^n}] \\ & \times \mathbb{E}_{U_A \sim \mathrm{Haar}(d)} \Big[\Big(U_A \,\mathrm{diag}\Big(\frac{\boldsymbol{n}}{n}\Big) U_A^\dagger\Big)^{\otimes n}\Big], \end{aligned} \tag{23}$$

where we used the fact that $\tau_A^{\otimes n}$ is permutation invariant and thus satisfies $\mathcal{R}_{A^n} \circ \mathcal{S}_{A^n}(\tau_A^{\otimes n}) = \tau_A^{\otimes n}$. We see that Eq. (23) takes a similar form to Eq. (3) but with different coefficients $f_q(n, d; \boldsymbol{n})$. For comparison, let us first reformulate the right-hand side of Eq. (23) using Lemma 1 as

$$\begin{aligned} \rho_{A^n} \le f_q(n, d) & \sum_{\boldsymbol{n} \in Y_n^d} \mathrm{Tr}[P_{\boldsymbol{n}} \rho_{A^n}] \\ & \times \mathbb{E}_{U_A \sim \mathrm{Haar}(d)} \Big[\Big(U_A \,\mathrm{diag}\Big(\frac{\boldsymbol{n}}{n}\Big) U_A^\dagger\Big)^{\otimes n}\Big]. \end{aligned} \tag{24}$$

Then, the right-hand side is in the form of a probabilistic mixture of i.i.d. states up to the factor $f_q(n, d)$. When $n \gg d$, which is of our interest, the function $f_q(n, d)$ scales as $n^{(d^2-1)/2}$ whereas the coefficient in the conventional bound (3) scales as n^{d^2-1}. Our bound thus achieves the square-root improvement in this case.

The $\boldsymbol{n}$-dependent coefficient of the bound in Eq. (23) may also be beneficial when there is a restriction on the probability mass function $\mathrm{Tr}[P_{\boldsymbol{n}} \rho_{A^n}]$ over $\boldsymbol{n} \in Y_n^d$ or when one performs a quantum measurement on the state that may depend on the spectrum of the i.i.d. state. For example, consider the case $\mathrm{Tr}[P_{\boldsymbol{n}} \rho_{A^n}] = \delta_{\boldsymbol{n}\boldsymbol{n}_0}$ with $\boldsymbol{n}_0 = (n, 0, \ldots, 0)$, which amounts to restricting ρ_{A^n} supported only on the symmetric subspace $\mathrm{Sym}^n(\mathcal{H}_A)$ of $\mathcal{H}_A^{\otimes n}$. In this case, we have $\dim \mathcal{U}_{\boldsymbol{n}_0} = \binom{n+d-1}{n}$, $\dim \mathcal{V}_{\boldsymbol{n}_0} = 1$, and $s_{\boldsymbol{n}_0}(\boldsymbol{n}_0/n) = 1$, and the density operator $\mathrm{diag}(\boldsymbol{n}_0/\boldsymbol{n})$ corresponds to a pure state. From Eq. (23) and Lemma 1, we then have

$$\rho_{A^n} \le \binom{n+d-1}{n} \int d\phi \, |\phi\rangle\langle\phi|_A^{\otimes n}, \tag{25}$$

where $d\phi$ denotes the Haar measure on the set of pure states in the system A. This inequality is exactly what was proved in Ref. [20]. Combining the above with the fact that any permutation-invariant state in the system A^n has a purification in the symmetric subspace $\mathrm{Sym}^n(\mathcal{H}_A \otimes \mathcal{H}_B)$ with $\mathcal{H}_B \cong \mathcal{H}_A$ [15, 20], they derived Eq. (3). Our Theorem 1 can thus be regarded as a generalization of the operator inequality obtained in Ref. [20].

The reason why we used the support projection in Theorem 1 is to make it composable when two or more parts of the system undergo separate permutations. Consider a system A^nB^m and a state $\rho_{A^nB^m}$ that is invariant under the permutations of n systems in A^n and also is invariant under the permutations of m systems in B^m. In such a case, a pair of inequalities like

$\rho_{A^nB^m} \le X_{A^n} \otimes I_{B^m}$ and $\rho_{A^nB^m} \le I_{A^n} \otimes X_{B^m}$ would not imply $\rho_{A^nB^m} \le X_{A^n} \otimes X_{B^m}$. On the other hand, in the case of support projections, it holds that $\mathrm{supp}(\rho_{A^nB^m}) \le \mathrm{supp}(\rho_{A^n}) \otimes \mathrm{supp}(\rho_{B^m})$. Thanks to this property, we can apply Theorem 1 with $(A, n, E) \Rightarrow (A, n, B^n)$ and $(A, n, E) \Rightarrow (B, m, A^n)$ to obtain

$$\mathcal{R}_{A^n} \circ \mathcal{S}_{A^n} \otimes \mathcal{R}_{B^m} \circ \mathcal{S}_{B^m}(\rho_{A^nB^m}) = \rho_{A^nB^m}, \tag{26}$$

and for any $\boldsymbol{n} \in Y_n^d$ and $\boldsymbol{m} \in Y_m^{d'}$, we have

$$\begin{aligned} &\mathcal{S}_{A^n} \otimes \mathcal{S}_{B^m}(P_{\boldsymbol{n}} \otimes P_{\boldsymbol{m}} \rho_{A^nB^m} P_{\boldsymbol{n}} \otimes P_{\boldsymbol{m}}) \\ &= P_{\boldsymbol{n}} \otimes P_{\boldsymbol{m}} \mathcal{S}_{A^n} \otimes \mathcal{S}_{B^m}(\rho_{A^nB^m}) P_{\boldsymbol{n}} \otimes P_{\boldsymbol{m}} \end{aligned} \tag{27}$$

and

$$\begin{aligned} &\mathrm{supp}(\mathcal{S}_{A^n} \otimes \mathcal{S}_{B^m}(P_{\boldsymbol{n}} \otimes P_{\boldsymbol{m}} \rho_{A^nB^m} P_{\boldsymbol{n}} \otimes P_{\boldsymbol{m}})) \\ &\le \mathrm{supp}(\mathcal{S}_{A^n}(P_{\boldsymbol{n}} \rho_{A^n} P_{\boldsymbol{n}})) \otimes \mathrm{supp}(\mathcal{S}_{B^m}(P_{\boldsymbol{m}} \rho_{B^m} P_{\boldsymbol{m}})) \end{aligned} \tag{28}$$

$$\begin{aligned} &\le f_q(n, d; \boldsymbol{n}) f_q(m, d'; \boldsymbol{m}) \\ &\quad \times \mathcal{S}_{A^n}\Big(\mathbb{E}_{U_A \sim \mathrm{Haar}(d)}\Big[\Big(U_A \,\mathrm{diag}\Big(\frac{\boldsymbol{n}}{n}\Big) U_A^\dagger\Big)^{\otimes n}\Big]\Big) \\ &\quad \otimes \mathcal{S}_{B^m}\Big(\mathbb{E}_{U_B \sim \mathrm{Haar}(d')}\Big[\Big(U_B \,\mathrm{diag}\Big(\frac{\boldsymbol{m}}{m}\Big) U_B^\dagger\Big)^{\otimes m}\Big]\Big), \end{aligned} \tag{29}$$

from which we can derive an inequality of the form $\rho_{A^nB^n} \le X_{A^n} \otimes X_{B^n}$ as we obtained Eq. (23).

Our next result is when a local system A has a (unitary) symmetry restriction. Suppose that a group G has a unitary action on $\mathcal{H}_A$ with $\dim \mathcal{H}_A = d$, namely, we have a group representation $\pi : G \to B(\mathcal{H}_A)$. Since any unitary representation of a group is completely decomposable, π is a direct sum of irreducible subrepresentations. Labeling the inequivalent representations appearing in the decomposition by integers $\{1, \dots, k\}$, the space $\mathcal{H}_A$ can be decomposed as

$$\mathcal{H}_A = \bigoplus_{j=1}^k \mathcal{H}_j^R \otimes \mathcal{H}_j^D, \tag{30}$$

under which the representation π is decomposed as

$$\pi(g) = \bigoplus_{j=1}^k \pi_j(g) \otimes I_j^D, \tag{31}$$

where $\pi_j : G \to B(H_j^R)$ is an irreducible representation with dimension $r_j = \dim \mathcal{H}_j^R$ and has degeneracy $d_j = \dim \mathcal{H}_j^D$ in π. These numbers are related to the total dimension d by

$$d = \sum_{j=1}^k r_j d_j. \tag{32}$$

We consider the case where each of the n systems in A^n is independently subject to the symmetry restriction associated with G. The symmetry of the whole n systems is then associated with the direct product $G^{\times n}$ and the tensor product representation $\pi^{\otimes n}$ of it. We say an operator X_{A^nE} acting on $\mathcal{H}_{A^n} \otimes \mathcal{H}_E$ is locally G-invariant when $\pi^{\otimes n}(\boldsymbol{g})\, X_{A^nE}\, \pi^{\otimes n}(\boldsymbol{g})^\dagger = X_{A^nE}$ holds for all $\boldsymbol{g} \in G^{\times n}$.

By decomposing the space of each subsystem in A^n via Eq. (30), the space $\mathcal{H}_A^{\otimes n}$ is decomposed into a sum over an n-tuple of integers $\boldsymbol{j} = (j_1, \dots, j_n)$ as

$$\mathcal{H}_A^{\otimes n} = \bigoplus_{\boldsymbol{j} \in \{1,\dots,k\}^{\times n}} \mathcal{H}_{\boldsymbol{j}}^R \otimes \mathcal{H}_{\boldsymbol{j}}^D \tag{33}$$

with

$$\mathcal{H}_{\boldsymbol{j}}^{R(D)} \coloneqq \bigotimes_{i=1}^n \mathcal{H}_{j_i}^{R(D)}. \tag{34}$$

The unitary $\pi^{\otimes n}(\boldsymbol{g})$ with $\boldsymbol{g} \in G^{\times n}$ acts on this space as

$$\pi^{\otimes n}(\boldsymbol{g}) = \bigoplus_{\boldsymbol{j} \in \{1,\dots,k\}^{\times n}} \pi_{\boldsymbol{j}}(\boldsymbol{g}) \otimes I_{\mathcal{H}_{\boldsymbol{j}}^D}, \tag{35}$$

where $\pi_{\boldsymbol{j}}(\boldsymbol{g}) = \bigotimes_{i=1}^n \pi_{j_i}(g_i)$ is an irreducible representation of $G^{\times n}$. We group the summation according to the frequency of each label $l \in \{1, \dots, k\}$ in the n-tuple $\boldsymbol{j}$, or the type of $\boldsymbol{j}$. For Z_n^k defined in Eq. (5), we define a map $\boldsymbol{\chi}$ as

$$\begin{aligned} \boldsymbol{\chi} : \{1, \dots, k\}^{\times n} &\to Z_n^k \\ \boldsymbol{j} &\mapsto (|\{i \in \{1, \dots, n\} : j_i = 1\}|, \\ &\qquad \dots, |\{i \in \{1, \dots, n\} : j_i = k\}|) \end{aligned} \tag{36}$$

and denote its preimage by

$$\mathcal{T}(\boldsymbol{m}) \coloneqq \boldsymbol{\chi}^{-1}(\boldsymbol{m}) \tag{37}$$

for $\boldsymbol{m} \in Z_n^k$. The cardinality $|\mathcal{T}(\boldsymbol{m})|$ of $\mathcal{T}(\boldsymbol{m})$ for $\boldsymbol{m} = (n_1, \dots, n_k)$ is given by

$$|\mathcal{T}(\boldsymbol{m})| = \frac{n!}{n_1! \cdots n_k!}. \tag{38}$$

Note that these sets give a grouping of the n-tuple $\boldsymbol{j}$, namely, $\{1,\dots,k\}^{\times n} = \bigsqcup_{\boldsymbol{m}\in Z_n^k} \mathcal{T}(\boldsymbol{m})$. Now, the space $\mathcal{H}_A^{\otimes n}$ is decomposed as

$$\mathcal{H}_A^{\otimes n} = \bigoplus_{\boldsymbol{m}\in Z_n^k} \bigoplus_{\boldsymbol{j}\in\mathcal{T}(\boldsymbol{m})} \mathcal{H}_{\boldsymbol{j}}^R \otimes \mathcal{H}_{\boldsymbol{j}}^D. \tag{39}$$

For $\boldsymbol{m} = (n_1,\dots,n_k) \in Z_n^k$, let us choose a representative $\boldsymbol{j}_0(\boldsymbol{m})$ of the set $\mathcal{T}(\boldsymbol{m})$ as

$$\boldsymbol{j}_0(\boldsymbol{m}) \coloneqq (\underbrace{1,\dots,1}_{n_1},\underbrace{2,\dots,2}_{n_2},\dots,\underbrace{k,\dots,k}_{n_k}). \tag{40}$$

For this particular sequence, the corresponding Hilbert space can simply be written as

$$\begin{aligned}
&\mathcal{H}_{\boldsymbol{j}_0(\boldsymbol{m})}^R \otimes \mathcal{H}_{\boldsymbol{j}_0(\boldsymbol{m})}^D \\
&= \bigotimes_{j=1}^k (\mathcal{H}_j^R \otimes \mathcal{H}_j^D)^{\otimes n_j} &(41)\\
&= \bigotimes_{j=1}^k (\mathcal{H}_j^R)^{\otimes n_j} \otimes \left(\bigoplus_{\boldsymbol{n}_j\in Y_{n_j}^{d_j}} \mathcal{U}_{\boldsymbol{n}_j} \otimes \mathcal{V}_{\boldsymbol{n}_j} \right), &(42)
\end{aligned}$$

where we applied the decomposition of Eq. (6) to $(\mathcal{H}_j^D)^{\otimes n_j}$. Other members of $\mathcal{T}(\boldsymbol{m})$ are connected to $\boldsymbol{j}_0(\boldsymbol{m})$ by permutations. That is to say, we may choose a permutation $\sigma_{\boldsymbol{j}}$ for each $\boldsymbol{j} \in \{1,\dots,k\}^{\times n}$ to satisfy

$$\sigma_{\boldsymbol{j}}(\boldsymbol{j}_0(\boldsymbol{\chi}(\boldsymbol{j}))) = \boldsymbol{j}. \tag{43}$$

Such a choice is not unique in general, but the following argument holds independent of the choice. Let $V_{\sigma_{\boldsymbol{j}}} = V_{\sigma_{\boldsymbol{j}}}^R \otimes V_{\sigma_{\boldsymbol{j}}}^D$ be a unitary representation of the permutation $\sigma_{\boldsymbol{j}}$, where $V_{\sigma_{\boldsymbol{j}}}^{R(D)}$ acts on the Hilbert space $(\bigoplus_{j=1}^k \mathcal{H}_j^{R(D)})^{\otimes n}$. Then, we have

$$\mathcal{H}_{\boldsymbol{j}}^{R(D)} = V_{\sigma_{\boldsymbol{j}}}^{R(D)} \mathcal{H}_{\boldsymbol{j}_0(\boldsymbol{\chi}(\boldsymbol{j}))}^{R(D)}. \tag{44}$$

We note here that the Hilbert space $(\bigoplus_{j=1}^k \mathcal{H}_j^R)^{\otimes n} \otimes (\bigoplus_{j=1}^k \mathcal{H}_j^D)^{\otimes n}$ on which the action of $V_{\sigma_{\boldsymbol{j}}}$ is defined is larger than the original Hilbert space $\mathcal{H}_A^{\otimes n}$. However, it keeps the subspace $\mathcal{H}_A^{\otimes n}$ invariant, and thus we can consider the restriction of $V_{\sigma_{\boldsymbol{j}}}$ onto $\mathcal{H}_A^{\otimes n}$. Combining Eqs. (33), (42), and (44), we arrive at the decomposition associated with the permutation group and the group $G^{\times n}$:

$$\mathcal{H}_A^{\otimes n} = \bigoplus_{\substack{\boldsymbol{m}=(n_1,\dots,n_k)\\ \in Z_n^k}} \bigoplus_{\substack{(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)\\ \in Y_{n_1}^{d_1}\cdots\times Y_{n_k}^{d_k}}} \mathcal{H}_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}, \tag{45}$$

with

$$\mathcal{H}_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k} \coloneqq \sum_{\boldsymbol{j}\in\mathcal{T}(\boldsymbol{m})} V_{\sigma_{\boldsymbol{j}}} \left[\bigotimes_{j=1}^k (H_j^R)^{\otimes n_j} \otimes (\mathcal{U}_{\boldsymbol{n}_j} \otimes \mathcal{V}_{\boldsymbol{n}_j}) \right]. \tag{46}$$

Let $P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}$ be the projection onto $\mathcal{H}_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}$. Our next theorem can then be stated as follows.

Theorem 2. *Let G be a group that has a unitary action on a Hilbert space $\mathcal{H}_A$ with $\dim \mathcal{H}_A = d$. Then, under the decomposition of Eq. (45), any density operator $\rho_{A^nE}^G$ that is invariant under permutation of n local systems of A^n and also locally G-invariant satisfies*

$$\rho_{A^nE}^G = \sum_{\boldsymbol{m}\in Z_n^k} \sum_{\substack{(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)\\ \in Y_{n_1}^{d_1}\cdots\times Y_{n_k}^{d_k}}} \rho_{A^nE}^G(\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k), \tag{47}$$

where

$$\rho_{A^nE}^G(\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k) \coloneqq P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k} \rho_{A^nE}^G P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}. \tag{48}$$

Furthermore, there exists a pair $(\mathcal{S}_{A^n}^G, \mathcal{R}_{A^n}^G)$ of CPTP maps that satisfies the following conditions (i), (ii), and (iii).

(i) For any density operator ρ_{A^nE} and for any $\boldsymbol{m} \in Z_n^k$ and $(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k) \in Y_{n_1}^{d_1} \times \cdots Y_{n_k}^{d_k}$, we have

$$\begin{aligned}
&\mathcal{S}_{A^n}^G \otimes \mathrm{Id}_E(P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k} \rho_{A^nE} P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}) \\
&= P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k} \mathcal{S}_{A^n}^G \otimes \mathrm{Id}_E(\rho_{A^nE}) P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}
\end{aligned} \tag{49}$$

(ii) For any $\rho_{A^nE}^G$ that is invariant under permutation of n local systems of A^n as well as locally G-invariant, we have

$$\mathcal{R}_{A^n}^G \circ \mathcal{S}_{A^n}^G \otimes \mathrm{Id}_E(\rho_{A^nE}^G) = \rho_{A^nE}^G. \tag{50}$$

(iii) For any $\rho_{A^n}^G$ that is invariant under permutation of n local systems of A^n as well as locally G-invariant and for any $\boldsymbol{m} \in Z_n^k$ and $(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k) \in Y_{n_1}^{d_1} \times \cdots Y_{n_k}^{d_k}$, we have

$$\begin{aligned}
&\mathrm{supp}\Big(\mathcal{S}_{A^n}^G(\rho_{A^n}^G(\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k))\Big) \\
&\le f_c(n,k;\boldsymbol{m}) \prod_{j=1}^k f_q(n_j,d_j;\boldsymbol{n}_j) \\
&\quad \times \mathcal{S}_{A^n}^G\big(\mathbb{E}_{\boldsymbol{U}\sim\times_{j=1}^k \mathrm{Haar}(d_j)}[\tau_{\boldsymbol{U}}(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)^{\otimes n}]\big),
\end{aligned} \tag{51}$$

where the density operator $\tau_{\boldsymbol{U}}(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)$ *is defined as*

$$\tau_{\boldsymbol{U}}(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k) \coloneqq \bigoplus_{j=1}^{k} \frac{I_{\mathcal{H}_j^R}}{r_j} \otimes \Big(U_j \operatorname{diag}\Big(\frac{\boldsymbol{n}_j}{n}\Big) U_j^\dagger\Big)_{\mathcal{H}_j^D}. \tag{52}$$

Again, as a direct consequence of the theorem above with the fact that

$$\begin{aligned}
&\mathcal{S}^G_{A^n}(\rho^G_{A^n}) \\
&= \sum_{\boldsymbol{m}\in Z_n^k} \sum_{\substack{(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)\\ \in Y_{n_1}^{d_1}\times\cdots\times Y_{n_k}^{d_k}}} \mathcal{S}^G_{A^n}(\rho^G_{A^n}(\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)) \qquad (53)\\
&\le \sum_{\boldsymbol{m}\in Z_n^k} \sum_{\substack{(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)\\ \in Y_{n_1}^{d_1}\times\cdots\times Y_{n_k}^{d_k}}} \operatorname{Tr}[\rho^G_{A^n}(\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)] \\
&\qquad \times \operatorname{supp}\Big(\mathcal{S}^G_{A^n}(\rho^G_{A^n}(\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k))\Big) \qquad (54)
\end{aligned}$$

always holds, we have

$$\begin{aligned}
\rho^G_{A^n} \le \sum_{\boldsymbol{m}\in Z_n^k} \sum_{\substack{(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)\\ \in Y_{n_1}^{d_1}\times\cdots\times Y_{n_k}^{d_k}}} & f_c(n,k;\boldsymbol{m}) \\
& \prod_{j=1}^{k} f_q(n_j,d_j;\boldsymbol{n}_j) \operatorname{Tr}\Big[P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}\rho^G_{A^n}\Big] \\
& \times \mathbb{E}_{\boldsymbol{U}\sim\times_{j=1}^k \operatorname{Haar}(d_j)}[\tau_{\boldsymbol{U}}(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)^{\otimes n}]. \qquad (55)
\end{aligned}$$

One may apply Lemma 1 to obtain an elementary expression for the prefactor.

