New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

CrossMark

OPEN ACCESS

RECEIVED
1 October 2019

REVISED
24 December 2019

ACCEPTED FOR PUBLICATION
13 January 2020

PUBLISHED
5 February 2020

PAPER

Application of an improved version of McDiarmid inequality in finite-key-length decoy-state quantum key distribution

H F Chau and K C Joseph Ng

Department of Physics, University of Hong Kong, Pokfulam Road, Hong Kong

E-mail: hfchau@hku.hk

Keywords: quantum key distribution, McDiarmid inequality, secret key rate

Abstract

In practical decoy-state quantum key distribution, the raw key length is finite. Thus, deviation of the estimated single photon yield and single photon error rate from their respective true values due to finite sample size can seriously lower the provably secure key rate *R*. Current method to obtain a lower bound of *R* follows an indirect path by first bounding the yields and error rates both conditioned on the type of decoy used. These bounds are then used to deduce the single photon yield and error rate, which in turn are used to calculate a lower bound of the key rate *R*. Here we report an improved version of McDiarmid inequality in statistics and show how use it to directly compute a lower bound of *R* via the so-called centering sequence. A novelty in this work is the optimization of the bound through the freedom of choosing possible centering sequences. The provably secure key rate of realistic 100 km long quantum channel obtained by our method is at least twice that of the state-of-the-art procedure when the raw key length ℓ_{raw} is $\approx 10^5$–10^6. In fact, our method can improve the key rate significantly over a wide range of raw key length from about 10^5 to 10^{11}. More importantly, it is achieved by pure theoretical analysis without altering the experimental setup or the post-processing method. In a boarder context, this work introduces powerful concentration inequality techniques in statistics to tackle physics problem beyond straightforward statistical data analysis especially when the data are correlated so that tools like the central limit theorem are not applicable.

1. Introduction

Quantum key distribution (QKD) enables two trusted parties Alice and Bob to share a provably secure secret key by preparing and measuring quantum states that are transmitted through a noisy channel controlled by an eavesdropper Eve. One of the major challenges to make QKD practical is to increase the number of secure bits generated per second [1]. That is why most QKD experiments to date use photons as the quantum information carriers; and these photons come from phase randomize Poissonian distributed sources instead of the much less efficient single photon sources. In addition, decoy state method is used to combat Eve's photon-number-splitting attack on multiple photon events emitted from the Poissonian sources [2, 3]. From the theoretical point of view, a more convenient figure of merit is the key rate, namely, the number of provably secure secret bits per average number of photon pulses prepared by Alice. This is because key rate measures the intrinsic performance of a QKD protocol (in other words, the software issue) without taking the frequency of the pulse (which is a hardware issue) into account. This is analogous to the use of time complexity measure rather than the actual runtime to gauge the performance of an algorithm in theoretical computer science.

Surely, provably secure lower bound of key rate *R* (which we simply call the key rate from now on) of a QKD scheme depends on various photon yields as well as error rates of those detected photons to be precisely defined in equations (1) and (2) below. The problem is that Alice and Bob can only transmit a finite number of photons in practice. Consequently, the yield and error rates estimated by any sampling technique may differ from their actual values. If Alice and Bob ignore these deviations, the actual number of bits of secret key they get could be smaller than that computed by the key rate *R*, posing a security threat.

Various key rate formulae which take the above finite-size statistical fluctuations into account for a few (decoy-state-based) QKD schemes had been reported in literature. For instance, Lim *et al* [4] computed the key rates of a certain implementation of the BB84 QKD scheme [5] using three types of decoy; recently, Chau [6] extended it to the case of using more than three types of decoys. Hayashi and Nakayama investigated the key rate for the BB84 scheme [7]. Brádler *et al* showed the key rate for a qudit-based QKD scheme using up to three mutually unbiased preparation and measurement bases [8]. And Wang *et al* proved that errors and fluctuations in the decoy photon intensities only have minor errors on the final key rate [9]. In brief, the provably secure key rate of a QKD scheme so far is found using the following three-step strategy. First, the yields $Q_{\mathtt{B},\mu_n}$ and error rates $E_{\mathtt{B},\mu_n}$ conditioned on the preparation and measurement basis $\mathtt{B}$ as well as the photon intensity parameter μ_n used are determined by comparing the relevant Bob's measurement outcomes, if any, with Alice's preparation states. The second step is to deduce yields and error rates conditioned on the number of photons emitted by the source. For a phase randomized Poissonian photon source,

$$Q_{\mathtt{B},\mu_n} = \sum_{m=0}^{+\infty} \frac{\mu_n^m Y_{\mathtt{B},m} \exp(-\mu_n)}{m!} \tag{1}$$

and

$$Q_{\mathtt{B},\mu_n} E_{\mathtt{B},\mu_n} = \sum_{m=0}^{+\infty} \frac{\mu_n^m Y_{\mathtt{B},m} e_{\mathtt{B},m} \exp(-\mu_n)}{m!}. \tag{2}$$

Here, $\mu_1 > \mu_2 > \cdots > \mu_k \geqslant 0$ are the photon intensities used in the decoy method with $k \geqslant 2$. Moreover, $Y_{\mathtt{B},m}$ is the probability of photon detection by Bob given that the photon pulse sent by Alice contains m photons and $e_{\mathtt{B},m}$ is the bit error rate for m photon emission events prepared in the $\mathtt{B}$ basis [2, 3, 10]. The key rate R depends on $Y_{\mathtt{B},0}$, $Y_{\mathtt{B},1}$ and $e_{\mathtt{B},1}$ [2–4, 10]. Nevertheless, the later quantities cannot be determined precisely because equations (1) and (2) are under-determined systems of equations given $Q_{\mathtt{B},\mu_n}$'s and $E_{\mathtt{B},\mu_n}$'s provided that the number of photon intensities used k is finite. To make things worse, in the finite-raw-key-length (FRKL) situation, the measured values of $Q_{\mathtt{B},\mu_n}$'s and $E_{\mathtt{B},\mu_n}$'s deviate from their true values due to finite sampling. Fortunately, effective lower bounds of $Y_{\mathtt{B},0}$ and $Y_{\mathtt{B},1}$ as well as upper bound of $e_{\mathtt{B},1}$ are available [2–4, 6, 10, 11]. In the FRKL situation, these bounds can be deduced with the help of Hoeffding's inequality [12]. (See, for example [4, 6], for details. Note that here we cannot assume the measurement outcomes are statistically independent and thus use more familiar tools such as central limit theorem because Eve may launch a coherent attack to all the photon pulses. In fact, we do not even know what kind of statistical distributions do $Q_{\mathtt{B},\mu_n}$'s and $E_{\mathtt{B},\mu_n}$'s follow.) The third step is to deduce R from these bounds [2–4, 8, 10].

Computing lower bound of R using this indirect strategy is not satisfactory in the FRKL situation because it is unlikely for each of the finite-size fluctuations in $Q_{\mathtt{B},\mu_n}$'s and $E_{\mathtt{B},\mu_n}$'s to decrease the value of the provably secure key rate. In fact, for a given security parameter, the worst case bounds on $Y_{\mathtt{B},0}$ and $Y_{\mathtt{B},1}$ cannot be not attained simultaneously if the raw key length is finite. (This is evident, say, from the bounds of $Y_{\mathtt{B},0}$ and $Y_{\mathtt{B},1}$ given by inequalities (2) and (3) in [4] or inequalities (12a) and (12b) in [6]. Note that there is a typo in inequality (12b) —the $Q_{\mathtt{B},\mu_i}^{\langle\langle k_0-i\rangle\rangle}$ there should be $Q_{\mathtt{B},\mu_i}^{\langle\langle k_0-i+1\rangle\rangle}$. In all cases, the finite-size statistical fluctuation that leads to the saturation of lower bound for $Y_{\mathtt{B},0}$ does not cause the saturation of the lower bound for $Y_{\mathtt{B},1}$ and vice versa.)

It is more effective if one could directly investigate the influence of finite-key-length on the key rate. To do so, one has to go beyond the use of Hoeffding's inequality to bound the statistical fluctuation, which only works for equally weighted sum of random variables that are either statistical independent or drawn from a finite population without replacement [12]. Here we use the computation of the key rate of a specific BB84 QKD protocol [5] that generates the raw key solely from $\mathtt{X}$ basis measurement results as an example to illustrate how to directly tackle statistical fluctuation in the FRKL situation by means of McDiarmid-type inequality [13] in statistics. The technique used here can be easily adapted to compute the key rates of other QKD schemes using finite-dimensional qudits in the FRKL situation. Our work here is based on an earlier preprint by one of the us [14]. Here we greatly extend and improve the original proposal by first proving a new and slightly extended McDiarmid-type of inequality on so-called centering sequences. (See definition 1 for the precise definition of a centering sequence.) Then we apply it through four different methods, each giving a separate provably secure key rate. We also optimize the provably secure key rate R by exploiting our freedom to pick the centering sequences. To our knowledge, this is the first time such an optimization is performed. In contrast, this type of optimization is not possible in previous approach that makes use of a less general inequality known as Hoeffding's inequality. It turns out that each method works best in different situations; and the best provably secure key rate among the four methods in realistic practical situation is at least about 10% better than the state-of-the-art method before [14]. Moreover, for raw key length $\ell_{\text{raw}} \approx 10^5$–$10^7$, this work almost double the secure key rate of the original proposal in [14] when four different photon intensities are used. From a broader perspective, the technique we introduce here is also applicable to bound the conclusion of a general physics experiment in the form of a real number due to finite-size statistical fluctuations of more than one type of measurement outcomes that are possibly statistically dependent.

2. The QKD scheme by Chau in [6] and the assumptions of the security proof

To illustrate how McDiarmid-type of inequality can be used to give a better key rate, we consider the QKD Scheme studied by Chau whose details can be found in [6]. Note that this scheme is a slight variation of the one studied by Lim *et al* in [4]. The only difference is that they use three different photon intensities while we consider the slightly more general case of using $k \geqslant 2$ different photon intensities. In essence, the Scheme in [6] is a decoy-state BB84 scheme with one-way classical communication using the $\mathtt{X}$-basis measurement results as the raw key and the $\mathtt{Z}$-basis measurement results for phase error estimation.

