# Modified BB84 quantum key distribution protocol robust to source imperfections

Margarida Pereira,[1,2,3,4,*] Guillermo Currás-Lorenzo,[1,2,3,4] Álvaro Navarrete,[1,2,3] Akihiro Mizutani,[5] Go Kato,[6]
Marcos Curty,[1,2,3] and Kiyoshi Tamaki[4]

[1]*Vigo Quantum Communication Center, University of Vigo, Vigo E-36315, Spain*
[2]*Escuela de Ingeniería de Telecomunicación, Department of Signal Theory and Communications, University of Vigo, Vigo E-36310, Spain*
[3]*atlanTTic Research Center, University of Vigo, Vigo E-36310, Spain*
[4]*Faculty of Engineering, University of Toyama, Gofuku 3190, Toyama 930-8555, Japan*
[5]*Mitsubishi Electric Corporation, Information Technology R&D Center, 5-1-1 Ofuna, Kamakura-shi, Kanagawa 247-8501, Japan*
[6]*National Institute of Information and Communications Technology, Nukui-kita, Koganei, Tokyo 184-8795, Japan*

(Received 24 October 2022; accepted 22 March 2023; published 27 April 2023)

The Bennett-Brassard 1984 (BB84) protocol is the most widely implemented quantum key distribution (QKD) scheme. However, despite enormous theoretical and experimental efforts in the past decades, the security of this protocol with imperfect sources has not yet been rigorously established. In this paper, we address this shortcoming and prove the security of the BB84 protocol in the presence of multiple source imperfections, including state preparation flaws and side channels, such as Trojan-horse attacks, mode dependencies and classical correlations between the emitted pulses. To do so, we consider a modified BB84 protocol that exploits the basis mismatched events, which are often discarded in standard security analyses of this scheme; and employ the reference technique, a powerful mathematical tool to accommodate source imperfections in the security analysis of QKD. Moreover, we compare the achievable secret-key rate of the modified BB84 protocol with that of the three-state loss-tolerant protocol, and show that the addition of a fourth state, while redundant in ideal conditions, significantly improves the estimation of the leaked information in the presence of source imperfections, resulting in a better performance. This paper demonstrates the relevance of the BB84 protocol in guaranteeing implementation security, taking us a step further towards closing the existing gap between theory and practice of QKD.

DOI: 10.1103/PhysRevResearch.5.023065

## I. INTRODUCTION

Quantum key distribution (QKD) enables two remote users, Alice and Bob, to securely establish cryptographic keys over an untrusted quantum channel [1–3]. Undoubtedly, the most widely used QKD scheme is the BB84 protocol, proposed by Bennett and Brassard in 1984 [4]. Almost four decades after its introduction, QKD has made an enormous progress both in theory and practice. However, despite its rigorous mathematical security proof, current physical implementations of QKD suffer from security loopholes due to inherent device imperfections and leakages of secret-key information.

Recent years have witnessed large efforts to reduce this discrepancy between theory and practice and guarantee the implementation security of QKD. A crucial breakthrough in this direction was the proposal of measurement-device-independent QKD (MDI-QKD) [5], which effectively closes all security loopholes on the detector side and is practical with existing hardware [6–10]. Moreover, a variant of MDI-QKD, called twin-field QKD [11], has been shown to provide a significant improvement on the achievable secret-key rate, allowing us to reach longer distances than ever before in fiber-based communications [12–14].

Having efficiently dealt with the measurement unit, the focus is now on securing the source. Essentially, source loopholes could arise from state preparation flaws (SPFs) [15–20] and side channels, such as Trojan-horse attacks (THAs) [20–25], mode dependencies [16,17,20,26–28], pulse correlations [29–34] and through changes in, for example, electromagnetic and acoustic radiation. Previous works have often looked at each of these imperfections individually, and developed experimental countermeasures and theoretical tools to minimize their impact on the secret-key rate and restore the security claim of QKD. For instance, the BB84 protocol has been shown to be secure in the presence of SPFs [18,19] and THAs [23–25,35]. However, one can only guarantee the implementation security of the source if all loopholes are taken into account simultaneously in the analysis, and thus the security of this protocol with imperfect sources has not yet been rigorously established.

The importance of achieving this level of security for the BB84 protocol cannot be overstated. In particular, many existing experiments [36–38], field-test QKD networks [39–45] and satellite-based quantum communication systems [46,47] employ the BB84 protocol. Hence, it is crucial to ensure the

[*] mpereira@com.uvigo.es

practical security of their transmitting units. Moreover, while in the absence of source imperfections the achievable secret-key rate is exactly the same when using three or four states [19], in the presence of imperfections the BB84 protocol may allow for a better estimation of the leaked secret information. As a consequence, this may lead to higher performances; a significant step towards attaining implementation security at an adequate level for practical QKD applications.

A recently proposed analytical tool for security proofs of QKD, the reference technique (RT) [31], is particularly well suited to address these issues, since it enables us to estimate the leaked secret-key information in the presence of multiple source imperfections. For this, one considers some reference states that are similar to the actual states emitted in the protocol, but whose simpler structure facilitates the estimation of some intermediate parameters. Then, since the two sets of states are close to each other, one can bound the maximum deviation between their detection probabilities using a Cauchy-Schwarz's type inequality (denoted in [31] as the $G$ function), and estimate the final parameters needed to guarantee the security of the actual protocol. The high flexibility and the high tolerance to source imperfections displayed by the RT with the $G$ function comes at a cost, however, as it requires the QKD protocol to be run sequentially, i.e., Alice only emits a particular pulse after Bob has measured the previous one [48]. Nonetheless, unlike other security proof approaches, the RT allows us to guarantee the security of QKD protocols with practical light sources against coherent attacks.

In this paper, we employ the RT with the $G$ function to prove the security of the BB84 protocol in the presence of multiple source imperfections; including SPFs and side channels, such as THAs, mode dependencies and classical pulse correlations. To do so, we consider a modified BB84 protocol that exploits the basis mismatched events, which are usually discarded in standard implementations of the protocol. Our security proof only requires an upper bound on a few parameters that quantify the quality of the source, i.e., no detailed information about the side-channel states is needed, thus facilitating the work of experimentalists. Additionally, we compare the achievable secret-key rate of this modified BB84 protocol with that of the three-state loss-tolerant protocol [19], and show that the emission of a fourth state, while redundant in ideal conditions, offers a significant improvement on the secret-key rate in the presence of source imperfections. This suggests that the modified BB84 protocol provides a clear performance advantage over the three-state protocol when dealing with imperfect sources.

The outline of the paper is as follows. In Sec. II, we describe in detail the emitted states in the modified BB84 protocol, and then list the assumptions imposed by our security proof on Alice's and Bob's devices. In Sec. III, we present our security analysis for the two different scenarios considered for the source, namely, the setting-independent and the setting-dependent scenarios. After that, in Sec. IV, we show the secret-key rate obtainable for the modified BB84 protocol in the presence of multiple source imperfections, and then compare it with that obtainable when using the three-state loss-tolerant protocol. Finally, in Sec. V, we summarize our findings.

## II. DESCRIPTION OF THE EMITTED STATES AND ASSUMPTIONS

For ease of discussion, we consider a modified BB84 protocol with an imperfect single-photon source (see Appendix A for a full protocol description). Nevertheless, our analysis could also be combined with the decoy-state method [49–51] to deal with phase-randomized coherent sources, as explained in [52]. In particular, we assume that the form of Alice's emitted states is affected by certain setting-choice-independent factors, such as temperature drifts or power fluctuations, that commonly arise in practical implementations of QKD. These factors can be modelled as a sequence of possibly-correlated random variables $\boldsymbol{\mathcal{G}} := \mathcal{G}_1, \dots, \mathcal{G}_N$, where $\mathcal{G}_k$ represents all the setting-independent factors that affect the state emitted in the $k$th round and $N$ is the total number of rounds. Since these factors are independent of Alice's encodings, for a given sequence of setting choices $j_1, \dots, j_N$, the global state emitted is mixed over the probability distribution of $\boldsymbol{\mathcal{G}}$. However, as shown in Appendix B, as long as one demonstrates that the protocol is secure for any particular value $\boldsymbol{g} := g_1, \dots, g_N$ of $\boldsymbol{\mathcal{G}}$, which can be assumed to have been fixed at the beginning of the protocol, then it is also secure for the actual case in which Alice emits mixed states. Thus, in the security proof, $\boldsymbol{g}$ can essentially be treated as a fixed parameter that affects the form of all the emitted states. Since the latter is mathematically equivalent to considering the mixed state case, here, we take this view for simplicity of presentation.

In addition, we investigate two different scenarios for the source. In the first one, the state of the emitted pulse on a particular round $k$ only depends on $g_k$ and on Alice's $k$th setting choice, i.e., it is independent of all her other setting choices. This is known as setting-independent pulse correlations and it was first modelled in [29]. The source model considered in this paper, however, goes beyond that presented in [29] even for this scenario, as the single-mode assumption is removed and the effect of side channels is incorporated. In the second scenario, the state of the emitted pulse on round $k$ may not only depend on $g_k$ and on Alice's $k$th setting choice, but also on Alice's previous $l_c$ setting choices, for some known correlation length $l_c$. This is often denoted as setting-dependent pulse correlations, and could arise, for instance, from memory effects in the electronic devices inside the transmitting unit [32]. The source model considered for this latter case is similar to that introduced in [31], but here we also take into account the dependence of the emitted pulses on the setting-independent factors described above, whose effect was disregarded in [31]. In both scenarios, we assume that, given a particular sequence of setting choices $j_1, \dots, j_N$, the global state emitted by Alice is a classical mixture of a tensor product of $N$ pure states. That is, we exclude the possibility of quantum correlations in which the states emitted on different rounds are entangled, which can hardly happen in typical implementations of QKD.

In the setting-independent scenario, for each round $k$ of the protocol, Alice chooses a setting $j \in \{0_Z, 1_Z, 0_X, 1_X\}$ and emits a state to Bob. This state can be expressed as

$$|\psi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} = \sqrt{1-\epsilon^{(k)}_{j,\boldsymbol{g}}}\,|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} + \sqrt{\epsilon^{(k)}_{j,\boldsymbol{g}}}\,|\phi^{\perp}_{j,\boldsymbol{g}}\rangle_{B_k,E_k}, \quad (1)$$

where $B_k$ is a two-dimensional system and $E_k$ includes any other systems that carry information about the $k$th pulse. In practice, $B_k$ is essentially the qubit system that Alice intends to send to Bob, while $E_k$ is a higher-dimensional system that has been unintentionally sent to Bob, such as the back-reflected light from a THA (see Appendix C for more details). Note that $|\psi_{j,\mathbf{g}}\rangle_{B_k,E_k}$ in Eq. (1) is a uniquely determined pure state once $j$ and $g_k$ are fixed. However, here we write $\mathbf{g}$, rather than $g_k$, because as explained above, in our security proof the parameter $\mathbf{g}$, which contains $g_k$ for any $k$, is fixed at the beginning of the protocol.