These two theorems enable us to reduce the problems of concentration inequalities on the permutation-invariant states to those of i.i.d. states up to a polynomial factor. This factor can be negligible for large n since a concentration inequality typically yields exponential decay over n. Concentration inequalities on the i.i.d. state are relatively simple and tight bounds are known for these cases, which is beneficial for the finite-size analysis of the permutation-invariant states.

Our final result considers a different scenario from the previous two results. Let us assume that an adversary prepares the quantum state ρ^n in the n-composite system. We perform independent and identical measurements $\mathcal{M}^{\otimes n}$ on the state. The following theorem holds for the (classical) probability mass function of the measurement outcomes.

Theorem 3. *Consider a k-outcome measurement on a quantum system associated with a Hilbert space $\mathcal{H}$, specified by the POVM $\{M(i)\}_{i\in\mathcal{X}}$ with $\mathcal{X} \coloneqq \{1,\dots,k\}$. Let $\mathcal{D}(\mathcal{H})$ be the set of all the density operators on the Hilbert space $\mathcal{H}$. For $\rho \in \mathcal{D}(\mathcal{H})$, let $\mathcal{M}(\rho) : i \mapsto \operatorname{Tr}[M(i)\rho]$ be the probability mass function for the outcome of the measurement. Consider a probability simplex $\mathcal{P} \subseteq \mathbb{R}^k$ and define the region $\mathcal{Q} \subseteq \mathcal{P}$ as*

$$\begin{aligned}
\mathcal{Q} \coloneqq \{(p_1,\dots,p_k) \in \mathcal{P} :& \\
&\exists \rho \in \mathcal{D}(\mathcal{H}) \text{ s.t. } \forall i \in \mathcal{X}, p_i = \mathcal{M}(\rho)(i)\}. \qquad (56)
\end{aligned}$$

Assume $\boldsymbol{p}_1 \coloneqq (1,0,\dots,0) \notin \mathcal{Q}$. Let $\mathcal{R} \subseteq \mathcal{P}$ be a region defined as

$$\mathcal{R} \coloneqq \{\boldsymbol{p} \in \mathcal{P} : \boldsymbol{u}\cdot\boldsymbol{p} \le 0\}, \tag{57}$$

with $\boldsymbol{u} \coloneqq (1,-b_2,\dots,-b_k)$, where $b_2,\dots,b_k$ are nonnegative numbers chosen to satisfy $\mathcal{Q} \subseteq \mathcal{R}$. With $\boldsymbol{v} \coloneqq (1,1,\dots,1)/\sqrt{k}$, define $\mathcal{R}'$ as

$$\mathcal{R}' \coloneqq \left\{\boldsymbol{p} \in \mathcal{P} : \boldsymbol{u}\cdot\boldsymbol{p} \le \frac{\sqrt{2}\|\boldsymbol{u} - (\boldsymbol{v}\cdot\boldsymbol{u})\boldsymbol{v}\|}{n}\right\}, \tag{58}$$

where $\|\boldsymbol{a}\|$ denotes the Euclidean norm of the vector $\boldsymbol{a}$.

Consider the measurement $\mathcal{M}^{\otimes n}$ on n systems A^n, which has the POVM $\{M^{\otimes n}(\boldsymbol{i})\}_{\boldsymbol{i}\in\mathcal{X}^{\times n}}$ with $M^{\otimes n}((i_1,\dots,i_n)) = M(i_1)\otimes\cdots\otimes M(i_n)$. Let $\mathcal{A} \subseteq \mathcal{P}$ be a non-empty closed convex region of the probability simplex. Then, for any quantum state $\rho \in \mathcal{D}(\mathcal{H}^{\otimes n})$, the probability mass function $\mathcal{M}^{\otimes n}(\rho) : \boldsymbol{i} \mapsto \operatorname{Tr}[M^{\otimes n}(\boldsymbol{i})\rho]$ satisfies

$$\begin{aligned}
&\sum_{\boldsymbol{i}:\boldsymbol{\chi}(\boldsymbol{i})/n\in\mathcal{A}} \mathcal{M}^{\otimes n}(\rho)(\boldsymbol{i}) \\
&\le \max_{\boldsymbol{p}\in\mathcal{A},\boldsymbol{q}\in\mathcal{R}'} f_c(n,k)\exp[-nD(\boldsymbol{p}\|\boldsymbol{q})],
\end{aligned} \tag{59}$$

where $D(\cdot\|\cdot)$ denotes the Kullback-Leibler divergence, and the functions $\boldsymbol{\chi}(\boldsymbol{i})$ and $f_c(n,k)$ are defined respectively in Eqs. (36) and (14).

Geometrical relations between the sets $\mathcal{Q}$, $\mathcal{R}$, $\mathcal{R}'$, and $\mathcal{A}$ are illustrated in Fig. 1. This theorem gives a Sanov-type bound [2, 3] even though the quantum state is prepared by an adversary. Since Sanov-type inequalities are in general tighter than Azuma-Hoeffding-type inequalities [28, 14], which can also be applied in this setup,

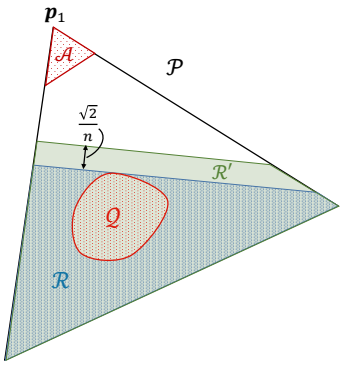

Figure 1: Geometrical relations between the (convex) subsets $\mathcal{Q}$, $\mathcal{R}$, $\mathcal{R}'$, and $\mathcal{A}$ of $\mathcal{P}$. The subset $\mathcal{A}$ is non-empty, closed, and convex. The subset $\mathcal{R}$ is bounded by a hyperplane and contains $\mathcal{Q}$ and all the extremal points of $\mathcal{P}$ except $\boldsymbol{p}_1 = (1, 0, \ldots, 0)$. The set $\mathcal{R}'$ consists of every point whose distance to $\mathcal{R}$ is no larger than $\sqrt{2}/n$.

it may be useful in, for example, finite-size analyses of quantum cryptography. Note that Theorem 3 also holds for a classical probability mass function $\mathcal{M}^n$ instead of $\mathcal{M}^{\otimes n}(\rho)$ as long as all the conditional probability mass function of $\mathcal{M}^n$ conditioned on $n-1$ outcomes are contained in a set $\mathcal{Q} \subseteq \mathcal{P}$, which replaces the precondition Eq. (56) in this case.

1.3 Organization of the paper

In Sec. 2, we prove Theorems 1–3 stated above, where each subsection is assigned for each theorem. In Sec. 3, we apply our results to simple examples and numerically demonstrate the tightness of our bounds by comparison with conventionally used bounds. The first example in Sec. 3.1 is an estimation task of measurement outcomes with two non-orthogonal projections, which is not directly related to real applications but is still informative. The second example is an application to the quantum key distribution protocol. Finally, in Sec. 4, we summarize our results and discuss possible applications.

2 Proofs of the main theorems

2.1 A permutation-invariant quantum state is bounded by a mixture of i.i.d. quantum states

In this section, we prove Lemma 1 and Theorem 1, which constructs an upper bound on a density operator of a permutation-invariant state with density operators of i.i.d. states up to a multiplication factor. This has been intensively studied in the context of the quantum de Finetti theorem (see e.g. [19] for a review) under the name of post-selection technique [15, 20, 21]. Ours can be regarded as a refinement and a generalization of the bound.

As defined prior to Theorem 1, let $\mathcal{H}_A$ be a d-dimensional Hilbert space, Y_n^d be the set of Young diagrams with n boxes and at most d rows whose elements can be labeled by $\boldsymbol{n} = (n_1, n_2, \ldots, n_d)$ with $n_1 \geq n_2 \geq \cdots n_d$ and $\sum_{i=1}^d n_i = n$. As shown in Eq. (6), $\mathcal{H}_A^{\otimes n}$ can be decomposed into the direct sum of the subspaces in the form $\mathcal{U}_{\boldsymbol{n}} \otimes \mathcal{V}_{\boldsymbol{n}}$ labeled by $\boldsymbol{n} \in Y_n^d$. Let $P_{\boldsymbol{n}}$ be a projection operator onto this subspace. Let ρ_{A^nE} be a state on $\mathcal{H}_A^{\otimes n} \otimes \mathcal{H}_E$ that is invariant under the permutation of n local systems of A^n. Then, from Schur's lemma, ρ_{A^nE} has the form

$$\rho_{A^nE} = \bigoplus_{\boldsymbol{n} \in Y_n^d} \mathrm{Tr}[P_{\boldsymbol{n}} \rho_{A^nE}] \, \rho_{\mathcal{U}_{\boldsymbol{n}}E} \otimes \frac{I_{\mathcal{V}_{\boldsymbol{n}}}}{\dim \mathcal{V}_{\boldsymbol{n}}}, \quad (60)$$

where $\rho_{\mathcal{U}_{\boldsymbol{n}}E}$ denotes a density operator on the subspace $\mathcal{U}_{\boldsymbol{n}}$ and E, and $I_{\mathcal{V}_{\boldsymbol{n}}}$ is an identity operator on the subspace $\mathcal{V}_{\boldsymbol{n}}$. From this, it is obvious that the state ρ_{A^nE} satisfies Eq. (7). We define a CPTP map $\mathcal{S}_{A^n}$ as follows:

$$\mathcal{S}_{A^n}(\tau_{A^n}) = \bigoplus_{\boldsymbol{n} \in Y_n^d} \mathrm{Tr}_{\mathcal{V}_{\boldsymbol{n}}}^{(\boldsymbol{n})}(P_{\boldsymbol{n}} \tau_{A^n} P_{\boldsymbol{n}}|_{\mathcal{U}_{\boldsymbol{n}} \otimes \mathcal{V}_{\boldsymbol{n}}}) \otimes |0\rangle\langle 0|_{\mathcal{V}_{\boldsymbol{n}}}, \quad (61)$$

where $\mathrm{Tr}^{(\boldsymbol{n})}$ denotes the trace acting on a matrix on $\mathcal{U}_{\boldsymbol{n}} \otimes \mathcal{V}_{\boldsymbol{n}}$, $\mathrm{Tr}_{\mathcal{V}_{\boldsymbol{n}}}^{(\boldsymbol{n})}$ denotes the partial trace over the space $\mathcal{V}_{\boldsymbol{n}}$, and $|0\rangle\langle 0|_{\mathcal{V}_{\boldsymbol{n}}}$ is an arbitrary pure state on $\mathcal{V}_{\boldsymbol{n}}$. We also define another CPTP map $\mathcal{R}_{A^n}$ as follows:

$$\mathcal{R}_{A^n}(\tau_{A^n}) = \bigoplus_{\boldsymbol{n} \in Y_n^d} \mathrm{Tr}_{\mathcal{V}_{\boldsymbol{n}}}^{(\boldsymbol{n})}(P_{\boldsymbol{n}} \tau_{A^n} P_{\boldsymbol{n}}|_{\mathcal{U}_{\boldsymbol{n}} \otimes \mathcal{V}_{\boldsymbol{n}}}) \otimes \frac{I_{\mathcal{V}_{\boldsymbol{n}}}}{\dim \mathcal{V}_{\boldsymbol{n}}}. \quad (62)$$

Now, let $\mathrm{diag}(\boldsymbol{t})$ be a density operator represented by a diagonal matrix with the spectrum $\boldsymbol{t} = (t_1, \ldots, t_d)$ in a fixed basis of $\mathcal{H}_A$. Without loss of generality, we assume $t_1 \geq \cdots \geq t_d (\geq 0)$. Then, it is known that [29, 30, 31]

$$\mathrm{Tr}[P_{\boldsymbol{n}} \mathrm{diag}(\boldsymbol{t})^{\otimes n}] = \dim \mathcal{V}_{\boldsymbol{n}} \; s_{\boldsymbol{n}}(\boldsymbol{t}), \quad (63)$$

where $s_{\boldsymbol{n}}$ is the Schur function defined in Eq. (8). Then, let us consider, instead of $\mathrm{diag}(\boldsymbol{t})^{\otimes n}$ in Eq. (63), the state $\mathbb{E}_{U_A\sim\mathrm{Haar}(d)}[(U_A\mathrm{diag}(\boldsymbol{t})U_A^\dagger)^{\otimes n}]$, where $\mathbb{E}_{U\sim\mathrm{Haar}(d)}[\cdot]$ denotes the Haar average over the d-dimensional unitary group $\mathrm{U}(d)$. This state commutes with any unitary in the form $U'^{\otimes n}$, $(U'\in\mathrm{U}(d))$, as well as any permutation of the systems, and thus satisfies

$$P_{\boldsymbol{n}}\mathbb{E}_{U_A\sim\mathrm{Haar}(d)}[(U_A\mathrm{diag}(\boldsymbol{t})U_A^\dagger)^{\otimes n}]P_{\boldsymbol{n}}=\alpha_{\boldsymbol{n}}(\boldsymbol{t})P_{\boldsymbol{n}}, \tag{64}$$

for coefficients $\alpha_{\boldsymbol{n}}(\boldsymbol{t})$ satisfying $\sum_{\boldsymbol{n}\in Y_n^d}\alpha_{\boldsymbol{n}}(\boldsymbol{t})=1$. (This is because $U'^{\otimes n}$ acts as $\pi_{\boldsymbol{n}}(U')\otimes I_{\mathcal{V}_{\boldsymbol{n}}}$ on $\mathcal{U}_{\boldsymbol{n}}\otimes\mathcal{V}_{\boldsymbol{n}}$, where $\pi_{\boldsymbol{n}}$ is an irreducible representation of $\mathrm{U}(d)$ labeled by $\boldsymbol{n}$.) By taking the trace of both sides, we have

$$\begin{aligned}&\mathrm{Tr}\Big[P_{\boldsymbol{n}}\mathbb{E}_{U_A\sim\mathrm{Haar}(d)}\Big[\big(U_A\mathrm{diag}(\boldsymbol{t})U_A^\dagger\big)^{\otimes n}\Big]\Big]\\&=\alpha_{\boldsymbol{n}}(\boldsymbol{t})\dim(\mathcal{U}_{\boldsymbol{n}}\otimes\mathcal{V}_{\boldsymbol{n}}).\end{aligned} \tag{65}$$

Since a unitary does not change the spectra, we have

$$\begin{aligned}&\mathrm{Tr}\Big(P_{\boldsymbol{n}}\mathbb{E}_{U_A\sim\mathrm{Haar}(d)}[(U_A\mathrm{diag}(\boldsymbol{t})U_A^\dagger)^{\otimes n}]\Big)\\&=\mathbb{E}_{U_A\sim\mathrm{Haar}(d)}\Big[\mathrm{Tr}(P_{\boldsymbol{n}}(U_A\mathrm{diag}(\boldsymbol{t})U_A^\dagger)^{\otimes n})\Big] &(66)\\&=\mathbb{E}_{U_A\sim\mathrm{Haar}(d)}[\dim\mathcal{V}_{\boldsymbol{n}}\ s_{\boldsymbol{n}}(\boldsymbol{t})] &(67)\\&=\dim\mathcal{V}_{\boldsymbol{n}}\ s_{\boldsymbol{n}}(\boldsymbol{t}). &(68)\end{aligned}$$

Therefore, from Eqs. (65) and (68), we have

$$\alpha_{\boldsymbol{n}}(\boldsymbol{t})=\frac{s_{\boldsymbol{n}}(\boldsymbol{t})}{\dim\mathcal{U}_{\boldsymbol{n}}}. \tag{69}$$

Now, from Eqs. (60) and (61), we have

$$\mathrm{supp}(\mathcal{S}_{A^n}(P_{\boldsymbol{n}}\rho_{A^n}P_{\boldsymbol{n}}))=\begin{cases}\mathrm{supp}(\rho_{\mathcal{U}_{\boldsymbol{n}}})\otimes|0\rangle\langle 0|_{\mathcal{V}_{\boldsymbol{n}}} & \text{on } \mathcal{U}_{\boldsymbol{n}}\otimes\mathcal{V}_{\boldsymbol{n}}\\ 0 & \text{otherwise}\end{cases}.$$

Combining the above with Eqs.(61), (64), and (69), we have

$$\begin{aligned}&\mathrm{supp}(\mathcal{S}_{A^n}(P_{\boldsymbol{n}}\rho_{A^n}P_{\boldsymbol{n}}))\\&\le\frac{\mathcal{S}_{A^n}(P_{\boldsymbol{n}})}{\dim\mathcal{V}_{\boldsymbol{n}}} &(70)\\&=\frac{1}{\alpha_{\boldsymbol{n}}(\boldsymbol{t})\dim\mathcal{V}_{\boldsymbol{n}}}\\&\quad\times\mathcal{S}_{A^n}(P_{\boldsymbol{n}}\mathbb{E}_{U_A\sim\mathrm{Haar}(d)}[(U_A\mathrm{diag}(\boldsymbol{t})U_A^\dagger)^{\otimes n}]P_{\boldsymbol{n}}) &(71)\\&\le\frac{\dim\mathcal{U}_{\boldsymbol{n}}}{s_{\boldsymbol{n}}(\boldsymbol{t})\dim\mathcal{V}_{\boldsymbol{n}}}\\&\quad\times\mathcal{S}_{A^n}(\mathbb{E}_{U_A\sim\mathrm{Haar}(d)}[(U_A\mathrm{diag}(\boldsymbol{t})U_A^\dagger)^{\otimes n}]). &(72)\end{aligned}$$

We thus obtain the following from Eqs. (60), (61), (62), and (72).