We assume that the light source is Poissonian distributed with intensities $\mu_1 > \mu_2 > ... > \mu_k \geqslant 0$ with $k \geqslant 2$. Using the result in [9], we simply our discussion by assuming that these photon intensities are accurately determined and fixed throughout the experiment. This is fine because fluctuation of photon intensity of a laser source is negligible in practice. Since our aim is to demonstrate our technique of using McDiarmid-type inequality in the simplest possible QKD implementation, we do not consider twin-field [15] or measurement device independent [16] setups although adaptation to these situations is straightforward though tedious. The measurement is performed using threshold photon detectors with random bit assignment in the event of multiple detector click. Last but not least, we assume both Alice and Bob have access to their own private perfect random number generators when choosing their preparation and measurement bases.

3. Finite-size decoy-state key rate

Recall that the error rate for this particular variation of the decoy-state BB84 QKD scheme using one-way classical communication is lower-bounded by [4, 6]

$$p_{\mathtt{X}}^2\left\{\langle\exp(-\mu)\rangle Y_{\mathtt{X},0} + \langle\mu\exp(-\mu)\rangle Y_{\mathtt{X},1}[1 - H_2(e_p)] - \Lambda_{\mathrm{EC}} - \frac{\langle Q_{\mathtt{X},\mu}\rangle}{\ell_{\mathrm{raw}}}\left[6\log_2\frac{\chi}{\epsilon_{\mathrm{sec}}} + \log_2\frac{2}{\epsilon_{\mathrm{cor}}}\right]\right\}, \tag{3}$$

where $p_{\mathtt{X}}$ denotes the probability that Alice (Bob) uses $\mathtt{X}$ as the preparation (measurement) basis, $\langle f(\mu)\rangle \equiv \sum_{n=1}^{k} p_{\mu_n} f(\mu_n)$ with p_{μ_n} being the probability for Alice to use photon intensity parameter μ_n. Furthermore, $H_2(x) \equiv -x\log_2 x - (1 - x)\log_2(1 - x)$ is the binary entropy function, e_p is the phase error rate of the single photon events in the raw key, and Λ_{EC} is the actual number of bits of information that leaks to Eve as Alice and Bob perform error correction on their raw bits. It is given by

$$\Lambda_{\mathrm{EC}} = \langle Q_{\mathtt{X},\mu} H_2(E_{\mathtt{X},\mu})\rangle \tag{4}$$

if they use the most efficient (classical) error correcting code to do the job. In addition, ℓ_{raw} is the raw sifted key length measured in bits, ϵ_{cor} is the upper bound of the chance that the final secret keys shared between Alice and Bob are different, and $\epsilon_{\mathrm{sec}} = (1 - p_{\mathrm{abort}})\|\rho_{\mathrm{AE}} - U_{\mathrm{A}} \otimes \rho_{\mathrm{E}}\|_1/2$. Here p_{abort} is the chance that the scheme aborts without generating a key, ρ_{AE} is the classical-quantum state describing the joint state of Alice and Eve, U_{A} is the uniform mixture of all the possible raw keys created by Alice, ρ_{E} is the reduced density matrix of Eve, and $\|\cdot\|_1$ is the trace norm [17–19]. Thus, Eve's information on the final key is at most ϵ_{sec}. Last but not least, χ is a QKD scheme specific factor which depends on the detailed security analysis used. In general, χ may also depend on other factors used in the QKD scheme such as the number of photon intensities k [4, 6].

For BB84, $e_p \to e_{\mathtt{Z},1}$ as $\ell_{\mathrm{raw}} \to +\infty$. More importantly, the best known bound on the difference between e_p and $e_{\mathtt{Z},1}$ due to finite sample size correction using properties of the hypergeometric distribution reported in given by [6, 20]

$$e_p \leqslant e_{\mathtt{Z},1} + \bar{\gamma}\left(\frac{\epsilon_{\mathrm{sec}}}{\chi}, e_{\mathtt{Z},1}, \frac{s_{\mathtt{Z}} Y_{\mathtt{Z},1}\langle\mu\exp(-\mu)\rangle}{\langle Q_{\mathtt{Z},\mu}\rangle}, \frac{s_{\mathtt{X}} Y_{\mathtt{X},1}\langle\mu\exp(-\mu)\rangle}{\langle Q_{\mathtt{X},\mu}\rangle}\right) \tag{5}$$

with probability at least $1 - \epsilon_{\mathrm{sec}}/\chi$, where

$$\bar{\gamma}(a, b, c, d) \equiv \sqrt{\frac{(c + d)(1 - b)b}{cd}\ln\left[\frac{c + d}{2\pi cd(1 - b)ba^2}\right]}, \tag{6}$$

and $s_{\mathtt{B}}$ is the number of bits that are prepared and measured in $\mathtt{B}$ basis. Clearly, $s_{\mathtt{X}} = \ell_{\mathrm{raw}}$ and $s_{\mathtt{Z}} \approx (1 - p_{\mathtt{X}})^2 s_{\mathtt{X}}\langle Q_{\mathtt{Z},\mu}\rangle/(p_{\mathtt{X}}^2\langle Q_{\mathtt{X},\mu}\rangle)$. (Note that $\bar{\gamma}$ becomes complex if a, c, d are too large. This is because in this case no $e_p \geqslant e_{\mathtt{Z},1}$ exists with failure probability a. We carefully picked parameters here so that $\bar{\gamma}$ is real.)

In the infinite-key-length limit, statistical fluctuations of $Q_{\mathtt{B},\mu_n}$ and $E_{\mathtt{B},\mu_n}$ can be ignored. Then based on the analysis in [6] with typos corrected, one has

$$Y_{\mathrm{B},0} \geqslant \max\left(0, \sum_{n=1}^{k} a_{0n} Q_{\mathrm{B},\mu_n}\right) \equiv \max\left(0, \sum_{n=k_0}^{k} \frac{-Q_{\mathrm{B},\mu_n}\exp[\mu_n]\hat{\prod}_{i\neq n}\mu_i}{\hat{\prod}_{j\neq n}[\mu_n-\mu_j]}\right), \tag{7a}$$

$$Y_{\mathrm{B},1} \geqslant \max\left(0, \sum_{n=1}^{k} a_{1n} Q_{\mathrm{B},\mu_n}\right) \equiv \max\left(0, \sum_{n=3-k_0}^{k} \frac{-Q_{\mathrm{B},\mu_n}\exp[\mu_n]\hat{S}_n}{\hat{\prod}_{j\neq n}[\mu_n-\mu_j]}\right) \tag{7b}$$

and

$$Y_{\mathrm{Z},1} e_{\mathrm{Z},1} \leqslant \min\left(\frac{Y_{\mathrm{Z},1}}{2}, \sum_{n=1}^{k} a_{2n} Q_{\mathrm{Z},\mu_n} E_{\mathrm{Z},\mu_n}\right) \equiv \min\left(\frac{Y_{\mathrm{Z},1}}{2}, \sum_{n=k_0}^{k} \frac{Q_{\mathrm{Z},\mu_n}E_{\mathrm{Z},\mu_n}\exp[\mu_n]\hat{S}_n}{\hat{\prod}_{j\neq n}[\mu_n-\mu_j]}\right), \tag{7c}$$

where $k_0 = 1$ (2) if k is even (odd), and $\hat{\prod}_{j\neq n}$ is over the dummy variable j from k_0 to k but skipping n. In addition, $\hat{S}_n = \sum'' \mu_{t_1}\mu_{t_2}\cdots\mu_{t_{k-k_0-1}}$ where the double primed sum is over $k_0 \leqslant t_1 < t_2 < \cdots < t_{k-k_0-1} \leqslant k$ with $t_1, t_2, \ldots, t_{k-k_0-1} \neq n$. (In other words, $a_{01} = a_{21} = 0$ if k is odd and $a_{11} = 0$ if k is even.) Note that in our subsequent analysis, we also need the following two inequalities, which can be proven using the same method as in inequality (7b):

$$Y_{\mathrm{Z},1}\bar{e}_{\mathrm{Z},1} \equiv Y_{\mathrm{Z},1}(1-e_{\mathrm{Z},1}) \geqslant \max\left(0, \sum_{n=1}^{k} a_{1n} Q_{\mathrm{Z},\mu_n}\bar{E}_{\mathrm{Z},\mu_n}\right) \equiv \max\left(0, \sum_{n=3-k_0}^{k} \frac{-Q_{\mathrm{Z},\mu_n}\bar{E}_{\mathrm{Z},\mu_n}\exp[\mu_n]\hat{S}_n}{\hat{\prod}_{j\neq n}[\mu_n-\mu_j]}\right) \tag{7d}$$

and

$$Y_{\mathrm{Z},1} e_{\mathrm{Z},1} \geqslant \max\left(0, \sum_{n=1}^{k} a_{1n} Q_{\mathrm{Z},\mu_n} E_{\mathrm{Z},\mu_n}\right) \equiv \max\left(0, \sum_{n=3-k_0}^{k} \frac{-Q_{\mathrm{Z},\mu_n}E_{\mathrm{Z},\mu_n}\exp[\mu_n]\hat{S}_n}{\hat{\prod}_{j\neq n}[\mu_n-\mu_j]}\right), \tag{7e}$$

where $\bar{E}_{\mathrm{Z},\mu_n} = 1 - E_{\mathrm{Z},\mu_n}$.

Substituting inequalities (5) and (7) into expression (3) gives the following lower bound of the key rate

$$\sum_{n=1}^{k} b_n Q_{\mathrm{X},\mu_n} - p_{\mathrm{X}}^2\left\{\Lambda_{\mathrm{EC}} + \frac{\langle Q_{\mathrm{X},\mu}\rangle}{\ell_{\mathrm{raw}}}\left[6\log_2\frac{\chi}{\epsilon_{\mathrm{sec}}} + \log_2\frac{2}{\epsilon_{\mathrm{cor}}}\right]\right\}, \tag{8}$$

where

$$b_n = p_{\mathrm{X}}^2\{\langle\exp(-\mu)\rangle a_{0n} + \langle\mu\exp(-\mu)\rangle a_{1n}[1 - H_2(e_p)]\} \tag{9}$$

provided that $Y_{\mathrm{X},0}, Y_{\mathrm{X},1} > 0$. (The cases of $Y_{\mathrm{X},0}$ or $Y_{\mathrm{X},1} = 0$ can be dealt with in the same way by changing the definition of b_n accordingly. But these cases are not interesting for they likely imply $R = 0$ in realistic channels.)