From construction, Eq. (1) is the most general description of the transmitted states within the framework of setting-independent correlations, since it is simply an expansion of the most general state $|\psi_{j,\mathbf{g}}\rangle_{B_k,E_k}$ in the basis $\{|\phi_{j,\mathbf{g}}\rangle_{B_k,E_k}, |\phi^{\perp}_{j,\mathbf{g}}\rangle_{B_k,E_k}\}$ [31,53]. In Eq. (1) the parameter $\epsilon^{(k)}_{j,\mathbf{g}} \in [0,1]$ quantifies the deviation of $|\psi_{j,\mathbf{g}}\rangle_{B_k,E_k}$ from the qubit state $|\phi_{j,\mathbf{g}}\rangle_{B_k,E_k} := |\omega_{j,\mathbf{g}}\rangle_{B_k}|\lambda_{\mathbf{g}}\rangle_{E_k}$, where $|\omega_{j,\mathbf{g}}\rangle_{B_k}$ is the state that Alice would send to Bob in the absence of side channels and $|\lambda_{\mathbf{g}}\rangle_{E_k}$ is a setting-independent state for the current round. Note that the state $|\omega_{j,\mathbf{g}}\rangle_{B_k}$ incorporates any imperfections in a qubit space, such as SPFs and phase fluctuations. The side channels are represented in Eq. (1) by the state $|\phi^{\perp}_{j,\mathbf{g}}\rangle_{B_k,E_k}$, which can live in a Hilbert space of arbitrary dimension and is orthogonal to $|\phi_{j,\mathbf{g}}\rangle_{B_k,E_k}$. In other words, the state $|\phi^{\perp}_{j,\mathbf{g}}\rangle_{B_k,E_k}$ corresponds to unwanted and possibly unknown modes, and it can incorporate side channels other than setting-dependent pulse correlations, such as THAs and mode dependencies.

In the setting-dependent scenario, the emitted state for each round $k$ can instead be expressed as

$$\left|\psi_{j,\mathbf{g}|j_{k-1},\dots,j_{k-l_c}}\right\rangle_{B_k,E_k} = \sqrt{1-\epsilon^{(k)}_{j,\mathbf{g}|j_{k-1},\dots,j_{k-l_c}}}\,|\phi_{j,\mathbf{g}}\rangle_{B_k,E_k} + \sqrt{\epsilon^{(k)}_{j,\mathbf{g}|j_{k-1},\dots,j_{k-l_c}}}\,|\phi^{\perp}_{j,\mathbf{g}|j_{k-1},\dots,j_{k-l_c}}\rangle_{B_k,E_k}, \tag{2}$$

where $j_{k-1},\dots,j_{k-l_c}$ represents the dependence of the $k$th pulse on Alice's previous $l_c$ setting choices. As before, Eq. (2) is simply an expansion of the state $|\psi_{j,\mathbf{g}|j_{k-1},\dots,j_{k-l_c}}\rangle_{B_k,E_k}$ in the basis $\{|\phi_{j,\mathbf{g}}\rangle_{B_k,E_k}, |\phi^{\perp}_{j,\mathbf{g}|j_{k-1},\dots,j_{k-l_c}}\rangle_{B_k,E_k}\}$, and within the framework of classical pulse correlations, this is the most general description of the transmitted states. Note that the state $|\phi^{\perp}_{j,\mathbf{g}|j_{k-1},\dots,j_{k-l_c}}\rangle_{B_k,E_k}$ in Eq. (2), besides incorporating all the side channels in $|\phi^{\perp}_{j,\mathbf{g}}\rangle_{B_k,E_k}$, also takes into account setting-dependent pulse correlations.

Importantly, due to the form of Eqs. (1) and (2), one can apply the RT to prove the security of the modified BB84 protocol as long as the following assumptions hold.

### A. Assumptions on Alice's transmitting unit

(A1) For all rounds of the protocol, Alice chooses the setting $j$ with a fixed probability $p_j$, with $p_{0_Z} = p_{1_Z}$.

Alice's setting selection in a given round is independent of those of other rounds, and Eve cannot tamper with her selection probabilities.

(A2) As described above, we consider two different scenarios for the source, which result in two security analyses with different assumptions:

(1) The emitted states do not depend on Alice's previous setting choices—Eq. (1).

We assume that an upper bound $\epsilon^{\mathrm{U}} \geqslant \epsilon^{(k)}_{j,\mathbf{g}}$ is known for all $k$, $j$ and $\mathbf{g}$. Note that, even in this case, the states emitted in different rounds of the protocol are not necessarily independent and identically distributed (IID) because the random variables $\mathcal{G}_1,\dots,\mathcal{G}_N$ that represent the setting-independent factors may be correlated between consecutive rounds. We show the security analysis under this scenario in Sec. III A.

(2) The emitted states depend on Alice's previous $l_c$ setting choices—Eq. (2).

We assume that an upper bound $\epsilon'^{\mathrm{U}} \geqslant \epsilon^{(k)}_{j,\mathbf{g}|j_{k-1},\dots,j_{k-l_c}}$ is known for all $k$, $j$, $\mathbf{g}$, and $j_{k-1},\dots,j_{k-l_c}$. Moreover, we assume that the state of the $k$th pulse is affected by $\mathbf{g}$ and Alice's previous $l_c$ setting choices, and that $l_c$ is a known parameter. The analysis under this scenario is given in Sec. III B. As we shall see, the data postprocessing in this case must be done differently. In particular, one needs to divide the sifted key in $(l_c+1)$ groups, and then perform the parameter estimation and privacy amplification separately for each group (see Appendix A). Note that, when $l_c = 0$, this scenario reduces to the setting-independent scenario described in Assumption (A2.a).

We emphasize that, while knowing the upper bound $\epsilon^{\mathrm{U}}$ ($\epsilon'^{\mathrm{U}}$) is a requirement to apply the RT, the characterization of the side-channel states $|\phi^{\perp}_{j,\mathbf{g}}\rangle_{B_k,E_k}$ ($|\phi^{\perp}_{j,\mathbf{g}|j_{k-1},\dots,j_{k-l_c}}\rangle_{B_k,E_k}$) is not needed. In other words, the inner products $\langle\phi^{\perp}_{j,\mathbf{g}}|\phi^{\perp}_{j',\mathbf{g}}\rangle_{B_k,E_k}$ ($\langle\phi^{\perp}_{j,\mathbf{g}|j_{k-1},\dots,j_{k-l_c}}|\phi^{\perp}_{j',\mathbf{g}|j_{k-1},\dots,j_{k-l_c}}\rangle_{B_k,E_k}$) and $\langle\phi_{j,\mathbf{g}}|\phi^{\perp}_{j',\mathbf{g}}\rangle_{B_k,E_k}$ ($\langle\phi_{j,\mathbf{g}}|\phi^{\perp}_{j',\mathbf{g}|j_{k-1},\dots,j_{k-l_c}}\rangle_{B_k,E_k}$) with $j \neq j'$ can be unknown. Importantly, this is not a necessary assumption but a fortunate consequence originating from the freedom to choose the reference states in the RT when using the particular inequality $G$ defined in Eq. (12). Since obtaining a full characterization of the side-channel states is very challenging in practice, previous theoretical works [31,53], as well as this paper, have exploited this advantage to consider device models that require minimal experimental characterization. Nonetheless, it is important to emphasize that if any information about the side channels is available it can be incorporated in the RT framework. This would most likely lead to higher performances because a better source characterization tends to result in a more accurate estimation of the phase-error rate. In fact, this has been recently shown for a particular time-dependent side channel in [28].

(A3) A partial characterization of the qubit state $|\omega_{j,\mathbf{g}}\rangle_{B_k}$ in Eqs. (1) and (2) can be obtained.

In the analysis presented in [31] (see also [35,53]), for simplicity, the qubit state $|\omega_{j,\mathbf{g}}\rangle_{B_k}$ in $|\phi_{j,\mathbf{g}}\rangle_{B_k,E_k} := |\omega_{j,\mathbf{g}}\rangle_{B_k}\,|\lambda_{\mathbf{g}}\rangle_{E_k}$ is assumed to be perfectly characterized and stable in time. Here, we go a step further and allow $|\omega_{j,\mathbf{g}}\rangle_{B_k}$ to vary slightly round by round, hence its dependence on $\mathbf{g}$. However, we assume that one can at least partially characterize this state, such that the upper bounds $c^{\mathrm{U}}_{\tau,j}$ and $p^{(\mathrm{vir})\mathrm{U}}_{\tau_X}$ on certain quantities that are defined later can be derived; see the discussion between Eqs. (10) and (11) for more details, including the definition of these parameters.

In Sec. IV A we show for illustration purposes that, for a typical phase-encoding setup in which the qubit component

$|\omega_{j,\boldsymbol{g}}\rangle_{B_k}$ of all the emitted states is in a standard basis plane (such as the $XZ$ plane) and the exact encoded phase $\theta_{j,\boldsymbol{g}}^{(k)}$ fluctuates over time, this requirement translates to being able to determine the range of these fluctuations, i.e., guaranteeing that $\theta_{j,\boldsymbol{g}}^{(k)} \in [\theta_j^{\mathrm{L}}, \theta_j^{\mathrm{U}}]$ for all $k$, $j$, and $\boldsymbol{g}$, where $\{\theta_j^{\mathrm{L}}, \theta_j^{\mathrm{U}}\}$ is known.

(A4) Alice only emits her $k$th pulse after Bob has performed his $k-1$th measurement [54].

This guarantees that the measurement operator $\hat{M}_{\gamma_X}^{(k)}$ in Eq. (6) [$\hat{D}_{\gamma_X}^{(k)}$ in Eq. (27)] satisfies $0 \leqslant \hat{M}_{\gamma_X}^{(k)} \leqslant \hat{\mathbb{1}}$ [$0 \leqslant \hat{D}_{\gamma_X}^{(k)} \leqslant \hat{\mathbb{1}}$] (see Appendices D and E2, respectively, for a proof of these statements), which is needed to apply the RT with the $G$ function defined in Eq. (12). We note that this assumption is also required when using the generalized entropy accumulation theorem [48] to prove the security of prepare-and-measure protocols against coherent attacks.

We remark that Assumption (A1) is present in all security proofs of QKD, Assumptions (A2) and (A4) are required to apply the RT, and Assumption (A3) is the necessary condition to take into account random fluctuations and setting-independent pulse correlations.

### B. Assumptions on Bob's measurement unit

(B1) For all rounds of the protocol, Bob chooses a measurement basis $\beta \in \{Z, X\}$ with probabilities $p_{Z_B}$ and $p_{X_B}$, respectively.

Bob's basis selection in a given round is independent of those of other rounds, and Eve cannot tamper with his selection probabilities.

(B2) Bob's measurements satisfy the basis-independent-efficiency condition.

We assume that Bob's measurements can be represented by the positive operator-valued measures (POVMs) $\{\hat{m}_{0_\beta}, \hat{m}_{1_\beta}, \hat{m}_f\}$ where $\hat{m}_{0_\beta}$ ($\hat{m}_{1_\beta}$) corresponds to Bob obtaining the bit value 0 (1) when selecting the basis $\beta$, and $\hat{m}_f$ is associated with an inconclusive outcome. That is, the detection efficiency of Bob's unit is independent of his measurement basis choice $\beta$. This assumption is required by many security proofs of QKD to remove detector side-channel attacks exploiting channel loss [56,57].

(B3) There are no side channels on Bob's device.

All these assumptions on Bob's device can be avoided by considering a MDI-type protocol, which removes all detector loopholes and to which our analysis could easily be extended (see [58]).

## III. SECURITY PROOF

Here, we show how the RT can be used to prove the security of the modified BB84 protocol against coherent attacks in the presence of multiple source imperfections. In particular, we explain how to estimate the phase-error rate, which bounds the amount of information leakage to a potential eavesdropper, Eve, and determines the amount of privacy amplification that is needed to guarantee a secure final key. We do this for two different security analyses that consider the two scenarios described previously for the transmitting unit. Namely, in Sec. III A, the emitted states only depend on setting-independent factors and in Sec. III B, in addition to these factors, the emitted states also depend on Alice's previous $l_c$ setting choices. Their corresponding assumptions are (A2.a) and (A2.b) in Sec. II A, respectively.