Lemma 2. *Consider a composite system A^nE associated with a Hilbert space $\mathcal{H}_A^n\otimes\mathcal{H}_E$ with $\dim\mathcal{H}_A=d$. Then, there exists a pair $(\mathcal{S}_{A^n},\mathcal{R}_{A^n})$ of CPTP maps that satisfies the following conditions (i), (ii), and (iii).*

(i) For any density operator ρ_{A^nE} and for any $\boldsymbol{n}\in Y_n^d$, we have

$$\mathcal{S}_{A^n}\otimes\mathrm{Id}_E(P_{\boldsymbol{n}}\rho_{A^nE}P_{\boldsymbol{n}})=P_{\boldsymbol{n}}\mathcal{S}_{A^n}\otimes\mathrm{Id}_E(\rho_{A^nE})P_{\boldsymbol{n}}. \tag{73}$$

(ii) For any ρ_{A^nE} that is invariant under permutation of n subsystems of A^n, the pair $(\mathcal{S}_{A^n},\mathcal{R}_{A^n})$ satisfies

$$\mathcal{R}_{A^n}\circ\mathcal{S}_{A^n}\otimes\mathrm{Id}_E(\rho_{A^nE})=\rho_{A^nE}. \tag{74}$$

(iii) For any ρ_{A^n} that is invariant under permutation of n subsystems of A^n and for any $\boldsymbol{n}\in Y_n^d$, we have

$$\begin{aligned}&\mathrm{supp}(\mathcal{S}_{A^n}(P_{\boldsymbol{n}}\rho_{A^n}P_{\boldsymbol{n}}))\\&\le\min_{\boldsymbol{t}=(t_1,\dots,t_d)\ge 0:\sum_{i=1}^d t_i=1}\frac{\dim\mathcal{U}_{\boldsymbol{n}}}{s_{\boldsymbol{n}}(\boldsymbol{t})\dim\mathcal{V}_{\boldsymbol{n}}}\\&\quad\times\mathcal{S}_{A^n}\Big(\mathbb{E}_{U_A\sim\mathrm{Haar}(d)}\Big[\big(U_A\,\mathrm{diag}(\boldsymbol{t})\,U_A^\dagger\big)^{\otimes n}\Big]\Big)\end{aligned} \tag{75}$$

Since $\boldsymbol{t}=\boldsymbol{n}/n$ is a suboptimal choice, Theorem 1 directly follows from Lemma 2. In fact, this choice is fairly good when $n\gg 1$, which is revealed in the following. Now we move to the proof of Lemma 1. First, it is known that the dimension of $\mathcal{V}_{\boldsymbol{n}}$ can be represented as [29, 30]

$$\dim\mathcal{V}_{\boldsymbol{n}}=\frac{n!}{\prod_{(i,j)}h_{\boldsymbol{n}}(i,j)}, \tag{76}$$

where $h_{\boldsymbol{n}}(i,j)$ denotes the hook length of the Young diagram at (i,j), i.e., i-th row and j-th column. On the other hand, the dimension of $\mathcal{U}_{\boldsymbol{n}}$ is given by [29, 30]

$$\dim\mathcal{U}_{\boldsymbol{n}}=\prod_{(i,j)}\frac{d-i+j}{h_{\boldsymbol{n}}(i,j)}, \tag{77}$$

and therefore,

$$\frac{\dim\mathcal{U}_{\boldsymbol{n}}}{\dim\mathcal{V}_{\boldsymbol{n}}}=\prod_{(i,j)}\frac{d-i+j}{n!}=\frac{n_1!\dots n_d!}{n!}\prod_{i=1}^d\binom{d-i+n_i}{d-i}, \tag{78}$$

which holds even when $n_j = \cdots = n_d = 0$ for $1 \le j \le d$. Next, we use an alternative expression for the Schur function $s_{\boldsymbol{n}}(\boldsymbol{t})$ defined in Eq. (8). Let $T_{\boldsymbol{n}}$ be the set of semistandard Young tableaux of shape $\boldsymbol{n}$, i.e., Young diagrams filled with integers from 1 to d non-decreasingly across each row and increasingly down each column. Then, we have [29, 30]

$$s_{\boldsymbol{n}}(\boldsymbol{t}) = \sum_{T_{\boldsymbol{n}}} t_1^{w_1} \cdots t_d^{w_d}, \tag{79}$$

where w_i, $i = 1, \ldots, d$, denotes the weight of a given Young tableau, i.e., the occurrences of the number i in the Young tableau. Note that we define $0^0 = 1$ here. The largest element in the above summation is given by filling i-th row with i, which then implies

$$s_{\boldsymbol{n}}(\boldsymbol{t}) \ge t_1^{n_1} \cdots t_d^{n_d} = \Pi_{i=1}^n t_i^{n_i}. \tag{80}$$

Combining this with Eq. (78), we have

$$\frac{\dim \mathcal{U}_{\boldsymbol{n}}}{\dim \mathcal{V}_{\boldsymbol{n}}\, s_{\boldsymbol{n}}(\boldsymbol{t})} \le \frac{1}{\mathrm{Mult}_{\boldsymbol{t}}(\boldsymbol{n})} \prod_{i=0}^{d-1} \binom{n+i}{i}, \tag{81}$$

where $\mathrm{Mult}_{\boldsymbol{t}}$ is defined in Eq. (12). Now, it is clear that the right-hand side is minimized by the choice $\boldsymbol{t} = \boldsymbol{n}/n$, which is the reason why Eq. (19) in Theorem 1 may be a fairly tight bound. Furthermore, $[\mathrm{Mult}_{\boldsymbol{n}/n}(\boldsymbol{n})]^{-1} = f_c(n, d; \boldsymbol{n})$ is bounded from above as

$$\begin{aligned} f_c(n, d; \boldsymbol{n}) &= [\mathrm{Mult}_{\boldsymbol{n}/n}(\boldsymbol{n})]^{-1} && (82)\\ &= \frac{n_1! \cdots n_d!}{n!} \left(\frac{n}{n_1}\right)^{n_1} \cdots \left(\frac{n}{n_d}\right)^{n_d} && (83)\\ &\le \frac{e^d \sqrt{n_1 \cdots n_d}}{\sqrt{2\pi n}} && (84)\\ &\le \frac{n^{(d-1)/2}}{\sqrt{2\pi (d/e^2)^d}} = f_c(n, d), && (85) \end{aligned}$$

where we used Stirling's inequality in the first inequality and used the relation between the arithmetic mean and the geometric mean in the second inequality. This proves Eq. (16) in Lemma 1. (Here, we prove Eq. (16) for $\boldsymbol{n} \in Y_n^d$, but this can obviously be generalized to $\boldsymbol{n} \in Z_n^d$ from the definition of $\mathrm{Mult}_{\boldsymbol{t}}(\boldsymbol{n})$ in Eq. (12).) We thus have

$$\begin{aligned} f_q(n, d; \boldsymbol{n}) &= \frac{\dim \mathcal{U}_{\boldsymbol{n}}}{\dim \mathcal{V}_{\boldsymbol{n}}\, s_{\boldsymbol{n}}(\boldsymbol{n}/n)}\\ &\le \frac{n^{\frac{d-1}{2}}}{\sqrt{2\pi (d/e^2)^d}} \prod_{i=0}^{d-1} \binom{n+i}{i} && (86)\\ &\le \frac{n^{\frac{d-1}{2}}}{\sqrt{2\pi (d/e^2)^d}} \prod_{i=0}^{d-1} \frac{(n+d-1)^i}{i!} && (87)\\ &\le \frac{(n+d-1)^{\frac{(d^2-1)}{2}}}{\sqrt{2\pi (d/e^2)^d} G(d+1)} = f_q(n, d), && (88) \end{aligned}$$

where $G(n)$ denotes the Barnes G-function defined in Eq. (11). Notice that we used the following trivial inequalities from Eq. (86) to Eq. (87)

$$\binom{n+i}{i} = \frac{\overbrace{(n+i)\cdots(n+1)}^{i}}{i!} \le (n+d-1)^i / i!. \tag{89}$$

(Note: the bound $\binom{n+i}{i} \le (n+1)^i$ is conventionally used for this context, but this leads to a looser bound in the case $n \gg d$, which is of our interest.) This proves Eqs. (15) in Lemma 1, which completes the proof of Theorem 1 and Lemma 1. □

Theorem 1 may have lots of applications. One immediate example is the following. Assume that a positive operator Π satisfies $\mathrm{Tr}(\Pi \rho^{\otimes n}) \le \mathcal{O}(\exp(-n))$ for any ρ. Such a positive operator Π frequently appears in a quantum key distribution or a quantum state verification. For such a scenario, we have, for any permutation-invariant state ρ_{A^n},

$$\begin{aligned} \mathrm{Tr}[\Pi \rho_{A^n}] &\le f_q(n, d) \sum_{\boldsymbol{n} \in Y_n^d} \mathrm{Tr}[P_{\boldsymbol{n}} \rho_{A^n}]\\ &\quad \times \mathrm{Tr}\left(\Pi\, \mathbb{E}_{U \sim \mathrm{Haar}(d)} \left[\left(U \mathrm{diag}\left(\frac{\boldsymbol{n}}{n}\right) U^\dagger\right)^{\otimes n}\right]\right), && (90)\\ &\le f_q(n, d) \sum_{\boldsymbol{n} \in Y_n^d} \mathrm{Tr}[P_{\boldsymbol{n}} \rho_{A^n}] \sup_\rho \mathrm{Tr}(\Pi \rho^{\otimes n}), && (91)\\ &\le f_q(n, d) \mathcal{O}(\exp(-n)), && (92) \end{aligned}$$

and thus almost the same exponential bound holds (up to a polynomial $f_q(n, d)$). The above inequality will be used later.

2.2 Generalization to the case with symmetry restrictions

In this section, we generalize the result of the previous section to the case in which a local d-dimensional system has an additional symmetry. There are lots of relevant situations in quantum information applications in which a local system has a symmetry. A classical-quantum state, for example, can be regarded as a quantum state having a U(1) symmetry in a part of a system. To make it more concrete, we here derive a tighter upper bound on a permutation-invariant state by i.i.d. states when a local system has an additional symmetry. Let G be a group that acts as a symmetry of local system A whose Hilbert space is $\mathcal{H}_A$. Then, without loss of generality, $\mathcal{H}_A$ can be decomposed into the direct sum given in Eq. (30) on which the representation π of the group G acts as in Eq. (31). Any state ρ^G_{AE} that is invariant under the action of the group G on the system A should have the form

$$\rho^G_{AE} = \bigoplus_{j=1}^{k} \mathrm{Tr}(Q_j \rho^G) \frac{I_{\mathcal{H}_j^R}}{r_j} \otimes \rho_{\mathcal{H}_j^D E}, \tag{93}$$

with a set of density operators $\{\rho_{\mathcal{H}_j^D E}\}_{j=1}^k$, where Q_j denotes the projection onto the subspace $\mathcal{H}_j^R \otimes \mathcal{H}_j^D$. Since a $\pi^{\otimes n}(G^{\times n})$-invariant state $\rho^G_{A^n E}$ also has this direct-sum structure in each local system A of A^n, it has a direct-sum decomposition into operators acting on $\mathcal{H}_{\boldsymbol{j}}^R \otimes \mathcal{H}_{\boldsymbol{j}}^D$ for $\boldsymbol{j} := (j_1, \ldots, j_n) \in \{1, \ldots, k\}^{\times n}$ given in Eq. (33). Let $Q_{\boldsymbol{j}}$ be a projection operator onto this subspace, i.e.,

$$Q_{\boldsymbol{j}} = Q_{j_1} \otimes \cdots \otimes Q_{j_n}, \tag{94}$$

where $Q_{\boldsymbol{j}} Q_{\boldsymbol{j}'} = \delta_{\boldsymbol{j}\boldsymbol{j}'} Q_{\boldsymbol{j}}$. Then, the $\pi^{\otimes n}(G^{\times n})$-invariant state $\rho^G_{A^n E}$ is decomposed as

$$\rho^G_{A^n E} = \sum_{\boldsymbol{j} \in \{1,\ldots,k\}^{\times n}} Q_{\boldsymbol{j}} \rho^G_{A^n E} Q_{\boldsymbol{j}} \tag{95}$$

$$= \bigoplus_{\boldsymbol{j} \in \{1,\ldots,k\}^{\times n}} \frac{\mathrm{Tr}[Q_{\boldsymbol{j}} \rho^G_{A^n E}]}{\dim \mathcal{H}_{\boldsymbol{j}}^R} I_{\mathcal{H}_{\boldsymbol{j}}^R} \otimes \rho_{\mathcal{H}_{\boldsymbol{j}}^D E} \tag{96}$$

with the set of density operators $\{\rho_{\mathcal{H}_{\boldsymbol{j}}^D E}\}$ for $\boldsymbol{j} \in \{1, \ldots, k\}^{\times n}$.

Now we assume that the state $\rho^G_{A^n E}$ is also invariant under the permutation of n systems of A^n. Due to the decomposition of the Hilbert space $\mathcal{H}_A$ as in Eq. (30), the unitary representation V_σ of the permutation $\sigma \in S_n$ acts as $V_\sigma = V_\sigma^R \otimes V_\sigma^D$, where $V_\sigma^{R(D)}$ acts on the Hilbert space $(\bigoplus_{j=1}^k \mathcal{H}_j^{R(D)})^{\otimes n}$. Note that the Hilbert space $(\bigoplus_{j=1}^k \mathcal{H}_j^R)^{\otimes n} \otimes (\bigoplus_{j=1}^k \mathcal{H}_j^D)^{\otimes n}$ on which V_σ acts is larger than $\mathcal{H}_A^{\otimes n}$, but it contains $\mathcal{H}_A^{\otimes n}$ as an invariant subspace. Now, we define $\boldsymbol{\chi}$, $\mathcal{T}(\boldsymbol{m})$, $\boldsymbol{j}_0(\boldsymbol{m})$, and $\sigma_{\boldsymbol{j}}$ as in Eqs. (36), (37), (40), and (43), respectively. Then, we have

$$Q_{\boldsymbol{j}_0(\boldsymbol{m})} V_{\sigma_{\boldsymbol{j}}}^\dagger V_{\sigma_{\boldsymbol{j}'}} Q_{\boldsymbol{j}_0(\boldsymbol{m})} = \delta_{\boldsymbol{j}\boldsymbol{j}'} Q_{\boldsymbol{j}_0(\boldsymbol{m})}, \tag{97}$$

where $\delta_{\boldsymbol{j}\boldsymbol{j}'}$ denotes the Kronecker delta. From $[V_{\sigma_{\boldsymbol{j}}}, \rho^G_{A^n E}] = 0$ and Eq. (96), we have

$$\rho_{\mathcal{H}_{\boldsymbol{j}}^D E} = V_{\sigma_{\boldsymbol{j}}}^D \rho_{\mathcal{H}_{\boldsymbol{j}_0(\boldsymbol{m})}^D E} (V_{\sigma_{\boldsymbol{j}}}^D)^\dagger, \tag{98}$$

for any $\boldsymbol{j} \in \mathcal{T}(\boldsymbol{m})$. This leads to the alternative expression of $\rho^G_{A^n E}$ as

$$\rho^G_{A^n E} = \bigoplus_{\boldsymbol{m} \in Z_n^k} \frac{\mathrm{Tr}[P_{\boldsymbol{m}} \rho^G_{A^n E}]}{r_{\boldsymbol{m}}} \bigoplus_{\boldsymbol{j} \in \mathcal{T}(\boldsymbol{m})} V_{\sigma_{\boldsymbol{j}}} \Big(I_{\mathcal{H}_{\boldsymbol{j}_0(\boldsymbol{m})}^R} \otimes \rho_{\mathcal{H}_{\boldsymbol{j}_0(\boldsymbol{m})}^D E} \Big) V_{\sigma_{\boldsymbol{j}}}^\dagger, \tag{99}$$

where

$$P_{\boldsymbol{m}} := \sum_{\boldsymbol{j} \in \mathcal{T}(\boldsymbol{m})} Q_{\boldsymbol{j}} \tag{100}$$

and

$$r_{\boldsymbol{m}} := |\mathcal{T}(\boldsymbol{m})| \prod_{j=1}^{k} r_j^{m_j}. \tag{101}$$

Note, for any $\boldsymbol{j} \in \mathcal{T}(\boldsymbol{m})$, we have $\dim \mathcal{H}_{\boldsymbol{j}}^R = \prod_{j=1}^k r_j^{m_j}$.

Now, we define a CP map $s[\boldsymbol{m}] : \mathcal{B}(\mathcal{H}_A^{\otimes n}) \to \mathcal{B}(\mathcal{H}_{\boldsymbol{j}_0(\boldsymbol{m})}^D)$ as follows:

$$s[\boldsymbol{m}](\rho) := \sum_{\boldsymbol{j} \in \mathcal{T}(\boldsymbol{m})} (V_{\sigma_{\boldsymbol{j}}}^D)^\dagger \mathrm{Tr}_{\mathcal{H}_{\boldsymbol{j}}^R}^{(\boldsymbol{j})} (Q_{\boldsymbol{j}} \rho Q_{\boldsymbol{j}}|_{\mathcal{H}_{\boldsymbol{j}}^R \otimes \mathcal{H}_{\boldsymbol{j}}^D}) V_{\sigma_{\boldsymbol{j}}}^D. \tag{102}$$

Moreover, let $r[\boldsymbol{m}] : \mathcal{B}(\mathcal{H}_{\boldsymbol{j}_0(\boldsymbol{m})}^D) \to \mathcal{B}(\mathcal{H}_A^{\otimes n})$ be a CPTP map defined as

$$r[\boldsymbol{m}](\rho) = \frac{1}{r_{\boldsymbol{m}}} \bigoplus_{\boldsymbol{j} \in \mathcal{T}(\boldsymbol{m})} I_{\mathcal{H}_{\boldsymbol{j}}^R} \otimes V_{\sigma_{\boldsymbol{j}}}^D \rho (V_{\sigma_{\boldsymbol{j}}}^D)^\dagger. \tag{103}$$

Then, for any state $\rho^G_{A^n E}$ that is invariant under $\pi^{\otimes n}(G^{\times n})$ as well as permutation of n systems of A^n, we have

$$s[\boldsymbol{m}] \otimes \mathrm{Id}_E(\rho^G_{A^n E}) = \mathrm{Tr}[P_{\boldsymbol{m}} \rho^G_{A^n E}] \rho_{\mathcal{H}_{\boldsymbol{j}_0(\boldsymbol{m})}^D E} \tag{104}$$

and

$$\bigoplus_{\boldsymbol{m}\in Z_n^k} (r[\boldsymbol{m}]\circ s[\boldsymbol{m}])\otimes \mathrm{Id}_E(\rho_{A^nE}^G)=\rho_{A^nE}^G. \quad (105)$$

For each $\boldsymbol{m}=(n_1,\ldots,n_k)\in Z_n^k$, let us pick up a permutation $\sigma\in S_{n_1}\times\cdots\times S_{n_k}\subset S_n$. Then, we have $[V_\sigma,\rho_{A^nE}^G]=0$ and $[V_\sigma,Q_{\boldsymbol{j}_0(\boldsymbol{m})}]$, which thus leads to

$$[V_\sigma,\rho_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}E}]=0. \quad (106)$$

Let us recall that $(\mathcal{H}_j^D)^{\otimes n_j}=\bigoplus_{\boldsymbol{n}_j\in Y_{n_j}^{d_j}}\mathcal{U}_{\boldsymbol{n}_j}\otimes\mathcal{V}_{\boldsymbol{n}_j}$. Then, we have

$$\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}=\bigoplus_{\substack{(\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k)\\ \in Y_{n_1}^{d_1}\times\cdots\times Y_{n_k}^{d_k}}}\bigotimes_{j=1}^k\mathcal{U}_{\boldsymbol{n}_j}\otimes\mathcal{V}_{\boldsymbol{n}_j}. \quad (107)$$

Let $P^D_{\boldsymbol{j}_0(\boldsymbol{m});\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k}$ be a projection operator onto the subspace $\bigotimes_{j=1}^k\mathcal{U}_{\boldsymbol{n}_j}\otimes\mathcal{V}_{\boldsymbol{n}_j}$. Then, from Eq. (106), we have for any $\boldsymbol{m}\in Z_n^k$ that

$$\begin{aligned}&\rho_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}E}\\ &=\bigoplus_{(\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k)} P^D_{\boldsymbol{j}_0(\boldsymbol{m});\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k}\rho_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}E}P^D_{\boldsymbol{j}_0(\boldsymbol{m});\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k}.\end{aligned} \quad (108)$$

Furthermore, from Theorem 1 and Eqs. (26), (27), and (29), there exists a pair $(\tilde{\mathcal{S}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}},\tilde{\mathcal{R}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}})$ of CPTP maps on $\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}$ that satisfies the following three conditions (i), (ii), and (iii).