Note that the worst case key rate corresponds to the situation that the spin flip and phase shift errors in the raw key are uncorrelated so that Alice and Bob cannot use the correlation information to increase the efficiency of entanglement distillation. Thus, we may separately consider statistical fluctuations in Q_{X,μ_n}'s and $e_{\mathrm{Z},1}$ in the FRKL situation.

4. An improved version of McDiarmid inequality

We now prove an improved version of a deep mathematical statistics result before applying it to improve the key rate R. Our insight is that statistical fluctuations in Q_{X,μ_n}'s and $e_{\mathrm{Z},1}$ can be bounded using McDiarmid-type inequality. Actually, the first inequality of this type was proven for the case of statistically independent random variables using martingale technique in [13]. The inequality we need here is a straightforward extension of theorem 6.7 in [13] and theorem 2.3 in [21] for statistically dependent random variables. (See also a closely related version in [22].)

We first introduce the concept of a centering sequence [21]. The definition below is written in a more apparent manner to physicists.

Definition 1. Let $\mathbf{W} = (W_1, W_2, \ldots, W_t)$ be a random real vector whose components W_i's are possibly statistically dependent random variables each taking values in the set $\mathcal{W}_i$. Let f_m be a real-valued bounded function of $\mathbf{W}$. Set $V_m = f_m(\mathbf{W})|_{B_m}$ where B_m denotes the conditions $W_j = w_j$ for $j = 1, 2, \ldots, m-1$. Then, the sequence of random variable $\{V_m\}_{m=1}^{t}$ is said to be *centering* if $E[U_m|V_{m-1} = v] \equiv E[V_m - V_{m-1}|V_{m-1} = v]$ is a decreasing functions of v for all $m = 1, 2, \ldots, t$. (Here we use the convention that $V_0 = 0$ and assume that all conditional expectation values $E[\cdot|\cdot]$ exist.)

Note that centering property implicitly depends on the distribution of $\mathbf{W}$ through the conditional expectation value of U_m. Moreover, $\{V_m\}$ is centering if $\{V_m\}$ is a martingale.

Theorem 1. *Using notations in definition 1, for a fixed $i = 1, 2, \ldots, t$, let $w_m \in \mathcal{W}_m$ and set*

$$\hat{r}_m(w_1, \ldots, w_{m-1}) = \operatorname{esssup}\{E[U_m(\mathbf{W})|W_m = w_m]\}_{w_m\in\mathcal{W}_m} - \operatorname{essinf}\{E[U_m(\mathbf{W})|W_m = w_m']\}_{w_m'\in\mathcal{W}_m} \equiv b_m(w_1, \ldots, w_{m-1}) - a_m(w_1, \ldots, w_{m-1}). \quad (10)$$

Here the symbols esssup *and* essinf *denote the essential supremum and infimum, respectively. Further set $\hat{r}^2 \equiv \hat{r}^2(w_1, \ldots, w_{t-1}) = \sum_{m=1}^{t}\hat{r}_m^2$. Then $f_t(\mathbf{w}) \equiv f_t(w_1, w_2, \ldots, w_t)$ obeys*

$$\Pr(f_t(\mathbf{w}) - E[f_t(\mathbf{W})] \geqslant \delta) \leqslant \exp\left[\frac{-2\delta^2}{\hat{r}^2(w_1, \ldots, w_{t-1})}\right] \quad (11a)$$

and

$$\Pr(f_t(\mathbf{w}) - E[f_t(\mathbf{W})] \leqslant -\delta) \leqslant \exp\left[\frac{-2\delta^2}{\hat{r}^2(w_1, \ldots, w_{t-1})}\right] \quad (11b)$$

for any $\delta > 0$, where $\Pr(\cdot)$ denotes the occurrence probability of the argument.

Remark 1. This version of McDiarmid inequality is slightly stronger than the one reported in [13] as we also utilize information of $\mathbf{w}$ in obtaining $\hat{r}$ whereas the original version in [13] made use of the worst case $\mathbf{w}$. The proof of this theorem is based on that of theorem 2.2 in [21].

Proof. Note that for any $h, \delta > 0$,

$$\begin{aligned}
&\Pr(V_t - E[V_t] \geqslant \delta)\\
&\leqslant E[\exp\{h(V_t - E[V_t] - \delta)\}] = e^{-h(\delta+E[V_t])}E[\exp(hV_t)] \quad \text{(by Bernstein's inequality)}\\
&= e^{-h(\delta+E[V_t])}E[\exp(hV_{t-1})E[\exp(hU_t)|V_{t-1}]]\\
&\leqslant e^{-h(\delta+E[V_t])}E\left[\exp(hV_{t-1})\left\{\frac{(b_t - E[U_t|V_{t-1}])e^{ha_t}}{b_t - a_t} + \frac{(E[U_t|V_{t-1}] - a_t)e^{hb_t}}{b_t - a_t}\right\}\right]\\
&\qquad \text{(since } a_t \leqslant E[U_t|V_{t-1}] \leqslant b_t \text{ and the line joining } (a_t, e^{ha_t}) \text{ and } (b_t, e^{hb_t}) \text{ is above the curve } y = e^{hx} \text{ for } x \in [a_t, b_t])\\
&\leqslant e^{-h(\delta+E[V_t])}E[\exp(hV_{t-1})]\left\{\frac{(b_t - E[U_t|V_{t-1}])e^{ha_t}}{b_t - a_t} + \frac{(E[U_t|V_{t-1}] - a_t)e^{hb_t}}{b_t - a_t}\right\}\\
&\qquad \text{(by Chebyshev's sum inequality on centering sequence)}\\
&\leqslant e^{-h(\delta+E[V_{t-1}])}e^{-hE[U_t|V_{t-1}]}\left\{\frac{(b_t - E[U_t|V_{t-1}])e^{ha_t}}{b_t - a_t} + \frac{(E[U_t|V_{t-1}] - a_t)e^{hb_t}}{b_t - a_t}\right\}. \quad \text{(by Jensen's inequality)}
\end{aligned} \quad (12)$$

To proceed, we consider the function $g(h) = -hx + \ln\{[(b_t - x)e^{ha_t} + (x - a_t)e^{hb_t}]/(b_t - a_t)\}$ for $x \in [a_t, b_t]$. It is straightforward to check that $g(0) = \mathrm{d}g/\mathrm{d}h|_{h=0} = 0$. Moreover,

$$\frac{\mathrm{d}^2g}{\mathrm{d}h^2} = \frac{(b_t - a_t)^2(b - x)(x - a)e^{h(b_t+a_t)}}{[(b - x)e^{ha_t} + (x - a)e^{hb_t}]^2} \leqslant \frac{(b_t - a_t)^2}{4} \quad (13)$$

with the equality holds whenever $(b - x)e^{ha_t} = (x - a)e^{hb_t}$. Therefore, Taylor's theorem gives $g(h) \leqslant h^2(b_t - a_t)^2/8$ for all $h \geqslant 0$. Applying this inequality with $x = E[U_t|V_{t-1}]$ to inequality (12), we have

$$\begin{aligned}
\Pr(f_t(\mathbf{w}) - E[f_t(\mathbf{W})] \geqslant \delta) &= \Pr(V_t - E[V_t] \geqslant \delta) \leqslant e^{-h(\delta+E[V_{t-1}])}e^{h^2(b_t-a_t)^2/8}\\
&\leqslant \exp\left[-\delta h + \frac{h^2\sum_{m=1}^{t}(b_m - a_m)^2}{8}\right] = \exp\left(\frac{h^2\hat{r}^2}{8} - \delta h\right)
\end{aligned} \quad (14)$$

for any $h > 0$. The rhs of inequality (14) is minimized by setting $h = 4\delta/\hat{r}^2$; and with this h, inequality (14) becomes inequality (11a).

Finally, by applying the same argument to $-f_m$'s instead of f_m's, we get inequality (11b). This completes our proof. □

Corollary 1. *Let* $\mathbf{W} = (W_1, \ldots, W_t)$ *be a random vector such that* W_m *takes on value from the same bounded set of real numbers* $\mathcal{W} = \{\alpha_j\}_{j=1}^{k}$ *for all* $m = 1, 2, \ldots, t$. *Suppose further that* W_m*'s are multivariate hypergeometrically distributed. Let* $f_m(\mathbf{W}) = \sum_{i=1}^{m} W_i$ *for all* $m = 1, 2, \ldots, t$. *Then, the sequence of random variables* $\{V_m\}_{m=1}^{t}$ *defined in definition 1 is centering provided that* $E[V_m - V_{m-1}|V_{m-1} = v]$ *is well-defined for all* v. *Besides, theorem 1 holds with* $\hat{r} = \sqrt{t}\,\mathrm{Width}(\mathcal{W})$, *where* $\mathrm{Width}(\mathcal{W}) \equiv \mathrm{esssup}\ \mathcal{W} - \mathrm{essinf}\ \mathcal{W}$.

Proof. This proof is adapted from example 1 in [21]. From definition 1, it suffices to show that $E[U_m|\sum_{i=1}^{m-1} W_i = v] = E[W_m|\sum_{i=1}^{m-1} W_i = v]$ is a decreasing function of v. Suppose W_i's are drawn from a collection of M objects out of which M_j of them take the value α_j for all j. Suppose further that among W_i's with $1 \leqslant i < m$, there are m_j of them taking the value of α_j for all j. Then, the probability that $W_m = \alpha_j$ is $(M_j - m_j)/(M - m + 1)$. Moreover, the condition $\sum_{i=1}^{m-1} W_i = v$ means that $\sum m_j \alpha_j = v$. As a result, $E[W_m|\sum_{i=1}^{m-1} W_i = v] = \sum_j (M_j - m_j)\alpha_j/(M - m + 1) = (\sum_j M_j \alpha_j - v)/(M - m + 1)$, which is a decreasing function of v whenever $\sum_{i=1}^{m-1} W_i = v$. Hence, $\{V_m\}$ is a centering sequence.

