Practical evaluation of security for quantum key distribution

Masahito Hayashi[*]

EARTO-SORST Quantum Computation and Information Project, JST, 5-28-3, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan, and Superrobust Computation Project, Information Science and Technology Strategic Core (21st Century COE by MEXT), Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

(Received 14 February 2006; revised manuscript received 10 April 2006; published 7 August 2006)

Many papers have proven the security of quantum key distribution (QKD) systems in the asymptotic framework. The degree of the security has not been discussed in the finite coding-length framework, sufficiently. However, to guarantee any implemented QKD system required, it is needed to evaluate a protocol with a finite coding length. For this purpose, we derive a tight upper bound of the eavesdropper's information. This bound is better than existing bounds. We also obtain the exponential rate of the eavesdropper's information. Further, we approximate our bound by using the normal distribution.

DOI: 10.1103/PhysRevA.74.022307 PACS number(s): 03.67.Dd, 03.67.Hk

I. INTRODUCTION

The quantum key distribution (QKD) was proposed by Bennett and Brassard in 1984 [1] as a protocol (BB84 protocol) sharing secret keys by using a quantum communication channel. Their original protocol assumes a noiseless quantum channel, but any quantum channel has noise in the realistic case. Hence, the security of the BB84 protocol in this realistic case had been an open problem for a long time and was proved by Mayers [2]. He showed that the protocol becomes secure when the protocol is constructed by combining classical error correction and randomly choosing a code for privacy amplification. In his proof, the secure generation key rate is $1-h(2p)-h(p)$ where p is the qubit error rate and $h(p)$ is the binary entropy $-p\log p-(1-p)\log(1-p)$ and the base of the logarithm is 2. He also gave a bound of Eve's information for a finite-length code. His discussion was extended to a more realistic framework by Inamori, Lütkenhaus, and Mayers [3].

After Mayers' proof, Shor and Preskill [4] proved the security based on the method of Calderbank-Shor-Steane (CSS) codes [5,6]. Then, they proved the existence of a code achieving the secure generation key rate $1-2h(2p)$ and pointed out the possibility of the secure generation key rate $1-2h(p)$. After their discussion, treating the reliability of CSS codes, Hamada [7] showed the existence of a code attaining the secure generation key rate $1-2h(p)$. He also derived a bound of Eve's information for a finite-length code, which yields the asymptotic secure generation key rate $1-2h(p)$. However, he did not discuss the complexity of the encoding and decoding [4,7], while the complexity of privacy amplification is not so large in Mayers' proof [2].

Following these researches, Christandl, Renner, and Ekert [8], Renner, Gisin, and Kraus [9], and Koashi [10] showed that the asymptotic secure generation key rate $1-2h(p)$ is attained when the protocol is constructed by combining classical error correction and randomly choosing. However, they did not give the bound of Eve's information of the finite-length code, explicitly. S. Watanabe, R. Matsumoto, and Uyematsu [11] considered Eve's information for a finite-length code based on random privacy amplification, which yields the asymptotic secure generation key rate $1-2h(p)$. Renner [12] obtained a similar tact in a more general framework. While Watanabe *et al.*'s bound goes to zero exponentially, his bound does only polynomially.

On the other hand, Stucki *et al.* [13] demonstrated a quantum key distribution over 67 km between Geneva and Lausanne. Kimura *et al.* [14] succeeded with a 150-km QKD transmission with an error rate of 8%–9%. Also Gobby *et al.* [15] produced a 122-km QKD transmission with an error rate of 8.9%. Tanaka *et al.* [16] demonstrated a continuous quantum key distribution over 16.3-km commercial use fiber over 14 days, and Yuan and Shields [17] did it over 20.3-km installed telecom fiber in 19 h. In these experiments, they succeeded in realizing a real system that could become truly secure if it had a coding system with infinite coding length. Hence, there is no implemented system whose security is guaranteed. Thus, it is required to realize the error-correcting code and privacy amplification for guaranteeing the security of the implemented QKD system.

However, the required sizes of the error-correcting code and random privacy amplification are not clarified for a given quantum bit error rate—e.g., 8%. Therefore, many QKD experimental researchers want to find a tighter upper bound of Eve's information for given sizes of the classical error-correcting code and random privacy amplification.

In this paper, we derive an upper bound of Eve's information satisfying the following conditions: (i) The upper bound depends only on the size of random privacy amplification. (ii) By using this bound, the key generation rate $1-2h(p)$ can be attained. In fact, Mayers' discussion [2] gives the upper bound in the finite-length case, but his discussion yields the rate $1-h(2p)-h(p)$ not the rate $1-2h(p)$. The discussion by Watanabe *et al.* [11] yields the rate $1-2h(p)$, but the bound depends on the error correction. Koashi's discussion [10] satisfies conditions (i) and (ii), but his discussion does not clearly give the bound in the finite-length case. Further, the

[*]Electronic address: masahito@qci.jst.go.jp

protocol in his paper [10] and his older paper [18] is slightly different from the simple combination of the classical error correction and random privacy amplification. Our upper bound is also better than that by Watanabe *et al.* [11].

Moreover, it is shown that our evaluation cannot be further improved in the sense of the exponential rate when the classical error-correcting code satisfies a specific condition. (For example, several degenerate codes do not satisfy this condition.) In this case, the exponential rate of our upper bound of Eve's information can be attained by a collective attack, which is realized by an individual operation to the channel and the collective operation to Eve's local memory, while our bound is valid even for the coherent attack, which includes any of Eve's attacks allowed by the physical principle. That is, any coherent attack cannot improve the best collective attack in the sense of the exponential rate of Eve's information. Indeed, Renner *et al.* [9] proved that it is sufficient to show the security for collective attacks for the treatment of the asymptotic key generation rate since any channel can be approximated by a separable channel by using random permutation. This result can be regarded as the extension of the result of Renner *et al.* to the exponential framework. Also, this implies that our evaluation gives the optimal (minimum) exponential rate of Eve's information.

There is another type of asymptotic treatment other than the exponential treatment. In statistics, when the variable obeys the independent and identical distribution, its distribution can be approximated by the normal distribution. We also succeeded in approximating our upper bound by using the normal distribution. In this approximation, we treat the asymptotic behavior when the size of the random privacy amplification is given as the form $2^{nh[\hat{p}_\times+\epsilon(\hat{p}_\times)]}$ for the estimate $\hat{p}_\times$ of the phase error rate while in the large-deviation case (the exponential-rate case) we treat it when the size is given as the form $2^{nh[\hat{p}_\times+\tilde{\epsilon}(\hat{p}_\times)/\sqrt{n}]}$, where ϵ and $\tilde{\epsilon}$ are functions of $\hat{p}_\times$.

Here, we should remark that our results cannot be obtained by a combination of existing results. The main technical point is the relation between Eve's information and the phase error probability, which is given in lemma 2. Owing to this lemma, Eve's information can be bounded without any discussion of the classical error-correcting code for bit errors. Further, in association with the error correction of the phase error, we obtain an upper bound of the average error probability of a modified random coding when minimum Hamming distance decoding is applied (lemma 1). Combining these techniques, we obtain the upper bounds (theorems 1 and 2) through a long careful derivation.

In the following, the organization of this paper is explained. First, we briefly explain classical error-correcting code and describe our protocol using this knowledge in Sec. II. In Sec. III, we give an upper bound of Eve's information per one code and that of Eve's information per one bit. The random privacy amplification corresponds to the random coding concerning the phase error. Hence, we treat the average error of random coding in Sec. IV. The generalized Pauli channel is known as an important class of noisy channels. In the quantum key distribution, the noisy channel does not necessarily belong to this class. However, if we use linear codes, we can treat any noisy channel as a generalized Pauli channel. We summarize the notations and properties of the generalized Pauli channel in Sec. V. In Sec. VI, we prove the main theorem by assuming an upper bound of Eve's information when Eve's attack is known. In Sec. VII, we derive a relation between the phase error and Eve's information. In Sec. VIII, the bound used in Sec. VI is proved by using the properties of a generalized Pauli channel, the bound of the average error, and the relation obtained in Sec. VII.

Further, we give the asymptotic behavior in the two asymptotic frameworks in Sec. III. Asymptotic formulas for a large deviation and limiting distribution are proved in Appendixes A and B, respectively. Based on this evaluation, we compare our large-deviation bound with the bound by Watanabe, Matsumoto, and Uyematsu [11]. Further, in Sec. IX, we prove that the exponential rate of our bound of Eve's information can be attained by a collective attack under a specific condition.

II. PROTOCOL

In this section, we describe our protocol. Since our protocol employs the method of the classical error-correcting code, we first explain the classical error-correcting code in preparation of a description of our protocol.

A. Classical error-correcting code

When the noise in a binary signal $\mathbf{F}_2=\{0,1\}$ is symmetric, the binary channel is described by a probability distribution $\{p,1-p\}$. In this case, when we send a binary string (in $\mathbf{F}_2^n$), the noise can be described by a binary string N and is characterized by the distribution P on $\mathbf{F}_2^n$. Then, when the input signal is described by the random variable X, the output signal is described by the random variable $X+N$. The error-correcting code is a method removing the difference N. In an error-correcting code with n bits, we prepare an m-dimensional linear subspace C of $\mathbf{F}_2^n$, and the sender (Alice) and the receiver (Bob) agree that only elements of C is sent before the communication. This linear subspace is called a code or a $[n,m]$ code. In this case, an encoding is given by a linear map $G(C)$ from $\mathbf{F}_2^m$ to C. Of course, the map $G(C)$ is given as an $m\times n$ matrix with 0,1 entries. Hence, when Bob receives an element out of C, he can find that there exists a noise and choose the most probable element among C based on the obtained binary string. Here, we can correct only one element among each equivalent class $[X]\in\mathbf{F}_2^n/C$. More precisely, we choose the most likely noise $\Gamma([X])$ among each equivalent class $[X]$. This element is often called the representative, and the set of representatives is denoted by Γ. More generally, the decoding process is described by a map $D:\mathbf{F}_2^n\rightarrow\mathbf{F}_2^m$.

Hence, when Bob receives $X+N$, he decodes it to $X+N-\Gamma([X+N])=X+N-\Gamma([N])$. Thus, the decoding error is described by the behavior of the random variable $N-\Gamma([N])$ and does not depend on the input signal X. When the noise belongs to the set Γ, we can properly correct the error. The error probability is equal to $1-P(\Gamma)$.

Suppose that there exists an eavesdropper (Eve) obtaining some information concerning the original signal X. In this case, we prepare a linear subspace C' of C and Alice sends the information as an element of C/C'. That is, when he sends a piece of information corresponding to $[X] \in C/C'$, he chooses one element among $[X]$ with equal probability and sends it. This operation is called privacy amplification.

B. Our protocol

Using this method, we can reduce Eve's information. However, it is not easy to evaluate how much information Eve has in this case. The purpose of this paper is evaluating Eve's information. In this case, the probability that Bob recovers the original information correctly is equal to $P(\Gamma + C')$, where $\Gamma + C' := \{\Gamma([X]) + X' \mid X \in \mathbf{F}_2^n, X' \in C'\}$. In addition, when we choose each linear subspace C' of C with equal probability and we regard C as $\mathbf{F}_2^m$ and C/C' as $\mathbf{F}_2^{m-\tilde{m}}$, the function from $\mathbf{F}_2^m$ to $\mathbf{F}_2^{m-\tilde{m}}$ is called the universal hashing function. For example, this function is can be constructed as an $(m-\tilde{m}) \times m$ matrix by choosing elements with a uniform distribution.

Using this preparation, we briefly describe our protocol for the quantum key distribution that can be realized by small complexity. After this description, we present it precisely. In our protocol, after quantum communication, Alice and Bob check their basis by using a public channel, announce a part of the obtained bits, and estimate the bit error rate p_+ and phase error rate $p_\times$. Here, we denote Alice's remaining bit string with the + basis and the × basis by X_+ and $X_\times$, respectively. Similarly, we denote Bob's remaining bit string by $\tilde{X}_+$ and $\tilde{X}_\times$. These bit strings are called raw keys. Hence, the rates of 1 in the difference $N_+ = X_+ - \tilde{X}_+$ and the difference $N_\times = X_\times - \tilde{X}_\times$ are almost equal to p_+ and $p_\times$, respectively.

Using the following process, Alice and Bob remove their errors and share the bit string with almost no error. Alice generates another bit string X' and sends the bit string $K := X' + X_+$ to Bob. Based on the information K, Bob obtains the information $X'' := K - \tilde{X}_+ = X' + N_+$. Using this method, we can realize a classical channel with input X' and output X''. The error rate of this channel is almost equal to p_+. By applying a classical error correction to this channel, Alice and Bob can share a bit string with almost zero error. In this case, Alice generates an element $X' \in \mathbf{F}_2^m \cong C$ and Bob recovers $X''' = D(X'')$. Then, X''' coincides with X' in a high probability. Finally, Alice and Bob perform the above-mentioned hashing function for their respective keys. That is, Alice generates a $(m-l) \times m$ matrix A with rank $m-l$ randomly and sends this matrix. Then, Alice and Bob obtain their final keys AX' and AX'''.

Therefore, the rate of the final key to the raw key is equal to $R = \frac{m-l}{n}$. Roughly speaking, it is suitable to choose m as an integer a little smaller than $[1 - h(p_+)]n$ and $\tilde{m}$ as an integer a little larger than $h(p_\times)n$. Then, the generation rate R is almost equal to $1 - h(p_+) - h(p_\times)$.

In the following, we describe our protocol more precisely. For this purpose, we need some mathematical notations. The quantum system of each quantum signal is the two-dimensional Hilbert space $\mathcal{H}_2$, which is spanned by $\{|a\rangle\}_{a \in \mathbf{F}_2}$. We need to fix the integers n_+, l_+, m_+, $n_\times$, $l_\times$, and $m_\times$, which describe the size of our code. For a classical error correction, we choose an m_+-dimensional classical code $C_{1,+}$ in $\mathbf{F}_2^{n_+}$ (an m-dimensional linear space $C_{1,+}$ of $\mathbf{F}_2^{n_+}$) and an $m_\times$-dimensional classical code $C_{1,\times}$ in $\mathbf{F}_2^{n_\times}$. We also fix the thresholds $\underline{k}_+$, $\bar{k}_+$, $\underline{k}_\times$, and $\bar{k}_\times$ and the allowable statistical fluctuation δ_k for each count k of error.