(i) For any $\rho\in\mathcal{D}(\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}\otimes\mathcal{H}_E)$ and for any $\boldsymbol{n}_j\in Y_{n_j}^{d_j}$ for $j=1,\ldots,k$, we have

$$\begin{aligned}&\tilde{\mathcal{S}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}\otimes\mathrm{Id}_E(P^D_{\boldsymbol{j}_0(\boldsymbol{m});\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k}\rho P^D_{\boldsymbol{j}_0(\boldsymbol{m});\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k})\\ &=P^D_{\boldsymbol{j}_0(\boldsymbol{m});\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k}\tilde{\mathcal{S}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}\otimes\mathrm{Id}_E(\rho)P^D_{\boldsymbol{j}_0(\boldsymbol{m});\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k}.\end{aligned} \quad (109)$$

(ii) For any $\rho\in\mathcal{D}(\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}\otimes\mathcal{H}_E)$ that is $(S_{n_1}\times\cdots\times S_{n_k})$-invariant on $\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}$, we have

$$(\tilde{\mathcal{R}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}\circ\tilde{\mathcal{S}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}})\otimes\mathrm{Id}_E(\rho)=\rho. \quad (110)$$

(iii) For any $\rho\in\mathcal{D}(\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})})$ that is $(S_{n_1}\times\cdots\times S_{n_k})$-invariant and for any $\boldsymbol{n}_j\in Y_{n_j}^{d_j}$ for $j=1,\ldots,k$, we have

$$\begin{aligned}&\mathrm{supp}(\tilde{\mathcal{S}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}(P^D_{\boldsymbol{j}_0(\boldsymbol{m});\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k}\rho P^D_{\boldsymbol{j}_0(\boldsymbol{m});\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k}))\\ &\le\prod_{j=1}^k f_q(n_j,d_j;\boldsymbol{n}_j)\,\tilde{\mathcal{S}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}\Big(\bigotimes_{j=1}^k\\ &\qquad\mathbb{E}_{U_j^D\sim\mathrm{Haar}(d_j)}\Big[\Big(U_j^D\mathrm{diag}\Big(\frac{\boldsymbol{n}_j}{n_j}\Big)(U_j^D)^\dagger\Big)^{\otimes n_j}_{\mathcal{H}_j^D}\Big]\Big).\end{aligned} \quad (111)$$

Now, define a pair $(\mathcal{S}_{A^n}^G,\mathcal{R}_{A^n}^G)$ of CPTP maps on $\mathcal{H}_A^{\otimes n}$ as

$$\mathcal{S}_{A^n}^G(\rho):=\bigoplus_{\boldsymbol{m}\in Z_n^k}|\varnothing\rangle\langle\varnothing|_{\mathcal{H}^R_{\boldsymbol{j}_0(\boldsymbol{m})}}\otimes(\tilde{\mathcal{S}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}\circ s[\boldsymbol{m}](\rho)), \quad (112)$$

$$\begin{aligned}\mathcal{R}_{A^n}^G(\rho):=\bigoplus_{\boldsymbol{m}\in Z_n^k} r[\boldsymbol{m}]\circ\tilde{\mathcal{R}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}\circ\mathrm{Tr}^{(\boldsymbol{j}_0(\boldsymbol{m}))}_{\mathcal{H}^R_{\boldsymbol{j}_0(\boldsymbol{m})}}\\ \Big(\sum_{\boldsymbol{j}\in\mathcal{T}(\boldsymbol{m})}V_{\sigma_{\boldsymbol{j}}}^\dagger Q_{\boldsymbol{j}}\rho Q_{\boldsymbol{j}}V_{\sigma_{\boldsymbol{j}}}\Big|_{\mathcal{H}^R_{\boldsymbol{j}_0(\boldsymbol{m})}\otimes\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}\Big),\end{aligned} \quad (113)$$

where $|\varnothing\rangle\langle\varnothing|_{\mathcal{H}^R_{\boldsymbol{j}_0(\boldsymbol{m})}}$ is an arbitrary pure state on $\mathcal{H}^R_{\boldsymbol{j}_0(\boldsymbol{m})}$. The map $\mathcal{S}_{A^n}^G$ is actually trace preserving since $\sum_{\boldsymbol{m}\in Z_n^k}s[\boldsymbol{m}]$ and $\tilde{\mathcal{S}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}$ are trace preserving. The map $\mathcal{R}_{A^n}^G$ is also trace preserving since $r[\boldsymbol{m}]$ and $\tilde{\mathcal{R}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}$ are trace preserving and, for any $\rho\in\mathcal{D}(\mathcal{H}_A^{\otimes n})$,

$$\mathrm{Tr}\left[\sum_{\boldsymbol{j}\in\mathcal{T}(\boldsymbol{m})}V_{\sigma_{\boldsymbol{j}}}^\dagger Q_{\boldsymbol{j}}\rho Q_{\boldsymbol{j}}V_{\sigma_{\boldsymbol{j}}}\right]=\mathrm{Tr}[P_{\boldsymbol{m}}\rho]. \quad (114)$$

Let $\rho_{A^nE}^G$ be $\pi^{\otimes n}(G^{\times n})$-invariant and permutation invariant over n systems of A^n. Then, from Eqs. (105), (110), (112), and (113), we have

$$\begin{aligned}&\mathcal{R}_{A^n}^G\circ\mathcal{S}_{A^n}^G\otimes\mathrm{Id}_E(\rho_{A^nE}^G)\\ &=\bigoplus_{\boldsymbol{m}\in Z_n^k}(r[\boldsymbol{m}]\circ\tilde{\mathcal{R}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}\circ\tilde{\mathcal{S}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}\circ s[\boldsymbol{m}])\otimes\mathrm{Id}_E(\rho_{A^nE}^G)\end{aligned} \quad (115)$$

$$=\rho_{A^nE}^G, \quad (116)$$

which proves Eq. (50) in Theorem 2. Define $P_{\boldsymbol{m};\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k}$ as

$$P_{\boldsymbol{m};\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k}=\bigoplus_{\boldsymbol{j}\in\mathcal{T}(\boldsymbol{m})}V_{\sigma_{\boldsymbol{j}}}\Big(I_{\mathcal{H}^R_{\boldsymbol{j}_0(\boldsymbol{m})}}\otimes P^D_{\boldsymbol{j}_0(\boldsymbol{m});\boldsymbol{n}_1,\ldots,\boldsymbol{n}_k}\Big)V_{\sigma_{\boldsymbol{j}}}^\dagger. \quad (117)$$

Then, combined with Eqs. (97), (99), and (108), we have

$$\rho^G_{A^nE} = \bigoplus_{\boldsymbol{m}\in Z_n^k} \bigoplus_{\substack{(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)\\ \in Y_{n_1}^{d_1}\times\cdots\times Y_{n_k}^{d_k}}} P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}\rho^G_{A^nE}P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}, \tag{118}$$

which proves Eq. (47). Now, from Eq. (102), we have

$$\begin{aligned} &s[\boldsymbol{m}]\otimes \mathrm{Id}_E(P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}\rho^G_{A^nE}P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}) \\ &= P^D_{\boldsymbol{j}_0(\boldsymbol{m});\boldsymbol{n}_1,\dots,\boldsymbol{n}_k} s[\boldsymbol{m}]\otimes \mathrm{Id}_E(\rho^G_{A^nE})P^D_{\boldsymbol{j}_0(\boldsymbol{m});\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}. \end{aligned} \tag{119}$$

Then, combining this with Eqs. (97), (109), and (112), we prove Eq. (49) in Theorem 2. Finally, by combining Eqs. (111) and (119), we have

$$\begin{aligned} &\mathrm{supp}\Big(\tilde{\mathcal{S}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}(s[\boldsymbol{m}](P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}\rho^G_{A^nE}P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}))\Big) \\ &\le \prod_{j=1}^k f_q(n_j,d_j;\boldsymbol{n}_j)\,\tilde{\mathcal{S}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}\Big(\bigotimes_{j=1}^k \\ &\quad \mathbb{E}_{U_j^D\sim\mathrm{Haar}(d_j)}\Big[\Big(U_j^D\mathrm{diag}\Big(\frac{\boldsymbol{n}_j}{n_j}\Big)(U_j^D)^\dagger\Big)^{\otimes n_j}_{\mathcal{H}_j^D}\Big]\Big). \end{aligned} \tag{120}$$

Let $\tau_{\boldsymbol{U}}(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)$ be as defined in Eq. (52). Then, from Eq. (102), we have

$$\begin{aligned} &s[\boldsymbol{m}]\big([\tau_{\boldsymbol{U}}(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)]^{\otimes n}\big) \\ &= \mathrm{Mult}_{\boldsymbol{m}/n}(\boldsymbol{m})\bigotimes_{j=1}^k\Big(U_j^D\mathrm{diag}\Big(\frac{\boldsymbol{n}_j}{n_j}\Big)(U_j^D)^\dagger\Big)^{\otimes n_j}_{\mathcal{H}_j^D}. \end{aligned} \tag{121}$$

Combining this with Eqs. (13), (112), and (120), we have

$$\begin{aligned} &\mathrm{supp}\big(S^G_{A^n}(P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k}\rho^G_{A^nE}P_{\boldsymbol{m};\boldsymbol{n}_1,\dots,\boldsymbol{n}_k})\big) \\ &\le f_c(n,k;\boldsymbol{m})\prod_{j=1}^k f_q(n_j,d_j;\boldsymbol{n}_j)\,|\varnothing\rangle\langle\varnothing|_{\mathcal{H}^R_{\boldsymbol{j}_0}} \\ &\quad\otimes \mathbb{E}_{\boldsymbol{U}\sim\times_{j=1}^k\mathrm{Haar}(d_j)}\Big[\tilde{\mathcal{S}}_{\mathcal{H}^D_{\boldsymbol{j}_0(\boldsymbol{m})}}\circ s[\boldsymbol{m}]\Big(\\ &\qquad\qquad \tau_{\boldsymbol{U}}(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)^{\otimes n}\Big)\Big] \\ &\le f_c(n,k;\boldsymbol{m})\prod_{j=1}^k f_q(n_j,d_j;\boldsymbol{n}_j) \\ &\quad\times S^G_{A^n}\Big(\mathbb{E}_{\boldsymbol{U}\sim\times_{j=1}^k\mathrm{Haar}(d_j)}[\tau_{\boldsymbol{U}}(\boldsymbol{n}_1,\dots,\boldsymbol{n}_k)^{\otimes n}]\Big), \end{aligned} \tag{122, 123}$$

which proves Eq. (51) in Theorem 2. □

Remark 1. *Instead of applying Eq. (19) in Theorem 1 to Eq. (111), one can apply Eq. (75) in Lemma 2 to obtain an even tighter bound. However, the inequality in Theorem 2 may be more convenient for practical use.*

A particularly interesting case is when $G = \mathrm{U}(1)$. Since all the irreducible representations of $\mathrm{U}(1)$ is one dimensional, $r_j = 1$ for any j. Let us further assume that $d_1 = \cdots = d_k = d/k$. This case can be regarded as a classical-quantum state with k being the dimension of the classical system. Then, the coefficient would be

$$f_c(n,k)\prod_{j=1}^k f_q(n_j,d_j) \simeq \mathcal{O}(n^{\frac{d^2/k-1}{2}}), \tag{124}$$

and thus we have approximately power-of-k improvement compared to the case when we would not consider the symmetry, i.e., when applying Theorem 1 naively.

2.3 Concentration inequality for an independent and identical measurement on an adversarial state

Now we leave from concentration inequality for a quantum state and consider a concentration inequality for quantum measurement outcomes. In this section, we consider the case where independent and identical measurements $\mathcal{M}^{\otimes n}$ are performed on an adversarial state ρ. Let $\mathcal{M}^{\otimes n}(\rho) : \mathcal{X}^{\times n} \to [0,1]$ be a probability mass function as defined in Theorem 3, and let $\mathrm{P}_\rho : Z_n^k \to [0,1]$ be another probability mass function defined as

$$\mathrm{P}_\rho(\boldsymbol{m}) \coloneqq \sum_{\boldsymbol{i}\in\mathcal{T}(\boldsymbol{m})} \mathcal{M}^{\otimes n}(\rho)(\boldsymbol{i}). \tag{125}$$

Since we will consider a concentration inequality for the empirical probability of the measurement outcomes, the relevant probability mass function is $\mathrm{P}_\rho(\boldsymbol{m})$ rather than $\mathcal{M}^{\otimes n}(\rho)$. Furthermore, since $\boldsymbol{m}$ is invariant under the permutation of measurement outcomes $\boldsymbol{i}$, we have

$$\mathrm{P}_\rho(\boldsymbol{m}) = \mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}), \tag{126}$$

where $\tilde{\rho}$ is the permutation-symmetrized version of ρ. Thus, we consider $\tilde{\rho}$ instead of ρ in the following analysis. Due to the permutation invariance, we can apply Theorem 2 to $\mathcal{M}^{\otimes n}(\tilde{\rho})$

with $r_j = d_j = 1$ for $j = 1, \dots, k$ and obtain

$$\mathcal{M}^{\otimes n}(\tilde{\rho})(\boldsymbol{i}) = \sum_{\boldsymbol{m} \in Z_n^k} \mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}) \sum_{\boldsymbol{j} \in \mathcal{T}(\boldsymbol{m})} \frac{\delta_{\boldsymbol{j}\boldsymbol{i}}}{|\mathcal{T}(\boldsymbol{m})|}, \tag{127}$$

for any $\boldsymbol{i} \in \mathcal{X}^{\times n}$. What we want to know is a restriction imposed on the probability $\mathrm{P}_{\tilde{\rho}}(\boldsymbol{m})$ from the fact that the independent and identical measurements are performed on the permutation-invariant state. To know this, let $\mathcal{P}$ be the probability simplex, i.e., the set of probability vectors $\boldsymbol{p} = (p_1, \dots, p_k)$ with $p_i \geq 0$ and $\sum_{i=1}^k p_i = 1$, and $\mathcal{Q} \subseteq \mathcal{P}$ be its subset defined as

$$\mathcal{Q} = \{\boldsymbol{p} \in \mathcal{P} : \exists \rho \in \mathcal{D}(\mathcal{H}) \text{ s.t. } \forall i \in \mathcal{X}, p_i = \mathcal{M}(\rho)(i)\}. \tag{128}$$

Then, for a particular sequence $\boldsymbol{i} = (i_1, \dots, i_n)$ of the measurement outcomes, a conditional probability of the n-th outcome i_n conditioned on any sequence $\boldsymbol{i}' = (i_1, \dots, i_{n-1})$ of the other $n-1$ outcomes should be in $\mathcal{Q}$ due to the definition of $\mathcal{Q}$. Let $\boldsymbol{l} = (n'_1, \dots, n'_k)$ with $n'_j = \chi'_j(\boldsymbol{i}')$, where $\boldsymbol{\chi}' : \mathcal{X}^{\times(n-1)} \to Z_{n-1}^k$ is defined similarly to Eq. (36) for $(n-1)$ tuple, and let $\boldsymbol{l}[j] \in Z_n^k$ be defined as a k-tuple that adds one to j-th element of $\boldsymbol{l}$, i.e., $\boldsymbol{l}[j] = (n'_1, \dots, n'_{j-1}, n'_j + 1, n'_{j+1}, \dots, n'_k)$. By applying the above restriction on the conditional probability to Eq. (127), we have

$$\left(\sum_{j=1}^k \frac{\mathrm{P}_{\tilde{\rho}}(\boldsymbol{l}[j])}{|\mathcal{T}(\boldsymbol{l}[j])|}\right)^{-1} \left(\frac{\mathrm{P}_{\tilde{\rho}}(\boldsymbol{l}[1])}{|\mathcal{T}(\boldsymbol{l}[1])|}, \dots, \frac{\mathrm{P}_{\tilde{\rho}}(\boldsymbol{l}[k])}{|\mathcal{T}(\boldsymbol{l}[k])|}\right) \in \mathcal{Q}. \tag{129}$$

(Note, if only this condition is satisfied, the following argument holds for any classical probability mass function as well.) This restriction on $\mathrm{P}_{\tilde{\rho}}(\boldsymbol{m})$ for each $\boldsymbol{m}$ is strict, but is not very convenient for the later analysis. So, we consider a looser version for this that is easier to analyze.

Let $\mathcal{R} \coloneqq \{(p_1, \dots, p_k) \in \mathcal{P} : \sum_{j=1}^k a_j p_j \leq b\}$ be a convex subset of $\mathcal{P}$ with an affine boundary such that $\mathcal{Q} \subseteq \mathcal{R}$ (see Fig. 1). *We restrict our attention to the case where the complement $\mathcal{P} \setminus \mathcal{R}$ of $\mathcal{R}$ contains only one probability vector that corresponds to having a deterministic outcome*, say $\boldsymbol{p}_1 \coloneqq (1, 0, \dots, 0)$. This may lead to a looser bound when applied to a specific problem or limit the applicability of the obtained bound, but as can be shown later, it still has a wide range of applications. The coefficients $\{a_j\}_{j=1}^k$ that defines the set $\mathcal{R}$ in this case should satisfy $a_1 > b$, $a_j \leq b$ for $j = 2, \dots, k$. By redefining $b_j = -(a_j - b)/(a_1 - b)(\geq 0)$ for $j = 2, \dots, k$, we have

$$\mathcal{R} = \left\{(p_1, \dots, p_k) \in \mathcal{P} : p_1 - \left(\sum_{j=2}^k b_j p_j\right) \leq 0\right\} \tag{130}$$

$$= \{\boldsymbol{p} = (p_1, \dots, p_k) \in \mathcal{P} : \boldsymbol{u} \cdot \boldsymbol{p} \leq 0\}, \tag{131}$$

where $\boldsymbol{u} \coloneqq (1, -b_2, \dots, -b_k)$ is as defined in the theorem statement. In this way, we obtain the expression of $\mathcal{R}$ in Eq. (57). In other words, Eq. (57) is defined to satisfy the condition on the complement $\mathcal{P} \setminus \mathcal{R}$ stated above. We also define $\mathcal{R}'$ as in the theorem statement, i.e., $\mathcal{R}' \coloneqq \{\boldsymbol{p} \in \mathcal{P} : \boldsymbol{u} \cdot \boldsymbol{p} \leq \frac{\sqrt{2}\|\boldsymbol{u} - (\boldsymbol{v}\cdot\boldsymbol{u})\boldsymbol{v}\|}{n}\}$, where $\boldsymbol{v} \coloneqq (1, \dots, 1)/\sqrt{k}$ is as defined in the theorem statement. Since $\mathcal{Q} \subseteq \mathcal{R}$ by assumption, we have from Eq. (129) that

$$\left(\sum_{j=1}^k \frac{\mathrm{P}_{\tilde{\rho}}(\boldsymbol{l}[j])}{|\mathcal{T}(\boldsymbol{l}[j])|}\right)^{-1} \left(\frac{\mathrm{P}_{\tilde{\rho}}(\boldsymbol{l}[1])}{|\mathcal{T}(\boldsymbol{l}[1])|}, \dots, \frac{\mathrm{P}_{\tilde{\rho}}(\boldsymbol{l}[k])}{|\mathcal{T}(\boldsymbol{l}[k])|}\right) \in \mathcal{R}. \tag{132}$$

Then, with this loosened constraint, we will find an upper bound on $\mathrm{P}_{\tilde{\rho}}(\boldsymbol{m})$ for an arbitrary frequency distribution $\boldsymbol{m} = (n_1, \dots, n_k) \in Z_n^k$ of the outcomes.