By applying theorem 1 to $\{U_m\}$, we have $r_m = \mathrm{esssup}\{W_m|V_{m-1} = v\} - \mathrm{essinf}\{W_m|V_{m-1} = v\} = \mathrm{esssup}\mathcal{W} - \mathrm{essinf}\ \mathcal{W} = \mathrm{Width}(\mathcal{W})$ for all m and V_{m-1}. Hence, it is proved. □

Remark 2. The above corollary was first proven by Hoeffding in [12] without using the concept of centering sequence. Actually, corollary 1 is more often referred to as the Hoeffding's inequality. In fact, Hoeffding's inequality has been used to compute the provably secure key rate R when the raw key length ℓ_{raw} is finite in previous works [4, 6–8]. In section 5 below, we use the above corollary to bound $e_{Z,1}$ in Methods A and B.

Corollary 2. *Let* $\mathbf{W} = (W_1, \ldots, W_t)$ *be a random vector where each* W_m *takes on value from a bounded set of real numbers* $\mathcal{W} = \{\alpha_j\}_{j=1}^{k}$. *Suppose* W_m*'s are multivariate hypergeometrically distributed in the sense that they are chosen without replacement from a collection of* M *objects out of which* M_j *of them take the value of* α_j *for all* j. *Let* $x \in [\mathrm{essinf}\ \mathcal{W}, \mathrm{esssup}\mathcal{W}]$ *and* $y > 0$ *be two fixed numbers. Let* $P\colon \{1, 2, \ldots, t\} \mapsto \{1, 2, \ldots, t\}$ *be an arbitrary but fixed permutation. Suppose*

$$y + \sum_{i=1}^{t} W_{P(i)} > \mathrm{esssup}\ \mathcal{W} - \mathrm{essinf}\ \mathcal{W} \geqslant 0. \tag{15}$$

Define

$$f_m(\mathbf{W}) = \frac{(t - m)x + \sum_{i=1}^{m} W_{P(i)}}{y + (t - m)x + \sum_{i=1}^{m} W_{P(i)}} \tag{16}$$

for $m = 1, 2, \ldots, t$. *Then, the sequence* $\{V_m\}_{m=1}^{t}$ *is centering provided that*

$$x \leqslant \min_{m=1}^{t} \frac{2\sum_{j=1}^{k} M_j \alpha_j - \sup \sum_{i=1}^{m-1} W_{P(i)} + y - \delta}{2M - t - m + 1} \tag{17}$$

where δ *is a small correlation term of the order of* $\mathrm{Width}(\mathcal{W})/(y + tx)^2$. *Furthermore, by picking* x *to be the rhs of inequality (17), then inequality (11) is true with*

$$\hat{r}^2 = \sum_{m=1}^{t} \left\{ \frac{y\mathrm{Width}(\mathcal{W})}{\left[y + (t - m)x + \mathrm{esssup}\mathcal{W} + \sum_{i=1}^{m-1} w_{P(i)}\right]\left[y + (t - m)x + \mathrm{essinf}\ \mathcal{W} + \sum_{i=1}^{m-1} w_{P(i)}\right]} \right\}^2, \tag{18}$$

where $\{w_{P(i)}\}$ *is a decreasing sequence.*

Proof. Since $(W_{P(1)}, W_{P(2)}, \ldots, W_{P(t)})$ is also a multivariate hypergeometrically distributed random vector, we only need to prove the case when P is an identity operator as the general case can be proven in the same way. From equation (15), f_m has a positive denominator and is an increasing function of W_m. So to prove that $\{V_m\}$ is centering, it suffices to show that $E[U_m|\sum_{i=1}^{m-1} W_i = (m - 1)w]$ is a decreasing function of $(m - 1)w = \sum_{j=1}^{k} m_j \alpha_j$ for all non-negative integers m_j's obeying $\sum_{j=1}^{k} m_j = m - 1$. Since α_j's are fixed, the only way to change w is to change m_j's but at the same time keeping $\sum_{j=1}^{n} m_j$ fixed. Clearly, w can only be changed if $m \geqslant 3$. More importantly, as m_j's are integers, any such change can be expressed as a composition of a series of elementary changes, each increases a certain m_{j_1} by one and decreasing a certain m_{j_2} by one with $1 \leqslant j_1 \neq j_2 \leqslant k$.

Observe that

$$\begin{aligned}
&E\left[U_m \middle| \sum_{i=1}^{m-1} W_i = (m-1)w = \sum_{j=1}^{k} m_j\alpha_j\right] \\
&= E\left[\frac{y(W_m - x)}{\{y + (t-m+1)x + (m-1)w\}\{y + (t-m)x + (m-1)w + W_m\}} \right. \\
&\qquad \left. \middle| \sum_{i=1}^{m-1} W_i = (m-1)w = \sum_{j=1}^{k} m_j\alpha_j\right] \\
&= \frac{y}{(M-m+1)[y + (t-m+1)x + (m-1)w]} \sum_{j=1}^{k} \frac{(M_j - m_j)(\alpha_j - x)}{y + (t-m)x + (m-1)w + \alpha_j} \\
&\equiv \frac{y}{(M-m+1)D} \sum_{j=1}^{k} \frac{(M_j - m_j)(\alpha_j - x)}{D + \alpha_j - x}.
\end{aligned} \tag{19}$$

Moreover, after the elementary change, $w \mapsto w + \alpha_{j_1} - \alpha_{j_2} \equiv w + \Delta w$. From inequality (15), $D > |\alpha_j - x|$. So by Taylor's theorem,

$$\begin{aligned}
&E\left[U_m \middle| \sum_{i=1}^{m-1} W_i = (m-1)w = \sum_{j=1}^{k} m_j\alpha_j\right] \\
&\longmapsto E\left[U_m \middle| \sum_{i=1}^{m-1} W_i = (m-1)w + \Delta w = \alpha_{j_1} - \alpha_{j_2} + \sum_{j=1}^{k} m_j\alpha_j\right] \\
&= \frac{y}{(M-m+1)(D+\Delta w)} \left\{ \sum_{j=1}^{k} \frac{(M_j - m_j)(\alpha_j - x)}{D + \Delta w} \left[1 - \frac{\alpha_j - x}{D + \Delta w} + \xi_1 \left(\frac{\alpha_j - x}{D + \Delta w}\right)^2\right] \right. \\
&\qquad \left. + \frac{-(\alpha_{j_1} - x)}{D + \Delta w + \alpha_{j_1}} + \frac{\alpha_{j_2} - x}{D + \Delta w + \alpha_{j_2}} \right\} \\
&= \frac{y}{(M-m+1)(D+\Delta w)^2} \left[\sum_{j=1}^{k} M_j\alpha_j - (M-m+1)x - (m-1)w \right. \\
&\qquad \left. - \frac{\sum_{j=1}^{n} (M_j - m_j)(\alpha_j - x)^2}{D + \Delta w} + \xi_1 \left(\frac{\alpha_j - x}{D + \Delta w}\right)^2 \right] \\
&\qquad - \frac{y\Delta w}{(M-m+1)(D + \Delta w + \alpha_{j_1} - x)(D + \Delta w + \alpha_{j_2} - x)}
\end{aligned} \tag{20}$$

with $\xi_1 \in [0, 1]$. As $x \in [\operatorname{essinf} \mathcal{W}, \operatorname{esssup} \mathcal{W}]$, we conclude that $0 \leqslant \sum_j (M_j - m_j)(\alpha_j - x)^2 \leqslant (M - m + 1)\mathrm{Width}(\mathcal{W})^2$ almost surely. From inequality (15), we may expand $1/(D + \Delta w)$, $1/(D + \Delta w + \alpha_{j_1} - x)$ and $1/(D + \Delta w + \alpha_{j_2} - x)$ as series of Δw via Taylor's theorem. In this way, the rhs of equation (20) can be expressed in the form $E[U_m | \sum_{i=1}^{m-1} W_i = (m-1)w = \sum_{j=1}^{k} m_j\alpha_j] + g_1\Delta w + g_2$ with

$$\begin{aligned}
g_1 &= -\frac{y}{(M-m+1)D^2} \left\{ \frac{2\left[\sum_{j=1}^{k} M_j\alpha_j - (M-m+1)x - (m-1)w\right]}{D} \right. \\
&\qquad \left. + \left[1 - \frac{\xi_2(\alpha_{j_1} - x)}{D}\right]\left[1 - \frac{\xi_3(\alpha_{j_2} - x)}{D}\right] - \frac{3\sum_{j=1}^{k} (M_j - m_j)(\alpha_j - x)^2}{D^2} \right\} \\
&\leqslant -\frac{y}{(M-m+1)D^2} \left\{ \frac{2\left[\sum_{j=1}^{k} M_j\alpha_j - (M-m+1)x - (m-1)w\right]}{D} \right. \\
&\qquad \left. + \left[1 - \frac{\mathrm{Width}(\mathcal{W})}{D}\right]^2 - \frac{3(M-m+1)\mathrm{Width}(\mathcal{W})^2}{D^2} \right\},
\end{aligned} \tag{21}$$

where $\xi_2, \xi_3 \in [0, 1]$. And the correlation term g_2 obeys $|g_2| \leqslant 3y[\sum_j M_j\alpha_j - (M - m + 1)x - (m - 1)w](\Delta w)^2/[(M - m + 1)D^4]$.