### A. Scenario in which the emitted states do not depend on Alice's previous setting choices

In this scenario, for each pulse emission, Alice sends a state $|\psi_{j,\boldsymbol{g}}\rangle_{B_k,E_k}$ given by Eq. (1) through the quantum channel to Bob, who then performs his POVM measurements. The secret key is distilled from the rounds in which both Alice and Bob select the $Z$ basis. As a basic framework to prove the security of these events, we use the complementarity approach [59,60]. First, note that from Eve's perspective, the key generation rounds are equivalently described by an entanglement-based scenario in which, after selecting the $Z$ basis, Alice prepares the following entangled state

$$\left|\Psi_{\boldsymbol{g}}^{Z}\right\rangle_{A_k,B_k,E_k} = \frac{1}{\sqrt{2}} \sum_{\tau\in\{0,1\}} |\tau_Z\rangle_{A_k} \left|\psi_{\tau_Z,\boldsymbol{g}}\right\rangle_{B_k,E_k}, \tag{3}$$

sends systems $B_k$, $E_k$ to Bob while keeping system $A_k$ in her laboratory, and then both Alice and Bob perform $Z$-basis measurements on their local and received systems, respectively.

To prove the security of these events, we consider the number of phase errors that Alice and Bob would have obtained if they had performed their measurements in the $X$ basis instead. This virtual scenario is equivalent to Alice sending Bob the fictitious virtual states

$$\left|\psi_{\tau_X,\boldsymbol{g}}^{(\mathrm{vir})}\right\rangle_{B_k,E_k} = \frac{\left|\psi_{0_Z,\boldsymbol{g}}\right\rangle_{B_k,E_k} + (-1)^{\tau}\left|\psi_{1_Z,\boldsymbol{g}}\right\rangle_{B_k,E_k}}{2\sqrt{\frac{\tilde{p}_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}}{p_{Z_A}}}}, \tag{4}$$

where $\tau \in \{0, 1\}$, with probabilities

$$\tilde{p}_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})} = \tfrac{1}{2} p_{Z_A}\big[1 + (-1)^{\tau}\Re\big(\langle\psi_{0_Z,\boldsymbol{g}}|\psi_{1_Z,\boldsymbol{g}}\rangle_{B_k,E_k}\big)\big], \tag{5}$$

who then performs $X$-basis measurements on the received systems. In Eq. (5), $\tilde{p}_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$ represents the joint probability that Alice chooses the $Z$ basis ($p_{Z_A}$) and prepares the virtual state $|\psi_{\tau_X,\boldsymbol{g}}^{(\mathrm{vir})}\rangle_{B_k,E_k}$.

A phase error occurs when Alice selects the virtual state associated to $1_X$ ($0_X$) and Bob obtains the bit value 0 (1) in his $X$-basis measurement. In Appendix D, we show that the probability of obtaining a phase error on round $k$, conditioned on all the previous outcomes, can be expressed as

$$P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Act}) := \sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}} \tilde{p}_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})} p_{Z_B} \mathrm{Tr}\big[\tilde{\sigma}_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})} \hat{M}_{\gamma_X}^{(k)}\big], \tag{6}$$

where $\tilde{\sigma}_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})} := |\psi_{\tau_X,\boldsymbol{g}}^{(\mathrm{vir})}\rangle\langle\psi_{\tau_X,\boldsymbol{g}}^{(\mathrm{vir})}|_{B_k,E_k}$ and $\{\hat{M}_{0_X}^{(k)}, \hat{M}_{1_X}^{(k)}, \hat{M}_f^{(k)}\}$ is a set of operators that can be regarded as the effective POVM that determines the $X$-basis detection statistics of the $k$th pulse; its specific form depends on Bob's actual POVM elements $\{\hat{m}_{0_X}, \hat{m}_{1_X}, \hat{m}_f\}$, on Eve's attack, and on the outcomes of the $k-1$ previous rounds [see Eq. (D7) in Appendix D for more details].

The detection probabilities $\mathrm{Tr}[\tilde{\sigma}_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}\hat{M}_{\gamma_X}^{(k)}]$ in Eq. (6) are not directly observed in the experiment because the virtual states are not actually emitted. Moreover, estimating them using the data collected in the protocol might be difficult due to the presence of multiple source imperfections. However, thanks to the RT, we can overcome this difficulty by estimating $P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Act})$ indirectly. For this, we first select some reference states that are similar to the actual states emitted in the protocol, and which allow an easy estimate of the phase-error probability that would be observed if they had been emitted: $P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})$. Then, by evaluating the deviation between the reference and actual states, we obtain $P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Act})$ from $P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})$. Finally, by applying concentration inequalities we derive an upper bound on the phase-error rate.

*Applying the reference technique.* First, we define four reference states. Even though our choice of states is unrestricted, higher secret-key rates are achieved if they are close to the actual states. Here, we pick the set of reference states to be $\{|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k}\}_{j\in\{0_Z,1_Z,0_X,1_X\}}$, which are defined in Eq. (1) as the qubit part of the actual states. Then, by replacing the actual $Z$-basis states by their corresponding reference states in Eq. (3), we define

$$\left|\Phi_{\boldsymbol{g}}^{Z}\right\rangle_{A_k,B_k,E_k}=\frac{1}{\sqrt{2}}\sum_{\tau\in\{0,1\}}|\tau_Z\rangle_{A_k}\left|\phi_{\tau_Z,\boldsymbol{g}}\right\rangle_{B_k,E_k}, \tag{7}$$

analogous to $|\Psi_{\boldsymbol{g}}^{Z}\rangle_{A_k,B_k,E_k}$. Similarly, we define the virtual states $|\phi_{\tau_X,\boldsymbol{g}}^{(\mathrm{vir})}\rangle_{B_k,E_k}$ and the probabilities $p_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$, which are analogous to $|\psi_{\tau_X,\boldsymbol{g}}^{(\mathrm{vir})}\rangle_{B_k,E_k}$ and $\tilde{p}_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$, respectively. This allows us to define the quantity

$$P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref}):=\sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}}p_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}p_{Z_B}\mathrm{Tr}\big[\sigma_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}\hat{M}_{\gamma_X}^{(k)}\big], \tag{8}$$

where $\sigma_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}:=|\phi_{\tau_X,\boldsymbol{g}}^{(\mathrm{vir})}\rangle\langle\phi_{\tau_X,\boldsymbol{g}}^{(\mathrm{vir})}|_{B_k,E_k}$. Here, $P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})$ could be interpreted as the probability of a phase error on the $k$th round when using the reference states. We emphasize that these replacements of actual states by their reference counterparts are purely mathematical. The reference states are never prepared nor sent in an actual implementation of the protocol.

A convenient feature of Eq. (8) over Eq. (6) is that the states $\sigma_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$ live in the same qubit space as the reference states $\sigma_{j,\boldsymbol{g}}^{(k)}:=|\phi_{j,\boldsymbol{g}}\rangle\langle\phi_{j,\boldsymbol{g}}|_{B_k,E_k}$, and therefore one can employ the idea of the loss-tolerant protocol [19] to write the former states as a linear function of the latter. For simplicity, here we assume that these states all lie in the $XZ$ plane of the Bloch sphere; see Appendix B in [61] for a more general treatment. Then, we have that

$$\sigma_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}=\sum_j c_{\tau,j,\boldsymbol{g}}^{(k)}\sigma_{j,\boldsymbol{g}}^{(k)}, \tag{9}$$

for $j\in\{0_Z,1_Z,0_X,1_X\}$, where $c_{\tau,j,\boldsymbol{g}}^{(k)}$ are real coefficients. To find these coefficients, one has to solve two systems of three linear equations with four unknowns. These systems have infinitely many solutions, and therefore one can choose the solutions that provide the tightest bound on the phase-error rate. This is the crucial difference with respect to the three-state protocol, and the reason why the modified BB84 protocol can provide higher secret-key rates (see Sec. IV C). We note that, in the case of the three-state protocol, the $1_X$ state is not emitted and thus $c_{1,1_X,\boldsymbol{g}}^{(k)}=c_{0,1_X,\boldsymbol{g}}^{(k)}=0$. This results in two systems of three linear equations with three unknowns, which have a unique solution each.

After substituting Eq. (9) in Eq. (8), we obtain

$$P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})=\sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}}p_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}p_{Z_B}\sum_j c_{\tau,j,\boldsymbol{g}}^{(k)}\mathrm{Tr}\big[\sigma_{j,\boldsymbol{g}}^{(k)}\hat{M}_{\gamma_X}^{(k)}\big], \tag{10}$$

where we have used the linearity of the trace operation. Note that since the reference states $\sigma_{j,\boldsymbol{g}}^{(k)}$ depend on $\boldsymbol{g}$, the coefficients $c_{\tau,j,\boldsymbol{g}}^{(k)}$ and the probabilities $p_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$ also depend on $\boldsymbol{g}$, and therefore their exact value is in general unknown. Nevertheless, our analysis only requires knowing upper bounds $c_{\tau,j}^{\mathrm{U}}\geqslant c_{\tau,j,\boldsymbol{g}}^{(k)}$ and $p_{\tau_X}^{(\mathrm{vir})\mathrm{U}}\geqslant p_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$ on each of these quantities [see Assumption (A3) in Sec. II A]. In Sec. IV A, we show how to derive these upper bounds in practice for a particular device model. Substituting them in Eq. (10), we obtain

$$P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})\leqslant\sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}}p_{\tau_X}^{(\mathrm{vir})\mathrm{U}}p_{Z_B}\sum_j c_{\tau,j}^{\mathrm{U}}\mathrm{Tr}\big[\sigma_{j,\boldsymbol{g}}^{(k)}\hat{M}_{\gamma_X}^{(k)}\big]. \tag{11}$$

The fictitious probabilities $\mathrm{Tr}[\sigma_{j,\boldsymbol{g}}^{(k)}\hat{M}_{\gamma_X}^{(k)}]$ in Eq. (11) are also unknown because, as mentioned before, the reference states are never sent in the actual protocol. However, by evaluating the deviation between the reference and the actual states we are able to bound these probabilities, and consequently estimate an upper bound on $P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})$. In doing so, we may adopt a number of different inequalities; but here, we choose the following inequality [31]:

$$\begin{aligned}&G_-(\mathrm{Tr}[|A\rangle\langle A|\hat{M}],|\langle A|R\rangle|)\\ &\quad\leqslant\mathrm{Tr}[|R\rangle\langle R|\hat{M}]\\ &\quad\leqslant G_+(\mathrm{Tr}[|A\rangle\langle A|\hat{M}],|\langle A|R\rangle|),\end{aligned} \tag{12}$$

where $|A\rangle$ and $|R\rangle$ are any two normalized states, and $\hat{M}$ is a measurement operator satisfying $0\leqslant\hat{M}\leqslant\hat{\mathbb{1}}$. The latter imposes a restriction on the repetition rate of the protocol [see Assumption (A4) in Sec. II A]. In Eq. (12), the functions $G_-(y,z)$ and $G_+(y,z)$ are defined for $0\leqslant y\leqslant 1$ and $0\leqslant z\leqslant 1$ as

$$G_-(y,z)=\begin{cases}g_-(y,z) & \text{if } y>1-z^2\\ 0 & \text{otherwise,}\end{cases} \tag{13}$$

and

$$G_+(y,z)=\begin{cases}g_+(y,z) & \text{if } y<z^2\\ 1 & \text{otherwise,}\end{cases} \tag{14}$$

with

$$g_\pm(y,z)=y+(1-z^2)(1-2y)\pm 2z\sqrt{(1-z^2)y(1-y)}. \tag{15}$$

Next, we apply the bound in Eq. (12) to each term in Eq. (11), by selecting the function $G_-(y,z)$ for the terms whose coefficient is negative, and the function $G_+(y,z)$ for the terms whose coefficient is positive, thus maximising $P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})$. We can then express $P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})$ as a function of the detection probabilities of the actual states,

$$P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref}) \leqslant \sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}} p_{\tau_X}^{(\mathrm{vir})\mathrm{U}} p_{Z_B}\left[\sum_{\substack{j\\ c_{\tau,j}^{\mathrm{U}}>0}} c_{\tau,j}^{\mathrm{U}} G_+\big(\mathrm{Tr}\big[\tilde{\sigma}_{j,\boldsymbol{g}}^{(k)}\hat{M}_{\gamma_X}^{(k)}\big], |\langle\psi_{j,\boldsymbol{g}}|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k}|\big) + \sum_{\substack{j\\ c_{\tau,j}^{\mathrm{U}}<0}} c_{\tau,j}^{\mathrm{U}} G_-\big(\mathrm{Tr}\big[\tilde{\sigma}_{j,\boldsymbol{g}}^{(k)}\hat{M}_{\gamma_X}^{(k)}\big], |\langle\psi_{j,\boldsymbol{g}}|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k}|\big)\right], \tag{16}$$

where $\tilde{\sigma}_{j,\boldsymbol{g}}^{(k)} := |\psi_{j,\boldsymbol{g}}\rangle\langle\psi_{j,\boldsymbol{g}}|_{B_k,E_k}$.