As a first step, we will show that given an arbitrary probability vector $\boldsymbol{p} \in \mathcal{R}$, any probability vector $\boldsymbol{q} \in \mathcal{P}$ with $\|\boldsymbol{p} - \boldsymbol{q}\| \leq \sqrt{2}/n$ is contained in $\mathcal{R}'$. From the Cauchy-Schwarz inequality, we have

$$\boldsymbol{u} \cdot (\boldsymbol{q} - \boldsymbol{p}) = (\boldsymbol{u} - (\boldsymbol{v} \cdot \boldsymbol{u})\boldsymbol{v}) \cdot (\boldsymbol{q} - \boldsymbol{p}) \tag{133}$$

$$\leq \|\boldsymbol{u} - (\boldsymbol{v} \cdot \boldsymbol{u})\boldsymbol{v}\| \|\boldsymbol{q} - \boldsymbol{p}\| \tag{134}$$

$$\leq \frac{\sqrt{2}\|\boldsymbol{u} - (\boldsymbol{v} \cdot \boldsymbol{u})\boldsymbol{v}\|}{n}, \tag{135}$$

where we used $\boldsymbol{v} \cdot \boldsymbol{p} = \sqrt{k} = \boldsymbol{v} \cdot \boldsymbol{q}$ in the first equality. (Note that $\boldsymbol{u} - (\boldsymbol{v} \cdot \boldsymbol{u})\boldsymbol{v}$ is a vector projection of $\boldsymbol{u}$ onto the affine hull of $\mathcal{P}$ in $\mathbb{R}^k$.) Since $\boldsymbol{p} \in \mathcal{R}$ implies $\boldsymbol{u} \cdot \boldsymbol{p} \leq 0$, we have from Eq. (135) that

$$\boldsymbol{u} \cdot \boldsymbol{q} \leq \boldsymbol{u} \cdot \boldsymbol{p} + \frac{\sqrt{2}\|\boldsymbol{u} - (\boldsymbol{v} \cdot \boldsymbol{u})\boldsymbol{v}\|}{n} \leq \frac{\sqrt{2}\|\boldsymbol{u} - (\boldsymbol{v} \cdot \boldsymbol{u})\boldsymbol{v}\|}{n}, \tag{136}$$

which then implies that $\boldsymbol{q} \in \mathcal{R}'$ from the definition of $\mathcal{R}'$.

Now, let $\tilde{\mathcal{P}}$ be the set of types for n outcomes defined as

$$\tilde{\mathcal{P}} \coloneqq \{(n_1, \dots, n_k)/n \in \mathcal{P} : \forall j \in \{1, \dots, k\}, n_j \in \mathbb{Z}\}, \tag{137}$$

and let $\tilde{\mathcal{R}}$ be its subset defined as

$$\tilde{\mathcal{R}} \coloneqq \tilde{\mathcal{P}} \cap (\mathcal{R}' \setminus \mathcal{R}). \tag{138}$$

We prove that every path in $\tilde{\mathcal{P}}$ from $\boldsymbol{p}_1 = (1, 0, \ldots, 0)$ to $(0, n_2, \ldots, n_k)/n \in \mathcal{R}$ (for any possible combination $(n_2, \ldots, n_k)$ with $\sum_{j=2}^k n_j = n$) by repeatedly subtracting $1/n$ from the first entry and adding $1/n$ to one of the other entries hits at least one element in $\tilde{\mathcal{R}}$. We define the neighbors of a type $\boldsymbol{p} \in \tilde{\mathcal{P}}$ to be those types that can be reached by subtracting $1/n$ from one entry of $\boldsymbol{p}$ and adding $1/n$ to another entry of $\boldsymbol{p}$. Then, the Euclidean distance between neighboring types is $\sqrt{2}/n$. If there exists a path from $(1, 0, \ldots, 0)$ to $(0, n_2, \ldots, n_k)/n$ that does not hit any element of $\tilde{\mathcal{R}}$, then that means there exists a pair $(\boldsymbol{p}, \boldsymbol{p}')$ of neighboring probability vectors such that $\boldsymbol{p} \in \mathcal{P} \setminus \mathcal{R}'$ and $\boldsymbol{p}' \in \mathcal{R}$. But this contradicts the fact that any $\boldsymbol{q} \in \mathcal{P}$ with $\|\boldsymbol{q} - \boldsymbol{p}'\| \leq \sqrt{2}/n$ is contained in $\mathcal{R}'$. This completes the proof by contradiction.

Now, let us consider the following subnormalized convex mixture σ of i.i.d. probability distributions over $\mathcal{X}^{\times n}$:

$$\sigma(\boldsymbol{i}) = \sum_{\substack{\boldsymbol{m}=(n_1,\ldots,n_k)\in Z_n^k \\ :\boldsymbol{m}/n\in\tilde{\mathcal{R}}}} \mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}) \prod_{l=1}^n \frac{n_{i_l}}{n} \tag{139}$$

$$= \sum_{\substack{\boldsymbol{m}=(n_1,\ldots,n_k)\in Z_n^k \\ :\boldsymbol{m}/n\in\tilde{\mathcal{R}}}} \mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}) \prod_{j=1}^k \left(\frac{n_j}{n}\right)^{\chi_j(\boldsymbol{i})}, \tag{140}$$

which is subnormalized since $\sum_{\boldsymbol{m}\in Z_n^k : \boldsymbol{m}/n\in\tilde{\mathcal{R}}} \mathrm{P}_{\tilde{\rho}}(\boldsymbol{m})$ may not be unity in general. Let $\mathrm{Q}_\sigma : Z_n^k \to [0,1]$ be a probability mass function associated with σ defined as

$$\mathrm{Q}_\sigma(\boldsymbol{m}) \coloneqq \sum_{\boldsymbol{i}\in\mathcal{T}(\boldsymbol{m})} \sigma(\boldsymbol{i}) \tag{141}$$

$$= \sum_{\boldsymbol{m}'/n\in\tilde{\mathcal{R}}} \mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}')\mathrm{Mult}_{\boldsymbol{m}'/n}(\boldsymbol{m}). \tag{142}$$

From Eqs. (14) and (85), we have, for any $\boldsymbol{m} \in Z_n^k$ with $\boldsymbol{m}/n \in \tilde{\mathcal{R}}$,

$$\mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}) \leq f_c(n,k)\mathrm{Q}_\sigma(\boldsymbol{m}). \tag{143}$$

In the following, we inductively prove that $\mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}) \leq f_c(n,k)\mathrm{Q}_\sigma(\boldsymbol{m})$ holds for any $\boldsymbol{m} \in Z_n^k$ with $\boldsymbol{m}/n \in \mathcal{P} \setminus \mathcal{R}$. Let us define $\tilde{\mathcal{R}}^{(0)} \coloneqq \tilde{\mathcal{R}}$. Then, for any frequency distribution $\boldsymbol{m}$ with its type in $\tilde{\mathcal{R}}^{(0)}$, Eq. (143) holds. Let $\boldsymbol{l} \in Z_{n-1}^k$ be an arbitrary frequency distribution such that $\boldsymbol{l}[1]/n \in \tilde{\mathcal{P}} \setminus \tilde{\mathcal{R}}^{(0)}$ and $\boldsymbol{l}[j]/n \in \tilde{\mathcal{R}}^{(0)}$ for $j = 2, \ldots, k$. (Note that $\boldsymbol{l}[1]/n \in \mathcal{P}\setminus\mathcal{R}'$ automatically holds if the above conditions are satisfied.) Such a sequence $\boldsymbol{l}$ always exists; if there does not exist such a sequence, then that means there exists j such that $\boldsymbol{l}[1]/n \in \tilde{\mathcal{P}} \setminus \tilde{\mathcal{R}}^{(0)}$ and $\boldsymbol{l}[j]/n \notin \tilde{\mathcal{R}}^{(0)}$. If $\boldsymbol{l}[j]/n \in \tilde{\mathcal{P}} \setminus \tilde{\mathcal{R}}^{(0)}$, then we can find a sequence $\boldsymbol{l}'$ that satisfies the imposed condition by reducing the first entry of $\boldsymbol{l}$ by a positive integer multiple of $1/n$ and add it to the j-th entry. The other possibility is that $\boldsymbol{l}[j]/n \in \mathcal{R}$, but this implies that we can find a path that goes from $\boldsymbol{p}_1$ through $\boldsymbol{l}[1]/n \in \tilde{\mathcal{P}} \setminus \tilde{\mathcal{R}}^{(0)}$ and $\boldsymbol{l}[j]/n \in \mathcal{R}$ to $(0, n_2, \ldots, n_k)/n$ (with $\sum_{j=2}^k n_j = n$) without intersecting $\tilde{\mathcal{R}}^{(0)}$, which contradicts the fact that every path in $\tilde{\mathcal{P}}$ from $\boldsymbol{p}_1$ to $(0, n_2, \ldots, n_k)/n$ hits at least one element in $\tilde{\mathcal{R}}$. From the constraint on $\mathrm{P}_{\tilde{\rho}}(\cdot)$ in Eq. (132) and the definition of $\mathcal{R}$ in Eq. (57), we have

$$\frac{\mathrm{P}_{\tilde{\rho}}(\boldsymbol{l}[1])}{|\mathcal{T}(\boldsymbol{l}[1])|} - \left(\sum_{j=2}^k b_j \frac{\mathrm{P}_{\tilde{\rho}}(\boldsymbol{l}[j])}{|\mathcal{T}(\boldsymbol{l}[j])|}\right) \leq 0. \tag{144}$$

Furthermore, since σ is a mixture of the i.i.d. probability distributions and is thus invariant under the permutation of n outcomes, it has the same form as the right-hand side of Eq. (127) with $\mathrm{P}_{\tilde{\rho}}$ replaced with Q_σ. Considering again a conditional probability mass function of the n-th outcome i_n conditioned on a sequence $\boldsymbol{i}' = (i_1, \ldots, i_{n-1})$ of the other $n-1$ outcomes of a particular sequence $\boldsymbol{i} = (i_1, \ldots, i_n)$ for σ, we notice that the resulting conditional probability mass function is an $\boldsymbol{i}'$ (or in fact $\boldsymbol{l}$) dependent mixture of probability distributions in $\tilde{\mathcal{R}}$ since σ is a mixture of i.i.d. probability distributions in $\tilde{\mathcal{R}}$. Combining these facts, we have

$$\left(\sum_{j=1}^k \frac{\mathrm{Q}_\sigma(\boldsymbol{l}[j])}{|\mathcal{T}(\boldsymbol{l}[j])|}\right)^{-1} \left(\frac{\mathrm{Q}_\sigma(\boldsymbol{l}[1])}{|\mathcal{T}(\boldsymbol{l}[1])|}, \ldots, \frac{\mathrm{Q}_\sigma(\boldsymbol{l}[k])}{|\mathcal{T}(\boldsymbol{l}[k])|}\right) \in \mathcal{R}' \setminus \mathcal{R} \subseteq \mathcal{P} \setminus \mathcal{R}, \tag{145}$$

and thus have from Eq. (131) that

$$\frac{\mathrm{Q}_\sigma(\boldsymbol{l}[1])}{|\mathcal{T}(\boldsymbol{l}[1])|} - \left(\sum_{j=2}^k b_j \frac{\mathrm{Q}_\sigma(\boldsymbol{l}[j])}{|\mathcal{T}(\boldsymbol{l}[j])|}\right) \geq 0. \tag{146}$$

Combining Eqs. (143), (144) and (146), we have

$$\mathrm{P}_{\tilde{\rho}}(\boldsymbol{l}[1]) \leq f_c(n,k)\mathrm{Q}_{\sigma}(\boldsymbol{l}[1]). \tag{147}$$

Thus, we have the same inequality as Eq. (143) for $\boldsymbol{l}[1]$. We repeatedly apply the above argument by replacing $\tilde{\mathcal{R}}^{(0)}$ with $\tilde{\mathcal{R}}^{(1)} := \tilde{\mathcal{R}}^{(0)} \cup \{\boldsymbol{l}[1]/n\}$ and so on, and finally obtains $\mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}) \leq f_c(n,k)\mathrm{Q}_{\sigma}(\boldsymbol{m})$ for any frequency distribution $\boldsymbol{m} \in Z_n^k$ with $\boldsymbol{m}/n \in \mathcal{P} \setminus \mathcal{R}$ by induction.

From Eq. (14) and (85), we also have

$$\mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}) \leq \sum_{\boldsymbol{m}':\boldsymbol{m}'/n\in\mathcal{R}} f_c(n,k)\,\mathrm{Mult}_{\boldsymbol{m}'/n}(\boldsymbol{m}) \tag{148}$$

for any $\boldsymbol{m}/n \in \mathcal{R}$. Thus, for any $\boldsymbol{m} \in Z_n^k$, we have

$$\begin{aligned}\mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}) \leq & \sum_{\boldsymbol{m}':\boldsymbol{m}'/n\in\tilde{\mathcal{R}}} f_c(n,k)\mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}')\,\mathrm{Mult}_{\boldsymbol{m}'/n}(\boldsymbol{m}) \\ & + \sum_{\boldsymbol{m}':\boldsymbol{m}'/n\in\mathcal{R}} f_c(n,k)\mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}')\,\mathrm{Mult}_{\boldsymbol{m}'/n}(\boldsymbol{m}) \qquad (149)\\ = & \sum_{\boldsymbol{m}':\boldsymbol{m}'/n\in\mathcal{R}'} f_c(n,k)\mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}')\,\mathrm{Mult}_{\boldsymbol{m}'/n}(\boldsymbol{m}). \qquad (150)\end{aligned}$$

This inequality immediately implies that for any subset B of Z_n^k, we have

$$\sum_{\boldsymbol{m}\in B} \mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}) \leq \max_{\boldsymbol{q}\in\mathcal{R}'} \sum_{\boldsymbol{m}\in B} f_c(n,k)\,\mathrm{Mult}_{\boldsymbol{q}}(\boldsymbol{m}). \tag{151}$$

From this result, we can obtain a concentration inequality for measurement outcomes. Let $D(\boldsymbol{p}\|\boldsymbol{q})$ be a Kullback-Leibler divergence given by

$$D(\boldsymbol{p}\|\boldsymbol{q}) := \begin{cases} \mathrm{Tr}[\boldsymbol{p}\ln\boldsymbol{p} - \boldsymbol{p}\ln\boldsymbol{q}] & \mathrm{supp}(\boldsymbol{p}) \subseteq \mathrm{supp}(\boldsymbol{q}) \\ +\infty & \text{Otherwise} \end{cases}, \tag{152}$$

where $\mathrm{supp}(\boldsymbol{r})$ denotes the support of the (discrete) probability distribution $\boldsymbol{r}$. Then, the following theorem is known [3].

Theorem 4 (Refined Sanov's theorem for a closed convex set [3])**.** *Let $\mathcal{A}$ be a non-empty closed convex subset of the $(k-1)$-simplex $\mathcal{P}$ of probability distributions. Let $\boldsymbol{q} \in \mathcal{P}$ be a probability distribution, and let Z_n^k be as defined in Eq. (5). Then, for a multinomial distribution $\mathrm{Mult}_{\boldsymbol{q}}$ of $\boldsymbol{q}$, one has*

$$\sum_{\boldsymbol{m}\in Z_n^k:\boldsymbol{m}/n\in\mathcal{A}} \mathrm{Mult}_{\boldsymbol{q}}(\boldsymbol{m}) \leq \max_{\boldsymbol{p}\in\mathcal{A}} \exp[-nD(\boldsymbol{p}\|\boldsymbol{q})]. \tag{153}$$

(In the above theorem, if $\mathcal{A}$ is not restricted to a closed convex subset, then a polynomial factor is multiplied to the right-hand side, which may be more conventional [2, 32]. However, for our purpose of applying it to quantum measurement scenario, we typically consider a subset with an affine boundary obtained through an operator inequality, and thus the above mentioned restriction on $\mathcal{A}$ may not lose generality.) Combining the above theorem with Eq. (151), we obtain

$$\begin{aligned}\sum_{\boldsymbol{m}/n\in\mathcal{A}} \mathrm{P}_{\tilde{\rho}}(\boldsymbol{m}) &\leq \max_{\boldsymbol{q}\in\mathcal{R}'} \sum_{\boldsymbol{m}/n\in\mathcal{A}} f_c(n,k)\,\mathrm{Mult}_{\boldsymbol{q}}(\boldsymbol{m}) \qquad (154)\\ &\leq \max_{\boldsymbol{p}\in\mathcal{A}} \max_{\boldsymbol{q}\in\mathcal{R}'} f_c(n,k)\exp[-nD(\boldsymbol{p}\|\boldsymbol{q})]. \qquad (155)\end{aligned}$$

Note that the simultaneous maximization over $\boldsymbol{p} \in \mathcal{A}$ and $\boldsymbol{q} \in \mathcal{R}'$ is a convex optimization problem, and thus the order of maximization does not matter. Combining this with Eqs. (125) and (126), we prove Theorem 3. □

Thus, to obtain a tight bound, the regions $\mathcal{R}'$ and $\mathcal{A}$ should be chosen so that the KL divergence between the two regions is maximized.

3 Numerical comparison of the performance with other concentration inequalities

The obtained bound in the previous section has only a polynomial factor compared to the i.i.d. case, and thus it is asymptotically tight. However, what one may have an interest in a practical setup is tightness in the finite-size regime. This is especially important for applications such as QKD. Here, we compare its tightness with other known concentration inequalities that can deal with non-commutative observables. One famous example is Azuma's inequality [28], which is conventionally used in the so-called phase error correction approach to prove the security of QKD protocols. Azuma's inequality does not require permutation invariance

unlike Theorem 1 and can be applied to arbitrary correlated random variables in combination with the so-called Doob decomposition theorem [33]. Various generalizations of Azuma's inequality have been studied [12, 13], most of which have tightened the inequality by using the bound on the higher moments of the true distribution. Recently, Kato has succeeded in tightening Azuma's inequality by using unconfirmed knowledge, i.e., a priori knowledge on the expectation value of random variables [14]. The inequality holds even when a priori knowledge is false, and it tightens the inequality to the level of i.i.d. case for a binary random variable when a priori knowledge is true, which is extremely useful in applications such as QKD. Since our method can also be applied to adversarial setups such as QKD, we compare the performance of our new method with these two conventionally used bounds.

In order to avoid the complexity of the analysis, we consider a simple setup. In the following, we first consider the toy model of estimating the outcomes of a rank-one projective measurement from the outcomes of a slightly tilted projective measurement. This setup allows us to compare the tightness of the estimation in the finite number of rounds when we apply different concentration inequalities. Next, we consider a simple quantum key distribution protocol, the three-state protocol, and compare the finite-size key rates obtained through these concentration inequalities.

3.1 Toy model for statistical estimation

For an N-qubit system and positive parameters ϵ and r, we consider two types of binary-outcome measurements performed on a qubit system, type I: $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ and type II: $\{|\phi_0\rangle\langle\phi_0|, |\phi_1\rangle\langle\phi_1|\}$, where $|\phi_0\rangle = \sqrt{1-r}\,|0\rangle + \sqrt{r}\,|1\rangle$ and $|\phi_1\rangle = \sqrt{r}\,|0\rangle - \sqrt{1-r}\,|1\rangle$. Both measurements outcome $j \in \{0,1\}$ for the corresponding POVM elements $|j\rangle\langle j|$ and $|\phi_j\rangle\langle\phi_j|$ (i.e., POVM elements for this measurement procedure is given by $\{\frac{1}{2}|0\rangle\langle 0|, \frac{1}{2}|1\rangle\langle 1|, \frac{1}{2}|\phi_0\rangle\langle\phi_0|, \frac{1}{2}|\phi_1\rangle\langle\phi_1|\}$). Then, the problem is stated as follows; given a state ρ on N-qubit system prepared by an adversary, Alice and Bob perform type I or type II measurement on each of them uniformly randomly. Alice obtains the results of the type I measurements and Bob obtains those of the type II. From the sum $\hat{B} \coloneqq \sum_i \hat{b}_i$ of outcomes $\hat{b}_i \in \{0,1\}$ of the type II measurements, Bob aims at obtaining an upper bound on the sum $\hat{A} \coloneqq \sum_i \hat{a}_i$ of outcomes $\hat{a}_i \in \{0,1\}$ of the type I measurements that Alice has, allowing the failure probability ϵ.