(i) The sender, Alice, and the receiver, Bob, repeat steps (ii)-(iv) for each i.

(ii) Alice chooses a random bit $\mathbf{a}_i$ and a random bit $\mathbf{b}_i$.

(iii) Bob chooses a random bit $\mathbf{c}_i$.

(iv) When $\mathbf{b}_i = 0$, Alice sends the quantum state $|\mathbf{a}_i\rangle$, otherwise the state $\frac{1}{\sqrt{2}}(|0\rangle + (-1)^{\mathbf{a}_i}|1\rangle$. In the following, $\{|0\rangle, |1\rangle\}$ is called the + basis and $\{\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)\}$ is called the × basis.

(v) Alice and Bob announce $\mathbf{b}_i$ and $\mathbf{c}_i$ and discard any results for $\mathbf{b}_i \neq \mathbf{c}_i$. They obtain $n_+ + l_+$ bits sequence with $\mathbf{b}_i = \mathbf{c}_i = 0$ and $n_\times + l_\times$ bits sequence with $\mathbf{b}_i = \mathbf{c}_i = 1$.

(vi) Alice randomly chooses l_+ check bits $X_{+,c,1}, \dots, X_{+,c,l_+}$ among $n_+ + l_+$ bits with the + basis and $l_\times$ check bits $X_{\times,c,1}, \dots, X_{\times,c,l_+}$ among $n_\times + l_\times$ bits with the + basis, announces the positions of these bits, and sends their information. They obtain the estimates $\hat{p}_+$ and $\hat{p}_\times$ with the respective basis. That is, they count the number of error bits $k_+ = |\{i | X_{+,c,i} \neq \tilde{X}_{+,c,i}\}|$ and $k_\times = |\{i | X_{\times,c,i} \neq \tilde{X}_{\times,c,i}\}|$, where $\tilde{X}_{+,c,i}$ and $\tilde{X}_{\times,c,i}$ are Bob's check bits. However, when k_+ is greater than the threshold $\bar{k}_+$, they discard their remaining bits with the × basis. When $k_\times$ is greater than the threshold $\bar{k}_\times$, they discard their remaining bits with the + basis. Further, when k_+ is less than the other threshold $\underline{k}_+$, they replace k_+ by $\underline{k}_+$. When $k_\times$ is less than the other threshold $\underline{k}_\times$, they replace $k_\times$ by $\underline{k}_\times$.

In the following, we treat only the bit string of the + basis. We denote Alice's (Bob's) remaining n_+-bit strings with the + basis by X_+ ($\tilde{X}_+$). After this process, they apply the same procedure to the remaining bit strings with the × basis.

(vii) Alice generates $Z_+ \in \mathbf{F}_2^{m_+}$ randomly and sends Bob $G(C_{1,+})Z_+ + X_+$.

(viii) Bob obtains the signal $G(C_{1,+})Z_+ + X_+ - \tilde{X}_+ \in \mathbf{F}_2^{n_+}$. Performing the decoding of the code $C_{1,+} \approx \mathbf{F}_2^{m_+}$, he obtains $\tilde{Z}_+ \in \mathbf{F}_2^{m_+}$.

(ix) Alice chooses $\tilde{m} := n_\times h(k_\times / l_\times + \delta_{k_\times})$ dimensional subcode $C_{2,+}(Y_+, k_\times) \subset \mathbf{F}_2^{m_+}$ based on random variables Y_+ such that any element $x \neq 0 \in \mathbf{F}_2^{m_+}$ belongs to $C_{2,+}(Y_+, k_\times)$ with the probability $\frac{2^{n_+ h(k_\times / l_\times + \delta_{k_\times})} - 1}{2^{m_+} - 1}$.

(x) Alice obtains the secret information $\bar{Z}_+ := [Z_+]_{C_{2,+}(Y_+, k_\times)} \in \mathbf{F}_2^{m_+} / C_{2,+}(Y_+, k_\times)$.

(xi) Bob obtains the secret information $\bar{Z}_{+,B} := [\tilde{Z}_+]_{C_{2,+}(Y_+, k_\times)} \in \mathbf{F}_2^{m_+} / C_{2,+}(Y_+, k_\times)$.

For example, an s-dimensional code $C_2(Y, s)$ in $\mathbf{F}_2^{m_+}$ is constructed based on k random variables $Y := (X_1, \dots, X_s)$ in $\mathbf{F}_2^{m_+}$ as $C_2(Y, s) := \langle X_1, \cdots, X_s \rangle$, where Y obeys the uniform distribution on the set $\{Y | X_1, \dots, X_s$ are linearly independent.$\}$.

C. Extension of our protocol

Indeed, in the realistic case, the bottleneck is often the estimation error of the error rate. Hence, in order to decrease the error of the estimation of the phase error rate $p_\times$, we propose the following the modified protocol for any integer a. In the modified protocol, we repace steps (v) and (vi) by the following and add step (xii).

(v) Alice and Bob announce $\mathbf{b}_i$ and $\mathbf{c}_i$ and discard any results for $\mathbf{b}_i \neq \mathbf{c}_i$. They obtain an $(an_+ + l_+)$-bit sequence with $\mathbf{b}_i = \mathbf{c}_i = 0$ and an $(an_\times + l_\times)$-bit sequence with $\mathbf{b}_i = \mathbf{c}_i = 1$.

(vi) Alice randomly chooses n_+ bits among remaining an_+ bits with + basis and obtain n_+ bit string X_+. She also sends the her positions to Bob. Bob obtains the n_+-bit string $\tilde{X}_+$. They do the same procedure for the × basis.

(xii) They repeat steps (vii)-(xi) a times.

In the above protocol, the estimation of the phase error $p_\times$ has the same accuracy as that of the first protocol with $al_\times$ check bits of the × basis.

III. SECURITY

In this section, we evaluate the security of our protocol. In the following, for simplicity, we abbreviate $l_\times$ and n_+ by l and n, respectively.

A. Finite-length case

The security of this protocol is evaluated by the mutual information $I(\bar{Z}_+, Z_E)$ between Alice's final key $\bar{Z}_+$ and eavesdropper (Eve)'s information Z_E. It is mathematically defined by

$$I(\bar{Z}_+, Z_E) := -\sum_{Z_E} P(Z_E)\log P(Z_E) + \sum_{\bar{Z}_+} P(\bar{Z}_+)\sum_{Z_E} P(Z_E|\bar{Z}_+)\log P(Z_E|\bar{Z}_+).$$

In order to evaluate this value, we have to treat the hypergeometric distribution

$$P_{hg}(k|n,l,j) := \frac{\binom{l}{k}\binom{n}{j-k}}{\binom{n+l}{j}}.$$

This is because the random sampling obeys the hypergeometric distribution. It is known that its average is $\frac{lj}{n+l}$ and its variance is $\frac{j\ln(n+l-j)}{(n+l)^2(n+l-1)}$. In this paper, we focus on the average of Eve's information $\mathrm{E}_{\mathrm{pos}_\times, k_\times, Y_+ | \mathrm{pos}_+, k_+, Y_\times}[I(\bar{Z}_+, Z_E)]$ for each n, l, where pos_+ and $\mathrm{pos}_\times$ are the random variables indicating the positions of the check bit of × basis and + basis, respectively. Some papers [2,4,10,11,18] guarantee the security by proving that for any $\epsilon_1 > 0$ and $\epsilon_2 > 0$ there exist integers n and l such that

$$P(I(\bar{Z}_+, Z_E) \geq \epsilon_2) \leq \epsilon_1. \tag{1}$$

Indeed, when $\mathrm{E}_{\mathrm{pos}_\times, k_\times, Y_+ | \mathrm{pos}_+, k_+, Y_\times}[I(\bar{Z}_+, Z_E)] \leq \epsilon_1\epsilon_2$, Markov's inequality guarantees the inequality (1). Hence, we can recover the probabilistic behavior (1) of Eve's information from an evaluation of the average of Eve's information. Therefore, in this paper, we concentrate the evaluation of the average of Eve's information.

Theorem 1. When R is the rate of the code C_1 and the threshold $\bar{k}$ is less than $\frac{n}{2}$, we have

$$\mathrm{E}_{\mathrm{pos}_\times, k_\times, Y_+ | \mathrm{pos}_+, k_+, Y_\times}[I(\bar{Z}_+, Z_E)] \leq P(\delta, n, l, \underline{k}, \bar{k}), \tag{2}$$

where

$$\begin{aligned} P(\delta,n,l,\underline{k},\bar{k}) := {} & \max_j \bar{h}\left(\sum_{k=0}^{\underline{k}} P_{hg}(k|n,l,j) f(j-\underline{k},\underline{k}|n,l,\delta_k) + \sum_{k=\underline{k}+1}^{\bar{k}} P_{hg}(k|n,l,j) f(j-k,k|n,l,\delta_k)\right) \\ & + \max_j \left[\sum_{k=0}^{\underline{k}} P_{hg}(k|n,l,j) f(j-\underline{k},\underline{k}|n,l,\delta_k) n[R - h(\underline{k}/l + \delta_k)] + \sum_{k=\underline{k}+1}^{\bar{k}} P_{hg}(k|n,l,j) f(j-k,k|n,l,\delta_k) n[R - h(k/l + \delta_k)]\right] \end{aligned}$$

and

$$\bar{h}(x) := \begin{cases} h(x) & x < 1/2, \\ 1 & x \geq 1/2, \end{cases} \qquad f(k',k|n,l,\delta) := \begin{cases} \min\{2^{n[h(k'/n) - h(k/l+\delta)]}, 1\} & \text{if } k' < n/2, \\ 1 & \text{if } k' \geq n/2. \end{cases}$$

Further, Eve's information per one bit is evaluated as follows.

Theorem 2. When R is the rate of the code C_1, we have

$$\mathrm{E}_{\mathrm{pos}_\times, k_\times, Y_+ | \mathrm{pos}_+, k_+, Y_\times}\left[\frac{I(\bar{Z}_+, Z_E)}{n[R - h(k_\times/l_\times + \delta_{k_\times})]}\right] \leq \tilde{P}(\delta, n, l, \underline{k}, \bar{k}),$$

where

$$\widetilde{P}(\delta,n,l,\underline{k},\bar{k}) := \max_j \frac{1}{n(R-h(\bar{k}/l+\delta_{\bar{k}}))}\bar{h}\left(\sum_{k=0}^{\underline{k}} P_{hg}(k|n,l,j)f(j-\underline{k},\underline{k}|n,l,\delta_k) + \sum_{k=\underline{k}+1}^{\bar{k}} P_{hg}(k|n,l,j)f(j-k,k|n,l,\delta_k)\right)$$
$$+\max_j\left[\sum_{k=0}^{\underline{k}} P_{hg}(k|n,l,j)f(j-\underline{k},\underline{k}|n,l,\delta_k) + \sum_{k=\underline{k}+1}^{\bar{k}} P_{hg}(k|n,l,j)f(j-k,k|n,l,\delta_k)\right].$$

The proofs of these theorems are divided into two parts: (i) The security of the known channel (Sec. VII) and (ii) the security of the unknown channel, which is given by estimating the channel and employing part (i) (Sec. VI). For a treatment of the quantum channel, we prepare the notations of the generalized Pauli channel in Sec. V. For a discussion of part (i), we derive a bound of the average error concerning the classical error-correcting code in Sec. IV and a bound of Eve's information using the phase error in Sec. VII.

B. Approximation using the normal distribution

In the following, we calculate the above value approximately. For this purpose, we choose two probabilities $\underline{p}<\bar{p}<\frac{1}{2}$ and a continuous function $p\mapsto\tilde{\epsilon}(p)$. When $\bar{k}=\bar{p}l$, $\underline{k}=\underline{p}l$, $\frac{n}{n+l}=r$, and $\delta_k=\frac{\tilde{\epsilon}(p)}{\sqrt{n+l}}$, as shown in Appendix A, we obtain

$$\lim_{n\to\infty}\widetilde{P}(\delta,n,l,\underline{k},\bar{k}) = \max_{p\in[\underline{p},\bar{p}]}\Phi\left(-\frac{\sqrt{r(1-r)}}{\sqrt{p(1-p)}}\tilde{\epsilon}(p)\right), \quad (3)$$

where the distribution function Φ is the standard Gaussian distribution:

$$\Phi(x) := \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx.$$

Hence, in order to keep the security level ε per one bit, it is suitable to choose δ_k to be

$$-\frac{1}{\sqrt{n+l}}\frac{\sqrt{\frac{k}{l}\left(1-\frac{k}{l}\right)}}{\sqrt{\frac{n}{n+l}\frac{l}{n+l}}}\Phi^{-1}(\varepsilon) = -\sqrt{\frac{n+l}{nl}}\sqrt{\frac{k}{l}\left(1-\frac{k}{l}\right)}\Phi^{-1}(\varepsilon)$$

when $\widetilde{P}(\delta,n,l,\underline{k},\bar{k})$ can be approximated by the right-hand side (RHS) of Eq. (3). That is, our upper bound is almost determined by $[\sqrt{nl/(n+l)}\sqrt{(k/l)(1-k/l)}]\delta_k$.

Now, we consider the case when we use a low-density parity-check (LDPC) code as the code C_1 [19]. In this case, the case of $R=0.5$ and $n=10\,000$ is one realistic case. As a realistic case, let us consider the case $l=1000$, $\bar{p}=0.075$, $\delta_{k_\times}=0.01$. Then, we have

$$-\frac{\sqrt{\frac{nl}{n+l}}}{\sqrt{\frac{k}{l}\left(1-\frac{k}{l}\right)}}\delta_k = -1.14.$$

The security level

$$\Phi\left(-\frac{\sqrt{\frac{nl}{n+l}}}{\sqrt{\frac{k}{l}\left(1-\frac{k}{l}\right)}}\delta_k\right) = 0.126$$

is not sufficient.