A sufficient condition for $\{V_m\}$ to be centering is $g_1 \Delta w + g_2 \leqslant 0$ for all m and w. Moreover, this condition is satisfied if

$$x \leqslant \frac{2\sum_{j=1}^{k} M_j \alpha_j - (m-1)w + y - \delta}{2M - t - m + 1} \tag{22}$$

for all $m = 1, \ldots, t$ and for all $(m-1)w = \sum_{i=1}^{m-1} W_i$, where the correlation term $\delta \leqslant |g_2| \Delta w + 2\mathrm{Width}(\mathcal{W}) + (3M - 3m + 2)\mathrm{Width}(\mathcal{W})^2/D$. (Note that inequality (22) is consistent with the constraint that $\operatorname{essinf} \mathcal{W} \leqslant x \leqslant \operatorname{esssup} \mathcal{W}$ because this inequality is trivially satisfied when $x = \operatorname{essinf} \mathcal{W}$.) Hence, $\{V_m\}$ is centering if inequality (17) holds.

We now switch back to consider the situation of an arbitrary but fixed permutation P. To optimize the bound in theorem 1, we use the freedom to pick a suitable permutation P to minimize $\hat{r}$. From theorem 1, $\hat{r}_m = y\ \mathrm{Width}(\mathcal{W})/\{[y + (t-m)x + \operatorname{esssup} \mathcal{W} + \sum_{i=1}^{m-1} w_{P(i)}][y + (t-m)x + \operatorname{essinf} \mathcal{W} + \sum_{i=1}^{m-1} w_{P(i)}]\}$, which is a decreasing function of both x and w. Hence, the optimal situation occurs when we pick the permutation so that $w_{P(i)}$ is a decreasing function of i. In this case, $\sum_{i=1}^{m} w_{P(i)}/m$ is a decreasing function of m. In this way, we arrive at $\hat{r}^2$ in equation (18). □

Remark 3. The ability to optimize $\hat{r}$ by means of picking the best possible permutation P and hence the best possible centering sequence is a novel feature of McDiarmid inequality. As far as we know, this feature has not been exploited before. In contrast, from the proof of corollary 1, it is clear that the value of $\hat{r}$ obtained from the Hoeffding's inequality does not depend on the choice of P. In section 5 below, we fully exploit this freedom of picking P to bound $e_{\mathrm{Z},1}$ in Method D. Note however that the above corollary requires the knowledge of M_j's. In addition, $\hat{r}^2$ is written as a rather involved sum. Let us replace every $w_{P(i)}$ in equation (18) by the average observed value, namely, $\sum_{i=1}^{t} w_i/t$. In this way, $\hat{r}$ would increase by a factor of $\mathrm{O}(t\,\mathrm{Width}(\mathcal{W})/D)$. Suppose that we fix $x = \sum_{i=1}^{t} w_i/t \equiv \langle w \rangle$ as well (without caring whether inequality (17) holds or not). Then $\hat{r}$ would change by a factor of $\mathrm{O}(\mathrm{Width}(\mathcal{W})/D\sqrt{t})$ most of the time due to statistical fluctuation. Thus, in practice, we may replace $\hat{r}$ in equation (18) by the following more convenient and useful expression

$$\hat{r} = \frac{\sqrt{t}\, y \mathrm{Width}(\mathcal{W})}{[y + (t-1)\langle w \rangle + \operatorname{essinf} \mathcal{W}][y + (t-1)\langle w \rangle + \operatorname{esssup} \mathcal{W}]}, \tag{23}$$

which does not depend on the knowledge of M_j's. This expression for $\hat{r}$ shall be used to bound $e_{\mathrm{Z},1}$ in Method C to be reported in section 5.

5. Application of the improved McDiarmid inequality in finding the key rate

There is a subtlety in applying theorem 1 to study the statistical fluctuation of $e_{\mathrm{Z},1}$. A naive way to do so is to use inequalities (5) and (7) to obtain the bound $e_{\mathrm{Z},1} \leqslant (\sum_{n=1}^{k} a_{2n} Q_{\mathrm{Z},\mu_n} E_{\mathrm{Z},\mu_n}) / (\sum_{n=1}^{k} a_{1n} Q_{\mathrm{Z},\mu_n})$. Then one could regard Q_{Z,μ_n}'s and $Q_{\mathrm{Z},\mu_n} E_{\mathrm{Z},\mu_n}$'s as random variables and directly apply theorem 1 and definition 1 to the rhs of the above inequality. Nonetheless, it does not work for the rhs of this inequality need not be bounded. Besides, the bound obtained is not strong enough even if we ignore the boundedness problem.

To proceed, we first write $Q_{\mathrm{Z},\mu_n} = \sum_j \tilde{W}_{nj} / \tilde{s}_{\mathrm{Z},\mu_n}$ where $\tilde{s}_{\mathrm{Z},\mu_n}$ is the number of photon pulses that Alice prepares using photon intensity μ_n and that Alice prepares and Bob tries to measure (but may or may not have detection) in Z basis. In addition, $\tilde{W}_{nj}$ denotes the possibly correlated random variable whose value is 1 (0) if the jth photon pulse among the $\tilde{s}_{\mathrm{Z},\mu_n}$ photon pulses is (not) detected by Bob. Clearly, $\tilde{s}_{\mathrm{Z},\mu_n} \approx T p_{\mathrm{Z}}^2 p_{\mu_n}$ with T being the total number of photon pulses sent by Alice and $p_{\mathrm{Z}} = 1 - p_{\mathrm{X}}$ is the probability for Alice (Bob) to prepare (measure) in the Z basis. Since $s_{\mathrm{Z}} \approx T p_{\mathrm{Z}}^2 \langle Q_{\mathrm{Z},\mu} \rangle$, I arrive at

$$Y_{\mathrm{Z},1} \geqslant \max\left(0, \sum_{n=1}^{k} a_{1n} Q_{\mathrm{Z},\mu_n}\right) = \max\left(0, \frac{\langle Q_{\mathrm{Z},\mu} \rangle}{s_{\mathrm{Z}}} \sum_{n=1}^{k} \left\{ \frac{a_{1n}}{p_{\mu_n}} \left[\sum_j \tilde{W}_{nj} \right] \right\} \right) = \max\left(0, \frac{\langle Q_{\mathrm{Z},\mu} \rangle}{s_{\mathrm{Z}}} \sum_{i=1}^{s_{\mathrm{Z}}} W_{\mathrm{Z},i} \right). \tag{24a}$$

Here $W_{\mathrm{Z},i}$ is the random variable that takes the value a_{1n}/p_{μ_n} if the ith photon pulse that are prepared by Alice and then successfully measured by Bob both in the Z basis is in fact prepared using photon intensity μ_n. Recall that Eve knows the number of photons in each pulse and may act accordingly. However, she does not know the photon intensity parameter used in each pulse and the preparation basis until the pulse is measured by Bob. Hence, $W_{\mathrm{Z},n}$'s may be correlated. Actually, the most general situation is that $W_{\mathrm{Z},n}$'s are drawn from a larger population without replacement. That is to say, these random variables obey the multivariate hypergeometric distribution. By the same argument, inequalities (7c) and (7d) gives

$$Y_{Z,1}e_{Z,1} \leqslant \min\left(\frac{Y_{Z,1}}{2}, \sum_{n=1}^{k} a_{2n} Q_{Z,\mu_n} E_{Z,\mu_n}\right) = \min\left(\frac{Y_{Z,1}}{2}, \frac{\langle Q_{Z,\mu}\rangle}{s_Z} \sum_{i=1}^{s_Z^e} W_{Z,i}^{e}\right) \tag{24b}$$

and

$$Y_{Z,1}\bar{e}_{Z,1} \geqslant \max\left(0, \sum_{n=1}^{k} a_{1n} Q_{Z,\mu_n} \bar{E}_{Z,\mu_n}\right) = \max\left(0, \frac{\langle Q_{Z,\mu}\rangle}{s_Z} \sum_{i=1}^{s_Z^{\bar{e}}} W_{Z,i}^{\bar{e}}\right), \tag{24c}$$

where $s_Z^e = s_Z \langle Q_{Z,\mu} E_{Z,\mu}\rangle / \langle Q_{Z,\mu}\rangle$ and $s_Z^{\bar{e}} = s_Z \langle Q_{Z,\mu} \bar{E}_{Z,\mu}\rangle / \langle Q_{Z,\mu}\rangle$) are the number of bits that are prepared and successfully measured in the Z basis such that the preparation by Alice and measurement result by Bob are unequal and equal, respectively. Moreover, $W_{Z,i}^e$'s ($W_{Z,i}^{\bar{e}}$'s) are multivariate hypergeometrically distributed random variables taking values in the set $\{a_{2n}/p_{\mu_n}\}_{n=1}^k$ ($\{a_{1n}/p_{\mu_n}\}_{n=1}^k$).

From inequalities (24a)–(24c), $e_{Z,1}$ obeys

$$e_{Z,1} \leqslant \max\left(0, \min\left(\frac{1}{2}, \frac{\sum_{i=1}^{s_Z^e} W_{Z,i}^e}{\sum_{j=1}^{s_Z} W_{Z,j}}\right)\right) \tag{25a}$$

and

$$e_{Z,1} \leqslant \max\left(0, \min\left(\frac{1}{2}, \frac{\sum_{i=1}^{s_Z^e} W_{Z,i}^e}{\sum_{i=1}^{s_Z^e} W_{Z,i}^e + \sum_{j=1}^{s_Z^{\bar{e}}} W_{Z,j}^{\bar{e}}}\right)\right). \tag{25b}$$

Interestingly, these two inequalities can be used to give four different bounds on the finite-size statistical fluctuations in $e_{Z,1}$. More importantly, these four bounds are