Since the setup of the problem above is invariant under the permutation of N-qubit system and uses independent and identical measurements, we can apply Theorems 1 and 3 as well as Azuma's and Kato's inequalities to this problem. For a positive parameter $\gamma > 0$ and $\Delta > 0$, we consider the following region $\mathcal{A}(\gamma,\Delta)$ as the "failure region" and will find an upper bound on the probability to fall in this region:

$$\begin{aligned}&\mathcal{A}(\gamma,\Delta)\\ &\coloneqq \{(p_1,p_2,p_3) : p_1,p_2,p_3 \geq 0,\ p_1+p_2+p_3 = 1,\\ &\qquad\qquad p_1 - \gamma p_2 \geq m(\gamma) + \Delta\},\end{aligned} \tag{156}$$

where $m(\gamma)$ denotes (an upper bound on) the expectation value and Δ denotes a deviation from it. If the probability that a triple $(\hat{A}/N, \hat{B}/N, 1-\hat{A}/N-\hat{B}/N)$ of outcomes lies in $\mathcal{A}(\gamma,\Delta)$ is bounded from above by ϵ, then we can say that the upper bound on $\hat{A}/N$ is given by $\gamma\hat{B}/N + m(\gamma) + \Delta$ allowing the failure probability ϵ. The term $m(\gamma)$ can be determined as follows. Let $M_c(\gamma)$ be a self-adjoint operator defined as

$$M_c(\gamma) = \frac{1}{2}|1\rangle\langle 1| - \frac{\gamma}{2}|\phi_1\rangle\langle\phi_1|. \tag{157}$$

Calculating the largest eigenvalue $m(\gamma)$ of $M_c(\gamma)$, we have the following inequality:

$$M_c(\gamma) \leq m(\gamma) I, \tag{158}$$

$$m(\gamma) \coloneqq \frac{1-\gamma+\sqrt{(1-\gamma)^2+4r\gamma}}{4}, \tag{159}$$

where I denotes the identity operator. This means that for any density operator ρ, we have

$$\mathbb{E}_\rho[\hat{a}_i - \gamma\hat{b}_i] \leq m(\gamma), \tag{160}$$

which then satisfies Eq. (156) in expectation for any $\gamma > 0$ and $\Delta > 0$. The reason why γ is taken to be positive is now clear. Since $|1\rangle$ and $|\phi_1\rangle$ has a large overlap, $\hat{A}$ is expected to get larger when $\hat{B}$ gets larger. Thus, the case $\gamma \leq 0$ is not expected to give a nontrivial bound.

For our purpose, we consider the case in which we fix the failure probability ϵ and find Δ as

the function of γ, N, and ϵ, which then satisfies $\Delta(\gamma, N, \epsilon) \to +0$ as $N \to \infty$, or we fix the deviation Δ and find ϵ as the function of γ, N, and Δ, which then satisfies $\epsilon(\gamma, N, \Delta) \to +0$ as $N \to \infty$. That means, we would like to find the function $\Delta(\gamma, N, \epsilon)$ that satisfies

$$\Pr\left[\left(\frac{\hat{A}}{N}, \frac{\hat{B}}{N}, 1 - \frac{\hat{A}}{N} - \frac{\hat{B}}{N}\right) \in \mathcal{A}(\gamma, \Delta(\gamma, N, \epsilon))\right] \leq \epsilon, \tag{161}$$

or find the function $\epsilon(\gamma, N, \Delta)$ that satisfies

$$\Pr\left[\left(\frac{\hat{A}}{N}, \frac{\hat{B}}{N}, 1 - \frac{\hat{A}}{N} - \frac{\hat{B}}{N}\right) \in \mathcal{A}(\gamma, \Delta)\right] \leq \epsilon(\gamma, N, \Delta). \tag{162}$$

First, we apply Azuma's inequality to the above problem. For $i \in \{1, \ldots, N\}$, let $\hat{c}_i$ be defined as $\hat{c}_i = \hat{a}_i - \gamma \hat{b}_i$. For the system i where the type I measurement is chosen, $\hat{b}_i$ is regarded as zero, and for the system i' where the type II measurement is chosen, $\hat{a}_{i'}$ is regarded as zero. Then, the operator $M_c(\gamma)$ gives the (expectation) value of the random variable $\hat{c}_i$ for a density operator ρ as

$$\mathbb{E}_\rho[\hat{c}_i] = \mathrm{Tr}[M_c(\gamma)\rho]. \tag{163}$$

Noticing that $\hat{c}_j \in \{0, 1, -\gamma\}$, we apply Azuma's inequality [28] (see also Proposition 2 in Ref. [34]) to $\hat{c}_j$ to obtain

$$\begin{aligned} &\Pr\left[\frac{1}{N}\sum_i (\hat{c}_i - \mathbb{E}[\hat{c}_i \mid \hat{c}_{<i}]) \geq t\right] \\ &\leq \exp[-2Nt^2/(1+\gamma)^2], \end{aligned} \tag{164}$$

where $\hat{c}_{<i} \coloneqq (\hat{c}_0, \ldots, \hat{c}_{i-1})$ with $\hat{c}_0 \coloneqq 0$. Since we have, for any $i \in \{1, \ldots, N\}$,

$$\mathbb{E}[\hat{c}_i \mid \hat{c}_{<i}] \leq \sup_\rho \mathrm{Tr}[M_c(\gamma)\rho] \leq m(\gamma), \tag{165}$$

we obtain, from Eqs. (164) and (165), the following bound:

$$\Pr\left[\frac{\hat{A}}{N} - \gamma\frac{\hat{B}}{N} \geq m(\gamma) + (1+\gamma)\sqrt{\frac{\ln(1/\epsilon)}{2N}}\right] \leq \epsilon, \tag{166}$$

$$\begin{aligned} &\Pr\left[\frac{\hat{A}}{N} - \gamma\frac{\hat{B}}{N} \geq m(\gamma) + \Delta\right] \\ &\leq \exp[-2N\Delta^2/(1+\gamma)^2]. \end{aligned} \tag{167}$$

Thus, setting

$$\Delta_{\mathrm{azuma}}(\gamma, N, \epsilon) = (1+\gamma)\sqrt{\frac{\ln(1/\epsilon)}{2N}} \tag{168}$$

achieves Eq. (161), and setting

$$\epsilon_{\mathrm{azuma}}(\gamma, N, \Delta) = \exp[-2N\Delta^2/(1+\gamma)^2] \tag{169}$$

achieves Eq. (162).

Next, for the application of Kato's inequality [14], we split the random variable $\hat{c}_i$ into $\hat{a}_i$ and $-\gamma\hat{b}_i$, and apply Kato's inequality to $\{\hat{b}_i\}_{i=1}^N$, which will eventually be observed by Bob. (Again, we regard $\hat{b}_i = 0$ when the type I measurement is chosen for i-th system.) For $\hat{a}_i$, we simply apply Azuma's inequality (since this cannot be observed by Bob). Using union bound, we finally obtain the concentration inequalities in the form Eqs. (161) and (162). In fact, by applying Kato's inequality under the filtration of $\{\hat{a}_{<i}, \hat{b}_{<i}\}_{j=1}^N$, we have [14]

$$\begin{aligned} &\Pr\left[\sum_{i=1}^N \mathbb{E}[\hat{b}_i \mid \hat{a}_{<i}, \hat{b}_{<i}] - \hat{b}_i \geq \left(\beta - \alpha + \frac{2\alpha}{N}\sum_{i=1}^N \hat{b}_i\right)\sqrt{N}\right] \\ &\leq \exp\left(-\frac{2(\beta^2 - \alpha^2)}{(1 + \frac{4\alpha}{3\sqrt{N}})^2}\right), \end{aligned} \tag{170}$$

where $\alpha, \beta \in \mathbb{R}$ and $\beta \geq 0$. We then need to optimize the parameters α and β in the inequality with the expected outcome of $\hat{B}$. In order to obtain a concentration inequality of the type Eq. (161), we optimize α and β so that the right-hand side of Eq. (170) is equal to $\epsilon/2$. Let B^* be the expected outcome of $\hat{B}$. Then, the good choices of parameters are [35]

$$\begin{aligned} &\alpha^*(B^*, N, \epsilon/2) \\ &\coloneqq \frac{27\sqrt{2}N(N-2B^*)\sqrt{\ln(2/\epsilon)[9B^*(N-B^*) + 2N\ln(2/\epsilon)]}}{4(9N + 8\ln(2/\epsilon))(9B^*(N-B^*) + 2N\ln(2/\epsilon))} \\ &\quad - \frac{54\sqrt{N}B^*(N-B^*)\ln(2/\epsilon) + 12N^{\frac{3}{2}}(\ln(2/\epsilon))^2}{(9N + 8\ln(2/\epsilon))(9B^*(N-B^*) + 2N\ln(2/\epsilon))}, \end{aligned} \tag{171}$$

$$\begin{aligned} &\beta^*(B^*, N, \epsilon/2) \\ &\coloneqq \sqrt{[\alpha^*(B^*, N, \epsilon/2)]^2 + \frac{\ln(2/\epsilon)}{2}\left(1 + \frac{4\alpha^*(B^*, N, \epsilon/2)}{3\sqrt{N}}\right)^2}. \end{aligned} \tag{172}$$

From Eq. (170) and applying Azuma's inequality

to $\{\hat{a}_i\}_{i=1}^N$, we have

$$\Pr\Bigg[\sum_{i=1}^N \hat{a}_i - \mathbb{E}[\hat{a}_i \mid \hat{a}_{<i}, \hat{b}_{<i}] - \gamma\left(\hat{b}_i - \mathbb{E}[\hat{b}_i \mid \hat{a}_{<i}, \hat{b}_{<i}]\right) \geq \gamma\sqrt{N}\left(\beta^* - \alpha^* + \frac{2\alpha^*}{N}\sum_{i=1}^N \hat{b}_i\right) + \sqrt{\frac{N\ln(2/\epsilon)}{2}}\Bigg] \leq \epsilon, \tag{173}$$

from the union bound, where we abbreviate the dependencies of α^* and β^* on B^*, N, and $\epsilon/2$. Since

$$\mathbb{E}[\hat{a}_i - \gamma\hat{b}_i \mid \hat{a}_{<i}, \hat{b}_{<i}] \leq \sup_\rho \mathrm{Tr}[M_c(\gamma)\rho] \leq m(\gamma) \tag{174}$$

again holds, we have

$$\Pr\left[\frac{\hat{A}}{N} - \gamma\frac{\hat{B}}{N} \geq m(\gamma) + \sqrt{\frac{\ln(2/\epsilon)}{2N}} + \frac{\gamma}{\sqrt{N}}\left(\beta^* - \alpha^* + \frac{2\alpha^*\hat{B}}{N}\right)\right] \leq \epsilon. \tag{175}$$

Thus, setting

$$\Delta_{\mathrm{kato}}(\gamma, N, \epsilon) = \frac{\gamma}{\sqrt{N}}\left[\beta^* - \alpha^*\left(1 - \frac{2\hat{B}}{N}\right)\right] + \sqrt{\frac{\ln(2/\epsilon)}{2N}} \tag{176}$$

achieves Eq. (161) in this case.

Due to the use of the union bound, the derivation of a concentration inequality of the type Eq. (162) is not carried out in the same way as the case of Azuma's inequality. With Corollary 1 in Ref. [14] applied to $\{\hat{b}\}_{i=1}^N$ and δ and ϵ in Corollary 1 set to $1 - 2B^*/N$ and $\Delta'\sqrt{N}$, respectively, we have

$$\Pr\left[\frac{1}{N}\sum_{i=1}^N (\mathbb{E}[\hat{b}_i \mid \hat{a}_{<i}, \hat{b}_{<i}] - \hat{b}_i) \geq \xi(\hat{B}, B^*, \Delta')\Delta'\right] \leq \exp\left[-\frac{2N\Delta'^2}{1 - (1 - 2\frac{B^*}{N} - \frac{4}{3}\Delta')^2}\right], \tag{177}$$

where $\xi(\hat{B}, B^*, \Delta')$ is defined as

$$\xi(\hat{B}, B^*, \Delta') \coloneqq \frac{1 - (1 - 2\frac{\hat{B}}{N})(1 - \frac{B^*}{N} - \frac{4}{3}\Delta')}{1 - (1 - 2\frac{B^*}{N})(1 - \frac{B^*}{N} - \frac{4}{3}\Delta')}. \tag{178}$$

When $\hat{B} = B^*$, it is unity, i.e., $\xi(B^*, B^*, \Delta') = 1$. Now, given Δ, choose Δ' so that $\gamma\xi(\hat{B}, B^*, \Delta')\Delta' \leq \Delta$. Then, with Azuma's inequality applied to $\{\hat{a}_i\}_{i=1}^N$, we have

$$\Pr\left[\frac{1}{N}\sum_{i=1}^N (\hat{a}_i - \mathbb{E}[\hat{a}_i \mid \hat{a}_{<i}, \hat{b}_{<i}]) \geq \Delta - \gamma\xi(\hat{B}, B^*, \Delta')\Delta'\right] \leq \exp\left[-2N(\Delta - \gamma\xi(\hat{B}, B^*, \Delta')\Delta')^2\right]. \tag{179}$$

Combining Eqs. (160), (177), and (179), we have

$$\Pr\left[\frac{\hat{A}}{N} - \gamma\frac{\hat{B}}{N} \geq m(\gamma) + \Delta\right] \leq \epsilon_{\mathrm{kato}}(\gamma, N, \Delta), \tag{180}$$

where $\epsilon_{\mathrm{kato}}(\gamma, N, \Delta)$ is given by

$$\epsilon_{\mathrm{kato}}(\gamma, N, \Delta) = \min_{0 \leq \gamma\xi(\hat{B}, B^*, \Delta')\Delta' \leq \Delta} \exp\left[-\frac{2N\Delta'^2}{1 - (1 - 2\frac{B^*}{N} - \frac{4}{3}\Delta')^2}\right] + \exp\left[-2N(\Delta - \gamma\xi(\hat{B}, B^*, \Delta')\Delta')^2\right]. \tag{181}$$

Third, we consider applying Theorem 1 to this problem. For that, we consider the case of i.i.d. quantum state first. Let $\mathcal{Q}$ be the convex set of probability mass functions that are realizable by this measurement setup. Since we estimate the frequency of the event $(\mathrm{I}, 1)$ from that of $(\mathbb{I}, 1)$, we need not to distinguish $(\mathrm{I}, 0)$ and $(\mathbb{I}, 0)$, and thus we group these outcomes as a single outcome $(\cdot, 0)$. Thus, the set $\mathcal{Q}$ can be explicitly given as

$$\mathcal{Q} \coloneqq \Bigg\{(p(\mathrm{I}, 1), p(\mathbb{I}, 1), p(\cdot, 0)) : \forall\rho,\ p(\mathrm{I}, 1) = \frac{\langle 1|\rho|1\rangle}{2}, p(\mathbb{I}, 1) = \frac{\langle\phi_1|\rho|\phi_1\rangle}{2}, p(\cdot, 0) = \frac{\langle 0|\rho|0\rangle + \langle\phi_0|\rho|\phi_0\rangle}{2}\Bigg\}. \tag{182}$$

Now, assume that the i.i.d. state $\rho(\boldsymbol{q})^{\otimes n}$ is prepared with $\rho(\boldsymbol{q})$ giving the probability mass function $\boldsymbol{q} \in \mathcal{Q}$. Then, the probability that the sums of the outcomes of these measurements $\hat{A}$ and $\hat{B}$ lies in the region $\mathcal{A}(\gamma, \Delta_{\mathrm{iid}})$ is given from the refined version of the Sanov's theorem [2, 3] as

$$\Pr\left[\left(\frac{\hat{A}}{N}, \frac{\hat{B}}{N}, 1 - \frac{\hat{A}}{N} - \frac{\hat{B}}{N}\right) \in \mathcal{A}(\gamma, \Delta_{\mathrm{iid}}) \middle| \rho(\boldsymbol{q})^{\otimes n}\right] \leq \max_{\boldsymbol{p} \in \mathcal{A}(\gamma, \Delta_{\mathrm{iid}})} \exp[-ND(\boldsymbol{p}\|\boldsymbol{q})]. \tag{183}$$

Then, for any i.i.d. state $\rho^{\otimes N}$, we have the following bound:

$$\Pr\left[\left(\frac{\hat{A}}{N},\frac{\hat{B}}{N},1-\frac{\hat{A}}{N}-\frac{\hat{B}}{N}\right)\in\mathcal{A}(\gamma,\Delta_{\mathrm{iid}})\middle|\rho^{\otimes N}\right] \le \max_{\boldsymbol{p}\in\mathcal{A}(\gamma,\Delta_{\mathrm{iid}})}\max_{\boldsymbol{q}\in\mathcal{Q}}\exp[-ND(\boldsymbol{p}\|\boldsymbol{q})]. \tag{184}$$

Now, we consider a permutation-invariant state ρ_{sym} and apply Theorem 1 to have

$$\Pr\left[\left(\frac{\hat{A}}{N},\frac{\hat{B}}{N},1-\frac{\hat{A}}{N}-\frac{\hat{B}}{N}\right)\in\mathcal{A}(\gamma,\Delta_q)\middle|\rho_{\mathrm{sym}}\right] \le \max_{\boldsymbol{p}\in\mathcal{A}(\gamma,\Delta_q)}\max_{\boldsymbol{q}\in\mathcal{Q}} f_q(N,2)\exp[-ND(\boldsymbol{p}\|\boldsymbol{q})]. \tag{185}$$

However, since the sum $\hat{A}$ of measurement outcomes is permutation invariant and thus the permutation symmetrization does not change the probability, we can apply the same bound for any state that is not necessarily permutation invariant. Finally, by choosing the parameter Δ_q so that the right-hand side of Eq. (185) is equal to ϵ, we obtain Eq. (161). (The parameter Δ_q is thus the function of γ, N, and ϵ.) On the other hand, if we fix $\Delta_q=\Delta$ and set

$$\epsilon_q(\gamma,N,\Delta)=\max_{\boldsymbol{p}\in\mathcal{A}(\gamma,\Delta)}\max_{\boldsymbol{q}\in\mathcal{Q}} f_q(N,2)\exp[-ND(\boldsymbol{p}\|\boldsymbol{q})], \tag{186}$$

then we obtain a concentration inequality of the type Eq. (162). For later use, we define $\Delta_{\mathrm{iid}}(\gamma,N,\epsilon)$ and $\epsilon_{\mathrm{iid}}(\gamma,N,\Delta)$ in the same way using Eq. (184), i.e.,

$$\epsilon_{\mathrm{iid}}(\gamma,N,\Delta)=\max_{\boldsymbol{p}\in\mathcal{A}(\gamma,\Delta)}\max_{\boldsymbol{q}\in\mathcal{Q}}\exp[-ND(\boldsymbol{p}\|\boldsymbol{q})]. \tag{187}$$

We finally consider applying Theorem 3 to this problem. As previously mentioned, an independent and identical measurement is performed on an unknown quantum state and the set of measurement outcomes is given by $\{(\mathrm{I},0),(\mathrm{I},1),(\mathrm{II},0),(\mathrm{II},1)\}$. Since we do not distinguish $(\mathrm{II},0)$ and $(\mathrm{II},1)$ as mentioned previously, we regard this measurement as three-outcome measurement and set $k=3$. For the set $\mathcal{Q}$ of probability mass functions that can be realized by this measurement given in Eq. (182), we consider the following set $\mathcal{R}(\gamma')\supseteq\mathcal{Q}$:

$$\mathcal{R}(\gamma')=\Big\{(p(\mathrm{I},1),p(\mathrm{II},1),p(\cdot,0)):\ p(\mathrm{I},1)-\gamma' p(\mathrm{II},1)\le m(\gamma')\Big\}. \tag{188}$$

The fact that this includes $\mathcal{Q}$ as a subset follows from Eqs. (157) and (159). Due to the requirement of Theorem 3 that the complement $\mathcal{P}\backslash\mathcal{R}(\gamma')$ of $\mathcal{R}(\gamma')$ contains only one probability mass function that corresponds to having a deterministic outcome, the parameter γ' is chosen to satisfy $0\le m(\gamma')<1$. Likewise, the region $\mathcal{R}'(\gamma')$ can be defined through Eq. (58) by

$$\mathcal{R}'(\gamma')=\left\{(p(\mathrm{I},1),p(\mathrm{II},1),p(\cdot,0)):\ p(\mathrm{I},1)-\gamma' p(\mathrm{II},1)\le m(\gamma')+\frac{\sqrt{2}}{N}\sqrt{\frac{2\gamma'^2+2\gamma'+2}{3}}\right\}. \tag{189}$$

Thus, from Eq. (155), we have

$$\Pr\left[\left(\frac{\hat{A}}{N},\frac{\hat{B}}{N},1-\frac{\hat{A}}{N}-\frac{\hat{B}}{N}\right)\in\mathcal{A}(\gamma,\Delta_c)\right] \le \max_{\boldsymbol{p}\in\mathcal{A}(\gamma,\Delta_c)}\min_{\gamma':0\le m(\gamma')<1}\max_{\boldsymbol{q}\in\mathcal{R}'(\gamma')} f_c(N,3)\exp[-ND(\boldsymbol{p}\|\boldsymbol{q})], \tag{190}$$

where minimization over γ' is inserted because it is a free parameter. Then, by setting the parameter Δ_c so that the right-hand side of Eq. (190) is equal to ϵ, we achieve Eq. (161). (The parameter Δ_c is thus the function of γ, N, and ϵ.) On the other hand, by setting $\Delta_c=\Delta$ and setting

$$\epsilon_c(\gamma,N,\Delta) = \max_{\boldsymbol{p}\in\mathcal{A}(\gamma,\Delta)}\min_{\gamma':0\le m(\gamma')<1}\max_{\boldsymbol{q}\in\mathcal{R}'(\gamma')} f_c(N,3)\exp[-ND(\boldsymbol{p}\|\boldsymbol{q})], \tag{191}$$

we obtain a concentration inequality of the type Eq. (162).