However, it is not easy to increase the size n. Hence, we adopt the modified protocol. In this case, we replace only l by the following values (in the case of $l=20\,000$, the security level is almost 0.001):

l	1000	10 000	20 000	30 000	40 000	50 000
$-\frac{\sqrt{\frac{nl}{n+l}}}{\sqrt{\frac{k}{l}\left(1-\frac{k}{l}\right)}}\delta_k$	−1.14	− 2.68	− 3.10	− 3.29	− 4.00	− 3.47
$\Phi\left(-\frac{\sqrt{\frac{nl}{n+l}}}{\sqrt{\frac{k}{l}\left(1-\frac{k}{l}\right)}}\delta_k\right)$	0.126	0.00363	0.000968	0.000505	0.000342	0.000264

C. Large deviation

Next, we focus on the large-deviation-type evaluation. Choose a function $p \in [\underline{p},\bar{p}] \mapsto \epsilon(p)$ and define

$$E(\epsilon,r,\underline{p},\bar{p}) := \min_{p\in[\underline{p},\bar{p}],\epsilon'\geqslant 0} (h\{p+r[\epsilon(p)-\epsilon']\}-(1-r)h(p) \\ -2rh[p+\epsilon(p)-\epsilon']+rh[p+\epsilon(p)]).$$

When $\bar{k}=\bar{p}l$, $r=\frac{n}{n+l}$, and $\delta_k=\epsilon\left(\frac{k}{l}\right)$, as shown in Appendix B, we obtain

$$E(\epsilon,r,\underline{p},\bar{p}) = \lim_{n\to\infty}\frac{-r}{n}\log P(\delta,n,l,\underline{k},\bar{k}). \tag{4}$$

Further,

$$P(\delta,n,l,\underline{k},\bar{k}) \leqslant \bar{k}(n+l+1)n[R-h(\underline{p}+\delta_{\underline{p}})]2^{(-n/r)E(\epsilon,r,\underline{p},\bar{p})} \\ +h[\bar{k}(n+l+1)2^{(-n/r)E(\epsilon,r,\underline{p},\bar{p})}]. \tag{5}$$

Hence, given a fixed real number E, it is suitable to choose $\epsilon(p)$ satisfying that

$$E=\min_{\epsilon'\geqslant 0}(h\{p+r[\epsilon(p)-\epsilon']\}-(1-r)h(p) \\ -2rh[p+\epsilon(p)-\epsilon']+rh[p+\epsilon(p)])$$

for any probability $p\in[\underline{p},\bar{p}]$. Further, when $\epsilon(p)$ is sufficiently small, using the relation $d(p\|q):=p\log\frac{p}{q}+(1-p)\log\frac{1-p}{1-q}\cong\frac{(p-q)^2}{p(1-p)\ln 2}$, we have the approximation

$$h\{p+r[\epsilon(p)-\epsilon']\}-(1-r)h(p)-rh\{p+[\epsilon(p)-\epsilon']\} \\ +rh[p+\epsilon(p)]-rh\{p+[\epsilon(p)-\epsilon']\} \\ =(1-r)d\{p\|p+r[\epsilon(p)-\epsilon']\}+rd\{p+\epsilon(p)\|p+r[\epsilon(p) \\ -\epsilon']\}+r(h[p+\epsilon(p)]-h\{p+[\epsilon(p)-\epsilon']\}) \\ \cong(1-r)\frac{r^2[\epsilon(p)-\epsilon']^2}{p(1-p)}+r\frac{(1-r)^2[\epsilon(p)-\epsilon']^2}{p(1-p)}+rh'(p)\epsilon'.$$

In this approximation, when $\epsilon(p)$ is small enough, the minimum is attained at $\epsilon'=0$. Hence,

$$\min_{\epsilon'\geqslant 0}(h\{p+r[\epsilon(p)-\epsilon']\}-(1-r)h(p)-rh[p+\epsilon(p)-\epsilon'] \\ +rh[p+\epsilon(p)]-rh[p+\epsilon(p)-\epsilon']) \\ =h[p+r\epsilon(p)]-(1-r)h(p)-rh[p+\epsilon(p)]. \tag{6}$$

The maximum value of $\epsilon(p)$ satisfying Eq. (6) corresponds to the critical rate in the classical channel-coding theory [20]. Therefore, when the number $\epsilon(p)$ is sufficiently small for each $p\in[\underline{p},\bar{p}]$, we obtain

$$E=h[p+r\epsilon(p)]-(1-r)h(p)-rh[p+\epsilon(p)] \\ \cong\frac{r(1-r)\epsilon(p)^2}{(\log 2)[p+r\epsilon(p)]\{1-[p+r\epsilon(p)]\}},\quad\forall\, p\in[\underline{p},\bar{p}]. \tag{7}$$

Hence, in this case, in order to keep the exponential rate E, we choose $\epsilon(p)$ as

$$\epsilon(p)=\frac{(\ln 2)Er(1-2p)}{2[r(1-r)+(\ln 2)Er^2]} \\ +\frac{\sqrt{(\ln 2)^2E^2r^2+4p(1-p)r(1-r)(\ln 2)E}}{2[r(1-r)+(\ln 2)Er^2]} \\ \cong\frac{\sqrt{p(1-p)}}{\sqrt{r(1-r)}}\sqrt{(\ln 2)E}\quad\text{as}\quad E\to 0.$$

Here, we compare our bound with that by Watanabe, Matsumoto, Uyematsu [11]. Since their protocol is different from our protocol, we compare our protocol with their protocol with the same size of code. This is because the size of the code almost corresponds to the cost of its realization. Then, their case corresponds to our case with $\bar{p}=\underline{p}=p$ and $l=n$. They derived the following upper bound (8) of the security in their protocol when the codes $C_2\subset C_1$ satisfy the following conditions: The codes C_1/C_2 and $C_2^{\perp}/C_1^{\perp}$ have the decoding error probability ε when the channel is the binary symmetric channel with error probability p:

$$\mathrm{E}_{\mathrm{pos}_{\times},k_{\times}|\mathrm{pos}_{+},k_{+}}[I(\bar{Z}_{+},Z_E)] \\ \leqslant h\left(2\left(\frac{n}{2}+1\right)^2\varepsilon+4(n+1)^2e^{-[\epsilon(p)^2/4]n}\right) \\ +4n\left(\frac{n}{2}+1\right)^2\varepsilon+8n(n+1)^2e^{-[\epsilon(p)^2/4]n}. \tag{8}$$

However, even if the error probability ε is zero, our evaluation (5) is better than their evaluation (8). In particular, when $\epsilon(p)$ is sufficiently small, we can use Eq. (6). From Pinsker's inequality $(\ln 2)d(p\|q)\geqslant(p-q)^2$, [20] our exponential rate is evaluated as

$$\frac{\ln 2}{r}\{h[p+r\epsilon(p)]-(1-r)h(p)-rh[p+\epsilon(p)]\} \\ =\frac{\ln 2}{r}\{(1-r)d[p\|p+r\epsilon(p)]+rd[p+\epsilon(p)\|p+r\epsilon(p)]\} \\ \geqslant(1-r)\epsilon(p)^2=\frac{\epsilon(p)^2}{2},$$

which is greater than their rate $\frac{\epsilon(p)^2}{4}$ even in the case of $\epsilon'=0$. Further, our coefficient is smaller than their coefficient in this case as follows:

$$\bar{k}(n+l+1)n[R-h(\underline{p}+\delta_{\underline{p}})] \leqslant pn(n+n+1)nR < 8n(n+1)^2,$$

$$\bar{k}(n+l+1) = pn(n+n+1) < 4(n+1)^2$$

because $p \leqslant 1/2$.

Hence, in order to obtain a tighter bound, it is better to use our formula (2).

IV. ERROR-CORRECTING CODE

A. Type method

In this section, we treat the classical error-correcting code. For this purpose, we review the type method for binary strings. For any element $x \in \mathbf{F}_2^n$, we define $|x| := |\{i | x_i = 1\}|$ and $T_n^k := \{x \in \mathbf{F}_2^n | |x| = k\}$. Further, the number of elements is evaluated by

$$\frac{1}{n+1} 2^{nh(k/n)} \leqslant |T_n^k| = \binom{n}{k} \leqslant |\cup_{k' \leqslant k} T_n^{k'}| \leqslant 2^{nh(k/n)} \qquad (9)$$

for $k \leqslant n/2$. For any distribution P on $\mathbf{F}_2^n$, we define the distribution $\widetilde{P}$ on $\{0, \ldots, n\}$ and P_k on T_n^k as

$$\widetilde{P}(k) := P(T_n^k), \quad P_k(x) := \begin{cases} \dfrac{P(x)}{\widetilde{P}(k)}, & \text{if } x \in T_n^k, \\ 0, & \text{otherwise}. \end{cases}$$

Hence, we have

$$P(x) = \sum_{k=0}^{n} \widetilde{P}(k) P_k(x).$$

B. Bound for random coding

In this paper, we focus on linear codes, which are defined as linear subspaces of $\mathbf{F}_2^n$. For the preoperation of the following section, we consider the error probability when the noise of the classical communication channel is given as a classical channel W (a stochastic transition matrix) on $\mathbf{F}_2^n$. If a channel W is written by a distribution P_W on $\mathbf{F}_2^n$ as

$$W(y|x) = P_W(y-x),$$

it is called an additive channel. For an additive channel W, we define the following distribution:

$$P_W(k) := P_W\{x||x| = k\}.$$

In order to protect our message from noise, we often restrict our message to be sent in a subset of $\mathbf{F}_2^n$. This subset is called a code. When the noise is given by an additive channel, a linear subspace C of $\mathbf{F}_2^n$ is suitable for our code because of the symmetry of the noise. Hence, in the following, we call a linear subspace C of $\mathbf{F}_2^n$ a code.

Now, for a preoperation of the following section, we consider the error-correcting code using a pair of codes $C_1 \subset C_2$. In order to send any information $[x_2]_1 \in C_2/C_1$, we send $x_1 + x_2$ by choosing $x_1 \in C_1$ with a uniform distribution, where $[x]_i$ denotes the equivalent class divided by C_i. In this case, the decoder is described by the map D from $\mathbf{F}_2^n$ to itself. When the channel is given by W, the average error probability is

$$P_{e,W}(D) = \frac{1}{|C_2/C_1|} \sum_{[x_2]_1 \in C_2/C_1} \frac{1}{|C_1|} \sum_{x_1 \in C_1} \sum_{D(y) \neq [x_2]} W(y|x_2 + x_1).$$

However, we often describe our decoder by the coset representative $\Gamma([x]_2)$ for each $[x]_2 \in \mathbf{F}_2^n / C_2$. That is, when the decoder receives the element y, he decodes it to $D^{\Gamma}(y) := [y - \Gamma([y]_2)]_1$. When the channel is given by a additive channel W, the error probability is

$$P_{e,W}(D^{\Gamma}) = 1 - P_W(\Gamma + C_1),$$

where $\Gamma := \{\Gamma([x]_1) | [x]_1 \in \mathbf{F}_2^n / C_2\}$ and $\Gamma + C_1 = \{x + x_1 | x \in \Gamma, x_1 \in C_1\}$. For example, when we choose the minimum Hamming distance decoding D_{C_2/C_1},

$$D_{C_2/C_1}(y) := \operatorname*{argmin}_{[x_2]_1 \in C_2/C_1} \min_{x_1 \in C_1} |y - (x_1 + x_2)|.$$

By using the map $\Gamma([x]_2)$,

$$\Gamma([x]_2) = x + \operatorname*{argmin}_{x_2 \in C_2} |x + x_2|,$$

it can be written as

$$D_{C_2/C_1}(y) = [y - \Gamma([y]_2)]_1.$$

In the following, we denote the above Γ by Γ_{C_2}.

Now, we consider the average error when we choose the larger code C_2 randomly.

Lemma 1. Let C_1 be a arbitrary $[n,t]$ code $(C_1 \subset \mathbf{F}_2^n)$. We randomly choose the $(t+l)$-dimensional code $C_2(X) \supset C_1$ such that any element $x \in \mathbf{F}_2^n \backslash C_1$ belongs to $C_2(X)$ with probability $\frac{2^{l+t}-2^t}{2^n-2^t}$. Then, any additive channel W satisfies

$$\begin{aligned} \mathrm{E}_X[P_{e,W}(D^{\Gamma_{C_2(X)}})] &= \mathrm{E}_X[1 - P_W(\Gamma_{C_2(X)} + C_1)] \\ &\leqslant \sum_{k=0}^{n} \widetilde{P}_W(k) g(2^{l+t-n}|n,k), \end{aligned}$$

where

$$g(x|n,k) := \begin{cases} \min\{2^{n\bar{h}(k/n)}x, 1\}, & k \leqslant \lfloor n/2 \rfloor, \\ 1, & k > \lfloor n/2 \rfloor. \end{cases}$$

Proof. Let T_k^n be the set $\{x \in \mathbf{F}_2^n | |x| = k\}$. Then, $P(x) = \Sigma_{k=0}^n \widetilde{P}(k) P_k(x)$. Hence, $P(\Gamma_{C_2(X)} + C_1) = \Sigma_{k=0}^n \widetilde{P}(k) P_k(\Gamma_{C_2(X)} + C_1)$.