A. Use an upper bound of $\sum_{i=1}^{s_Z^e} W_{Z,i}^e$ and a lower bound of $\sum_{j=1}^{s_Z} W_{Z,j}$ to deduce an upper bound of $e_{Z,1}$. Specifically, from corollary 1, we conclude that the true value of $\sum_{j=1}^{s_Z} W_{Z,j}$ is less than the observed value by $[s_Z \ln(1/\epsilon_Z)/2]^{1/2}$ Width($\{a_{1n}/p_{\mu_n}\}_{n=1}^k$) with probability at most ϵ_Z. (Recall that Width($\mathcal{W}$) of a bounded set $\mathcal{W}$ of real numbers is defined as esssup$\mathcal{W}$ − essinf $\mathcal{W}$.) And the true value of $\sum_{i=1}^{s_Z^e} W_{Z,i}^e$ is greater than its observed value by $[s_Z^e \ln(1/\epsilon_Z^e)/2]^{1/2}$ Width($\{a_{2n}/p_{\mu_n}\}_{n=1}^k$) = $[s_Z \langle Q_{Z,\mu} E_{Z,\mu}\rangle \ln(1/\epsilon_Z^e)/2\langle Q_{Z,\mu}\rangle]^{1/2}$ Width($\{a_{2n}/p_{\mu_n}\}_{n=1}^k$) with probability at most ϵ_Z^e. Since $W_{Z_i}^e$ and $W_{Z,j}$ are positively correlated, from inequalities (24a), (24b) and (25a), we have

$$e_{Z,1} \leqslant \max\left(0, \min\left(\frac{1}{2}, \frac{\sum_{n=1}^{k} a_{2n} Q_{Z,\mu_n} E_{Z,\mu_n} + \Delta Y_{Z,1}e_{Z,1}}{\sum_{n=1}^{k} a_{1n} Q_{Z,\mu_n} - \Delta Y_{Z,1}}\right)\right) \tag{26a}$$

with probability at least $1 - \epsilon_Z - \epsilon_Z^e$, where

$$\Delta Y_{Z,1}e_{Z,1} = \left[\frac{\langle Q_{Z,\mu}\rangle \langle Q_{Z,\mu} E_{Z,\mu}\rangle \ln(1/\epsilon_Z^e)}{2 s_Z}\right]^{1/2} \text{Width}\left(\left\{\frac{a_{2n}}{p_{\mu_n}}\right\}_{n=1}^{k}\right) \tag{26b}$$

and

$$\Delta Y_{Z,1} = \langle Q_{Z,\mu}\rangle \left[\frac{\ln(1/\epsilon_Z)}{2 s_Z}\right]^{1/2} \text{Width}\left(\left\{\frac{a_{1n}}{p_{\mu_n}}\right\}_{n=1}^{k}\right). \tag{26c}$$

Incidentally, this is the method reported in the preprint by one of us in [14]. Moreover, similar bounds on statistical fluctuations of $Q_{B,n}$'s and $Q_{B,1}E_{B,1}$ have been obtained using Hoeffding's inequality in [4, 6]. That method is not as effective as the one reported here since they indirectly deal with finite sampling statistical fluctuation of $Y_{Z,1}$ and $Y_{Z,1}e_{Z,1}$.

B. Alternatively, we may use inequality (25b) and corollary 1 to bound $e_{Z,1}$. Specifically, the true value of $\sum_{j=1}^{s_Z^{\bar{e}}} W_{Z,j}^{\bar{e}}$ is less than the observed value by $[s_Z \langle Q_{Z,\mu} \bar{E}_{Z,\mu}\rangle \ln(1/\epsilon_Z^{\bar{e}})/2\langle Q_{Z,\mu}\rangle]^{1/2}$ Width($\{a_{1n}/p_{\mu_n}\}_{n=1}^k$) with probability at most $\epsilon_Z^{\bar{e}}$. Note that $W_{Z,j}^e$'s and $W_{Z,j}^{\bar{e}}$'s are statistically independent. Therefore, from inequalities (24b), (24c) and (25b), we have

$$e_{Z,1} \leqslant \max\left(0, \min\left(\frac{1}{2}, \frac{\sum_{n=1}^{k} a_{2n} Q_{Z,\mu_n} E_{Z,\mu_n} + \Delta Y_{Z,1} e_{Z,1}}{\sum_{n=1}^{k} a_{1n} Q_{Z,\mu_n} \bar{E}_{Z,\mu_n} + \sum_{n=1}^{k} a_{2n} Q_{Z,\mu_n} E_{Z,\mu_n} - \Delta Y_{Z,1} \bar{e}_{Z,1} + \Delta Y_{Z,1} e_{Z,1}}\right)\right) \tag{26d}$$

with probability at least $1 - \epsilon_Z^{e} - \epsilon_Z^{\bar{e}}$, where $\Delta Y_{Z,1} e_{Z,1}$ is given by equation (26b) and

$$\Delta Y_{Z,1} \bar{e}_{Z,1} = \left[\frac{\langle Q_{Z,\mu}\rangle \langle Q_{Z,\mu} \bar{E}_{Z,\mu}\rangle \ln(1/\epsilon_Z^{e})}{2 s_Z}\right]^{1/2} \text{Width}\left(\left\{\frac{a_{1n}}{p_{\mu_n}}\right\}_{n=1}^{k}\right). \tag{26e}$$

C. An even more interesting way to bound $e_{Z,1}$ is to use inequality (25b), corollary 2 and remark 3. Since $\langle w \rangle$ in this case is the measured $Y_{Z,1} e_{Z,1}/s_Z^{e}$, which is lower-bounded by inequality (7e), remark 3 gives

$$e_{Z,1} \leqslant \max\left(0, \min\left(\frac{1}{2}, \frac{\sum_{n=1}^{k} a_{2n} Q_{Z,\mu_n} E_{Z,\mu_n}}{\sum_{n=1}^{k} a_{1n} Q_{Z,\mu_n} \bar{E}_{Z,\mu_n} + \sum_{n=1}^{k} a_{2n} Q_{Z,\mu_n} E_{Z,\mu_n} - \Delta Y_{Z,1} \bar{e}_{Z,1}} + \Delta e_{Z,1}\right)\right) \tag{26f}$$

with probability at least $1 - \epsilon_Z^{\bar{e}} - \epsilon_Z^{e}$, where

$$\begin{aligned} \Delta e_{Z,1} = {} & \left[\frac{\langle Q_{Z,\mu}\rangle \langle Q_{Z,\mu} E_{Z,\mu}\rangle \ln(1/\epsilon_Z^{e})}{2 s_Z}\right]^{1/2} \left(\sum_{n=1}^{k} a_{1n} Q_{Z,\mu_n} \bar{E}_{Z,\mu_n} - \Delta Y_{Z,1} \bar{e}_{Z,1}\right) \text{Width}\left(\left\{\frac{a_{2n}}{p_{\mu_n}}\right\}_{n=1}^{k}\right) \\ & \times \left[\sum_{n=1}^{k} a_{1n} Q_{Z,\mu_n} \bar{E}_{Z,\mu_n} - \Delta Y_{Z,1} \bar{e}_{Z,1} + \left(1 - \frac{\langle Q_{Z,\mu}\rangle}{s_Z \langle Q_{Z,\mu} E_{Z,\mu}\rangle}\right) \sum_{n=1}^{k} a_{1n} Q_{Z,\mu_n} E_{Z,\mu_n} \right. \\ & \left. + \frac{\langle Q_{Z,\mu}\rangle^2}{s_Z^2 \langle Q_{Z,\mu} E_{Z,\mu}\rangle} \max_{n=1}^{k}\left\{\frac{a_{2n}}{p_{\mu_n}}\right\}\right]^{-1} \\ & \times \left[\sum_{n=1}^{k} a_{1n} Q_{Z,\mu_n} \bar{E}_{Z,\mu_n} - \Delta Y_{Z,1} \bar{e}_{Z,1} + \left(1 - \frac{\langle Q_{Z,\mu}\rangle}{s_Z \langle Q_{Z,\mu} E_{Z,\mu}\rangle}\right) \sum_{n=1}^{k} a_{1n} Q_{Z,\mu_n} E_{Z,\mu_n} \right. \\ & \left. + \frac{\langle Q_{Z,\mu}\rangle^2}{s_Z^2 \langle Q_{Z,\mu} E_{Z,\mu}\rangle} \min_{n=1}^{k}\left\{\frac{a_{2n}}{p_{\mu_n}}\right\}\right]^{-1} \end{aligned} \tag{26g}$$

provided that inequalities (15) and (17) hold. (See [23] for an alternative proof of this result.)

D. There is an alternative way to apply inequality (25b) and corollary 2 to find $\Delta e_{Z,1}$ in inequality (26f), which is quite aggressive. Since $\sum_{i=1}^{t} w_i/t$ is an estimate of $Y_{Z,1} e_{Z,1} = \langle w \rangle$, we know from corollary 1 and inequality (7e) that $\langle w \rangle \geqslant \sum_{n=1}^{k} a_{1n} Q_{Z,\mu_n} E_{Z,\mu_n} - \Delta Y_{Z,1} e_{Z,1}$ with probability at least $1 - \epsilon_Z^{e}$. In other words, by fixing $t = s_Z^{e}$ and $x = (\sum_{n=1}^{k} a_{1n} Q_{Z,\mu_n} E_{Z,\mu_n} - \Delta Y_{Z,1} e_{Z,1})/s_Z^{e}$, we conclude that inequality (22) is satisfied with probability at least $1 - \epsilon_Z^{e}$. Next, we could upper-bound the rhs of equation (18) by approximating the sum over m there by an integral. Specifically, set $y = \sum_{n=1}^{k} a_{1n} Q_{Z,\mu_n} \bar{E}_{Z,\mu_n} - \Delta Y_{Z,1} \bar{e}_{Z,1}$, then

$$\begin{aligned} \hat{r}^2 \approx {} & \sum_{m=1}^{t} \int_0^1 \left\{\frac{y}{y + (t-m+1)x + \operatorname{essinf} \mathcal{W} + \sum_{i=1}^{m-1} w_{P(i)} + \mu[w_{P(m)} - x]} \right. \\ & \left. - \frac{y}{y + (t-m+1)x + \operatorname{esssup} \mathcal{W} + \sum_{i=1}^{m-1} w_{P(i)} + \mu[w_{P(m)} - x]}\right\}^2 d\mu \\ = {} & y^2 \sum_{m=1}^{t} \frac{1}{w_{P(m)} - x}\left(-\frac{1}{y + (t-m+1)x + \operatorname{essinf} \mathcal{W} + \sum_{i=1}^{m-1} w_{P(i)} + \mu[w_{P(m)} - x]} \right. \\ & - \frac{1}{y + (t-m+1)x + \operatorname{esssup} \mathcal{W} + \sum_{i=1}^{m-1} w_{P(i)} + \mu[w_{P(m)} - x]} \\ & \left. \left. + \frac{2}{\text{Width}(\mathcal{W})} \ln\left\{\frac{y + (t-m+1)x + \operatorname{esssup} \mathcal{W} + \sum_{i=1}^{m-1} w_{P(i)} + \mu[w_{P(m)} - x]}{y + (t-m+1)x + \operatorname{essinf} \mathcal{W} + \sum_{i=1}^{m-1} w_{P(i)} + \mu[w_{P(m)} - x]}\right\}\right)\right|_{\mu=0}^{1}. \end{aligned} \tag{26h}$$