When applied to this specific problem, we can contrast the bounds Eqs. (186) and (191). In Eq. (186), the polynomial factor depends on the dimension of the local quantum system that is eventually measured, but the region with which the maximization of the Kullback-Leibler divergence is taken is the convex set $\mathcal{Q}$ of all the possible probability mass functions realized by the measurement $\{\frac{1}{2}|1\rangle\langle 1|,\frac{1}{2}|\phi_1\rangle\langle\phi_1|,\frac{1}{2}(|0\rangle\langle 0|+$

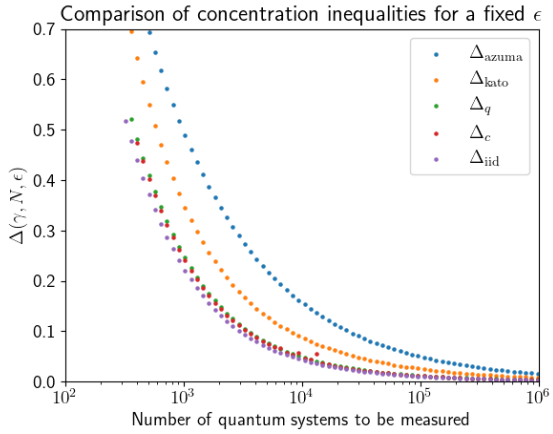

Figure 2: Comparison of the deviation $\Delta(\gamma, N, \epsilon)$ from (an upper bound on) the expectation value in Eqs. (156) and (161) when allowing the failure probability ϵ. The parameters are chosen to be $\epsilon = 10^{-30}$, $r = 0.01$, and $B^* = 0.01N$, and γ is optimized to be 1.662 so that an unconfirmed upper bound $\gamma B^*/N + m(\gamma)$ on $\hat{A}/N$ is minimized to be 0.02873. We further assume that the random variable $\hat{B}$ is equal to the expected value B^*. The functions Δ_{azuma} and Δ_{kato} are defined respectively in Eqs. (168) and (176), and the functions Δ_q, and Δ_c are defined implicitly through Eqs. (185) and (190). For comparison, we also plot the deviation Δ_{iid} when the i.i.d. quantum state is assumed to be given, where Δ_{iid} is defined implicitly through Eq. (184).

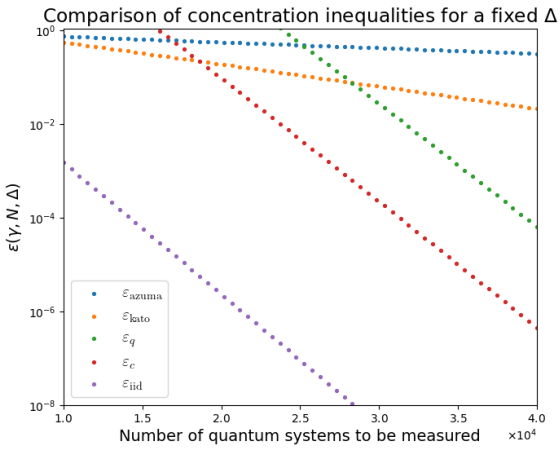

Figure 3: Comparison of the failure probability $\epsilon(\gamma, N, \Delta)$ in Eq. (162) against the number N of measured quantum systems for a fixed value of the deviation Δ. The parameters are chosen to be $\Delta = 0.01$, $r = 0.01$, and $B^* = 0.01N$, and γ is optimized to be 1.662 so that an unconfirmed upper bound $\gamma B^*/N + m(\gamma)$ on $\hat{A}/N$ is minimized to be 0.02873. We further assume that the random variable $\hat{B}$ is equal to the expected value B^*. The functions $\epsilon_{\mathrm{azuma}}$, ϵ_q, ϵ_c, and ϵ_{iid} are defined in Eqs. (169), (186), (191), and (187), respectively.

$|\phi_0\rangle\langle\phi_0|)\}$ on all the possible local quantum states. In Eq. (191) on the other hand, the polynomial factor depends on the number of outcomes of the measurement and is independent of the dimension of the underlying quantum system. The region with which the maximization of the Kullback-Leibler divergence is taken is, however, larger than $\mathcal{Q}$. This may be a price to pay. When the dimension of the quantum system is small but the number of outcomes of the measurement is large, Eq. (186) may be smaller. On the contrary, if the dimension of the quantum system is large (even infinite) but the number of outcomes of the measurement is small, Eq. (191) may be smaller.

We numerically demonstrate the tightness of these bounds in some cases. The first example is the case that is relevant to the finite-size analysis of the QKD. In QKD, the parameter ϵ is fixed and its value is typically chosen to be $\sim 10^{-30}$. In the QKD application, we have the expected value B^* of $\hat{B}$ from the channel that we use for the transmission, which is the reason why Kato's inequality [14] typically gives a tighter bound in the finite-size regime. For a model, we set $\epsilon = 10^{-30}$, $r = 0.01$, and $B^* = 0.01N$. The parameter γ in $\mathcal{A}(\gamma, \Delta)$ is chosen so that the expected (unconfirmed) upper bound $\gamma B^*/N + m(\gamma)$ on $\hat{A}/N$ in the limit $N \to \infty$ is minimized. In the choice of parameters above, we have $\gamma = 1.662$ to give the expected asymptotic upper bound 0.02873 on $\hat{A}/N$. We further assume the case in which $\hat{B}$ is exactly equal to B^*. Under these assumptions, we derive the deviation from the above asymptotic upper bound for finite N allowing the failure probability ϵ with the four methods described above. Figure 2 shows the deviation $\Delta(\gamma, N, \epsilon)$ when varying the number N of quantum systems to be measured. As the figure suggests, Δ_c defined through Eq. 190 is the smallest among four bounds in an adversarial setup (i.e., excluding Δ_{iid}) and with the above particular choices of parameters, except for some points that may be caused by the insufficient optimization. We can also see that Δ_q defined through Eq. (185) gives the comparable bound, and these two (Δ_c and Δ_q) are as small as that of the i.i.d. case given by Δ_{iid}. They are smaller than Δ_{kato} defined in Eq. (176), while Δ_{kato} is still smaller than the Δ_{azuma} defined in Eq. (168).

The next example is the case when we fix the deviation Δ from the expectation value (that is, an upper bound that holds in the asymptotic limit as shown in Eq. (165)) to a small value, and see how quickly the failure rate $\epsilon(\gamma, N, \Delta)$ in Eq. (162) goes to zero. We set the deviation $\Delta = 0.01$, $r = 0.01$ and $B^* = 0.01N$, and choose $\gamma = 1.662$ to give the asymptotically optimal upper bound 0.02873 on $\hat{A}/N$ as was done in the previous figure. We also consider the case in which the observed value $\hat{B}$ is exactly equal to B^*. With these conditions, we plot ϵ_{azuma} in Eq. (169), ϵ_{kato} in Eq. (181), ϵ_q in Eq. (186), ϵ_c in Eq. (191), and ϵ_{iid} in Eq. (187) in Fig. 3. As the figure suggests, the failure probability exponentially decays, but the exponents are very different between Azuma's inequality, Kato's inequality, and ours. In fact, the convergence speed of our bounds is much quicker than that of Azuma's or Kato's inequality and as quick as that of i.i.d.. Furthermore, ϵ_q, ϵ_c, and ϵ_{iid} appear to have almost constant gaps in the logarithmic plot. This gap exactly comes from the difference of the factors $f_q(N, d = 2) \sim N^{\frac{3}{2}}$, $f_c(N, k = 3) \sim N$, and 1. We should note that there are regions in which Azuma's and Kato's inequality gives a better bound than ours when N is small and the failure probability is large. When we alter the parameters r, M, and Δ in our model, then this region varies. However, our bound overall gives a faster exponential decay.

3.2 Application to quantum key distribution

For the next application, we consider a QKD protocol with qubits. Even though Bennett-Brassard 1984 protocol [36] may be the most common QKD protocol, its finite-size security analysis can be reduced to a classical random sampling problem [37, 38, 39] in which one can apply a better concentration inequality than Azuma's and our inequalities. Therefore, we alternatively consider Phoenix-Barnett-Chefles 2000 (PBC00) protocol [40], which uses three quantum states forming an equilateral triangle on the X-Z plane in the Bloch sphere. The security of this protocol is proved in the finite-size case against general attacks in Ref. [11] using Azuma's inequality. In order to make the paper as self-contained as possible, we list the protocol below. Let $\{|0\rangle, |1\rangle\}$ be the Z basis of a qubit. Let $\{|+\rangle, |-\rangle\}$ with $|\pm\rangle := (|0\rangle \pm |1\rangle)/\sqrt{2}$ be the X basis. The three states used in PBC00 protocol are $|\psi_1\rangle := \cos(2\pi/3)|+\rangle + \sin(2\pi/3)|-\rangle$, $|\psi_2\rangle := \cos(4\pi/3)|+\rangle + \sin(4\pi/3)|-\rangle$, and $|\psi_3\rangle := |+\rangle$. Let us also define $|\bar{\psi}_i\rangle$ for $i = 0, 1, 2$ that satisfies $\langle\psi_i|\bar{\psi}_i\rangle = 0$. Then, the PBC00 protocol is described as follows.

— **PBC00** —

1. For each of N rounds, Alice uniformly randomly generates a trit $\hat{r}_i$ and a bit $\hat{a}_i$, where $i \in \{1, \ldots, N\}$ denotes the label of the round. She chooses the pair $\{|\psi_1\rangle, |\psi_2\rangle\}$ if $\hat{r}_i = 0$, $\{|\psi_2\rangle, |\psi_3\rangle\}$ if $\hat{r}_i = 1$, and $\{|\psi_3\rangle, |\psi_1\rangle\}$ if $\hat{r}_i = 2$. If $\hat{a}_i = 0(1)$, she sends the first (second) state of the chosen pair to Bob.

2. For each qubit received in the i-th round, Bob performs a measurement described by the POVM $\{\frac{2}{3}|\bar{\psi}_1\rangle\langle\bar{\psi}_1|, \frac{2}{3}|\bar{\psi}_2\rangle\langle\bar{\psi}_2|, \frac{2}{3}|\bar{\psi}_3\rangle\langle\bar{\psi}_3|\}$.

3. After the N rounds of communication, Alice announces the sequence of trits $\{\hat{r}_i\}_{i=1}^N$. Bob defines $\hat{b}_i = 0(1)$ if $\hat{r}_i = 0$ and his measurement outcome in the i-th round is $|\bar{\psi}_2\rangle$ ($|\bar{\psi}_1\rangle$), if $\hat{r}_i = 1$ and his measurement outcome in the i-th round is $|\bar{\psi}_3\rangle$ ($|\bar{\psi}_2\rangle$), and if $\hat{r}_i = 2$ and his measurement outcome in the i-th round is $|\bar{\psi}_1\rangle$ ($|\bar{\psi}_3\rangle$). All other events are regarded as inconclusive. Bob announces whether his measurement is inconclusive or not for each i. Alice and Bob keep $\hat{a}_i$ and $\hat{b}_i$, respectively, when Bob's outcome in the i-th round is conclusive. Let $\hat{N}_{\text{con}}$ be the number of conclusive rounds.

4. For each element of the $\hat{N}_{\text{con}}$-bit sequence that Alice keeps in the previous step, Alice uniformly randomly chooses the label "test" or "sift" and announces it. Let $\hat{N}_{\text{test}}$ and $\hat{N}_{\text{sift}}$ be the number of test and sift bits, respectively, which satisfies $\hat{N}_{\text{con}} = \hat{N}_{\text{test}} + \hat{N}_{\text{sift}}$. Alice and Bob announce all the bits labeled by "test" and estimate the bit error rate. From the estimated bit error rate, they perform a bit error correction on the sifted key, i.e., $\hat{N}_{\text{sift}}$-bit sequence labeled by "sift". Let $\hat{N}_{\text{KC}}$ be the number of pre-shared secret keys consumed during the bit error correction.

5. Alice and Bob agree on the amount $\hat{N}_{\text{PA}}$ of

privacy amplification and perform the linear hash function to obtain the final key.

As can be seen from the protocol, the net key gain $\hat{G}$ per communication rounds of this protocol can be given by

$$\hat{G} = (\hat{N}_{\text{sift}} - \hat{N}_{\text{KC}} - \hat{N}_{\text{PA}})/N. \tag{192}$$

In Ref. [11], the protocol is shown to be secure in the finite-size regime by the so-called phase error correction approach [7, 6, 8]. In fact, if we consider the purified version of the protocol defined above in which Alice instead prepares an entangled state $|\Psi\rangle_{AB} \coloneqq (|0\rangle_A \otimes R_y(2\pi\hat{r}_i/3)\,|\psi_1\rangle_B + |1\rangle_A \otimes R_y(2\pi(\hat{r}_i+1)/3)\,|\psi_1\rangle_B)\sqrt{2}$, where $R_y(\theta)$ denotes the θ rotation around y axis in the Bloch sphere, and sends the system B to Bob, then the security of the key can be reduced to how much maximally entangled state $|\Phi\rangle_{AB} \coloneqq (|0\rangle_A\,|0\rangle_B + |1\rangle_A\,|1\rangle_B)/\sqrt{2}$ Alice and Bob can extract in this purified protocol. Thus, the bit error operator Π_{bit} and phase error operator Π_{ph} can be defined, which correspond to the event that they obtain bit and phase flipped maximally entangled states, respectively, in the purified protocol. Explicitly, Π_{bit} and Π_{ph} are given by [11]

$$\begin{aligned}
&3\Pi_{\text{bit}} \\
&= \sum_{r=0}^{2}\Big[|0\rangle\langle 0|_A \otimes R_y\Big(\frac{2\pi r}{3}\Big)F\,|1\rangle\langle 1|_B\,FR_y\Big(-\frac{2\pi r}{3}\Big) \\
&\quad + |1\rangle\langle 1|_A \otimes R_y\Big(\frac{2\pi r}{3}\Big)F\,|0\rangle\langle 0|_B\,FR_y\Big(-\frac{2\pi r}{3}\Big)\Big],
\end{aligned} \tag{193}$$

$$\begin{aligned}
&3\Pi_{\text{ph}} \\
&= \sum_{r=0}^{2}\Big[|+\rangle\langle +|_A \otimes R_y\Big(\frac{2\pi r}{3}\Big)F\,|-\rangle\langle -|_B\,FR_y\Big(-\frac{2\pi r}{3}\Big) \\
&\quad + |-\rangle\langle -|_A \otimes R_y\Big(\frac{2\pi r}{3}\Big)F\,|+\rangle\langle +|_B\,FR_y\Big(-\frac{2\pi r}{3}\Big)\Big],
\end{aligned} \tag{194}$$

where the filter operator $F \coloneqq |0\rangle\langle 0| + \frac{1}{\sqrt{3}}\,|1\rangle\langle 1|$. From these, we can derive the following [11]:

$$\Pi_{\text{ph}} \le \frac{5}{4}\Pi_{\text{bit}}. \tag{195}$$

Since the bit error rate e_{bit} can directly be observed in the "test" rounds, the asymptotic key rate $\hat{G}_{\text{aympt}}$ of this protocol can be given by

$$\hat{G}_{\text{asympt}} = \frac{1}{4}\left(1 - h(e_{\text{bit}}) - h\Big(\frac{5}{4}e_{\text{bit}}\Big)\right), \tag{196}$$

where $h(x) \coloneqq -x\log_2 x - (1-x)\log_2(1-x)$ denotes the binary-entropy function. In the above, the prefactor $1/4$ comes from the fact that the measurement is conclusive with the probability $1/2$ (without Eve's attack) and that the "sift" is chosen with the probability $1/2$. In the finite-size case, however, we need to take into account the statistical fluctuations. Let us consider the purified protocol again and $\hat{x}^{(i)}_{\text{bit}}$ be a random variable at i-th round that takes the value one when "test" is chosen and the bit error is observed while takes zero otherwise. Then, $\hat{N}_{\text{bit}} \coloneqq \sum_{i=1}^{N}\hat{x}^{(i)}_{\text{bit}}$ corresponds to the number of bit error observed in the protocol. Let $\hat{x}^{(i)}_{\text{ph}}$ be another random variable at i-th round that takes the value one when "sift" is chosen and the phase error occurs while takes zero otherwise. Then, $\hat{N}_{\text{ph}} \coloneqq \sum_{i=1}^{N}\hat{x}^{(i)}_{\text{ph}}$ corresponds to the number of phase errors in the sifted key. It is known that if we can show

$$\Pr[\hat{N}_{\text{ph}} \ge U(\hat{N}_{\text{bit}})] \le \epsilon, \tag{197}$$

for a function $U(\hat{N}_{\text{bit}})$ of $\hat{N}_{\text{bit}}$, and by setting

$$\hat{N}_{\text{PA}} = \hat{N}_{\text{sift}}h(U(\hat{N}_{\text{bit}})/\hat{N}_{\text{sift}}) + s, \tag{198}$$

then the protocol is $\sqrt{2(\epsilon + 2^{-s})}$-secret [8, 38, 9]. Furthermore, we assume that the secret-key consumption in the bit error correction at Step 4 of the protocol is given by

$$\hat{N}_{\text{KC}} = \hat{N}_{\text{sift}}h(\hat{N}_{\text{bit}}/\hat{N}_{\text{sift}}) + s', \tag{199}$$

for sending an $\hat{N}_{\text{sift}}h(\hat{N}_{\text{bit}}/\hat{N}_{\text{sift}})$-bit syndrome string and s'-bit verification string to ensure $2^{-s'}$-correctness. Thus, if we can obtain the function $U(\cdot)$ to satisfy Eq. (197), we can obtain a key rate in the finite-size case as well satisfying $2^{-s'}$-correctness and $\sqrt{2(\epsilon + 2^{-s})}$-secrecy.