Indeed, if $y \in T_k^n \subset \mathbf{F}_2^n$ does not belong to $\Gamma_{C_2(X)} + C_1$, there exists an element $x \in C_2(X) \backslash C_1$ such that $|y - x| \leqslant k$. Hence, the probability that at least one element belongs to the set $\{x||x-y| \leqslant k\}$ is less than $2^{nh(k/n)} \frac{2^{l+t}-2^t}{2^n-2^t}$ for $k \leqslant n/2$ because $|\{x||x-y| \leqslant k\}| = |\{z||z| \leqslant k\}| \leqslant 2^{nh(k/n)}$ [see Eq. (9)]. Therefore,

$$\begin{aligned}\mathrm{E}_X[1 - P_k(\Gamma_{C_2(X)} + C_1)] &\leq \sum_{y \in T_n^k} P_k(y) 2^{nh(k/n)} \frac{2^{l+t} - 2^t}{2^n - 2^t} \\ &\leq 2^{nh(k/n)} \frac{2^{l+t} - 2^t}{2^n - 2^t} \\ &\leq 2^{nh(k/n)} \frac{2^{l+t}}{2^n}\end{aligned}$$

for $k \leq n/2$, where the last inequality follows from $l+t \leq n$. This value is also bounded by 1. Hence,

$$\begin{aligned}\mathrm{E}_X[1 - P(\Gamma_{C_2(X)} + C_1)] &= \sum_{k=0}^{n} \tilde{P}(k) \mathrm{E}_X[1 - P_k(\Gamma_{C_2(X)} + C_1)] \\ &\leq \sum_{k=0}^{n} \tilde{P}(k) g(2^{l+t-n} | n,k). \qquad \blacksquare\end{aligned}$$

V. GENERALIZED PAULI CHANNEL

In this section, for the preparation of our proof, we give some notations concerning generalized Pauli channels. In order to describe it, for any two elements $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n) \in \mathbf{F}_2^n$, we use the product

$$x \cdot y := \sum_{i=1}^{n} x_i y_i .$$

Thus, the space $\mathcal{H}_2^{\otimes n} = (\mathbb{C}^2)^{\otimes n}$ is spanned by the $\{|x\rangle\}_{x \in \mathbf{F}_2^n}$. Now, we define the unitary matrices $\mathbf{X}^x$ and $\mathbf{Z}^z$ for x, $z \in \mathbf{F}_2^n$ as

$$\mathbf{X}^x |x'\rangle = |x' - x\rangle,$$

$$\mathbf{Z}^z |x'\rangle = (-1)^{x' \cdot z} |x'\rangle.$$

From the definition, we have the relation [21]

$$(\mathbf{X}^x \mathbf{Z}^z)(\mathbf{X}^{x'} \mathbf{Z}^{z'}) = (-1)^{x \cdot z' - x' \cdot z} (\mathbf{X}^{x'} \mathbf{Z}^{z'})(\mathbf{X}^x \mathbf{Z}^z).$$

When the channel Λ has the form

$$\Lambda(\rho) = \sum_{x,z \in \mathbf{F}_2^n} P_\Lambda(x,z) (\mathbf{X}^x \mathbf{Z}^z) \rho (\mathbf{X}^x \mathbf{Z}^z)^\dagger ,$$

it is called a generalized Pauli channel. Indeed, a generalized Pauli channel is the quantum analog of an additive channel. In fact, it is known [22,23] that the channel Λ is a generalized Pauli channel if and only if

$$\Lambda(\rho) = (\mathbf{X}^x \mathbf{Z}^z)^\dagger \Lambda((\mathbf{X}^x \mathbf{Z}^z) \rho (\mathbf{X}^x \mathbf{Z}^z)^\dagger)(\mathbf{X}^x \mathbf{Z}^z), \qquad \forall\, x,z \in \mathbf{F}_2^n. \tag{10}$$

For any channel Λ, we often focus on its twirling Λ_t defined as

$$\Lambda_t(\rho) := \frac{1}{2^{2n}} \sum_{x,z \in \mathbf{F}_2^n} \Lambda^{x,z}(\rho),$$

$$\Lambda^{x,z}(\rho) := (\mathbf{X}^x \mathbf{Z}^z)^\dagger \Lambda((\mathbf{X}^x \mathbf{Z}^z) \rho (\mathbf{X}^x \mathbf{Z}^z)^\dagger)(\mathbf{X}^x \mathbf{Z}^z).$$

From Eq. (10), the twirling Λ_t is always a generalized Pauli channel.

In the treatment of generalized Pauli channels, the distribution $P_\Lambda(x,z)$ is important. Hence, we introduce some notations for this distribution. We define the distributions $P_{\Lambda,X}(x)$ and $P_{\Lambda,Z}(z)$ as

$$P_{\Lambda,X}(x) := \sum_{z \in \mathbf{F}_2^n} P_\Lambda(x,z), \quad P_{\Lambda,Z}(z) := \sum_{x \in \mathbf{F}_2^n} P_\Lambda(x,z).$$

These are called marginal distributions. We also define the conditional distribution as

$$P_{\Lambda,Z|X}(z|x) := \frac{P_\Lambda(x,z)}{P_{\Lambda,X}(x)}.$$

Next, we treat a generalized Pauli channel Λ on the tensor product system $(\mathbb{C}^2)^{\otimes n_1} \otimes (\mathbb{C}^2)^{\otimes n_2}$. In this case, we use the following notation:

$$P_{\Lambda,1}(x_1,z_1) := \sum_{x_2,z_2 \in \mathbf{F}_2^{n_2}} P_\Lambda(x_1 x_2, z_1 z_2),$$

$$P_{\Lambda,2}(x_2,z_2) := \sum_{x_1,z_1 \in \mathbf{F}_2^{n_1}} P_\Lambda(x_1 x_2, z_1 z_2),$$

$$P_{\Lambda,X,i}(x_i) := \sum_{z_i \in \mathbf{F}_2^{n_i}} P_{\Lambda,i}(x_i,z_i),$$

$$P_{\Lambda,Z,i}(z_i) := \sum_{x_i \in \mathbf{F}_2^{n_i}} P_{\Lambda,i}(x_i,z_i),$$

$$\tilde{P}_{\Lambda,Z,1,2}(k_1,k_2) := \sum_{x_i := \mathbf{F}_2^{n_i}} P_{\Lambda,Z}(T_{n_1}^{k_1} \times T_{n_2}^{k_2}), \tag{11}$$

$$P_{\Lambda,1|Z,2}(x_1,z_1|z_2) := \frac{\sum_{x_2 := \mathbf{F}_2^{n_2}} P_{\Lambda,1|Z,2}(x_1 x_2, z_1 z_2)}{P_{\Lambda,Z,2}(z_2)},$$

$$P_{\Lambda,Z,1|Z,2}(z_1|z_2) := \sum_{x_1 := \mathbf{F}_2^{n_1}} P_{\Lambda,1|Z,2}(x_1,z_1|z_2).$$

Note that $\tilde{P}_{\Lambda,Z,1,2}$ is different from $\tilde{P}_{\Lambda,Z}$. These notations will be used in the following sections.

VI. PROOF OF THE MAIN THEOREM

A. Modified protocol

In this section, we prove theorem 1 by treating the security of the following protocol. In the following protocol, we fix the generalized Pauli channel Λ from an n-qubit system to itself.

(i) Alice generates $Z_+ \in \mathbf{F}_2^m$ randomly and sends Bob $G(C_1)Z_+ \in \mathbf{F}_2^n$ with the $+$ basis through the n-qubit generalized Pauli channel Λ.

(ii) Bob measures the received n qubits with the $+$ basis. Performing the decoding of the code $C_1 \approx \mathbf{F}_2^m$, he obtains $\tilde{Z}_+ \in \mathbf{F}_2^m$.

(iii) They do processes (ix)—(xi) of the previous protocol. In this case, we assume that the dimension of the code C_1 [the subcode $C_{2,+}(Y_+)$] is $t(s)$.

This protocol is the special case that the channel is known.

For any channel Λ from the system $\mathcal{H}$ to itself, the state of the environment system can be described by using its Stinespring representation $(\mathcal{H}_E, U, |0\rangle_E \in \mathcal{H}_E)$:

$$\Lambda(\rho) = \mathrm{Tr}_{\mathcal{H}_E} U\rho \otimes |0\rangle_{EE}\langle 0|U^*.$$

That is, the state of the environment system is characterized by another channel $\Lambda_E(\rho) := \mathrm{Tr}_{\mathcal{H}_2^n} U\rho \otimes |0\rangle_{EE}\langle 0|U^*$.

In this above protocol, the distribution of Eve's signal Z_E is described by a positive-operator-valued measure (POVM) M_{Z_E} on $\mathcal{H}_E$ as $P(Z_E|Z) = \mathrm{Tr}\, M_{Z_E}\Lambda_E(\rho_Z)$. Therefore, in order to evaluate the classical mutual information $I(\bar{Z}, Z_E)$ it is sufficient to evaluate the quantum mutual information (Holevo information)

$$\begin{aligned} I([z] \in C_1/C_2(Y), \rho_{\Lambda,E}^{C_1/C_2(Y)}([z])) \\ := \frac{1}{2^{m-s}} \sum_{[z]\in C_1/C_2(Y)} \mathrm{Tr}\rho_{\Lambda,E}^{C_1/C_2(Y)}([z])(\log \rho_{\Lambda,E}^{C_1/C_2(Y)}([z]) \\ - \log \rho_{\Lambda,E}^{C_1/C_2(Y)}), \end{aligned} \tag{12}$$

where

$$\rho_{\Lambda,E}^{C_1/C_2(Y)}([z]) := \sum\nolimits_{z_2 \in C_2(Y)} \Lambda_E(|z+z_2\rangle\langle z+z_2|)$$

and

$$\rho_{\Lambda,E}^{C_1/C_2(Y)} := \tfrac{1}{2^{t-s}} \sum\nolimits_{[z]\in C_1/C_2(Y)} \rho_{\Lambda,E}^{C_1/C_2(Y)}([z]).$$

In the following, we often abbreviate (12) as $I_H(\bar{Z}, Z_E)$.

Theorem 3. We can evaluate Eve's information as follows:

$$\begin{aligned} \mathrm{E}_{Y_+}[I([z] \in C_1/C_2(Y_+,s), \rho_{\Lambda,E}^{C_1/C_2(Y)}([z]))] \\ \leqslant \eta_{m-s}\left(\sum_{k=0}^{n} \tilde{P}_{\Lambda,Z}(k) g(2^{-s}|n,k)\right), \end{aligned}$$

where $m = \dim C_1$ and η_k is defined as

$$\eta_k(x) := \bar{h}(x) + kx.$$

This theorem will be proved in Sec. VIII.

B. Proof of theorem 1

Now, we back to our main protocol. First, we fix the random variables $\mathrm{pos}_+, k_+, Y_\times$. Then, it is sufficient to treat the quantum system of $n_+ + l_\times$ qubits. In the following, we characterize the system of raw keys $\mathbb{C}^n$ by the subscript k and the other system of check qubits $\mathbb{C}^{l\times}$ by the subscript c.

Hence, we denote the quantum channel of this system by Λ. Note that Λ is not necessarily a generalized Pauli channel. In the following, we abbreviate $l_\times, \mathrm{pos}_\times, k_\times, Y_+$ by l, pos, k, Y, respectively.

In this case, the variable pos takes a subset of l elements $\{i_1, \ldots, i_l\} \subset \{1, \ldots, n+l\}$, where $i_1 < \cdots < i_l$. Then, we define the unitary matrix U_{pos} as

$$U_{\mathrm{pos}}(u_{i_1} \otimes \cdots \otimes u_{i_l} \otimes u_{j_1} \otimes \cdots \otimes u_{j_n}) = u_1 \otimes \cdots \otimes u_{n+l},$$

where $\{j_1, \ldots, j_n\} = \{i_1, \ldots, i_l\}^c$ and $j_1 < \cdots < j_n$. Every subset is chosen with the probability $1/\binom{n+l}{l}$. We also define the channel Λ^{pos} for any channel Λ as

$$\Lambda^{\mathrm{pos}}(\rho) := U_{\mathrm{pos}}^\dagger(\Lambda(U_{\mathrm{pos}}\rho U_{\mathrm{pos}}^\dagger))U_{\mathrm{pos}}.$$

Then, we can show that

$$(\Lambda^{\mathrm{pos}})_t = (\Lambda_t)^{\mathrm{pos}}. \tag{13}$$

Hence, any generalized Pauli channel Λ satisfies

$$\mathrm{E}_{\mathrm{pos}}[\tilde{P}_{\Lambda^{\mathrm{pos}},Z,k,c}(k_k,k_c)] = \tilde{P}_{\Lambda,Z}(k_k+k_c)P_{hg}(k_c|n,l,k_k+k_c), \tag{14}$$

where we used the notation given in Eq. (11).

Now, we consider the case where Alice and Bob choose a variable pos and obtain the difference z_c between their check bit with the $\times$ basis. When $\underline{k} \leqslant |z_c| \leqslant \bar{k}$, the average of Eve's final information is evaluated as

$$\begin{aligned} \mathrm{E}_{Y_+}\left[I\left([z] \in C_1/C_2\left(Y_+, nh\left(\frac{|z_c|}{l} + \delta_{|z_c|}\right)\right), \rho_{(\Lambda_t)^{\mathrm{pos},z},E}^{C_1/C_2(Y)}([z])\right)\right] \\ \leqslant \bar{h}\left(\sum_{k=0}^{n} \tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z,k|Z,c}(k|z_c) f(k,|z_c||n,l,\delta_k)\right) \\ + n[R - h(|z_c|/l + \delta_k)] \sum_{k=0}^{n} \tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z,k|Z,c}(k|z_c) \\ \times f(k,|z_c||n,l,\delta_k). \end{aligned} \tag{15}$$

When $|z_c| < \underline{k}$, we obtain

$$\begin{aligned} \mathrm{E}_{Y_+}\left[I\left([z] \in C_1/C_2\left(Y_+, nh\left(\frac{\underline{k}}{l} + \delta_{\underline{k}}\right)\right), \rho_{(\Lambda_t)^{\mathrm{pos},z},E}^{C_1/C_2(Y)}([z])\right)\right] \\ \leqslant \bar{h}\left(\sum_{k=0}^{n} \tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z,k|Z,c}(k|z_c) f(k,\underline{k}|n,l,\delta_{\underline{k}})\right) \\ + n[R - h(\underline{k}/l + \delta_{\underline{k}})] \sum_{k=0}^{n} \tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z,k|Z,c}(k|z_c) \\ \times f(k,\underline{k}|n,l,\delta_{\underline{k}}). \end{aligned} \tag{16}$$

Of course, when $|z_c| > \underline{k}$, the average of Eve's final information is equal to zero because any information is discarded in this case. Inequalities (15) and (16) will be shown in Appendix C by using theorem 3.