Using the definitions of the actual and reference states, we find that the inner products in Eq. (16) have the form $|\langle\psi_{j,\boldsymbol{g}}|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k}| = \sqrt{1-\epsilon_{j,\boldsymbol{g}}^{(k)}}$. For simplicity, we now use the upper bound $\epsilon^{\mathrm{U}} \geqslant \epsilon_{j,\boldsymbol{g}}^{(k)}$ for all $k$, $j$ and $\boldsymbol{g}$ [see Assumption (A2.a) in Sec. II A]. Using the fact that the functions $G_+(y,z)$ and $-G_-(y,z)$ are decreasing with respect to $z$ and that $\sqrt{1-\epsilon_{j,\boldsymbol{g}}^{(k)}} \geqslant \sqrt{1-\epsilon^{\mathrm{U}}}$, Eq. (16) can then be upper bounded by

$$P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref}) \leqslant \sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}} p_{\tau_X}^{(\mathrm{vir})\mathrm{U}} p_{Z_B}\left[\sum_{\substack{j\\ c_{\tau,j}^{\mathrm{U}}>0}} c_{\tau,j}^{\mathrm{U}} G_+\left(\frac{P_{\boldsymbol{g}}^{(k)}(j,\gamma_X|\mathrm{Act})}{p_j p_{X_B}}, \sqrt{1-\epsilon^{\mathrm{U}}}\right) + \sum_{\substack{j\\ c_{\tau,j}^{\mathrm{U}}<0}} c_{\tau,j}^{\mathrm{U}} G_-\left(\frac{P_{\boldsymbol{g}}^{(k)}(j,\gamma_X|\mathrm{Act})}{p_j p_{X_B}}, \sqrt{1-\epsilon^{\mathrm{U}}}\right)\right], \tag{17}$$

where $P_{\boldsymbol{g}}^{(k)}(j,\gamma_X|\mathrm{Act}) := p_j p_{X_B}\mathrm{Tr}[\tilde{\sigma}_{j,\boldsymbol{g}}^{(k)}\hat{M}_{\gamma_X}^{(k)}]$ is the joint probability that Alice prepares the actual state $|\psi_{j,\boldsymbol{g}}\rangle_{B_k,E_k}$, Bob chooses the $X$ basis and his measurement outcome is $\gamma$. Importantly, these probabilities are related to quantities directly observed in the protocol, and all the other parameters in Eq. (17) are known.

Now that we have an upper bound on $P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})$, the next step is to obtain an upper bound on the probability of a phase error in the actual protocol: $P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Act})$. For this, first note that $P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})$ in Eq. (8) can be written as

$$P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref}) = p_{Z_A}p_{Z_B}\mathrm{Tr}\big[\big|\Phi_{\boldsymbol{g}}^{Z}\big\rangle\big\langle\Phi_{\boldsymbol{g}}^{Z}\big|_{B_k,E_k}\hat{M}_{\mathrm{ph}}^{(k)}\big], \tag{18}$$

where

$$\hat{M}_{\mathrm{ph}}^{(k)} = \hat{P}\left(\frac{|0_Z\rangle_{A_k} - |1_Z\rangle_{A_k}}{\sqrt{2}}\right)\otimes\hat{M}_{0_X}^{(k)} + \hat{P}\left(\frac{|0_Z\rangle_{A_k} + |1_Z\rangle_{A_k}}{\sqrt{2}}\right)\otimes\hat{M}_{1_X}^{(k)}, \tag{19}$$

with $\hat{P}(|\cdot\rangle) = |\cdot\rangle\langle\cdot|$. Similarly, $P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Act})$ in Eq. (6) can be written as

$$P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Act}) = p_{Z_A}p_{Z_B}\mathrm{Tr}\big[\big|\Psi_{\boldsymbol{g}}^{Z}\big\rangle\big\langle\Psi_{\boldsymbol{g}}^{Z}\big|_{B_k,E_k}\hat{M}_{\mathrm{ph}}^{(k)}\big]. \tag{20}$$

Hence, one can simply employ the bound in Eq. (12) again to obtain the following expression:

$$\begin{aligned} P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Act}) &\leqslant p_{Z_A}p_{Z_B}G_+\left(\frac{P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})}{p_{Z_A}p_{Z_B}}, \big|\big\langle\Psi_{\boldsymbol{g}}^{Z}\big|\Phi_{\boldsymbol{g}}^{Z}\big\rangle_{B_k,E_k}\big|\right)\\ &\leqslant p_{Z_A}p_{Z_B}G_+\left(\frac{P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})}{p_{Z_A}p_{Z_B}}, \sqrt{1-\epsilon^{\mathrm{U}}}\right). \end{aligned} \tag{21}$$

In the last inequality of Eq. (21) we have used the fact that $G_+(y,z)$ is a decreasing function with respect to $z$, and the fact that $|\langle\Psi_{\boldsymbol{g}}^{Z}|\Phi_{\boldsymbol{g}}^{Z}\rangle_{B_k,E_k}| = (\sqrt{1-\epsilon_{0_Z,\boldsymbol{g}}^{(k)}} + \sqrt{1-\epsilon_{1_Z,\boldsymbol{g}}^{(k)}})/2 \geqslant \sqrt{1-\epsilon^{\mathrm{U}}}$.

The only missing step in the security analysis is to convert Eq. (21) into an expression in terms of observables. We start by taking the average over all transmitted rounds $N$ on both sides of Eq. (21) and then applying Jensen's inequality [62] to the right-hand side,

$$\begin{aligned} \frac{1}{N}\sum_{k}^{N} P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Act}) &\leqslant \frac{1}{N}\sum_{k}^{N} p_{Z_A}p_{Z_B}G_+\left(\frac{P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})}{p_{Z_A}p_{Z_B}}, \sqrt{1-\epsilon^{\mathrm{U}}}\right)\\ &\leqslant p_{Z_A}p_{Z_B}G_+\left(\frac{1}{N}\sum_{k}^{N}\frac{P_{\boldsymbol{g}}^{(k)}(\mathrm{ph}|\mathrm{Ref})}{p_{Z_A}p_{Z_B}}, \sqrt{1-\epsilon^{\mathrm{U}}}\right), \end{aligned} \tag{22}$$

where in the second inequality we have used the concavity of the function $G_+(y,z)$ with respect to its argument $y$. Then, by applying Jensen's inequality to this argument we have that

$$\frac{1}{N}\sum_k^N P_{\boldsymbol{g}}^{(k)}(\text{ph}|\text{Ref}) \leqslant \sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}} p_{\tau_X}^{(\text{vir})\text{U}} p_{Z_B}\left[\sum_{\substack{j\\ c^{\text{U}}_{\tau,j}>0}} c^{\text{U}}_{\tau,j}\, G_+\left(\frac{1}{N}\sum_k^N \frac{P_{\boldsymbol{g}}^{(k)}(j,\gamma_X|\text{Act})}{p_j p_{X_B}}, \sqrt{1-\epsilon^{\text{U}}}\right) + \sum_{\substack{j\\ c^{\text{U}}_{\tau,j}<0}} c^{\text{U}}_{\tau,j}\, G_-\left(\frac{1}{N}\sum_k^N \frac{P_{\boldsymbol{g}}^{(k)}(j,\gamma_X|\text{Act})}{p_j p_{X_B}}, \sqrt{1-\epsilon^{\text{U}}}\right)\right], \tag{23}$$

where we have used Eq. (17) and the concavity of the functions $G_+(y,z)$ and $-G_-(y,z)$.

Now, we apply Azuma's inequality [63] or Kato's inequality [64] to substitute each sum of probabilities into its corresponding observable. In particular, for the asymptotic case where $N\to\infty$, we have that $\sum_k^N P_{\boldsymbol{g}}^{(k)}(\text{ph}|\text{Act}) \simeq N_{\text{ph}}$, where $N_{\text{ph}}$ is the number of phase errors. Similarly, for $N\to\infty$, we find that $\sum_k^N P_{\boldsymbol{g}}^{(k)}(j,\gamma_X|\text{Act}) \simeq N_{j,\gamma_X}$, where $N_{j,\gamma_X}$ is the number of observed events in which Alice's setting choice is $j$, Bob selects the $X$ basis and his measurement outcome is $\gamma$. Therefore, by using the fact that $G_+(y,z)$ is an increasing function with respect to $y$, we combine Eqs. (22) and (23), then we multiply by $N$ on both sides, and apply Azuma's or Kato's inequality assuming $N\to\infty$, thus arriving at the following bound:

$$N_{\text{ph}} \leqslant N p_{Z_A} p_{Z_B} G_+\left(\sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}} \frac{p_{\tau_X}^{(\text{vir})\text{U}}}{p_{Z_A}}\left[\sum_{\substack{j\\ c^{\text{U}}_{\tau,j}>0}} c^{\text{U}}_{\tau,j}\, G_+\left(\frac{N_{j,\gamma_X}}{N p_j p_{X_B}}, \sqrt{1-\epsilon^{\text{U}}}\right) + \sum_{\substack{j\\ c^{\text{U}}_{\tau,j}<0}} c^{\text{U}}_{\tau,j}\, G_-\left(\frac{N_{j,\gamma_X}}{N p_j p_{X_B}}, \sqrt{1-\epsilon^{\text{U}}}\right)\right], \sqrt{1-\epsilon^{\text{U}}}\right) =: N_{\text{ph}}^{\text{U}}. \tag{24}$$

Finally, we can calculate an upper bound on the phase-error rate of the actual protocol by using

$$e_{\text{ph}}^{\text{U}} := \frac{N_{\text{ph}}^{\text{U}}}{N_{\text{det}}^{(Z)}}, \tag{25}$$

where $N_{\text{det}}^{(Z)}$ is the number of detected rounds in which both Alice and Bob selected the $Z$ basis, i.e., the length of the sifted key. Importantly, both Azuma's and Kato's inequalities allow us to take into account the fact that, under a coherent attack by Eve, the probabilities $P_{\boldsymbol{g}}^{(k)}(\text{ph}|\text{Act})$ and $P_{\boldsymbol{g}}^{(k)}(j,\gamma_X|\text{Act})$ may depend on the outcomes of the previous $k-1$ rounds.

Note that it is straightforward to modify Eq. (24) to apply it to the finite-key regime by simply including the deviation terms of the concentration inequality employed, taking into account that the functions $G_+(y,z)$ and $G_-(y,z)$ are increasing with respect to $y$. Moreover, Kato's inequality has been shown to provide tight estimations for practical values of $N$ [35,61,65,66], and therefore the performance of the protocol should not be significantly affected in the finite-key regime. We remark, however, that due to the restriction imposed by Assumption (A4) on the repetition rate of the protocol, the time required to reach these values of $N$ will increase.