The derivation of Eq. (197) is very similar to the one in the previous section. When Azuma's or Kato's is applied to the problem, the operator inequality (195) can be used as an inequality between the conditional expectations of $\hat{x}^{(i)}_{\text{bit}}$ and $\hat{x}^{(i)}_{\text{ph}}$. By applying Azuma's inequality to $\hat{x}^{(i)}_{\text{ph}} - \frac{5}{4}\hat{x}^{(i)}_{\text{bit}}$, we have

$$\Pr\left[\frac{\hat{N}_{\text{ph}}}{N} - \frac{5}{4}\frac{\hat{N}_{\text{bit}}}{N} \ge \left(1 + \frac{5}{4}\right)\sqrt{\frac{\ln(1/\epsilon)}{2N}}\right] \le \epsilon. \tag{200}$$

In this case, the function U_{azuma} to satisfy Eq. (197) should be

$$U_{\text{azuma}}(\hat{N}_{\text{bit}}) = \frac{5}{4}\hat{N}_{\text{bit}} + \frac{9}{4}\sqrt{\frac{N\ln(1/\epsilon)}{2}}. \quad (201)$$

Similarly, by applying Kato's inequality to $\hat{x}^{(i)}_{\text{bit}}$ while applying Azuma's inequality to $\hat{x}^{(i)}_{\text{ph}}$, we have

$$\Pr\left[\frac{\hat{N}_{\text{ph}}}{N} - \frac{5}{4}\frac{\hat{N}_{\text{bit}}}{N} \geq \sqrt{\frac{\ln(2/\epsilon)}{2N}} + \frac{5}{4\sqrt{N}}\left[\beta - \alpha\left(1 - \frac{2\hat{N}_{\text{bit}}}{N}\right)\right]\right] \leq \epsilon, \quad (202)$$

with

$$\alpha := \alpha^*(N^*_{\text{bit}}, N, \epsilon/2), \quad (203)$$
$$\beta := \beta^*(N^*_{\text{bit}}, N, \epsilon/2), \quad (204)$$

where the functions α^* and β^* are defined in Eqs. (171) and (172) with N^*_{bit} being an unconfirmed knowledge of $\hat{N}_{\text{bit}}$ prior to the protocol. The function U_{kato} to satisfy Eq. (197) should thus be

$$U_{\text{kato}}(\hat{N}_{\text{bit}}) = \frac{5}{4}\hat{N}_{\text{bit}} + \sqrt{\frac{N\ln(2/\epsilon)}{2}} + \frac{5\sqrt{N}}{4}\left(\beta - \alpha\left(1 - \frac{2\hat{N}_{\text{bit}}}{N}\right)\right). \quad (205)$$

Now we apply Theorem 1 to this protocol. Let $\mathcal{X}$ be a sample space defined as

$$\mathcal{X} := \{\text{phase error in "sift"}, \text{bit error in "test"}, \text{others}\}, \quad (206)$$

and $\boldsymbol{p} = (p_1, p_2, p_3)$ be a probability mass function on $\mathcal{X}$. Then, from Eq. (195), the set $\mathcal{Q}$ of allowed probability mass function for any quantum state between Alice and Bob can be given by

$$\mathcal{Q} = \left\{\boldsymbol{p} = (p_1, p_2, p_3) : p_1 \leq \frac{5}{4}p_2\right\}, \quad (207)$$

where we used the fact that "sift" and "test" rounds are chosen with equal probability. Thus, for any i.i.d. quantum state $\rho_{AB}^{\otimes N}$ between Alice and Bob, we have

$$\Pr\left[\left(\frac{\hat{N}_{\text{ph}}}{N}, \frac{\hat{N}_{\text{bit}}}{N}, 1 - \frac{\hat{N}_{\text{ph}}}{N} - \frac{\hat{N}_{\text{bit}}}{N}\right) \in \mathcal{A}(\delta)\middle|\rho_{AB}^{\otimes N}\right] \leq \max_{\boldsymbol{p}\in\mathcal{A}(\delta)} \max_{\boldsymbol{q}\in\mathcal{Q}} \exp[-ND(\boldsymbol{p}\|\boldsymbol{q})], \quad (208)$$

from Sanov's theorem [2, 3], where $\mathcal{A}(\delta)$ is defined as

$$\mathcal{A}(\delta) = \left\{\boldsymbol{p} = (p_1, p_2, p_3) : p_1 \geq \frac{5}{4}p_2 + \delta\right\}. \quad (209)$$

Now, since the number of bit and phase errors is invariant under the permutation of rounds, we can virtually permute the four-dimensional quantum system between Alice and Bob in the purified protocol across the rounds. Thus, we can apply Theorem 1 to the above and obtain

$$\Pr\left[\left(\frac{\hat{N}_{\text{ph}}}{N}, \frac{\hat{N}_{\text{bit}}}{N}, 1 - \frac{\hat{N}_{\text{ph}}}{N} - \frac{\hat{N}_{\text{bit}}}{N}\right) \in \mathcal{A}(\delta)\right] \leq \max_{\boldsymbol{p}\in\mathcal{A}(\delta)} \max_{\boldsymbol{q}\in\mathcal{Q}} f_q(N, 4)\exp[-ND(\boldsymbol{p}\|\boldsymbol{q})]. \quad (210)$$

We should optimize the value δ in the above so that the right-hand side is smaller than or equal to ϵ. If we write such an optimized δ in this case as δ_q, the function U_q to satisfy Eq. (197) is given by

$$U_q(\hat{N}_{\text{bit}}) = \frac{5}{4}\hat{N}_{\text{bit}} + \delta_q. \quad (211)$$

Finally, we apply Theorem 3 to this problem. We can regard the measurement in the purified protocol as the three-outcome measurement channel as described above. Since the set of probability mass functions after measurement is included in $\mathcal{Q}$ in Eq. (207), which is already linear and contains all the extremal points of the simplex of probability mass functions except $(1, 0, 0)$, we can choose $\mathcal{R} = \mathcal{Q}$, where $\mathcal{R}$ is defined in Theorem 3. The region $\mathcal{R}'$ in Theorem 3 is thus given by

$$\mathcal{R}' = \left\{\boldsymbol{p} = (p_1, p_2, p_3) : p_1 - \frac{5}{4}p_2 \leq \frac{\sqrt{2}}{N}\sqrt{\left(\frac{13}{12}\right)^2 + \left(-\frac{14}{12}\right)^2 + \left(\frac{1}{12}\right)^2}\right\}. \quad (212)$$

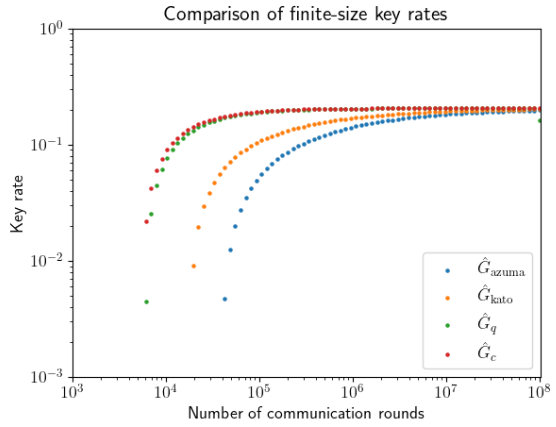

Figure 4: Comparison of key rates against the total number N of rounds when the bit error rate is one percent. The key rates are given by the formula (215), where the function U is either $U_{\rm azuma}$ in Eq. (201), $U_{\rm kato}$ in Eq. (205), U_q in Eq. (211), or U_c in Eq. (214) with the parameter $\epsilon = 2^{-102}$. The parameters s and s' in Eq. (215) are chosen to be $s = 102$ and $s' = 50$ so that the correctness and secrecy parameter of the protocol are both 2^{-50}.

Therefore, from Theorem 3, we have

$$\Pr\left[\left(\frac{\hat{N}_{\rm ph}}{N}, \frac{\hat{N}_{\rm bit}}{N}, 1 - \frac{\hat{N}_{\rm ph}}{N} - \frac{\hat{N}_{\rm bit}}{N}\right) \in \mathcal{A}(\delta)\right] \le \max_{\boldsymbol{p}\in\mathcal{A}(\delta)} \max_{\boldsymbol{q}\in\mathcal{R}'} f_c(N,3)\exp[-ND(\boldsymbol{p}\|\boldsymbol{q})]. \tag{213}$$

In the same way as the previous discussion, the value δ should be optimized so that the right-hand side above is smaller than or equal to ϵ. Let δ_c be such a value. Then, the function U_c to satisfy Eq. (197) in this case is given by

$$U_c(\hat{N}_{\rm bit}) = \frac{5}{4}\hat{N}_{\rm bit} + \delta_c. \tag{214}$$

We compare the key rates obtained through the different concentration inequalities that give the functions Eqs. (201), (205), (211), and (214), respectively. We set $\epsilon = 2^{-102}$, $s = 102$, and $s' = 50$ so that the secrecy and correctness parameters are 2^{-50}, which are conventional values. We also assume that the number $\hat{N}_{\rm con}$ of conclusive events is equal to its expectation value $N/2$ and that the length $\hat{N}_{\rm sift}$ of the sifted key is equal to its expectation value $N/4$. The observed bit error rate is one percent, i.e., $\hat{N}_{\rm bit} = (N_{\rm con} - \hat{N}_{\rm sift})/100 = N/400$. Under these assumptions, we computed the key rate $\hat{G}$ with respect to the total number N of rounds, where $\hat{G}$ is given from Eqs. (192), (198), and (199) by:

$$\hat{G} = \frac{\hat{N}_{\rm sift}[1 - h(\hat{N}_{\rm bit}/\hat{N}_{\rm sift}) - h(U(\hat{N}_{\rm bit})/\hat{N}_{\rm sift})] - s - s'}{N}. \tag{215}$$

In the above, the function U is either $U_{\rm azuma}$, $U_{\rm kato}$, U_q, or U_c. Figure 4 shows the comparison of the key rates against N with the above four different concentration inequalities applied. The figure shows that our concentration inequalities derived through Theorem 1 and 3 are better than those obtained through Azuma's and Kato's inequalities. Therefore, also in this illustrated example, the Sanov-type quick convergence predominates over the polynomial factors in Eqs. (211) and (214).

4 Discussion

In this paper, we developed concentration inequalities for a permutation-invariant quantum state possibly with an additional symmetry and a sequence of outcomes of independent and identical quantum measurements performed on an adversarial quantum state. For the former case, we used the technique from the representation theory of the permutation group, which is well-known in the community of quantum information theory. With this rather conventional technique, we can upper-bound the permutation-invariant state by i.i.d. quantum states. Compared to the previous de-Finetti-type statement [15, 20, 21], our improvement for this case comes from the fact that one can adaptively choose these i.i.d. states depending on the irreducible subspace of the permutation group.

For the latter case, we found an upper bound on the probability of obtaining each sequence of quantum measurement outcomes with a specific type and then obtained a concentration inequality for the probability of obtaining sequences with the atypical types. For this, we developed, to the best of the author's knowledge, a completely new technique that uses a geometric structure of the set of probability mass functions. Both bounds we obtained are Sanov-type and are thus expected to be tighter than the conventional bound obtained via Azuma's inequality. In specific models we discussed in Sec. 3, our bounds even outperform the recently developed refinement of Azuma's inequality [14].

There remain a few questions on our results. One of the questions is the range of applicability of our results compared to the conventional concentration inequalities such as Azuma's inequality. As mentioned, our results require either permutation symmetry in quantum system or independent and identical measurements while Azuma's inequality does not. There may be a few QKD protocols in which neither condition appears to hold such as the differential-phase-shift (DPS) protocol. However, the state-of-the-art finite-size security proof of it decomposes the pulse train into blocks with three pulses to ease the analysis [41], which may then make it possible to apply our results by actively imposing a permutation symmetry to the classical data. Thus, detailed classification of the cases in which one can or cannot apply our results is left to a future work.

Another question is when our concentration inequalities outperform Kato's inequality [14]. From Fig. 3, it seems that our concentration inequalities can achieve smaller ϵ more quickly than Azuma's or Kato's inequality, and thus the use of our inequalities may be better in the case we need to ensure a small failure probability. However, our concentration inequalities have a dependency either on the dimension of the quantum system (Theorem 1) or the number of the measurement outcomes (Theorem 3) while Azuma's and Kato's inequalities do not. Thus, which inequality performs better in a given setup is an open problem.

Acknowledgments

Note that in the latest preprint, Nahar *et al.* considered another direction of refinement on the de Finetti theorem for a jointly permutation-symmetric extension of a fixed i.i.d. reference state with and without local symmetry restrictions. Their result reproduces the original post-selection technique [20] as the special case when a reference system is trivial and there is no symmetry restriction. As a result, a similar bound as Eqs. (55) in this paper can be obtained through (independently studied) their result by setting the reference system trivial and taking a symmetry restriction into account. This work was supported by the Ministry of Internal Affairs and Communications (MIC), R&D of ICT Priority Technology Project (grant number JPMI00316); Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), "Promoting the application of advanced quantum technology platforms to social issues" (Funding agency: QST); JSPS Overseas Research Fellowships; FoPM, WINGS Program, the University of Tokyo; JSPS Grant-in-Aid for Early-Career Scientists No. JP22K13977.

References

[1] Wassily Hoeffding. "Probability Inequalities for Sums of Bounded Random Variables". Journal of the American Statistical Association **58**, 13–30 (1963).

[2] Ivan N. Sanov. "On the probability of large deviations of random magnitudes". Mat. Sb. (N.S.) **84**, 11–44 (1957).

[3] Imre Csiszar. "Sanov Property, Generalized I-Projection and a Conditional Limit Theorem". The Annals of Probability **12**, 768 – 793 (1984).

[4] Yuki Takeuchi and Tomoyuki Morimae. "Verification of Many-Qubit States". Phys. Rev. X **8**, 021060 (2018).

[5] Omar Fawzi, Richard Kueng, Damian Markham, and Aadil Oufkir. "Learning properties of quantum states without the IID assumption". Nature Communications **15**, 9677 (2024).

[6] Peter W. Shor and John Preskill. "Simple Proof of Security of the BB84 Quantum Key Distribution Protocol". Phys. Rev. Lett. **85**, 441–444 (2000).

[7] Hoi-Kwong Lo and H. F. Chau. "Unconditional Security of Quantum Key Distribution over Arbitrarily Long Distances". Science **283**, 2050–2056 (1999).

[8] M Koashi. "Simple security proof of quantum key distribution based on complementarity". New Journal of Physics **11**, 045018 (2009).

[9] Takaya Matsuura. "Digital quantum information processing with continuous-variable systems". Springer Nature. (2023).

[10] Kiyoshi Tamaki, Masato Koashi, and Nobuyuki Imoto. "Unconditionally secure key distribution based on two nonorthogonal states". Phys. Rev. Lett. **90**, 167904 (2003).

[11] J.-C. Boileau, K. Tamaki, J. Batuwantudawe, R. Laflamme, and J. M. Renes. "Unconditional Security of a Three State Quantum Key Distribution Protocol". Phys. Rev. Lett. **94**, 040503 (2005).

[12] Maxim Raginsky and Igal Sason. "Concentration of Measure Inequalities in Information Theory, Communications, and Coding". Foundations and Trends® in Communications and Information Theory **10**, 1–246 (2013).

[13] Colin McDiarmid. "Concentration". Pages 195–248. Springer Berlin Heidelberg. Berlin, Heidelberg (1998).

[14] Go Kato. "Concentration inequality using unconfirmed knowledge" (2020). arXiv:2002.04357.

[15] RENATO RENNER. "SECURITY OF QUANTUM KEY DISTRIBUTION". International Journal of Quantum Information **06**, 1–127 (2008).

[16] Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack. "Unknown quantum states: The quantum de Finetti representation". Journal of Mathematical Physics **43**, 4537–4559 (2002).

[17] Robert König and Renato Renner. "A de Finetti representation for finite symmetric quantum states". Journal of Mathematical Physics **46**, 122108 (2005).

[18] Matthias Christandl, Robert König, Graeme Mitchison, and Renato Renner. "One-and-a-Half Quantum de Finetti Theorems". Communications in Mathematical Physics **273**, 473–498 (2007).

[19] Renato Renner. "Symmetry of large physical systems implies independence of subsystems". Nature Physics **3**, 645–649 (2007).

[20] Matthias Christandl, Robert König, and Renato Renner. "Postselection Technique for Quantum Channels with Applications to Quantum Cryptography". Phys. Rev. Lett. **102**, 020504 (2009).

[21] Omar Fawzi and Renato Renner. "Quantum Conditional Mutual Information and Approximate Markov Chains". Communications in Mathematical Physics **340**, 575–611 (2015).

[22] Frédéric Dupuis, Omar Fawzi, and Renato Renner. "Entropy Accumulation". Communications in Mathematical Physics **379**, 867–913 (2020).

[23] Tony Metger, Omar Fawzi, David Sutter, and Renato Renner. "Generalised entropy accumulation". In 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS). Pages 844–850. (2022).

[24] Tony Metger and Renato Renner. "Security of quantum key distribution from generalised entropy accumulation". Nature Communications **14**, 5272 (2023).

[25] Anthony Leverrier and Nicolas J. Cerf. "Quantum de finetti theorem in phase-space representation". Phys. Rev. A **80**, 010102 (2009).

[26] Anthony Leverrier. "$SU(p, q)$ coherent states and a Gaussian de Finetti theorem". Journal of Mathematical Physics **59**, 042202 (2018).

[27] David Gross, Sepehr Nezami, and Michael Walter. "Schur–Weyl Duality for the Clifford Group with Applications: Property Testing, a Robust Hudson Theorem, and de Finetti Representations". Communications in Mathematical Physics **385**, 1325–1393 (2021).

[28] Kazuoki Azuma. "Weighted sums of certain dependent random variables". Tohoku Mathematical Journal **19**, 357 – 367 (1967).

[29] William Fulton and Joe Harris. "Representation Theory: A First Course". Graduate texts in mathematics. Springer. (1991).

[30] William Fulton. "Young Tableaux: With Applications to Representation Theory and Geometry". London Mathematical Society Student Texts. Cambridge University Press. (1996).

[31] Masahito Hayashi. "Group Representation for Quantum Theory". Springer Cham. (2017).

[32] Thomas M. Cover and Joy A. Thomas. "Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)". Wiley-Interscience. USA (2006).

[33] Joseph L. Doob. "Stochastic Processes". John Wiley & Sons. USA (1953). url: https://www.wiley.com/en-us/Stochastic+Processes-p-9780471523697.

[34] Takaya Matsuura, Shinichiro Yamano, Yui Kuramochi, Toshihiko Sasaki, and Masato

Koashi. "Refined finite-size analysis of binary-modulation continuous-variable quantum key distribution". Quantum **7**, 1095 (2023).

[35] Guillermo Currás-Lorenzo, Álvaro Navarrete, Koji Azuma, Go Kato, Marcos Curty, and Mohsen Razavi. "Tight finite-key security for twin-field quantum key distribution". npj Quantum Information **7**, 22 (2021).

[36] Charles H. Bennett and Gilles Brassard. "Quantum cryptography: Public key distribution and coin tossing". Theoretical Computer Science **560**, 7–11 (2014).

[37] Marco Tomamichel, Charles Ci Wen Lim, Nicolas Gisin, and Renato Renner. "Tight finite-key analysis for quantum cryptography". Nature Communications **3**, 634 (2012).

[38] Masahito Hayashi and Toyohiro Tsurumaru. "Concise and tight security analysis of the Bennett-Brassard 1984 protocol with finite key lengths". New Journal of Physics **14**, 093014 (2012).

[39] Charles Ci Wen Lim, Marcos Curty, Nino Walenta, Feihu Xu, and Hugo Zbinden. "Concise security bounds for practical decoy-state quantum key distribution". Phys. Rev. A **89**, 022307 (2014).

[40] Stephen M. Barnett Simon J. D. Phoenix and Anthony Chefles. "Three-state quantum cryptography". Journal of Modern Optics **47**, 507–516 (2000).

[41] Akihiro Mizutani, Yuki Takeuchi, and Kiyoshi Tamaki. "Finite-key security analysis of differential-phase-shift quantum key distribution". Phys. Rev. Res. **5**, 023132 (2023).