Finally, we take the expectation concerning z_c and pos:

$$\begin{aligned}
&\mathrm{E}_{\mathrm{pos}}\mathrm{E}_{z_c}\mathrm{E}_{Y_+}[I([z]\in C_1/C_2(Y_+,nh(|z_c|/l+\delta_{|z_c|})),\rho^{C_1/C_2(Y)}_{(\Lambda_t)^{\mathrm{pos},z},E}([z]))] \\
&\quad\leqslant \bar{h}\left(\max_j\left[\sum_{k_c=0}^{\underline{k}} P_{hg}(k_c|n,l,j)f(j-\underline{k},\underline{k}|n,l,\delta_{\underline{k}})+\sum_{k_c=\underline{k}+1}^{\bar{k}} P_{hg}(k_c|n,l,j)f(j-k_c,k_c|n,l,\delta_{k_c})\right]\right) \\
&\qquad+\max_j\left[\sum_{k_c=0}^{\underline{k}} P_{hg}(k_c|n,l,j)n[R-h(\underline{k}/l+\delta_{\underline{k}})]f(j-\underline{k},\underline{k}|n,l,\delta_{\underline{k}})\right. \\
&\qquad\qquad\left.+\sum_{k_c=\underline{k}+1}^{\bar{k}} P_{hg}(k_c|n,l,j)n[R-h(k_c/l+\delta_{k_c})]f(j-k_c,k_c|n,l,\delta_{k+c})\right]. \qquad (17)
\end{aligned}$$

This inequality will be proved in Appendix D. Hence, we obtain theorem 1. Similarly, we have

$$\begin{aligned}
&\mathrm{E}_{\mathrm{pos}}\mathrm{E}_{z_c}\mathrm{E}_{Y_+}\left[\frac{I([z]\in C_1/C_2(Y_+,nh(|z_c|/l+\delta_{|z_c|})),\rho^{C_1/C_2(Y)}_{(\Lambda_t)^{\mathrm{pos},z},E}([z]))}{n[R-h(k_\times/l_\times+\delta_{k_\times})]}\right] \\
&\leqslant \frac{1}{n[R-h(\bar{k}/l_\times+\delta_{\bar{k}})]}\bar{h}\left(\max_j\left[\sum_{k_c=0}^{\underline{k}} P_{hg}(k_c|n,l,j)f(k_k,\underline{k}|n,l,\delta_{\underline{k}})+\sum_{k_c=\underline{k}+1}^{\bar{k}} P_{hg}(k_c|n,l,j)f(k_k,k_c|n,l,\delta_{k_c})\right]\right) \\
&\quad+\max_j\left[\sum_{k_c=0}^{\underline{k}} P_{hg}(k_c|n,l,j)f(k_k,\underline{k}|n,l,\delta_{\underline{k}})+\sum_{k_c=\underline{k}+1}^{\bar{k}} P_{hg}(k_c|n,l,j)f(k_k,k_c|n,l,\delta_{k+c})\right], \qquad (18)
\end{aligned}$$

This inequality will be proved in Appendix D. Hence, we obtain theorem 2.

VII. SECURITY AND PHASE ERROR

In this section, we treat the relation between Eve's information and the phase error. This relation is one of essential parts for theorem 3. The purpose of this section is proving the following lemmas [29].

Lemma 2. Let Λ be a generalized Pauli channel on the system $(\mathbb{C}^2)^{\otimes n}$. Then, we have

$$I(x\in\mathbf{F}_2^n,\Lambda_E(|x\rangle\langle x|))\leqslant\eta_n[1-P_{\Lambda,Z}(0)]. \qquad (19)$$

Since $1-P_{\Lambda,Z}(0)$ can be regarded as the phase error, this lemma gives a relation between the phase error and Eve's information.

Proof. The Stinespring representation of Λ is given as $((\mathbb{C}^2)^{\otimes 2n},U,|\phi\rangle)$:

$$|\phi\rangle:=\sum_{x,z\in\mathbf{F}_2^n}\sqrt{P_\Lambda(x,z)}|x,z\rangle,$$

$$U:=\sum_{x,z\in\mathbf{F}_2^n}\mathbf{X}^x\mathbf{Z}^z\otimes|x,z\rangle\langle x,z|.$$

Since

$$\begin{aligned}
U|x'\rangle\otimes|\phi\rangle&=\sum_{x,z\in\mathbf{F}_2^n}\sqrt{P_\Lambda(x,z)}(-1)^{x'z}|x'-x\rangle\otimes|x,z\rangle \\
&=\sum_{x\in\mathbf{F}_2^n}|x'-x\rangle\otimes|\phi_{x,x'}\rangle\otimes\sqrt{P_{\Lambda,X}(x)}|x\rangle,
\end{aligned}$$

Eve's state can be written as

$$\Lambda_E(|x'\rangle\langle x'|)=\sum_{x\in\mathbf{F}_2^n}P_{\Lambda,X}(x)|\phi_{x,x'}\rangle\langle\phi_{x,x'}|\otimes|x\rangle\langle x|,$$

where $|\phi_{x,x'}\rangle:=\sum_{z\in\mathbf{F}_2^n}\sqrt{P_{\Lambda,Z|X}(z|x)}(-1)^{x'z}|z\rangle$. Since x' obeys a uniform distribution,

$$\begin{aligned}
I(x\in\mathbf{F}_2^n,\Lambda_E(|x\rangle\langle x|))&=\sum_{x\in\mathbf{F}_2^n}P_{\Lambda,X}(x)H\left(\frac{1}{2^n}\sum_{x'\in\mathbf{F}_2^n}|\phi_{x,x'}\rangle\langle\phi_{x,x'}|\right) \\
&=\sum_{x\in\mathbf{F}_2^n}P_{\Lambda,X}(x)H(P_{\Lambda,Z|X}(\cdot|x))\leqslant H(P_{\Lambda,Z}).
\end{aligned}$$

Hence, using lemma 3, we obtain (19). ■

Lemma 3. Let $P=\{P(i)\}$ be a distribution on $\{0,\ldots,d-1\}$. Then, $H(P)\leqslant h[1-P(0)]+\log(d-1)[1-P(0)]$.

Proof:

$$\begin{aligned} H(P) &= -P(0)\log P(0) - [1-P(0)]\log[1-P(0)] \\ &\quad -(1-P(0))\sum_{i=1}^{d-1}\frac{P(i)}{[1-P(0)]}\log\frac{P(i)}{[1-P(0)]} \\ &\leqslant h[1-P(0)] + \log(d-1)[1-P(0)]. \end{aligned}$$

■

VIII. SECURITY OF THE KNOWN CHANNEL

In this section, we treat the security when the channel is known—i.e., prove theorem 3 using lemmas 2 and 1. To prove it, for any code $C \subset \mathbf{F}_2^n$ and any elements $[z] \in \mathbf{F}_2^n/C^{\perp}$ and $[x] \in \mathbf{F}_2^n/C$, we define

$$|x,z\rangle_C := \frac{1}{\sqrt{|C|}}\sum_{x' \in C}(-1)^{zx'}|x+x'\rangle. \tag{20}$$

Note that this definition does not depend on the choice of the coset representative elements z (x) of $[z]$ ($[x]$). When we choose $\mathbf{F}_2^n$ as C, the above is the discrete Fourier transform. Then, we have the following lemma.

Lemma 4. When two codes C_1 and C_2 satisfy $C_2 \subset C_1$, any elements $x \in \mathbf{F}_2^n$, $[z_1] \in C_2^{\perp}/C_1^{\perp}$, and $[z_2] \in \mathbf{F}_2^n/C_2^{\perp}$ satisfy

$$|x,z_1+z_2\rangle_{C_1} = \frac{1}{\sqrt{|C_1/C_2|}}\sum_{[x_1]\in C_1/C_2}(-1)^{(z_1+z_2)x_1}|x+x_1,z_2\rangle_{C_2}. \tag{21}$$

Note that the RHS does not depend of the choice of the coset representative elements x_1 of $[x_1]$.

Proof:

$$\begin{aligned} &\frac{1}{\sqrt{|C_1/C_2|}}\sum_{[x_1]\in C_1/C_2}(-1)^{(z_1+z_2)x_1}|x+x_1,z_2\rangle_{C_2} \\ &\quad = \frac{1}{\sqrt{|C_1/C_2|}}\sum_{[x_1]\in C_1/C_2}\sum_{x_2\in C_2}\frac{(-1)^{(z_1+z_2)x_1+z_2x_2}}{\sqrt{|C_2|}}|x+x_1+x_2\rangle. \end{aligned}$$

Since $(z_1+z_2)(x_1+x_2)=(z_1+z_2)x_1+z_2x_2$, we obtain Eq. (21). ■

Lemma 5:

$$\sum_{x_1\in C_1}|x+x_1\rangle\langle x+x_1| = \sum_{[z_1]\in\mathbf{F}_2^n/C_1^{\perp}}|x,z_1\rangle_{C_1\,C_1}\langle x,z_1|. \tag{22}$$

Proof. From the definition of $|x,z_1\rangle_{C_1}$, we have

$$\begin{aligned} \sum_{[z_1]\in\mathbf{F}_2^n/C_1^{\perp}}|x,z_1\rangle_{C_1\,C_1}\langle x,z_1| &= \frac{1}{|C_1|}\sum_{[z_1]\in\mathbf{F}_2^n/C_1^{\perp}}\sum_{x'\in C_1}\sum_{x''\in C_1}(-1)^{z_1(x'+x'')}|x+x''\rangle\langle x+x'| \\ &= \frac{1}{|C_1|}\sum_{[z_1]\in\mathbf{F}_2^n/C_1^{\perp}}\sum_{x'\in C_1}\sum_{x''\in C_1}(-1)^{z_1[x'+x'+(x''-x')]}|x+x'+x''-x'\rangle\langle x+x'| \\ &= \frac{1}{|C_1|}\sum_{[z_1]\in\mathbf{F}_2^n/C_1^{\perp}}\sum_{x'\in C_1}\sum_{y\in C_1}(-1)^{z_1y}|x+x'+y\rangle\langle x+x'| \\ &= \sum_{x'\in C_1}|x+x'\rangle\langle x+x'|, \end{aligned}$$

because $y \in C_1$ satisfies

$$\frac{1}{|C_1|}\sum_{[z_1]\in\mathbf{F}_2^n/C_1^{\perp}}(-1)^{z_1y} = \begin{cases} 1 & \text{if } y=0, \\ 0 & \text{if } y\neq 0. \end{cases}$$

■

Now, we define the minimum error

$$P([z_1]\in C_2^{\perp}/C_1^{\perp},\Lambda(|0,z_1+z_2\rangle_{C_1C_1}\langle 0,z_1+z_2|)) := \min_M\left(1-\sum_{[z_1]\in C_2^{\perp}/C_1^{\perp}}\frac{\mathrm{Tr}M_{[z_1]}\Lambda(|0,z_1+z_2\rangle_{C_1\ C_1}\langle 0,z_1+z_2|))}{|C_2^{\perp}/C_1^{\perp}|}\right),$$

where M is a POVM $\{M_{[z_1]}\}_{[z_1]\in C_2^{\perp}/C_1^{\perp}}$. Then, we have the following evaluation.

Lemma 6:

$$I\left([x_1]\in C_1/C_2,\Lambda_E\left(\frac{1}{|C_2|}\sum_{x_2\in C_2}|x_1+x_2\rangle\langle x_1+x_2|\right)\right) \leqslant \eta_{m-s}\sum_{[z_2]\in\mathbf{F}_2^n/C_2^{\perp}}\frac{1}{|C_2|}P(z_1\in C_2^{\perp}/C_1^{\perp},\Lambda(|0,z_1+z_2\rangle_{C_1C_1}\langle 0,z_1+z_2|)), \tag{23}$$

where $m = \dim C_1$ and $s = \dim C_2$.

Proof. Using lemma 5 and the convexity of mutual information, we have

$$\begin{aligned} I\left([x_1] \in C_1/C_2, \Lambda_E\left(\frac{1}{|C_2|}\sum_{x_2 \in C_2} |x_1 + x_2\rangle\langle x_1 + x_2|\right)\right) &= I\left([x_1] \in C_1/C_2, \Lambda_E\left(\frac{1}{|C_2|}\sum_{[z_2] \in \mathbf{F}_2^n/C_2^\perp} |x_1, z_2\rangle_{C_2 C_2}\langle x_1, z_2|\right)\right) \\ &\leqslant \frac{1}{|C_2|}\sum_{[z_2] \in \mathbf{F}_2^n/C_2^\perp} I([x_1] \in C_1/C_2, \Lambda_E(|x_1, z_2\rangle_{C_2 C_2}\langle x_1, z_2|)). \end{aligned} \tag{24}$$

Applying lemma 2, we have

$$I(x_1 \in C_1/C_2, \Lambda_E(|x_1, z_2\rangle_{C_2 C_2}\langle x_1, z_2|)) \leqslant \eta_{m-s} P(z_1 \in C_2^\perp/C_1^\perp, \Lambda(|0, z_1 + z_2\rangle_{C_1 C_1}\langle 0, z_1 + z_2|)).$$

From Eq. (24), the concavity of η_{m-s} implies

$$\begin{aligned} I\left([x_1] \in C_1/C_2, \Lambda_E\left(\frac{1}{|C_2|}\sum_{x_2 \in C_2} |x_1 + x_2\rangle\langle x_1 + x_2|\right)\right) &\leqslant \frac{1}{|C_2|}\sum_{[z_2] \in \mathbf{F}_2^n/C_2^\perp} \eta_{m-s} P(z_1 \in C_2^\perp/C_1^\perp, \Lambda(|0, z_1 + z_2\rangle_{C_1 C_1}\langle 0, z_1 + z_2|)) \\ &\leqslant \eta_{m-s}\frac{1}{|C_2|}\sum_{[z_2] \in \mathbf{F}_2^n/C_2^\perp} P(z_1 \in C_2^\perp/C_1^\perp, \Lambda(|0, z_1 + z_2\rangle_{C_1 C_1}\langle 0, z_1 + z_2|)). \end{aligned}$$

■

Since Λ is a generalized Pauli channel, any coset $[x_0] \in \mathbf{F}_2^n/C_1$ satisfies

$$\Lambda(|x_0, z_1 + z_2\rangle_{C_1\ C_1}\langle x_0, z_1 + z_2|) = \mathbf{X}^{x_0}\Lambda(|0, z_1 + z_2\rangle_{C_1\ C_1}\langle 0, z_1 + z_2|)(\mathbf{X}^{x_0})^\dagger.$$

Hence,

$$P([z_1] \in C_2^\perp/C_1^\perp, \Lambda(|0, z_1 + z_2\rangle_{C_1 C_1}\langle 0, z_1 + z_2|)) = P([z_1] \in C_2^\perp/C_1^\perp, \Lambda(|x_0, z_1 + z_2\rangle_{C_1 C_1}\langle x_0, z_1 + z_2|)).$$