### B. Scenario in which the emitted states depend on Alice's previous $l_c$ setting choices

In this scenario, the state emitted on each round $k$ is $|\psi_{j,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,E_k}$, defined in Eq. (2), which depends on setting-independent factors and on Alice's previous $l_c$ setting choices. Note that, due to these setting-dependent pulse correlations, information about Alice's $k$th setting choice is leaked to the subsequent pulses. However, this leakage of information can essentially be regarded as a side channel to the $k$th pulse [31]. In Appendix E 1, we show that one can consider the states emitted in round $k$ to be $|\tilde{\psi}_{j,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,\mathbf{E}'_k}$ instead, where $\mathbf{E}'_k$ includes not only system $E_k$ but also the systems of all rounds after the $k$th round.

As before, in the key generation rounds, we can then consider that Alice prepares the entangled state

$$\left|\Psi^Z_{\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\right\rangle_{A_k,B_k,\mathbf{E}'_k} = \frac{1}{\sqrt{2}}\sum_{\tau\in\{0,1\}} |\tau_Z\rangle_{A_k}\left|\tilde{\psi}_{\tau_Z,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\right\rangle_{B_k,\mathbf{E}'_k}, \tag{26}$$

and measures her ancilla system $A_k$ in the $Z$ basis. To prove the security in this scenario, however, one cannot directly apply the analysis presented in Sec. III A. The reason is the following. To derive the state $|\tilde{\psi}_{j,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,\mathbf{E}'_k}$ and Eq. (26), we need to assume that, if any round in $\{k-1,\ldots,k-l_c\}$ was a key-generation round, Alice measured her corresponding ancilla system in the $Z$ basis. However, in Sec. III A, we considered a virtual protocol in which, in all key-generation rounds, Alice measures her ancilla system in the $X$ basis. Thus, the description in Eq. (26) is not valid for this virtual protocol.

To avoid this contradiction, we follow a similar approach to that in [31,67], and assume that the users assign to each round $k\in\{1,\ldots,N\}$ a tag $w\in\{0,\ldots,l_c\}$ according to the value $w = k \bmod (l_c+1)$, and define the $w$th sifted key as the subset of the total sifted key that originates from rounds with tag $w$ (see the protocol description in Appendix A). Then, to estimate the information leakage of the $w$th sifted key, we consider a virtual scenario in which Alice and Bob use the $X$ basis to check for phase errors only for the key rounds with tag $w$, while for the rest of the key rounds they use the $Z$ basis as in the actual protocol. We refer to this scenario as the $w$th virtual protocol, and we define the phase-error rate of the $w$th sifted key, $e_{\text{ph}}^{w}$, as the fraction of phase errors that Alice and Bob would observe if they had run this $w$th virtual protocol. By obtaining a bound on $e_{\text{ph}}^{w}$, Alice and Bob can determine the amount of privacy amplification that they need to apply to the $w$th sifted key to turn it into a secret key. Importantly, note that the $(l_c+1)$ virtual protocols are not compatible with each other, as Alice and Bob could not have run them at the same time due to the noncommutativity of the $X$ and $Z$ basis measurements. However, it turns out that the $w$th virtual protocol allows us to prove the security of the $w$th key, and the security of the total key is ensured by the universal composability of each individual security proof, as shown in [67].

In the $w$th virtual protocol, if any round in $\{k-1,\ldots,k-l_c\}$ is a key round, Alice uses the $Z$ basis to measure her ancilla system. Thus, the description in Eq. (26) is indeed valid for this alternative scenario. As shown in Appendix E 2, it follows that the probability that, in the $w$th virtual protocol, Alice and Bob obtain a phase error on some round $k$ with tag $w$, conditioned on all the previous outcomes of the $w$th virtual protocol, can be expressed as

$$P_{\boldsymbol{g}}^{(k)}(\text{ph}|\text{Act}) := \sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}} \tilde{p}^{(k,\text{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}} p_{Z_B} \times \text{Tr}\big[\tilde{\sigma}^{(k,\text{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}} \hat{D}^{(k)}_{\gamma_X}\big], \tag{27}$$

where

$$\tilde{p}^{(k,\text{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}} = \tfrac{1}{2} p_{Z_A}\big[1+(-1)^{\tau} \times \Re\big(\langle\tilde{\psi}_{0_Z,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}|\tilde{\psi}_{1_Z,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k\mathbf{E}'_k}\big)\big]. \tag{28}$$

For simplicity of notation, in Eq. (27) we do not include a subscript $w$ to indicate that it refers to the $w$th virtual protocol. Also, in this equation, $\hat{D}^{(k)}_{\gamma_X}$ with $\gamma\in\{0,1\}$ is Bob's effective POVM element for the $k$th pulse after Eve's coherent attack [see Eq. (E15) in Appendix E 2 for more details], $\tilde{\sigma}^{(k,\text{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}} := \hat{P}(|\psi^{(\text{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,\mathbf{E}'_k})$ where $|\psi^{(\text{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,\mathbf{E}'_k}$ is defined as

$$\frac{|\tilde{\psi}_{0_Z,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,\mathbf{E}'_k} + (-1)^{\tau}|\tilde{\psi}_{1_Z,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,\mathbf{E}'_k}}{2\sqrt{\frac{\tilde{p}^{(k,\text{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}{p_{Z_A}}}}. \tag{29}$$

Importantly, note that Eq. (27) is very similar to Eq. (6), and thus one can apply the RT in the same way as in Sec. III A to derive an upper bound on the phase-error rate, as explained below.

*Applying the reference technique.* In this case, we choose the reference states to be $\{|\phi_{j,\boldsymbol{g}}\rangle_{B_k,\mathbf{E}'_k}\}_{j\in\{0_Z,1_Z,0_X,1_X\}}$, defined in Eq. (E16), which live in a qubit space. Then, one can follow a similar analysis to that described between Eqs. (7) and (21) by making the following substitutions: $|\psi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} \to |\tilde{\psi}_{j,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,\mathbf{E}'_k}$, $|\Psi^Z_{\boldsymbol{g}}\rangle_{A_k,B_k,E_k} \to |\Psi^Z_{\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{A_k,B_k,\mathbf{E}'_k}$, $|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} \to |\phi_{j,\boldsymbol{g}}\rangle_{B_k,\mathbf{E}'_k}$ and $|\Phi^Z_{\boldsymbol{g}}\rangle_{A_k,B_k,E_k} \to |\Phi^Z_{\boldsymbol{g}}\rangle_{A_k,B_k,\mathbf{E}'_k}$, where $|\Phi^Z_{\boldsymbol{g}}\rangle_{A_k,B_k,\mathbf{E}'_k}$ is defined in Eq. (E17). In the derivations, the inner products $|\langle\tilde{\psi}_{j,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}|\phi_{j,\boldsymbol{g}}\rangle_{B_k,\mathbf{E}'_k}|$ and $|\langle\Psi^Z_{\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}|\Phi^Z_{\boldsymbol{g}}\rangle_{A_k,B_k,\mathbf{E}'_k}|$ will now appear. In Appendix E 3, we show that these inner products are both upper bounded by $\sqrt{1-\epsilon^{\text{U}}}$, where $\epsilon^{\text{U}} = 1-(1-\epsilon'^{\text{U}})^{l_c+1}$. Using this result, we obtain Eqs. (17) and (21), which now apply to all rounds with tag $w$ in the $w$th virtual protocol. Then, taking the average over all rounds with tag $w$ on both sides of Eq. (21), and following a similar procedure as in the previous subsection, we arrive at

$$N_{\text{ph},w} \leqslant N_w p_{Z_A} p_{Z_B} G_+\left(\sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}} \frac{p^{(\text{vir})\text{U}}_{\tau_X}}{p_{Z_A}}\left[\sum_{\substack{j\\ c^{\text{U}}_{\tau,j}>0}} c^{\text{U}}_{\tau,j} G_+\left(\frac{N_{j,\gamma_X,w}}{N_w p_j p_{X_B}}, \sqrt{1-\epsilon^{\text{U}}}\right) + \sum_{\substack{j\\ c^{\text{U}}_{\tau,j}<0}} c^{\text{U}}_{\tau,j} G_-\left(\frac{N_{j,\gamma_X,w}}{N_w p_j p_{X_B}}, \sqrt{1-\epsilon^{\text{U}}}\right)\right], \sqrt{1-\epsilon^{\text{U}}}\right) =: N^{\text{U}}_{\text{ph},w}, \tag{30}$$

where $N_w = \frac{N}{l_c+1}$ is the number of rounds with a tag $w$ and $N_{j,\gamma_X,w}$ is the number of events with a tag $w$ in which Alice's setting choice is $j$, Bob selects the $X$ basis and his measurement outcome is $\gamma$.

Finally, we obtain the upper bound on the phase-error rate of the $w$th sifted key,

$$e^{\text{U}}_{\text{ph},w} := \frac{N^{\text{U}}_{\text{ph},w}}{N^{(Z)}_{\text{det},w}}, \tag{31}$$

where $N^{(Z)}_{\text{det},w}$ is the number of detected rounds with a tag $w$ in which both Alice and Bob selected the $Z$ basis. This means that the users should now apply privacy amplification to the $w$th sifted key sacrificing a fraction $h(e^{\text{U}}_{\text{ph},w})$ of its bits, where $h(x) = -x\log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function.

The fraction of the total sifted key that Alice and Bob sacrifice satisfies

$$\sum_{w=0}^{l_c} q_w h(e^{\text{U}}_{\text{ph},w}) \leqslant h\left(\sum_{w=0}^{l_c} q_w e^{\text{U}}_{\text{ph},w}\right) \leqslant h(e^{\text{U}}_{\text{ph}}), \tag{32}$$

where $q_w := N^{(Z)}_{\text{det},w}/N^{(Z)}_{\text{det}}$ and $e^{\text{U}}_{\text{ph}}$ is the bound obtained using Eq. (25) in Sec. III A, which in this case can be regarded as an upper bound on the average phase-error rate. The first inequality in Eq. (32) is due to the concavity of the function $h(x)$ and the second inequality is proved in Appendix F. This result is useful if one is simply interested in computing a lower bound on the asymptotic secret-key rate of the protocol. However, we remark that, in practice, one cannot simply compute $e_{\mathrm{ph}}^{\mathrm{U}}$ and then apply privacy amplification to the total sifted key at once. One needs to compute each $e_{\mathrm{ph},w}^{\mathrm{U}}$, and apply privacy amplification separately to each of the $(l_c + 1)$ sifted keys (see the protocol description in Appendix A).

## IV. RESULTS AND DISCUSSION

In this section, as an example, we apply our security proof to the modified BB84 protocol with a phase-encoding scheme in the presence of multiple source imperfections. In particular, we show how to derive upper bounds on the coefficients $c_{\tau,j,\boldsymbol{g}}^{(k)}$ and on the probabilities $p_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$. Then, we simulate the secret-key rate that we would obtain in a practical implementation of the protocol. Finally, we compare its performance with that of the three-state loss-tolerant protocol [19] under the same parameter regimes. We remark that the results and discussion presented here apply to both security analyses in Sec. III.

### A. Particular device model

In general, the state of the emitted pulses for each round $k$ is in the form of Eqs. (1) and (2), which take into account the main source imperfections. As a particular example, we assume here that the state $|\omega_{j,\boldsymbol{g}}\rangle_{B_k}$ in the qubit part of these equations: $|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} = |\omega_{j,\boldsymbol{g}}\rangle_{B_k} \, |\lambda_{\boldsymbol{g}}\rangle_{E_k}$, satisfies

$$|\omega_{j,\boldsymbol{g}}\rangle_{B_k} = \cos\left(\frac{\theta_{j,\boldsymbol{g}}^{(k)}}{2}\right)|0_Z\rangle_{B_k} + \sin\left(\frac{\theta_{j,\boldsymbol{g}}^{(k)}}{2}\right)|1_Z\rangle_{B_k}, \tag{33}$$

where $\theta_{j,\boldsymbol{g}}^{(k)} \in [0, 2\pi)$ is the encoding phase, which depends on the round $k$. Recall that the state in Eq. (33) incorporates any imperfection in the qubit space, such as SPFs and phase fluctuations.