As $w_{P(m)}$'s are arranged in descending order and $\mathcal{W}$ in our case is the set $\{\langle Q_{Z,\mu}\rangle a_{2n}/(p_{\mu_n}s_Z)\}_{n=1}^{k}$ of at most k elements, rhs of the above inequality can be simplified to a big sum of at most k terms To be more explicit, suppose the descending sequence $\{w_{P(m)}\}_{m=1}^{t}$ contains $n^{(1)}$ copies of $w^{(1)}$, followed by $n^{(2)}$ copies of $w^{(2)}$, and so on until ending with $n^{(k)}$ copies of $w^{(k)}$. Surely, $\sum_{i=1}^{k} n^{(i)} = t = s_Z^{e}$ and $\sum_{i=1}^{k} n^{(i)}w^{(i)}/t$ is the observed $\langle w\rangle$. Then, equation (26*h*) becomes

$$\hat{r}^2 \approx y^2\sum_{m=1}^{k}\frac{1}{w^{(m)}-x}\Bigg(-\frac{1}{y+(t-\sum_{i<m}n^{(i)}+1)x+\text{essinf}\,\mathcal{W}+\sum_{i<m}n^{(i)}w^{(i)}+\mu[w^{(m)}-x]} - \frac{1}{y+(t-\sum_{i<m}n^{(i)}+1)x+\text{esssup}\mathcal{W}+\sum_{i<m}n^{(i)}w^{(i)}+\mu[w^{(m)}-x]} + \frac{2}{\text{Width}(\mathcal{W})}\ln\left\{\frac{y+(t-\sum_{i<m}n^{(i)}+1)x+\text{esssup}\mathcal{W}+\sum_{i<m}n^{(i)}w^{(i)}+\mu[w^{(m)}-x]}{y+(t-\sum_{i<m}n^{(i)}+1)x+\text{essinf}\mathcal{W}+\sum_{i<m}n^{(i)}w^{(i)}+\mu[w^{(m)}-x]}\right\}\Bigg)\Bigg|_{\mu=0}^{n^{(m)}}, \tag{26i}$$

which is efficient to compute. (Note that the sum in rhs of equation (18) is a decreasing function of x and $w_{P(i)}$'s, the above integral approximation is accurate up to a correction term of at most $\text{Width}(\mathcal{W})^2 = O(1/s_Z^2)$. Surely, this correction can be safely ignored in practice provided that $t = s_Z^{e} \gtrsim 10^4$.) In this way, inequality (26*f*) holds with probability at least $1-\epsilon_Z^{\bar{e}}-2\epsilon_Z^{e}$ with

$$\Delta e_{Z,1} = \hat{r}\left[\frac{\ln(1/\epsilon_Z^{e})}{2}\right]^{1/2}, \tag{26j}$$

where $\hat{r}$ is given by the rhs of inequality (26*i*).

In reality, we use the minimum of the above four methods to upper-bound the value of $e_{Z,1}$. To study the statistical fluctuation of R, it remains to consider the fluctuation of Q_{X,μ_n} in the first term in expression (8). (Although the second term also depends on Q_{X,μ_n}'s implicitly through Λ_{EC}, statistical fluctuation is absent from this term. This is because Λ_{EC} is the amount of information leaking to Eve during classical post-processing of the measured raw bits. Thus, it depends on the observed values of Q_{X,μ_n}'s and E_{X,μ_n}'s instead of their true values.) Using the same technique as in the estimation of statistical fluctuation in $e_{Z,1}$, the first term of Expression (8) can be rewritten as $\langle Q_{X,\mu}\rangle\sum_{i=1}^{s_X} W_{X,i}$ where $W_{X,i}$'s are multivariate hypergeometrically distributed random variables each taken values in the set $\{b_n/p_{\mu_n}\}_{n=1}^{k}$. Here b_n is given by equation (9) with e_p equals the R.H.S. of inequality (5) where $e_{Z,1}$ is given by any one of the following four equations depending on which of the four methods we use: equation (26*a*), (26*d*), (26*f*) and (26*j*). Corollary 1 implies that due to statistical fluctuation, the true value of the first term in Expression (8) is lower than the observed value by $\langle Q_{X,\mu}\rangle[\ln(1/\epsilon_X)/(2s_X)]^{1/2}\,\text{Width}(\{b_n/p_{\mu_n}\}_{n=1}^{k})$ with probability at most ϵ_X. We remark that this way of finding a lower bound for $\sum_n b_n Q_{X,\mu_n}$ is more direct than the standard one that separately bounds $Y_{X,0}$ and $Y_{X,1}$ [4, 6, 8, 11].

Putting everything together and by setting $\epsilon_X = \epsilon_Z = \epsilon_Z^{e} = \epsilon_Z^{\bar{e}} = \epsilon_{\bar{\gamma}} = \epsilon_{\text{sec}}/\chi$, we conclude that the secret key rate R satisfies

$$R = \sum_{n=1}^{k} b_n Q_{X,\mu_n} - \langle Q_{X,\mu}\rangle\left\{\frac{\ln[\chi/\epsilon_{\text{sec}}]}{2s_X}\right\}^{1/2}\text{Width}\left(\left\{\frac{b_n}{p_{\mu_n}}\right\}_{n=1}^{k}\right) - p_X^2\left\{\langle Q_{X,\mu}H_2(E_{X,\mu})\rangle + \frac{\langle Q_{X,\mu}\rangle}{s_X}\left[6\log_2\frac{\chi}{\epsilon_{\text{sec}}}+\log_2\frac{2}{\epsilon_{\text{cor}}}\right]\right\}, \tag{27}$$

where $b_n = b_n(e_p)$ is given by equation (9). Here e_p equals the rhs of inequality (5) with $e_{Z,1}$ given by equations (26*a*), (26*d*), (26*f*) or (26*j*). Note that $\chi = 9 = 4+1+4$ for Methods A to C and $\chi = 10$ for Method D. (Here the first number 4 comes from the generalized chain rule for smooth entropy in [4], the number 1 comes from the finite-size correction of the raw key in equation (B1) of [4], and the last number 4 comes from $\epsilon_{\bar{\gamma}}$, ϵ_X, ϵ_Z^{e} as well as either ϵ_Z or $\epsilon_Z^{\bar{e}}$. Moreover, χ for Method D is larger than the rest by 1 because of the extra condition on the statistical fluctuation of a lower bound of $Y_{Z,1}e_{Z,1}$.) Interestingly, unlike the schemes used in [4, 6, 7], the number χ in our scheme is independent on the number of photon intensities k used. This is because we directly tackle the finite sample statistical fluctuations of quantities like $Y_{B,1}$. Note however that even though χ does not depend on k, it does not mean that one could use arbitrarily large number of photon intensities as decoys (so as to obtain better bounds on quantities like $Y_{B,1}$) without adversely affecting the key rate for a fixed finite s_X. The reason is that $\text{Width}(\{a_{1n}/p_{\mu_n}\}_{n=1}^{k})$, $\text{Width}(\{a_{2n}/p_{\mu_n}\}_{n=1}^{k})$ and $\text{Width}(\{b_n/p_{\mu_n}\}_{n=1}^{k})$ diverge as $k\to+\infty$ due to divergence of a_{1n}, a_{2n} and b_n [6] as well as the decrease in $\min\{p_{\mu_n}\}_{n=1}^{k}$. Recall that

Table 1. Comparison between the state-of-the-art key rate $R^{\mathrm{O}} \equiv R^{\mathrm{O}}_{-5} \times 10^{-5}$ in [6] with the key rates in equation (27) (or more precisely $R^{\mathrm{I}}_{-5} \equiv \max(0, R^{\mathrm{I}} \times 10^{-5})$) for the dedicated quantum channel used in [4, 6] via Method I. These rate are optimized using the method stated in the main text.

	$k = 3$					$k = 4$					$k = 5$					$k = 6$				
s_{X}	R^{O}_{-5}	R^{A}_{-5}	R^{B}_{-5}	R^{C}_{-5}	R^{D}_{-5}	R^{O}_{-5}	R^{A}_{-5}	R^{B}_{-5}	R^{C}_{-5}	R^{D}_{-5}	R^{O}_{-5}	R^{A}_{-5}	R^{B}_{-5}	R^{C}_{-5}	R^{D}_{-5}	R^{O}_{-5}	R^{A}_{-5}	R^{B}_{-5}	R^{C}_{-5}	R^{D}_{-5}
10^5	0.052	0.300	0.326	0.470	0.254	0.027	0.270	0.291	0.487	0.747	0.000	0.152	0.156	0.160	0.142	0.000	0.052	0.076	0.076	0.000
10^6	0.294	0.743	0.789	0.835	0.637	0.194	0.727	0.763	0.829	1.41	0.100	0.660	0.694	0.516	0.484	0.055	0.404	0.434	0.407	0.966
10^7	0.687	1.18	1.23	1.22	1.04	0.573	1.27	1.30	1.27	1.84	0.421	1.21	1.20	1.20	1.12	0.259	0.949	1.01	0.823	1.57
10^8	1.11	1.43	1.48	1.45	1.64	1.04	1.60	1.63	1.59	2.18	0.929	1.66	1.68	1.63	1.81	0.624	1.32	1.34	1.33	2.00
10^9	1.51	1.70	1.75	1.72	2.05	1.57	1.91	1.94	1.90	2.40	1.46	2.04	2.10	2.06	2.37	1.08	1.74	1.75	1.71	2.38
10^{10}	1.87	1.98	2.02	1.99	2.32	1.97	2.20	2.22	2.19	2.58	1.94	2.40	2.42	2.40	2.72	1.72	2.16	2.18	2.14	2.63
10^{11}	2.20	2.25	2.29	2.26	2.43	2.32	2.46	2.48	2.45	2.81	2.46	2.67	2.69	2.69	2.88	2.18	2.50	2.52	2.48	2.86

computing a_{1n}, a_{2n} and b_n is numerically stable and with minimal lost in precision if $\mu_n - \mu_{n+1} \gtrsim 0.1$ for $n = 1, 2, ..., k - 1$ [6]. This means the number of photon intensities k used in practice should be $\lesssim 10$.