Thus,

$$\begin{aligned} &P(z_1 \in C_2^\perp/C_1^\perp, \Lambda(|0, z_1 + z_2\rangle_{C_1 C_1}\langle 0, z_1 + z_2|)) \\ &\quad = \frac{|C_1|}{2^n}\sum_{[x_0] \in \mathbf{F}_2^n/C_1} P(z_1 \in C_2^\perp/C_1^\perp, \Lambda(|x_0, z_1 + z_2\rangle_{C_1 C_1}\langle x_0, z_1 + z_2|)) \\ &\quad \leqslant P\left(z_1 \in C_2^\perp/C_1^\perp, \frac{|C_1|}{2^n}\sum_{[x_0] \in \mathbf{F}_2^n/C_1} \Lambda(|x_0, z_1 + z_2\rangle_{C_1 C_1}\langle x_0, z_1 + z_2|)\right) \\ &\quad = P\left(z_1 \in C_2^\perp/C_1^\perp, \Lambda\left(\frac{|C_1|}{2^n}\sum_{z_0 \in C_1^\perp} |z_0 + z_1 + z_2\rangle_{\mathbf{F}_2^n \mathbf{F}_2^n}\langle z_0 + z_1 + z_2|\right)\right). \end{aligned} \tag{25}$$

Now, we focus on the step (ix) and the subcode $G(C_1)C_2(Y,s) \subset C_1$ and abbreviate $G(C_1)C_2(Y,s)$ to $C_2(Y,s)$. Then, the dual code $C_2(Y,s)^\perp$ satisfies $C_1^\perp \subset C_2(Y,s)^\perp$ and the condition of $C_2(X)$ in lemma 1 when t, l, and C_1 in lemma 1 are given by $n-v$, $v-s$, and $C_1^\perp$, respectively. Then, $n-(l+t)$ in lemma 1 is given by s. Since the generalized Pauli channel can be regarded as the additive channel, we can apply lemma 1. Hence,

$$\mathrm{E}_Y\left[\frac{1}{|C_2(Y,s)|}\sum_{[z_2] \in \mathbf{F}_2^n/C_2(Y,s)^\perp} P\left(z_1 \in C_2(Y,s)^\perp/C_1^\perp, \Lambda_E\left(\frac{|C_1|}{2^n}\sum_{z_0 \in C_1^\perp} |z_0 + z_1 + z_2\rangle_{\mathbf{F}_2^n \mathbf{F}_2^n}\langle z_0 + z_1 + z_2|\right)\right)\right] \leqslant \sum_{k=0}^{n} \widetilde{P}_{W_t}(k) g(2^{-s}|n,k). \tag{26}$$

From (25), (26), and (23), the convexity of η_{m-s} yields that

$$\mathrm{E}_Y\left[\frac{1}{|C_2|}\sum_{[z_2]\in\mathbf{F}_2^n/C_2^\perp} I([x_1]\in C_1/C_2,\Lambda_E(|x_1,z_2\rangle_{C_2C_2}\langle x_1,z_2|))\right]$$
$$\leqslant \eta_{m-s}\left(\mathrm{E}_Y\left[\frac{1}{|C_2|}\sum_{[z_2]\in\mathbf{F}_2^n/C_2^\perp} P(z_1\in C_2^\perp/C_1^\perp,\Lambda(|0,z_1+z_2\rangle_{C_1C_1}\langle 0,z_1+z_2|))\right]\right)\leqslant \eta_{m-s}\left(\sum_{k=0}^{\lfloor n/2\rfloor}\tilde{P}_{W_t}(k)g(2^{-s}|n,k)\right).$$

Therefore, from (24), we obtain theorem 3.

IX. OPTIMAL ATTACK

In this section, we prove that there exists a collective attack attaining the the exponential rate (4) under a condition. Indeed, it is not so easy to evaluate $\max I(\bar{Z},Z_E)$. Hence, we treat $I_H(\bar{Z},Z_E)$ instead of $I(\bar{Z},Z_E)$.

Lemma 7. Assume that the sequence of codes C_{1,n,k_+} satisfies

$$\max_{k\leqslant[\bar{p}+\epsilon(\bar{p})]n} P_{e,W_{k,n}}(C_{1,n}^\perp)\to 0, \tag{27}$$

where the channel $W_{k,n}$ is defined on $\mathbf{F}_2^n$ as $P_{W_{k,n}}(j)=\delta_{k,j}$. Then, we have

$$\lim\frac{-r}{n}\log \mathrm{E}_{k_\times|\mathrm{pos}_\times,Y_+,\mathrm{pos}_+,k_+,Y_\times}[\max_{\mathcal{E}} I_H(\bar{Z},Z_E)]$$
$$\leqslant \min_{p\in[\underline{p},\bar{p}]} h[p+r\epsilon(p)]-(1-r)h(p)-rh[p+\epsilon(p)], \tag{28}$$

where the maximum is taken concerning Eve's operation $\mathcal{E}$. Note that the above inequality holds for any fixed variable Y_+.

Hence, if $\epsilon(p)$ is sufficiently small and the sequence of codes $C_{1,n}$ satisfies the condition (27), we have

$$\lim\frac{-r}{n}\log \mathrm{E}_{\mathrm{pos}_\times,k_\times,Y_+|\mathrm{pos}_+,k_+,Y_\times}[\max I_H(\bar{Z},Z_E)]$$
$$=h[p+r\epsilon(p)]-(1-r)h(p)-rh[p+\epsilon(p)]. \tag{29}$$

This indicates that the method of randomly choosing the code C_2 is optimal in the sense of large deviation.

In this lemma, we assume the condition (27). Indeed, we need some conditions in lemma 7. For example, consider the code C_1, which consists of the elements x whose first $n-m$ components are zero. In this case, the following proof is not valid. Indeed, when the limit $\lim_{n\to\infty}\frac{1}{n}\log|C_{1,n}|$ is greater than $h[\bar{p}+\epsilon(\bar{p})]$ and we choose $C_{1,n}$ randomly, the condition (27) holds. Hence, the condition (27) is not so unnatural. However, a more natural condition is needed.

As is shown later, the exponential rate $\min_{p\in[\underline{p},\bar{p}]}h[p+r\epsilon(p)]-(1-r)h(p)-rh[p+\epsilon(p)]$ can be attained by a collective attack, in which Eve's is allowed only individual unitary operations to quantum states sent by Alice and any global generalized measurement on the Eve's local states. Hence, the exponential rate of Eve's information cannot be improved by any collective attack, in which Eve's is allowed to use any unitary operation to all quantum states sent by Alice.

Now, we construct Eve's strategy attaining the bound $\min_{p\in[\underline{p},\bar{p}]}h[p+r\epsilon(p)]-(1-r)h(p)-rh[p+\epsilon(p)]$ and prove (28). Choose $p_0:=\mathrm{argmin}_{p\in[\underline{p},\bar{p}]}h[p+r\epsilon(p)]-(1-r)h(p)-rh[p+\epsilon(p)]$. Eve performs a unitary action $U_{p_0+r\epsilon(p_0)}$,

$$U_p|x\rangle\otimes|0\rangle_E:=\sqrt{p}|x\rangle\otimes|0\rangle_E+(-1)^x\sqrt{1-p}|x\rangle\otimes|1\rangle_E,$$

for a every qubit, where $|x\rangle_E$ is Eve's state.

We define the unitary $\tilde{U}_k$:

$$\tilde{U}_k^n|x\rangle\otimes|0\rangle_E:=\sqrt{\frac{1}{\binom{n}{k}}}\sum_{y\in\mathbf{F}_2^n:|x|=k}(-1)^{xy}|x\rangle\otimes|y\rangle_E.$$

We can easily show that

$$H\left(\Lambda_{E,k}^k\left(\sum_{x\in C_{1,n}}|x\rangle\langle x|\right)\right)$$
$$=H\left(\Lambda_{E,k}^k\left(\sum_{x\in C_{1,n}}|y+x\rangle\langle y+x|\right)\right)\quad\text{for } y\in\mathbf{F}_2^n.$$

Hence, applying lemma 6 to the case of $C_1=\mathbf{F}_2^n, C_2=C_{1,n}$, we have

$$\log\binom{n}{k}-H\left(\Lambda_{E,k}^n\left(\sum_{x\in C_{1,n}}|x\rangle\langle x|\right)\right)=H\left(\Lambda_{E,k}^n\left(\sum_{x\in\mathbf{F}_2^n}|x\rangle\langle x|\right)\right),$$
$$-H\left(\Lambda_{E,k}^n\left(\sum_{x\in C_{1,n}}|x\rangle\langle x|\right)\right)$$
$$\leqslant\bar{h}[P_{e,W_{k,n}}(C_{1,n}^\perp)]+\log|C_{1,n}|P_{e,W_{k,n}}(C_{1,n}^\perp),$$

where

$$\Lambda_{E,k}^n(\rho):=\mathrm{Tr}_B\tilde{U}_k^n(\rho\otimes|0\rangle_{EE}\langle 0|)(\tilde{U}_k^n)^\dagger.$$

Now, we evaluate Eve's information. In this case, the subcode $C_{2,n,k_\times}$ depends on the outcome $k_\times$. Taking the pinching map $\rho\mapsto\sum_k P_{n,k}\rho P_{n,k}$ ($P_{n,k}$ is the projection to the space spanned by $\{|x\rangle\}_{|x|=k}$), we have

$$H\left(\mathrm{E}_{k_\times}\left[\Lambda_E^{\otimes n}\left(\sum_{x\in C_{1,n}}|x\rangle\langle x|\right)\right]\right)-\sum_{[x]_2\in C_{1,n}/C_{2,n,k_\times}}H\left(\Lambda_E^{\otimes n}\left(\sum_{y\in C_{2,n,k_\times}}|x+y\rangle\langle x+y|\right)\right)$$
$$\geqslant H\left(\mathrm{E}_{k_\times,k}\left[\Lambda_{E,k}^{n}\left(\sum_{x\in C_{1,n}}|x\rangle\langle x|\right)\right]\right)-\sum_{[x]_2\in C_{1,n}/C_{2,n,k_\times}}H\left(\Lambda_{E,k}^{n}\left(\sum_{y\in C_{2,n,k_\times}}|x+y\rangle\langle x+y|\right)\right),$$

where k is the random variable with distribution $\tilde{P}(k):=\binom{n}{k}[p_0+r\epsilon(p_0)]^k[1-p_0-r\epsilon(p_0)]^{n-k}$.

When $k=n[p_0+\epsilon(p_0)+\epsilon]$, $k_\times=p_0$, we have

$$\frac{1}{n}H\left(\Lambda_{E,k}^{n}\left(\sum_{x\in C_{1,n}}|x\rangle\langle x|\right)\right)-\frac{1}{n}\sum_{[x]_2\in C_{1,n}/C_{2,n,k_\times}}H\left(\Lambda_{E,k}^{n}\left(\sum_{y\in C_{2,n,k_\times}}|x+y\rangle\langle x+y|\right)\right)$$
$$\geqslant\frac{1}{n}\left[\log\binom{n}{k}-\bar{h}[P_{e,W_{k,n}}(C_{1,n}^{\perp})]-\log|C_{1,n}|P_{e,W_{k,n}}(C_{1,n}^{\perp})-\log|C_{2,n,k_\times}|\right]$$
$$=\frac{1}{n}\left\{\log\binom{n}{k}-nh\left[\frac{k_\times}{n}+\epsilon\left(\frac{k_\times}{n}\right)\right]-\bar{h}[P_{e,W_{k,n}}(C_{1,n}^{\perp})]-\log|C_{1,n}|P_{e,W_{k,n}}(C_{1,n}^{\perp})\right\}$$
$$\to h[p_0+\epsilon(p_0)+\epsilon]-h[p_0+\epsilon(p_0)]\quad\text{as } n\to\infty.$$

Hence, Eve's information can be bounded as

$$\mathrm{E}_{k_\times}\left[H\left(\Lambda_E^{\otimes n}\left(\sum_{x\in C_{1,n}}|x\rangle\langle x|\right)\right)-\sum_{[x]_2\in C_{1,n}/C_{2,n,k_\times}}H\left(\Lambda_E^{\otimes n}\left(\sum_{y\in C_{2,n,k_\times}}|x+y\rangle\langle x+y|\right)\right)\right]$$
$$\geqslant\binom{n}{n(p_0+\epsilon(p_0)+\epsilon)}[p_0+r\epsilon(p_0)]^{n[p_0+\epsilon(p_0)+\epsilon]}[1-p_0-r\epsilon(p_0)]^{n-n[p_0+\epsilon(p_0)+\epsilon]}$$
$$\times\binom{l}{lp_0}[p_0+r\epsilon(p_0)]^{lp_0}[1-p_0-r\epsilon(p_0)]^{l(1-p_0)}(n\{h[p_0+\epsilon(p_0)+\epsilon]-h[p_0+\epsilon(p_0)]\}+o(n))$$
$$\geqslant\frac{n\{h[p_0+\epsilon(p_0)+\epsilon]-h[p_0+\epsilon(p_0)]\}+o(n)}{(n+1)^2}2^{-nd[p_0+\epsilon(p_0)+\epsilon\|p_0+r\epsilon(p_0)]-ld[p_0\|p_0+r\epsilon(p_0)]}.$$

Thus, we obtain

$$\lim\frac{-r}{n}\log\mathrm{E}_{k_\times|\mathrm{pos}_\times,Y_+,\mathrm{pos}_+,k_+,Y_\times}[\max I_H(\bar{Z}_+,Z_E)]\leqslant h[p_0+r\epsilon(p_0)]-(1-r)h(p_0)-rh[p_0+\epsilon(p_0)+\epsilon].$$

Taking the limit $\epsilon\to 0$, we obtain (28).

X. CONCLUSION

In this paper, we obtained a practical evaluation of security of the quantum key distribution. This bound improves existing bounds. In order to guarantee the security of the implemented QKD system, we need a tighter bound in the finite coding length. Hence, our bound is useful for guaranteeing the security of the quantum key distribution with a perfect single-photon source. However, for a precise evaluation, we have to treat hypergeometric distributions, because our bound contains hypergeometric distributions. Hence, it is needed to calculate these bounds by a numerical analysis based on several calculations of hypergeometric distributions.

We also derived the exponential rate of our bound as Eq. (4) and proved its optimality with in the sense of Holevo information with a class of one-way communication when C_p is less than the critical case. However, our condition for our code is not sufficiently natural. Hence, it is required to prove this optimality under a more natural condition. One candidate of a more natural condition is

$$\max_{k\leqslant\bar{h}^{-1}\{1-h[\bar{p}+\epsilon(\bar{p})]\}n}P_{e,W_{k,n}}(C_{1,n}^{\perp})\to 0. \tag{30}$$

Hence, it is a future problem to show optimality under the above condition.