When applying the RT to prove the security of this protocol, we relate the virtual states $\sigma_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$ and the reference states $\sigma_{j,\boldsymbol{g}}^{(k)}$ through Eq. (9). Then, for both virtual states, we choose the unique solution of Eq. (9) such that $c_{1,0_X,\boldsymbol{g}}^{(k)} = c_{0,1_X,\boldsymbol{g}}^{(k)} = 0$, which provides the best numerical results for our device model. We elaborate on this point in Sec. IV C. Using the definition of $|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k}$ and Eq. (33), we can solve the resulting systems of linear equations to find analytical expressions for the other coefficients $c_{\tau,j,\boldsymbol{g}}^{(k)}$ as a function of the encoding phases, i.e., $c_{1,j,\boldsymbol{g}}^{(k)}(\theta_{0_Z,\boldsymbol{g}}^{(k)}, \theta_{1_Z,\boldsymbol{g}}^{(k)}, \theta_{1_X,\boldsymbol{g}}^{(k)})$ for $j \in \{0_Z, 1_Z, 1_X\}$ and $c_{0,j,\boldsymbol{g}}^{(k)}(\theta_{0_Z,\boldsymbol{g}}^{(k)}, \theta_{1_Z,\boldsymbol{g}}^{(k)}, \theta_{0_X,\boldsymbol{g}}^{(k)})$ for $j \in \{0_Z, 1_Z, 0_X\}$ (see Appendix G for full expressions). Similarly, we can express $p_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$ as a function of $\theta_{0_Z,\boldsymbol{g}}^{(k)}$ and $\theta_{1_Z,\boldsymbol{g}}^{(k)}$. While the exact phases for a particular round are unknown, we assume that one can guarantee that they always fall in a known range, i.e., $\theta_{j,\boldsymbol{g}}^{(k)} \in [\theta_j^{\mathrm{L}}, \theta_j^{\mathrm{U}}]$ for $j \in \{0_Z, 1_Z, 0_X, 1_X\}$. Then, we can obtain upper bounds $c_{\tau,j}^{\mathrm{U}}$ and $p_{\tau_X}^{(\mathrm{vir})\mathrm{U}}$ on each individual coefficient $c_{\tau,j,\boldsymbol{g}}^{(k)}$ and probability $p_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$ by considering the worst case scenario for the encoding phases. In Appendix G, we provide their analytical solutions in Eqs. (G2) and (G4), respectively.

Finally, by substituting these bounds into Eq. (25), we obtain $e_{\mathrm{ph}}^{\mathrm{U}}$. A lower bound on the asymptotic secret-key rate can then be expressed as

$$R \geqslant Y_Z\big[1 - h\big(e_{\mathrm{ph}}^{\mathrm{U}}\big) - fh(e_{\mathrm{bit}})\big], \tag{34}$$

where $Y_Z$ is the joint probability that both Alice and Bob select the $Z$ basis and Bob obtains a detection event, $f$ is the error correction efficiency, and $e_{\mathrm{bit}}$ is the bit-error rate. Note that $Y_Z$ and $e_{\mathrm{bit}}$ would be directly observed in a practical implementation of the protocol. The lower bound in Eq. (34) applies both to the case in which the emitted states do not depend on Alice's previous setting choices, and to the case in which they do depend on the previous $l_c$ setting choices; the latter is due to Eq. (32). However, we emphasize that this lower bound depends on the value of $\epsilon^{\mathrm{U}}$, whose definition differs in the two cases. Namely, in the former, $\epsilon^{\mathrm{U}}$ is an upper bound on $\epsilon_{j,\boldsymbol{g}}^{(k)}$ in Eq. (1); while in the latter, $\epsilon^{\mathrm{U}} = 1 - (1 - \epsilon'^{\mathrm{U}})^{l_c+1} \approx (l_c + 1)\epsilon'^{U}$, where $\epsilon'^{\mathrm{U}}$ is an upper bound on $\epsilon_{j,\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}^{(k)}$ in Eq. (2).

### B. Simulation of the secret-key rate

To apply our analysis to this particular device model, one needs to experimentally measure the parameter $\epsilon^{\mathrm{U}}$, and $\{\theta_j^{\mathrm{L}}, \theta_j^{\mathrm{U}}\}$ for all $j$, which represents the uncertainty on the phase of the qubit component. Since there are no experimental works quantifying $\epsilon^{\mathrm{U}}$, in our simulations, we consider a range of values for this parameter. Also, as an example, we take the uncertainty on the phase to be the same for all $j$, i.e., $\theta_j^{\mathrm{L}} = \hat{\theta}_j - \Delta$ and $\theta_j^{\mathrm{U}} = \hat{\theta}_j + \Delta$ for some known $\hat{\theta}_j$ and $\Delta$. Moreover, we assume that $\hat{\theta}_j$ deviates from the ideal encoding angles due to SPFs; in particular, we assume that $\hat{\theta}_{0_Z} = 0, \hat{\theta}_{1_Z} = \kappa\pi, \hat{\theta}_{0_X} = \kappa\pi/2$, and $\hat{\theta}_{1_X} = \kappa 3\pi/2$, where $\kappa = 1 + \delta/\pi$ and $\delta \geqslant 0$ quantifies this deviation.

To simulate the data that one would obtain in an actual experiment, we use the channel model described in [20]. In particular, the model assumes that Alice sends a qubit state with an exact phase $\hat{\theta}_j$ when she selects the setting $j$. This is because $\Delta$ and $\epsilon^{\mathrm{U}}$ are both small, and do not result in significantly different experimental results compared with the ideal case. This does not contradict the assumptions of our security proof: these imperfections could still exist and allow Eve to learn some secret-key information, which our security proof takes into account.

For the simulations, we assume the following channel, device, and protocol parameters: $\Delta = 0.03$ [16], $\delta = 0.063$ [15,16], $f = 1.16$, and the dark count probability of Bob's detectors $p_d = 10^{-8}$ [2,12]. Here, we do not consider a specific value for the detection efficiency of Bob's detectors because we represent the secret-key rate as a function of the overall system loss. Moreover, we assume the efficient QKD scheme [68], where $p_{X_A} = p_{X_B} \to 0$ in the asymptotic scenario.

The results for the modified BB84 protocol are shown in Fig. 1. This figure shows that, as $\epsilon^{\mathrm{U}}$ increases, the secret-key rate $R$ decreases. This is expected because a higher value of $\epsilon^{\mathrm{U}}$ means that the emitted states may be more distinguishable, and thus more vulnerable to, for example, an unambiguous state discrimination (USD) attack [69,70], which could allow Eve to learn key information without introducing any errors.

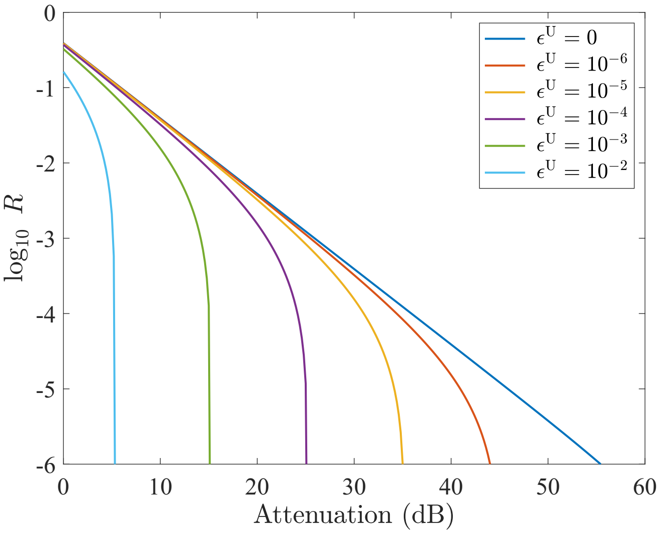

FIG. 1. Secret-key rate $R$ as a function of the overall system loss (dB) for $\delta = 0.063$ and different values of $\epsilon^{\mathrm{U}}$ when applying the RT to the modified BB84 protocol.

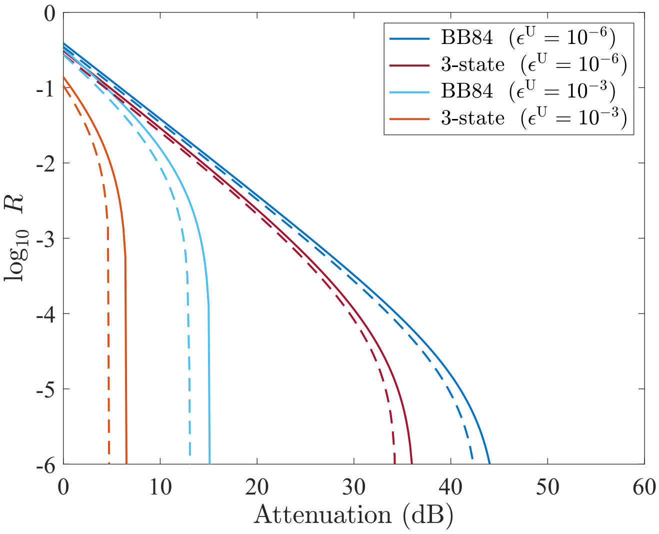

FIG. 2. Secret-key rate $R$ as a function of the overall system loss (dB) for different parameter regimes when applying the RT to the modified BB84 protocol (BB84) and to the three-state protocol (3-state). The solid lines correspond to $\delta = 0.063$ and the dashed lines correspond to $\delta = 0.126$.

To ensure implementation security with practical secret-key rates, it is therefore essential that experimentalists not only quantify $\epsilon^{\mathrm{U}}$ but also make an effort to minimize its magnitude, for example, by employing additional optical components in QKD setups, such as isolators and attenuators.

In addition, our security proof takes into account variable modulation flaws and setting-independent pulse correlations, allowing each emitted pulse to be different, i.e., non-IID, even in the absence of side channels. We note that there are other proof techniques that can also be employed to deal with source loopholes, such as those using semi-definite-programming (SDP) [66,71,72] or other convex optimization approaches [73,74]. However, so far, none of them is able to guarantee this level of implementation security, and therefore it is difficult to make a fair comparison. In particular, the analyses in [66,71–74] assume the emission of IID pulses. This assumption is unrealistic since in practice the emission of non-IID pulses is unavoidable due to fluctuations and pulse correlations.

### C. Comparison between the three-state and the modified BB84 protocols

For the comparison with the three-state protocol to be fair, here we consider the same particular device model (see Sec. IV A). There is, however, a difference in the security proof: for the three-state protocol, Eq. (9) can only be expressed as

$$\sigma_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})} = \sum_j c_{\tau,j,\boldsymbol{g}}^{(k)} \sigma_{j,\boldsymbol{g}}^{(k)}, \tag{35}$$

for $j \in \{0_Z, 1_Z, 0_X\}$. In this case, the coefficients $c_{\tau,j,\boldsymbol{g}}^{(k)}$ for $\tau \in \{0, 1\}$ are both functions of the same encoding phases: $c_{\tau,j,\boldsymbol{g}}^{(k)}(\theta_{0_Z,\boldsymbol{g}}^{(k)}, \theta_{1_Z,\boldsymbol{g}}^{(k)}, \theta_{0_X,\boldsymbol{g}}^{(k)})$. The form of these coefficients as well as their upper bounds are defined in Appendix H. For the simulations, we take $\epsilon^{\mathrm{U}} \in \{10^{-6}, 10^{-3}\}$ and $\delta \in \{0.063, 0.126\}$ [15,16]. All the other experimental parameters are the same as above.