6. Performance analysis

We study the following quantum channel, which models a commonly used 100 km long optical fiber in QKD experiments, to test the performance of this new key rate formula in realistic situation. The findings here are generic as the general trend and performance improvement are also found in other situations including using the same fiber of different lengths as well as other randomly generated quantum channels. The yield and error rate of that quantum channel is given by $Q_{\mathrm{B},\mu} = (1 + p_{\mathrm{ap}})d_\mu$ and $Q_{\mathrm{B},\mu}E_{\mathrm{B},\mu} = p_{\mathrm{dc}} + e_{\mathrm{mis}}[1 - \exp(-\eta_{\mathrm{ch}}\mu)] + p_{\mathrm{ap}}d_\mu/2$, where $d_\mu = 1 - (1 - 2p_{\mathrm{dc}})\exp(-\eta_{\mathrm{sys}}\mu)$. Here we fix after pulse probability $p_{\mathrm{ap}} = 4 \times 10^{-2}$, dark count probability $p_{\mathrm{dc}} = 6 \times 10^{-7}$, error rate of the optical system $e_{\mathrm{mis}} = 5 \times 10^{-3}$. In addition, the transmittance of the system $\eta_{\mathrm{sys}} = 0.1\eta_{\mathrm{ch}}$, and the transmittance of the fiber is given by $\eta_{\mathrm{ch}} = 10^{-0.2L/10}$ with L is the length of the fiber in km. These parameters are obtained from optical fiber experiment on a 100 km long fiber in [24]; and have been used in [4, 6] to study the performance of decoy-state QKD in the FRKL situation. We also follow [4, 6] by using the following security parameters: $\epsilon_{\mathrm{cor}} = \kappa = 10^{-15}$, where $\epsilon_{\mathrm{sec}} = \kappa\ell_{\mathrm{final}}$ with $\ell_{\mathrm{final}} \approx Rs_{\mathrm{X}}/(p_{\mathrm{X}}^2\langle Q_{\mathrm{X},\mu}\rangle)$ is the length of the final key measured in bits. Note that κ can be interpreted as the secrecy leakage per final secret bit.

Table 1 compares the optimized key rates for the state-of-the-art method reported recently equation (3) of [6] with equation (27) for various s_{X} and k. (This is the best provably secure key rate obtained before the posting of the original proposal using McDiarmid inequality by one of us in [14].) The optimized rates are found by fixing the minimum photon intensity to 1×10^{-6}, while maximizing over p_{X} as well as all other photon intensities μ_n's and all the p_{μ_n}'s. This optimization is done by Monte Carlo method plus simulated annealing with a sample size of at least 10^{10} for each data entry in table 1. For Method D, the optimized key rate depends on the actual Z-basis measurement results. Here we simply fix $n^{(i)}$'s to their expectation values.

The table clearly shows that using McDiarmid inequality improves the optimized key rates in almost all cases. It also shows that for any method used, the provably secure key rate increases as the raw key length s_{X} increases. And they all gradually converge to the same infinite-size key rate. Besides, the asymptotic key rate generally increases with k. These are natural as longer s_{X} implies smaller finite-size statistical fluctuation and larger number of decoys k used allows better estimation of the bounds of various $Y_{\mathrm{B},m}$'s and $Y_{\mathrm{B},1}e_{\mathrm{B},1}$'s.

Among the four methods introduced here, Method A almost always gives the least provably secure key rate. This implies that it is more effective to estimate a lower bound for $Y_{\mathrm{Z},1}$ via estimating an upper bound for $Y_{\mathrm{Z},1}e_{\mathrm{Z},1}$ plus a lower bound for $Y_{\mathrm{Z},1}\bar{e}_{\mathrm{Z},1}$. Method B is slightly better than Method C for large s_{X} (say when $\gtrsim 10^8$, the improvement is about a few percent). Method D is about 5%–15% or so better than Method C when $10^8 \lesssim s_{\mathrm{X}} \lesssim 10^{11}$. This is not unexpected for the following reason. Although Method D is more aggressive than Method C in estimating the statistical fluctuation of $e_{\mathrm{Z},1}$ and hence the key rate, it requires an additional condition for lower-bounding $\langle w\rangle$. Thus the value of χ for Method D is 1 greater than that of Method C. As a result, for small raw key length, the improvement in estimating $e_{\mathrm{Z},1}$ for Method D may not be able to compensate the need to control the statistical fluctuation of one more variable. Table 1 also depicts that Method D is about 5%–20% better than Method B when $10^8 \lesssim s_{\mathrm{X}} \lesssim 10^{11}$. Furthermore, for fixed s_{Z} and κ and a fixed method to compute bound for $e_{\mathrm{Z},1}$, the provably secure key rate reaches a maximum at a finite k. This is not unexpected because even though the χ we deduce is independent of the number photon intensities k used, $\mathrm{Width}(\mathcal{W})$ diverges as $k \to +\infty$. Last but not least, in the case of $k = 4$, Method D always gives the best key rate. We do not have a good answer to this observation. It is instructive to study why in future.

7. Summary and outlook

To summarize, for $s_{\mathrm{X}} \approx 10^5 - 10^6$, at least one of the four methods reported here could produce a provably secure key rate that is at least twice that of the state-of-the-art method. And for $s_{\mathrm{X}} \approx 10^8$, Method D is at least 40% better than the state-of-the-art method. These improvements are of great value in practical QKD because the computational and time costs for classical post-processing can be quite high when the raw key length s_{X} is long. More importantly, the McDiarmid inequality method reported here is effective to increase the key rate of real or close to real time on demand generation of the secret key—an application that is possible in near future with the advancement of laser technology. It is instructive to extend our McDiarmid inequality method to handle the case of FRKL decoy-state measurement-device-independent QKD and compare it with existing methods in literature, such as the one that uses the Chernoff bound [25] and its extension specifically for decoys with four different intensities [26].

In addition to QKD, powerful concentration inequalities in statistics such as McDiarmid inequality could also be used beyond straightforward statistical data analysis. One possibility is to use it to construct model independent test for physics experiments that involve a large number of parameters but with relatively few data points.

Acknowledgments

This work is supported by the RGC grants 17304716 and 17302019 of the Hong Kong SAR Government. We would like to thank K-B Luk for his discussion on potential applications of McDiarmid inequality in physics.

References

[1] Diamanti E, Lo H-K, Qi B and Yuan Z 2016 *NPJ Quantum Inf.* **2** 16025
[2] Wang X-B 2005 *Phys. Rev. Lett.* **94** 230503
[3] Lo H-K, Ma X and Chen K 2005 *Phys. Rev. Lett.* **94** 230504
[4] Lim C C W, Curty M, Walenta N, Xu F and Zbinden H 2014 *Phys. Rev.* A **89** 022307
[5] Bennett C H and Brassard G 1984 *Proc. IEEE Int. Conf. on Computers, Systems and Signal Processing* (IEEE) pp 175–9
[6] Chau H F 2018 *Phys. Rev.* A **97** 040301(R)
[7] Hayashi M and Nakayama R 2014 *New J. Phys.* **16** 063009
[8] Brádler K, Mirhosseini M, Fickler R, Broadbent A and Boyd R 2016 *New J. Phys.* **18** 073030
[9] Wang X-B, Peng C-Z, Zhang J, Yang L and Pan J-W 2008 *Phys. Rev.* A **77** 042311
[10] Ma X, Qi B, Zhao Y and Lo H-K 2005 *Phys. Rev.* A **72** 012326
[11] Hayashi M 2007 *New J. Phys.* **9** 284
[12] Hoeffding W 1963 *J. Am. Stat. Assoc.* **58** 13
[13] McDiarmid C 1989 On the method of bounded differences *Surverys in Combinatorics 1989* (*Lond. Math. Soc. Lect. Notes Series* vol 141) ed J Siemons (Cambridge: Cambridge University Press) pp 148–88
[14] Chau H F 2018 Application of McDiarmid inequality in finite-key-length decoy-state quantum key distribution arXiv:1806.05063
[15] Lucamarini M, Yuan Z L, Dynes J F and Shields A J 2018 *Nature* **557** 400
[16] Lo H-K, Curty M and Qi B 2012 *Phys. Rev. Lett.* **108** 130503
[17] Renner R 2005 Security of QKD *PhD Thesis* ETH Diss. ETH No. 16242, arXiv:quant-ph/0512258
[18] Kraus B, Gisin N and Renner R 2005 *Phys. Rev. Lett.* **95** 080501
[19] Renner R, Gisin N and Kraus B 2005 *Phys. Rev.* A **72** 012332
[20] Fung C-H F, Ma X and Chau H F 2010 *Phys. Rev.* A **81** 012318
[21] McDiarmid C 1997 *Combin. Prob. Comput.* **6** 79
[22] McDiarmid C 1998 Concentration *Probabilistic Methods for Algorithmic Discrete Mathematics* (*Algorithms and Combinatorics* vol 16) ed M Habib *et al* (Berlin: Springer) pp 195–248
[23] Ng K C J 2019 Improved secure key rate for the decoy state protocol in the finite key regime *Master's Thesis* University of Hong Kong
[24] Walenta N, Lunghi T, Guinnard O, Houlmann R, Zbinden H and Gisin N 2012 *J. Appl. Phys.* **112** 063106
[25] Zhou Y-H, Yu Z-W and Wang X-B 2016 *Phys. Rev.* A **93** 042324
[26] Mao C-C, Zhou X-Y, Zhu J-R, Zhang C-H, Zhang C-M and Wang Q 2018 *Opt. Express* **26** 13289