Further, we assumed a perfect single-photon source. One idea for the weak coherent case is the decoy method [24], which is based on the observation of security with imperfect devices [25]. However, any existing paper [26–28] of the decoy method does not discuss the degree of Eve's information in the framework of a finite coding length, precisely-.Hence, it is required to extend our result to the weak coherent case with the decoy method.

ACKNOWLEDGMENTS

The author would like to thank Professor Hiroshi Imai of the ERATO-SORST, QCI project for support. He is grateful to Professor Hiroshi Imai, Dr. Akihisa Tomita, Professor Keiji Matsumoto, and Jun Hasewaga for useful discussions.

APPENDIX A: DERIVATION OF (3)

Now, we prove (3). In the following, we denote $n+l$ by m and fix $p \in [\underline{p}, \overline{p}]$. We treat the case of $j=pm$ and define the number $k_p(m) := \max\{k \,|\, h(\frac{pm-k}{n}) - h(\frac{k}{l}+\delta_k) \geqslant 0\}$. In this case, the first term of $\widetilde{P}(\delta, n, l, \underline{k}, \overline{k})$ goes to 0. Hence, we focus on the second term of $\widetilde{P}(\delta, n, l, \underline{k}, \overline{k})$, which is divided as

$$\begin{aligned}\widetilde{P}_2(\delta,n,l,\underline{k},\overline{k}) &:= \sum_{k=0}^{\underline{k}} P_{hg}(k|n,l,j) f(j-\underline{k},\underline{k}|n,l,\delta_{k_\times}) \\ &\quad + \sum_{k=\underline{k}+1}^{\overline{k}} P_{hg}(k|n,l,j) f(j-k,k|n,l,\delta_{k_\times}) \\ &= \sum_{k=0}^{k_p(m)} P_{hg}(k|n,l,j) \\ &\quad + \sum_{k=k_p(m)+1}^{\overline{k}} P_{hg}(k|n,l,j) f(j-k,k|n,l,\delta_{k_\times}).\end{aligned}$$

Since

$$h\left(\frac{pm-k_p(m)}{n}\right) = h\left(\frac{k_p(m)}{l} + \delta_k\right),$$

we have $k_p(m) = pl - \frac{nl}{m}\delta_k$. Using the relation $\delta_k = \frac{\widetilde{\epsilon}_{k/l}}{\sqrt{m}}$ and the continuity of C_p, we have

$$k_p(m) = (1-r)pm - r(1-r)\widetilde{\epsilon}(p)\sqrt{m} + o(\sqrt{m}).$$

The average of k is $\frac{lj}{n+l} = (1-r)pm$ and the variance of k is

$$\frac{jln(n+l-j)}{(n+l)^2(n+l-1)} = \frac{r(1-r)p(1-p)m}{1-1/m}.$$

Hence,

$$\frac{k_p(m)-(1-r)pm}{\sqrt{\dfrac{r(1-r)p(1-p)m}{1-\dfrac{1}{m}}}} \to -\frac{\sqrt{r(1-r)}}{\sqrt{p(1-p)}}\widetilde{\epsilon}(p).$$

Thus, we have

$$\sum_{k=0}^{k_p(m)} P_{hg}(k|n,l,j) = \Phi\left(-\frac{\sqrt{r(1-r)}}{\sqrt{p(1-p)}}\widetilde{\epsilon}(p)\right).$$

When $k \geqslant k_p(m)$, we can approximate the difference as

$$h\left(\frac{j-k}{n}\right) - h\left(\frac{k}{l}+\delta_k\right) \cong -h'(p)\frac{l+n}{\ln}[k - k_p(m)].$$

Hence,

$$\begin{aligned}&\sum_{k=k_p(m)+1}^{\overline{k}} P_{hg}(k|n,l,j) f(j-k,k|n,l,\delta_{k_\times}) \\ &\cong \frac{1}{\sqrt{2\pi\dfrac{r(1-r)p(1-p)m}{1-\dfrac{1}{m}}}} \int_{k_p(m)}^{\overline{k}} \exp\left(-\frac{(x-(1-r)pm)^2}{2\dfrac{r(1-r)p(1-p)m}{1-\dfrac{1}{m}}}\right) 2^{-nh'(p)[(l+n)/\ln][x-k_p(m)]} dx \\ &\cong \frac{1}{2\pi}\int_{-[\sqrt{r(1-r)}/\sqrt{p(1-p)}]\widetilde{\epsilon}(p)}^{+\infty} e^{-x^2/2} 2^{-\sqrt{m}h'(p)[\sqrt{rp(1-p)}/(\sqrt{1-r})\{y+[\sqrt{r(1-r)}/\sqrt{p(1-p)}]\widetilde{\epsilon}(p)\}} \\ &\quad \times \sqrt{r(1-r)p(1-p)m}\,dy \to 0 \quad \text{as } m \to \infty,\end{aligned}$$

where

$$y = \frac{x-(1-r)pm}{\sqrt{r(1-r)p(1-p)m}}.$$

Next, we consider the case when $\frac{j}{m}$ is strictly smaller than $\underline{p}$. The value $-h(\frac{j-k}{n})+h(\frac{k}{l}+\delta_{\underline{k}})$ is strictly positive and $-h(\frac{j-k}{n})+h(\frac{k}{l}+\delta_{k})$ is smaller than this value if $k \geqslant \underline{k}$. Hence, $\tilde{P}_2(\delta,n,l,\underline{k},\bar{k})$ goes to 0.

Finally, we consider the case when $\frac{j}{m}$ is strictly greater than $\bar{p}$. In this case, as is mentioned in Appendix B, the probability that k is greater than $\bar{k}$ exponentially goes to 0. Hence, in this case $\tilde{P}(\delta,n,l,\underline{k},\bar{k})$ goes to 0. Therefore, we obtain (3).

APPENDIX B: DERIVATION OF (4)

From (9), we have

$$\frac{1}{(n+1)(l+1)}2^{lh(k/l)+nh[(j-k/n)]-(n+l)h[(j)/(n+l)]} \leqslant P_{hg}(k|n,l,j) = \frac{\binom{l}{k}\binom{n}{j-k}}{\binom{n+l}{j}} \leqslant (n+l+1)2^{lh(k/l)+nh[(j-k)/n]-(n+l)h[j/(n+l)]}.$$

Hence,

$$\begin{aligned}&\max_j \sum_{k=0}^{\underline{k}} \left[P_{hg}(k|n,l,j)f(j-\underline{k},\underline{k}|n,l,\delta_{k_\times}) + \sum_{k=\underline{k}+1}^{\bar{k}} P_{hg}(k|n,l,j)f(j-k,k|n,l,\delta_{k_\times}) \right] \\ &\leqslant \bar{k}(n+l+1)2^{\max_{j,k} lh(k/l)+nh[(j-k)/n]-(n+l)h[j/(n+l)]-n\{h(k/l\,+\,\delta_k)\,-\,h[(j\,-\,k)/n]\}_+}.\end{aligned}$$

Further,

$$\max_j \left[\sum_{k=0}^{\underline{k}} P_{hg}(k|n,l,j)f(j-\underline{k},\underline{k}|n,l,\delta_{k_\times})n\left[R-h\left(\frac{\underline{k}}{l}+\delta_{k_\times}\right)\right] + \sum_{k=\underline{k}+1}^{\bar{k}} P_{hg}(k|n,l,j)f(j-k,k|n,l,\delta_{k_\times})n\left[R-h\left(\frac{k}{l}+\delta_{k_\times}\right)\right]\right],$$

$$\leqslant n\{R-h[\underline{p}+\epsilon(\underline{p})]\}\bar{k}(n+l+1)2^{\max_{j,k} lh(k/l)+nh[(j-k)/n]-(n+l)h[j/(n+l)]-n\{h(k/l\,+\,\delta_k)\,-\,h[(j\,-\,k)/n]\}_+}.$$

Thus, substituting $p=\frac{k}{l}$, $r=\frac{n}{n+l}$, $\epsilon(p)=\delta_k$, and $\epsilon'=\frac{k}{l}-\delta_k-\frac{j-k}{n}$, we obtain Eq. (5). Since

$$\frac{-r}{n}\max_{j,k}\left\{lh\left(\frac{k}{l}\right)+nh\left(\frac{j-k}{n}\right)-(n+l)h\left(\frac{j}{n+l}\right)-n\left[h\left(\frac{k}{l}+\delta_k\right)-h\left(\frac{j-k}{n}\right)\right]_+\right\} \leqslant E(\epsilon,r,\underline{p},\bar{p}), \tag{B1}$$

we obtain the part $\leqslant$ in (4).

Conversely,

$$\begin{aligned}&\max_j \left[\sum_{k=0}^{\underline{k}} P_{hg}(k|n,l,j)f(j-\underline{k},\underline{k}|n,l,\delta_{k_\times}) + \sum_{k=\underline{k}+1}^{\bar{k}} P_{hg}(k|n,l,j)f(j-k,k|n,l,\delta_{k_\times}) \right] \\ &\geqslant \frac{2^{\max_{j,k} lh(k/l)+nh[(j-k)/n]-(n+l)h[j/(n+l)]-n\{h(k/l\,+\,\delta_k)\,-\,h[(j\,-\,k)/n]\}_+}}{(n+1)(l+1)}.\end{aligned}$$

Since the equality in (B1) holds in the limit $n \to \infty$, we obtain the part $\geqslant$ in Eq. (4).

APPENDIX C: PROOF OF (15) and (16)

When Alice sends the classical information $x+X_+$ $[x=G(C_1)Z]$, the probability that Bob obtains the local signal $x_b := x + X_+ - \tilde{X}_+$ is

$$
\begin{aligned}
&\frac{\mathrm{Tr}\frac{1}{2^{n+l}}\sum_{x_k'\in\mathbf{F}_2^n}\sum_{z_c'\in\mathbf{F}_2^l}\Lambda^{\mathrm{pos}}(|x_k'\rangle\langle x_k'|\otimes|z_c'\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle z_c'|)|x_k'+x-x_b\rangle\langle x_k'+x-x_b|\otimes|z_c'-z\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle z_c'-z|}{\mathrm{Tr}\frac{1}{2^l}\sum_{x_c'\in\mathbf{F}_2^l}\Lambda^{\mathrm{pos}}(\rho_{\mathrm{mix},n}\otimes|z_c'\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle z_c'|)I\otimes|z_c'-z\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle z_c'-z|}\\
&=\frac{\mathrm{Tr}\frac{1}{2^{n+l}}\sum_{x_k''\in\mathbf{F}_2^n}\sum_{z_c'\in\mathbf{F}_2^l}\Lambda^{\mathrm{pos}}(|x_k''-x_b\rangle\langle x_k''-x_b|\otimes|z_c'\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle z_c'|)|x_k''-x_b\rangle\langle x_k''-x_b|\otimes|z_c'-z\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle z_c'-z|}{\mathrm{Tr}\frac{1}{2^l}\sum_{x_c'\in\mathbf{F}_2^l}\Lambda^{\mathrm{pos}}(\rho_{\mathrm{mix},n}\otimes|z_c'\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle z_c'|)I\otimes|z_c'-z\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle z_c'-z|}\\
&=\frac{\mathrm{Tr}\frac{1}{2^{n+l}}\sum_{x_k''\in\mathbf{F}_2^n}\sum_{z_c'\in\mathbf{F}_2^l}(\Lambda^{\mathrm{pos}})^{(x_k''0,0z_c')}(|-x_b\rangle\langle -x_b|\otimes|0\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle 0|)|-x_b\rangle\langle -x_b|\otimes|-z\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle -z|}{\mathrm{Tr}\frac{1}{2^l}\sum_{x_c'\in\mathbf{F}_2^l}\Lambda^{\mathrm{pos}}(\rho_{\mathrm{mix},n}\otimes|0\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle 0|)I\otimes|-z\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle -z|}\\
&=\frac{\mathrm{Tr}\frac{1}{2^{2(n+l)}}\sum_{x'',z''\in\mathbf{F}_2^{n+l}}(\Lambda^{\mathrm{pos}})^{(x'',z'')}(|-x_b\rangle\langle -x_b|\otimes|0\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}|0\rangle_{\mathbf{F}_2^n}\langle 0|)|-x_b\rangle\langle -x_b|\otimes|-z\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle -z|}{\mathrm{Tr}\frac{1}{2^{2(n+l)}}\sum_{x'',z''\in\mathbf{F}_2^{n+l}}(\Lambda^{\mathrm{pos}})^{(x'',z'')}(\rho_{\mathrm{mix},n}\otimes|0\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle 0|)I\otimes|-z\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle -z|}\\
&=\frac{\mathrm{Tr}(\Lambda_t)^{\mathrm{pos}}(|-x\rangle\langle -x|\otimes|0\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle 0|)|-x_b\rangle\langle -x_b|\otimes|-z\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle -z|}{\mathrm{Tr}(\Lambda_t)^{\mathrm{pos}}(\rho_{\mathrm{mix},n}\otimes|0\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle 0|)I\otimes|-z\rangle_{\mathbf{F}_2^n\mathbf{F}_2^n}\langle -z|}=\mathrm{Tr}(\Lambda_t)^{\mathrm{pos},z}(|-x\rangle\langle -x|)|-x_b\rangle\langle -x_b|, \qquad \text{(C1)}
\end{aligned}
$$

where

$$
(\Lambda_t)^{\mathrm{pos},z_c}(\rho):=\sum_{x_k,z_k\in\mathbf{F}_2^n}P_{(\Lambda_t)^{\mathrm{pos}},k|Z,c}(x_k,z_k|z_c)\mathbf{X}^{x_k}\mathbf{Z}^{z_k}\rho(\mathbf{X}^{x_k}\mathbf{Z}^{z_k})^{\dagger}.
$$

In the derivation of (C1), we use (13).

In this case, we can regard that Bob measures the state $(\Lambda_t)^{\mathrm{pos},z}(|-x\rangle\langle -x|)$. Hence, Eve's state can be regarded as $((\Lambda_t)^{\mathrm{pos},z})_E(|-x\rangle\langle -x|)$. Hence, applying theorem 3, we obtain (15) and (16).