The results are presented in Fig. 2. This figure shows that the modified BB84 protocol achieves better secret-key rates in all parameter regimes investigated. For instance, in Fig. 2, the blue lines are significantly better than the red lines for $\epsilon^{\mathrm{U}} \in \{10^{-6}, 10^{-3}\}$. This indicates that using four states rather than three, while redundant in an idealized scenario [19], is advantageous in the presence of multiple source imperfections.

As pointed out in Sec. III A, this improvement comes from the extra choice in the coefficients of Eq. (9) when adding a fourth state. Loosely speaking, for each $\tau \in \{0, 1\}$, the estimation should be tightest when $\sum_j |c_{\tau,j,\boldsymbol{g}}^{(k)}|$ is minimized, since higher absolute values for these coefficients result in increased multiplicative factors for the deviation terms introduced by the application of the bound in Eq. (12). Thus, since $\sigma_{1_X,\boldsymbol{g}}^{(k)}$ and $\sigma_{1_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$ ($\sigma_{0_X,\boldsymbol{g}}^{(k)}$ and $\sigma_{0_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$) are close in the Bloch sphere, if we take $c_{1,0_X,\boldsymbol{g}}^{(k)} = 0$ ($c_{0,1_X,\boldsymbol{g}}^{(k)} = 0$), then $c_{1,1_X,\boldsymbol{g}}^{(k)}$ ($c_{0,0_X,\boldsymbol{g}}^{(k)}$) is close to one, while $c_{1,0_Z,\boldsymbol{g}}^{(k)}$ and $c_{1,1_Z,\boldsymbol{g}}^{(k)}$ ($c_{0,0_Z,\boldsymbol{g}}^{(k)}$ and $c_{0,1_Z,\boldsymbol{g}}^{(k)}$) are close to zero, minimizing the sum of absolute values. We have numerically tested all possible combinations in which $\sigma_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$ can be expressed as a function of $\sigma_{j,\boldsymbol{g}}^{(k)}$, for $j \in \{0_Z, 1_Z, 0_X, 1_X\}$, and confirmed that the choices $c_{1,0_X,\boldsymbol{g}}^{(k)} = c_{0,1_X,\boldsymbol{g}}^{(k)} = 0$ result in the highest secret-key rates. This implies that the key-rate advantage of the modified BB84 protocol with respect to the three-state protocol comes from the ability to express $\sigma_{1_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$ as a function of $\sigma_{1_X,\boldsymbol{g}}^{(k)}$ rather than $\sigma_{0_X,\boldsymbol{g}}^{(k)}$.

We remark that, recently, a similar conclusion has been reached by employing a SDP-type security proof [28]. Namely, the MDI BB84 protocol has been shown to provide higher secret-key rates than the MDI three-state protocol in the presence of a particular time-dependent side channel. In this paper, however, we show that the modified BB84 protocol achieves a better performance than the three-state protocol in the presence of multiple source imperfections, including classical pulse correlations, when using the RT.

Additionally, one can see in Fig. 2 that, as the SPFs increase, the secret-key rate is roughly the same in all cases investigated. That is, for both analyses the solid lines (corresponding to $\delta = 0.063$) and the dashed lines (corresponding to $\delta = 0.126$) are close to one another for any value of $\epsilon^{\mathrm{U}}$. This indicates that, when using the RT, an increase in $\delta$ has a much smaller impact on the achievable secret-key rate than an increase in $\epsilon^{\mathrm{U}}$. Intuitively, this is expected since SPFs are imperfections in the qubit space, and therefore they do not necessarily increase the protocol's vulnerability to an USD attack.

## V. CONCLUSIONS

The best known and most widely implemented quantum key distribution (QKD) scheme is the Bennett-Brassard 1984 (BB84) protocol [4]. Since its introduction, several rigorous security proofs have been proposed [18,60,75–77], but the security of the BB84 protocol with imperfect sources has not yet been fully established. In this paper, we have considered a modified BB84 protocol that does not discard the basis mismatched events, and we have used the reference technique (RT) to prove its security in the presence of most source imperfections.

Although we do not consider quantum pulse correlations, we believe that this imperfection is very unlikely to be relevant in practice, given the fragile nature of entanglement. Indeed, it is highly improbable that a source spontaneously emits states that are entangled between rounds. Moreover, even if during a THA Eve's injected light is entangled between rounds, one expects this entanglement to almost completely vanish from the output light, especially if the source employs optical isolators. Having said that, however, these quantum correlations are interesting from a theoretical perspective and we plan to address them in future works.

In this paper, we have also compared the performance offered by the modified BB84 protocol with that offered by the three-state loss-tolerant protocol when using the RT. We have shown that, in the presence of source imperfections, the addition of a fourth state allows us to obtain a tighter estimation of the phase-error rate, and consequently, higher secret-key rates. This indicates that the modified BB84 protocol offers a clear advantage in guaranteeing the practical security of QKD sources.

For completeness, we note that besides modifying the RT in [31] to deal with four states, which resulted in a significant improvement in its performance; in this paper, we have also made the RT more flexible and more experimentally friendly than before. In particular, here we incorporate setting-independent pulse correlations together with all the other imperfections previously considered. Moreover, we use the most general model to describe setting-dependent pulse correlations, rather than the simplified model considered in Eq. (10) of [31]. Finally, we relax the assumption on the qubit part of the emitted states made in [31], by allowing them to fluctuate in time.

In short, this paper proves the security of the modified BB84 protocol with practical sources and shows its robustness to source imperfections, taking us a step closer towards ensuring the implementation security of QKD at a level that is suitable for practical applications.

## ACKNOWLEDGMENTS

This work was supported by Cisco Systems Inc., the Galician Regional Government (consolidation of Research Units: AtlantTIC), the Spanish Ministry of Economy and Competitiveness (MINECO), the Fondo Europeo de Desarrollo Regional (FEDER) through Grant No. PID2020-118178RB-C21 and MICIN with funding from the European Union NextGenerationEU (PRTR-C17.I1) and the Galician Regional Government with own funding through the "Planes Complementarios de I+D+I con las Comunidades Autónomas" in Quantum Communication. M.P. and G.C.-L. acknowledge support from JSPS Postdoctoral Fellowships for Research in Japan. A.M. acknowledges support from JST, ACT-X Grant No. JPMJAX210O. G.K. acknowledges support from JSPS Kakenhi (C) No. 20K03779 and No. 21K03388. K.T. acknowledges support from JSPS KAKENHI Grant No. JP18H05237 and JST-CREST JPMJCR 1671.

## APPENDIX A: DESCRIPTION OF THE MODIFIED BB84 QKD PROTOCOL

(1) *State preparation.* For each round $k \in \{1, \dots, N\}$, Alice selects the setting $j \in \{0_Z, 1_Z, 0_X, 1_X\}$ with probability $p_j$, generates an encoded pulse and sends it to Bob through the quantum channel.

(2) *Detection*. Bob measures each incoming pulse in the basis $\beta \in \{Z, X\}$ with probabilities $p_{Z_B}$ and $p_{X_B}$, respectively.

(3) *Sifting*. Bob announces which rounds were detected, and Alice and Bob reveal their basis choices in these rounds. Let $\mathcal{K}$ be the set of detected rounds in which both users employed the $Z$ basis and $\mathcal{T}$ be the set of detected rounds in which Bob employed the $X$ basis. Then, Alice and Bob define their own sifted keys as the bit values associated with her emissions and his measurement results on rounds $\mathcal{K}$, respectively. As for rounds $\mathcal{T}$, Alice and Bob announce their respective bit values.

(4) *Parameter estimation.* Alice quantifies the number of events $N_{j,\gamma_X}$ on rounds $\mathcal{T}$ for $j \in \{0_Z, 1_Z, 0_X, 1_X\}$ and $\gamma \in \{0, 1\}$, where $\gamma$ denotes Bob's measurement outcome. Then, using these quantities she obtains an upper bound $N_{\mathrm{ph}}^{\mathrm{U}}$ on $N_{\mathrm{ph}}$, the number of phase errors in her sifted key, using Eq. (24).

(5) *Data postprocessing*. For error correction, Alice sends Bob encrypted syndrome information [78] about her sifted key through an authenticated public channel, which Bob uses to correct his sifted key. For error verification, Alice and Bob compute a hash of their corrected keys using a random two-universal hash function and check if they are identical. If not, they abort the protocol; otherwise, Alice and Bob perform privacy amplification on their corrected keys. For this, they use a random two-universal hash function to extract an identical secret key pair.

In the presence of setting-dependent pulse correlations, the protocol described above requires slight modifications. First, before step (1), Alice and Bob need to assign tags to each round:

(0) *Tag assignment*. For each round $k \in \{1, \ldots, N\}$, Alice and Bob assign a tag $w \in \{0, \ldots, l_c\}$ according to the value $w = k \bmod (l_c + 1)$, where $l_c$ is the correlation length.

Then, Alice and Bob perform the steps (1)–(3), with $\mathcal{K}_w$ ($\mathcal{T}_w$) defined as the set of detected rounds with tag $w$ in which both users employed the $Z$ basis (Bob employed the $X$ basis). Hence, the $w$th sifted key is defined as the subset of the total sifted key originating from the rounds in $\mathcal{K}_w$. After that, the users perform parameter estimation and employ data postprocessing to obtain a final secure key:

(1) *Parameter estimation*. Alice quantifies the number of events $N_{j,\gamma_X,w}$ on rounds $\mathcal{T}_w$ for $j \in \{0_Z, 1_Z, 0_X, 1_X\}$ and $\gamma \in \{0, 1\}$. Then, using these quantities she obtains an upper bound $N^{\text{U}}_{\text{ph},w}$ on $N_{\text{ph},w}$, the number of phase errors in her $w$th sifted key, using Eq. (30).

(2) *Data postprocessing*. For error correction, Alice sends Bob encrypted syndrome information about her $w$th sifted key through an authenticated public channel, which Bob uses to correct his $w$th sifted key. For error verification, Alice and Bob compute a hash of their $w$th corrected keys using a random two-universal hash function and check if they are identical. If not, they discard their $w$th corrected keys; otherwise, Alice and Bob perform privacy amplification. For this, they use a random two-universal hash function to extract the $w$th secret key pair.

(3) *Key concatenation*. After repeating the steps (3)–(5) for all tags $w$, Alice and Bob define their final secret key pair as the concatenation of all the $w$th secret keys.