APPENDIX D: PROOF OF (17) and (18)

First, we evaluate $\mathrm{E}_{\mathrm{pos}}\mathrm{E}_{z_c}\mathrm{E}_{Y_+}[I([z]\in C_1/C_2(Y_+,nh(|z_c|/l+\delta_{|z_c|})),\rho^{C_1/C_2(Y)}_{(\Lambda_t)^{\mathrm{pos},z},E}([z]))]$ as

$$
\begin{aligned}
&\mathrm{E}_{\mathrm{pos}}\mathrm{E}_{z_c}\mathrm{E}_{Y_+}[I([z]\in C_1/C_2(Y_+,nh(|z_c|/l+\delta_{|z_c|})),\rho^{C_1/C_2(Y)}_{(\Lambda_t)^{\mathrm{pos},z},E}([z]))]\\
&\leqslant \mathrm{E}_{\mathrm{pos}}\Bigg[\sum_{|z_c|<\underline{k}}P_{(\Lambda_t)^{\mathrm{pos}},Z,c}(z_c)\bar{h}\Bigg(\sum_{k_k=0}^{n}\tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z}(k_k|z_c)f(k_k,\underline{k}|n,l,\delta_{\underline{k}})\Bigg)\\
&\qquad+\sum_{\underline{k}\leqslant|z_c|\leqslant\bar{k}}P_{(\Lambda_t)^{\mathrm{pos}},Z,c}(z_c)\bar{h}\Bigg(\sum_{k_k=0}^{n}\tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z}(k_k|z_c)f(k_k,|z_c||n,l,\delta_{|z_c|})\Bigg)\Bigg]\\
&\quad+\mathrm{E}_{\mathrm{pos}}\Bigg[\sum_{|z_c|<\underline{k}}P_{(\Lambda_t)^{\mathrm{pos}},Z,c}(z_c)n[R-h(\underline{k}/l+\delta_{\underline{k}})]\sum_{k_k=0}^{n}\tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z}(k_k|z_c)f(k_k,\underline{k}|n,l,\delta_k)\\
&\qquad+\sum_{\underline{k}\leqslant|z_c|\leqslant\bar{k}}P_{(\Lambda_t)^{\mathrm{pos}},Z,c}(z_c)n[R-h(|z_c|/l+\delta_{|z_c|})]\sum_{k_k=0}^{n}\tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z}(k_k|z_c)f(k_k,|z_c||n,l,\delta_{|z_c|})\Bigg] \qquad \text{(D1)}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \bar{h}\Bigg(\mathrm{E}_{\mathrm{pos}}\Bigg[\sum_{|z_c|<\underline{k}} P_{(\Lambda_t)^{\mathrm{pos}},Z,c}(z_c) \sum_{k_k=0}^{n} \tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z}(k_k|z_c) f(k_k,\underline{k}|n,l,\delta_{\underline{k}}) \\
&\qquad + \sum_{\underline{k}\leq |z_c| \leq \bar{k}} P_{(\Lambda_t)^{\mathrm{pos}},Z,c}(z_c) \sum_{k_k=0}^{n} \tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z}(k_k|z_c) f(k_k,|z_c||n,l,\delta_{|z_c|}) \Bigg]\Bigg) \\
&\quad + \mathrm{E}_{\mathrm{pos}}\Bigg[\sum_{|z_c|<\underline{k}} P_{(\Lambda_t)^{\mathrm{pos}},Z,c}(z_c) n[R - h(\underline{k}/l + \delta_{\underline{k}})] \sum_{k_k=0}^{n} \tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z}(k_k|z_c) f(k_k,\underline{k}|n,l,\delta_k) \\
&\qquad + \sum_{\underline{k}\leq |z_c| \leq \bar{k}} P_{(\Lambda_t)^{\mathrm{pos}},Z,c}(z_c) n[R - h(|z_c|/l + \delta_{|z_c|})] \sum_{k_k=0}^{n} \tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z}(k_k|z_c) f(k_k,|z_c||n,l,\delta_k) \Bigg] \qquad \text{(D2)}
\end{aligned}
$$

$$
\begin{aligned}
&= \bar{h}\left(\sum_{k_k=0}^{n} \sum_{k_c=0}^{\underline{k}} \mathrm{E}_{\mathrm{pos}}[\tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z,k,c}(k_k,k_c) f(k_k,\underline{k}|n,l,\delta_{\underline{k}})] + \sum_{k_k=0}^{n} \sum_{k_c=\underline{k}+1}^{\bar{k}} \mathrm{E}_{\mathrm{pos}}[\tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z,k,c}(k_k,k_c) f(k_k,k_c|n,l,\delta_{k_c})] \right) \\
&\quad + \sum_{k_k=0}^{n} \sum_{k_c=0}^{\underline{k}} \mathrm{E}_{\mathrm{pos}}\{\tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z,k,c}(k_k,k_c) n[R - h(\underline{k}/l + \delta_{\underline{k}})] f(k_k,\underline{k}|n,l,\delta_{\underline{k}})\} \\
&\quad + \sum_{k_k=0}^{n} \sum_{k_c=\underline{k}+1}^{\bar{k}} \mathrm{E}_{\mathrm{pos}}\{\tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z,k,c}(k_k,k_c) n[R - h(k_c/l + \delta_{k_c})] f(k_k,k_c|n,l,\delta_{k+c})\}. \qquad \text{(D3)}
\end{aligned}
$$

Further, the RHS of (D3) is evaluated as

$$
\begin{aligned}
[\text{RHS of (D3)}] &= \bar{h}\left(\sum_{k_k=0}^{n} \sum_{k_c=0}^{\underline{k}} \tilde{P}_{(\Lambda_t),Z}(k_k+k_c) P_{hg}(k_c|n,l,k_k+k_c) f(k_k,\underline{k}|n,l,\delta_{\underline{k}}) \right. \\
&\qquad \left. + \sum_{k_k=0}^{n} \sum_{k_c=\underline{k}+1}^{\bar{k}} \tilde{P}_{(\Lambda_t),Z}(k_k+k_c) P_{hg}(k_c|n,l,k_k+k_c) f(k_k,k_c|n,l,\delta_{k_c}) \right) \\
&\qquad + \sum_{k_k=0}^{n} \sum_{k_c=0}^{\underline{k}} \tilde{P}_{(\Lambda_t),Z}(k_k+k_c) P_{hg}(k_c|n,l,k_k+k_c) n[R - h(\underline{k}/l + \delta_{\underline{k}})] f(k_k,\underline{k}|n,l,\delta_{\underline{k}}) \\
&\qquad + \sum_{k_k=0}^{n} \sum_{k_c=\underline{k}+1}^{\bar{k}} \tilde{P}_{(\Lambda_t),Z}(k_k+k_c) P_{hg}(k_c|n,l,k_k+k_c) n[R - h(k_c/l + \delta_{k_c})] f(k_k,k_c|n,l,\delta_{k+c}) \qquad \text{(D4)}
\end{aligned}
$$

$$
\begin{aligned}
&\leq \bar{h}\left(\max_j \left[\sum_{k_c=0}^{\underline{k}} P_{hg}(k_c|n,l,j) f(k_k,\underline{k}|n,l,\delta_{\underline{k}}) + \sum_{k_c=\underline{k}+1}^{\bar{k}} P_{hg}(k_c|n,l,j) f(k_k,k_c|n,l,\delta_{k_c}) \right] \right) \\
&\quad + \max_j \left[\sum_{k_c=0}^{\underline{k}} P_{hg}(k_c|n,l,j) n[R - h(\underline{k}/l + \delta_{\underline{k}})] f(k_k,\underline{k}|n,l,\delta_{\underline{k}}) \right. \\
&\qquad \left. + \sum_{k_c=\underline{k}+1}^{\bar{k}} P_{hg}(k_c|n,l,j) n[R - h(k_c/l + \delta_{k_c})] f(k_k,k_c|n,l,\delta_{k+c}) \right]. \qquad \text{(D5)}
\end{aligned}
$$

In the above relations, (D1) follows from (15), (16), and (D2) follows from the convexity of $\bar{h}$, (D4) follows from (14) and (D5) follows by replacing k_k+k_c by j. Hence, we obtain (17).

Similarly, we have

$$
\begin{aligned}
&\mathrm{E}_{\mathrm{pos}}\mathrm{E}_{z_c}\mathrm{E}_{Y_+}\left[\frac{I([z]\in C_1/C_2(Y_+,nh(|z_c|/l+\delta_{|z_c|})),\rho^{C_1/C_2(Y)}_{(\Lambda_t)^{\mathrm{pos}},z,E}([z]))}{n[R-h(k_\times/l_\times+\delta_{k_\times})]}\right]\\
&\leqslant \frac{1}{n[R-h(\bar{k}/l_\times+\delta_{\bar{k}})]}\mathrm{E}_{\mathrm{pos}}\Bigg[\sum_{|z_c|<\underline{k}} P_{(\Lambda_t)^{\mathrm{pos}},Z,c}(z_c)\bar{h}\left(\sum_{k_k=0}^{n}\tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z}(k_k|z_c)f(k_k,\underline{k}|n,l,\delta_{\underline{k}})\right)\\
&\qquad\qquad + \sum_{\underline{k}\leqslant|z_c|\leqslant\bar{k}} P_{(\Lambda_t)^{\mathrm{pos}},Z,c}(z_c)\bar{h}\left(\sum_{k_k=0}^{n}\tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z}(k_k|z_c)f(k_k,|z_c||n,l,\delta_{|z_c|})\right)\Bigg]\\
&\quad + \mathrm{E}_{\mathrm{pos}}\left[\sum_{|z_c|<\underline{k}} P_{(\Lambda_t)^{\mathrm{pos}},Z,c}(z_c)\sum_{k_k=0}^{n}\tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z}(k_k|z_c)f(k_k,\underline{k}|n,l,\delta_{k}) + \sum_{\underline{k}\leqslant|z_c|\leqslant\bar{k}} P_{(\Lambda_t)^{\mathrm{pos}},Z,c}(z_c)\sum_{k_k=0}^{n}\tilde{P}_{(\Lambda_t)^{\mathrm{pos}},Z}(k_k|z_c)f(k_k,|z_c||n,l,\delta_{|z_c|})\right]\\
&\leqslant \frac{1}{n[R-h(\bar{k}/l_\times+\delta_{\bar{k}})]}\bar{h}\left(\max_j\left[\sum_{k_c=0}^{\underline{k}} P_{hg}(k_c|n,l,j)f(k_k,\underline{k}|n,l,\delta_{\underline{k}}) + \sum_{k_c=\underline{k}+1}^{\bar{k}} P_{hg}(k_c|n,l,j)f(k_k,k_c|n,l,\delta_{k_c})\right]\right)\\
&\quad + \max_j\left[\sum_{k_c=0}^{\underline{k}} P_{hg}(k_c|n,l,j)f(k_k,\underline{k}|n,l,\delta_{\underline{k}}) + \sum_{k_c=\underline{k}+1}^{\bar{k}} P_{hg}(k_c|n,l,j)f(k_k,k_c|n,l,\delta_{k+c})\right].
\end{aligned}
$$

Hence, we obtain (18).

[1] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India, 1984* (IEEE, New York, 1984), p. 175.

[2] D. Mayers, in *Advances in Cryptology—Proceedings of Crypto'96*, Vol. 1109 of *Lecture Notes in Computer Science*, edited by N. Koblitz (Springer-Verlag, New York, 1996), p. 343; J. ACM **48** 351 (2001).

[3] H. Inamori, N. Lütkenhaus, and D. Mayers, e-print quant-ph/0107017.

[4] P. W. Shor and J. Preskill, Phys. Rev. Lett. **85**, 441 (2000).

[5] A. R. Calderbank and P. W. Shor, Phys. Rev. A **54**, 1098 (1996).

[6] M. Steane, Proc. R. Soc. London, Ser. A **452**, 2551 (1996).

[7] M. Hamada, J. Phys. A **37** (2004).

[8] M. Christandl, R. Renner, and A. Ekert, e-print quant-ph/0402131.

[9] R. Renner, N. Gisin, and B. Kraus, Phys. Rev. A **72**, 012332 (2005).

[10] M. Koashi, e-print quant-ph/0505108.

[11] S. Watanabe, R. Matsumoto, and T. Uyematsu, e-print quant-ph/0412070.

[12] R. Renner, e-print quant-ph/0512258.

[13] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden, New J. Phys. **4**, 41 (2002).

[14] T. Kimura, Y. Nambu, T. Hatanaka, A. Tomita, H. Kosaka, and K. Nakamura, Jpn. J. Appl. Phys., Part 2 **43**, L1217 (2004).

[15] C. Gobby, Z. L. Yuan, and A. J. Shields, Appl. Phys. Lett. **84**, 3762 (2004).

[16] A. Tanaka, W. Maeda, A. Tajima, and S. Takahashi, in *Proceedings of the 18th Annual Meeting of the IEEE Lasers and Electro-Optics Society, Sidney, Australia, 2005* (IEEE, New York, 2005), p. 557.

[17] Z. L. Yuan and A. J. Shields, Opt. Express **13**, 660 (2005).

[18] M. Koashi and J. Preskill, Phys. Rev. Lett. **90**, 057902 (2003).

[19] Y. Watanabe, W. Matsumoto, and Hideki Imai, in *Proceedings of the International Symposium on Information Theory and Its Applications, Parma, Italy, October 2004* (Sita, Tokyo, 2004), p. 1265.

[20] Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Academic, New York, 1981).

[21] H. Weyl, *Gruppentheorie und Quantenmechanik* (Verlag von S. Hirzel, Leipzig, 1928). English translation: *The Theory of Groups and Quantum Mechanics*, 2nd ed. (Dover, New York, 1950).

[22] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A **54**, 3824 (1996).

[23] M. Hamada, Phys. Rev. A **68**, 012301 (2003).

[24] W-Y. Hwang, Phys. Rev. Lett. **91**, 057901 (2003).

[25] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, Quantum Inf. Comput. **5**, 325 (2004).

[26] H.-K. Lo, in *Proceedings of the 2004 IEEE International Symposium on Information Theory, Chicago, 2004* (IEEE, New York, 2004), 17.

[27] H. K. Lo, X.-F. Ma, and K. Chen, Phys. Rev. Lett. **94**, 230504 (2005).

[28] X.-B. Wang, Phys. Rev. Lett. **94**, 230503 (2005).

[29] A similar lemma has been obtained independently by T. Miyadera and Hideki Imai, Phys. Rev. A **73**, 042317 (2006).