## APPENDIX B: TREATMENT OF THE SETTING-INDEPENDENT FACTORS IN THE SECURITY PROOF

Let $\boldsymbol{\mathcal{G}} = \mathcal{G}_1, \ldots, \mathcal{G}_N$ denote the random variables that represent the setting-independent factors affecting the form of Alice's emitted states, and let $\boldsymbol{g} = g_1, \ldots, g_N$ denote a particular value of these random variables. To prove the security of the actual protocol, we consider an equivalent entanglement-based scenario in which Alice prepares a entangled state with ancillary systems $\mathbf{A} := A_1, \ldots, A_N$ that she can measure to learn her setting choices. Since in the actual protocol Alice's emitted states depend on $\boldsymbol{g}$, this entangled state can be expressed as

$$\int_{\boldsymbol{g}\in\text{dom}(f_{\boldsymbol{\mathcal{G}}})} f_{\boldsymbol{\mathcal{G}}}(\boldsymbol{g}) |\Psi_{\boldsymbol{g}}\rangle\langle\Psi_{\boldsymbol{g}}|_{\mathbf{A},\mathbf{B},\mathbf{E}} d\boldsymbol{g}, \tag{B1}$$

where $f_{\boldsymbol{\mathcal{G}}}(\boldsymbol{g})$ is the probability density function of $\boldsymbol{\mathcal{G}}$, $\text{dom}(f_{\boldsymbol{\mathcal{G}}})$ is the domain of $f_{\boldsymbol{\mathcal{G}}}(\boldsymbol{g})$, and $\mathbf{B}, \mathbf{E} := B_1, E_1, \ldots, B_N, E_N$. In the setting-independent scenario, the state $|\Psi_{\boldsymbol{g}}\rangle_{\mathbf{A},\mathbf{B},\mathbf{E}}$ in Eq. (B1) takes the form

$$|\Psi_{\boldsymbol{g}}\rangle_{\mathbf{A},\mathbf{B},\mathbf{E}} = \bigotimes_{k=1}^{N} \sum_{j_k} \sqrt{p_{j_k}} |j_k\rangle_{A_k} |\psi_{j_k,\boldsymbol{g}}\rangle_{B_k,E_k}, \tag{B2}$$

where $|\psi_{j_k,\boldsymbol{g}}\rangle_{B_k,E_k}$ is defined in Eq. (1). Note that, for ease of discussion, in this Appendix we explicitly write the setting choice on the $k$th round as $j_k$, rather than $j$. Alternatively, in the setting-dependent scenario, the state $|\Psi_{\boldsymbol{g}}\rangle_{\mathbf{A},\mathbf{B},\mathbf{E}}$ in Eq. (B1) takes the form

$$|\Psi_{\boldsymbol{g}}\rangle_{\mathbf{A},\mathbf{B},\mathbf{E}} = \bigotimes_{k=1}^{N} \sum_{j_k} \sqrt{p_{j_k}} |j_k\rangle_{A_k} \left|\psi_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\right\rangle_{B_k,E_k}, \tag{B3}$$

where $|\psi_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,E_k}$ is defined in Eq. (2).

However, note that Eve cannot distinguish the scenario in which Alice prepares the mixed state in Eq. (B1) from the scenario in which she prepares its purification,

$$\int_{\boldsymbol{g}\in\text{dom}(f_{\boldsymbol{\mathcal{G}}})} \sqrt{f_{\boldsymbol{\mathcal{G}}}(\boldsymbol{g})} |\boldsymbol{g}\rangle_G |\Psi_{\boldsymbol{g}}\rangle_{\mathbf{A},\mathbf{B},\mathbf{E}} d\boldsymbol{g}, \tag{B4}$$

and then measures system $G$ in the very beginning of the protocol, obtaining some outcome $\boldsymbol{g}$ and post-measurement state $|\Psi_{\boldsymbol{g}}\rangle_{\mathbf{A},\mathbf{B},\mathbf{E}}$. For simplicity of presentation, in our security proof, we consider this equivalent scenario, in which the value of $\boldsymbol{g}$ has effectively become fixed before Alice emits any pulses to Bob. Then, in Sec. III A, we derive a bound $N^{\text{U}}_{\text{ph}}$ on the number of phase errors $N_{\text{ph}}$ that is in fact conditional on this value of $\boldsymbol{g}$, such that

$$\Pr[N_{\text{ph}} > N^{\text{U}}_{\text{ph}} | \boldsymbol{\mathcal{G}} = \boldsymbol{g}] \leqslant \varepsilon, \tag{B5}$$

where $\varepsilon$ is the failure probability. Importantly, however, this bound is valid for all possible values $\boldsymbol{g}$, as neither $N^{\text{U}}_{\text{ph}}$ nor $\varepsilon$ depend on $\boldsymbol{g}$. This implies that the bound in Eq. (B5) is also valid for the scenario in which Alice prepares the mixed state in Eq. (B1), since

$$\begin{aligned} \Pr\left[N_{\text{ph}} > N^{\text{U}}_{\text{ph}}\right] &= \int_{\boldsymbol{g}\in\text{dom}(f_{\boldsymbol{\mathcal{G}}})} f_{\boldsymbol{\mathcal{G}}}(\boldsymbol{g}) \Pr\left[N_{\text{ph}} > N^{\text{U}}_{\text{ph}} | \boldsymbol{\mathcal{G}} = \boldsymbol{g}\right] \\ &\leqslant \int_{\boldsymbol{g}\in\text{dom}(f_{\boldsymbol{\mathcal{G}}})} f_{\boldsymbol{\mathcal{G}}}(\boldsymbol{g}) \, \varepsilon = \varepsilon. \end{aligned} \tag{B6}$$

By a similar argument, we conclude that the phase-error rate bounds derived in Sec. III B also apply to the scenario in which Alice prepares the mixed state in Eq. (B1).

## APPENDIX C: TROJAN-HORSE ATTACKS

Here, we explicitly show how to incorporate THAs in our security analysis. In particular, we assume that Eve's injected light does not alter Alice's prepared signals, other than adding extra modes of light that contain information about Alice's setting choices. That is, in the presence of a THA, the total emitted state consists of a tensor product between the state that Alice would have emitted in Eve's absence and the back-reflected light from the THA. Moreover, we assume that the back-reflected light on round $k$ only contains information about the $k$th setting choice, and is in tensor product form with the back-reflected light from all the other rounds. In this context, we consider two cases for the THA: (1) the back-reflected light is a pure state, and (2) the back-reflected light is a mixed state. For simplicity of discussion, in what follows we assume that THAs are the only side channel present. However, one can easily combine these results with the other side channels considered in this paper, as shown below.

### 1. Pure output light

In this case, the state emitted on round $k$ can be written as $|\psi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} = |\omega_{j,\boldsymbol{g}}\rangle_{B_k} \otimes |E_{j,\boldsymbol{g}}\rangle_{E_k}$, where $|\omega_{j,\boldsymbol{g}}\rangle_{B_k}$ is a qubit state and

$$|E_{j,\boldsymbol{g}}\rangle_{E_k} = \sqrt{1-\epsilon^{(k)}}|v\rangle_{E_k} + \sqrt{\epsilon^{(k)}}|e_{j,\boldsymbol{g}}\rangle_{E_k}, \tag{C1}$$

is the back-reflected light from the THA. Here, $\epsilon^{(k)} \in [0, 1]$ quantifies the deviation of $|E_{j,\boldsymbol{g}}\rangle_{E_k}$ from the vacuum state $|v\rangle_{E_k}$, and $|e_{j,\boldsymbol{g}}\rangle_{E_k}$ is a setting-dependent state orthogonal to $|v\rangle_{E_k}$, i.e., it belongs to the Fock subspace $\{|1\rangle, |2\rangle, \ldots\}$. Note that $|E_{j,\boldsymbol{g}}\rangle_{E_k}$ could also represent other side channels, such as electromagnetic radiation. Using Eq. (C1), it is straightforward to write the $k$th emitted state in the form of Eq. (1),

$$|\psi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} = \sqrt{1-\epsilon^{(k)}}|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} + \sqrt{\epsilon^{(k)}}|\phi^{\perp}_{j,\boldsymbol{g}}\rangle_{B_k,E_k}, \tag{C2}$$

where the qubit state $|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} := |\omega_{j,\boldsymbol{g}}\rangle_{B_k}|v\rangle_{E_k}$ and the side-channel state $|\phi^{\perp}_{j,\boldsymbol{g}}\rangle_{B_k,E_k} := |\omega_{j,\boldsymbol{g}}\rangle_{B_k}|e_{j,\boldsymbol{g}}\rangle_{E_k}$; and consequently, apply our security analysis, given that an upper bound on $\epsilon^{(k)}$ is known.

We remark that the inclusion of side channels other than THAs can be readily accommodated as follows. Suppose that, in the absence of a THA, the state of Alice's prepared signal on round $k$ is the following:

$$|\Omega_{j,\boldsymbol{g}}\rangle_{B_k} = \sqrt{1-\tilde{\epsilon}^{(k)}_{j,\boldsymbol{g}}}|\omega_{j,\boldsymbol{g}}\rangle_{B_k} + \sqrt{\tilde{\epsilon}^{(k)}_{j,\boldsymbol{g}}}|\omega^{\perp}_{j,\boldsymbol{g}}\rangle_{B_k}, \tag{C3}$$

where $|\omega^{\perp}_{j,\boldsymbol{g}}\rangle_{B_k}$ is a side-channel state that lives in any Hilbert space orthogonal to $|\omega_{j,\boldsymbol{g}}\rangle_{B_k}$. Note that the state in Eq. (C3) includes any kind of mode dependencies with $|\omega_{j,\boldsymbol{g}}\rangle_{B_k}$ indicating the desired mode. Then, in the presence of a THA, the $k$th emitted state would simply become $|\psi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} = |\Omega_{j,\boldsymbol{g}}\rangle_{B_k} \otimes |E_{j,\boldsymbol{g}}\rangle_{E_k}$, which can be written in the form of Eq. (1),

$$\begin{aligned}|\psi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} =& \sqrt{1-\tilde{\epsilon}^{(k)}_{j,\boldsymbol{g}}}\sqrt{1-\epsilon^{(k)}}|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k}\\ &+ \sqrt{1-\big(1-\tilde{\epsilon}^{(k)}_{j,\boldsymbol{g}}\big)(1-\epsilon^{(k)})}|\phi^{\perp}_{j,\boldsymbol{g}}\rangle_{B_k,E_k},\end{aligned} \tag{C4}$$

where $|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} := |\omega_{j,\boldsymbol{g}}\rangle_{B_k}|v\rangle_{E_k}$ and

$$\begin{aligned}|\phi^{\perp}_{j,\boldsymbol{g}}\rangle_{B_k,E_k} :=& \Big[\sqrt{1-\tilde{\epsilon}^{(k)}_{j,\boldsymbol{g}}}\sqrt{\epsilon^{(k)}}|\omega_{j,\boldsymbol{g}}\rangle_{B_k}|e_{j,\boldsymbol{g}}\rangle_{E_k}\\ &+\sqrt{\tilde{\epsilon}^{(k)}_{j,\boldsymbol{g}}}|\omega^{\perp}_{j,\boldsymbol{g}}\rangle_{B_k}|E_{j,\boldsymbol{g}}\rangle_{E_k}\Big]\Big/\sqrt{1-\big(1-\tilde{\epsilon}^{(k)}_{j,\boldsymbol{g}}\big)(1-\epsilon^{(k)})}.\end{aligned} \tag{C5}$$

In a similar way, if in the absence of a THA, Alice's prepared signal on round $k$ is the following:

$$\begin{aligned}|\Omega_{j,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k} =& \sqrt{1-\tilde{\epsilon}^{(k)}_{j,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}|\omega_{j,\boldsymbol{g}}\rangle_{B_k}\\ &+ \sqrt{\tilde{\epsilon}^{(k)}_{j,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}\big|\omega^{\perp}_{j,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\big\rangle_{B_k},\end{aligned} \tag{C6}$$

due to setting-dependent pulse correlations and mode dependencies; then, in the presence of a THA, the $k$th emitted state would become $|\psi_{j,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,E_k} = |\Omega_{j,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k} \otimes |E_{j,\boldsymbol{g}}\rangle_{E_k}$, which can be written in the form of Eq. (2).

*Experimental method to upper bound $\epsilon^{(k)}$.* Here, we show how one could upper bound $\epsilon^{(k)}$ by using only information about the output light intensity (average number of photons). To obtain this information, one needs to first determine the maximum light intensity that Eve could inject into Alice's source without being detected. Then, by characterizing the value of the attenuation in Alice's setup [23], one can bound the maximum intensity of the output light $\nu_{\max}$. Mathematically, this bound can be expressed as $\langle E_{j,\boldsymbol{g}}|\hat{N}|E_{j,\boldsymbol{g}}\rangle_{E_k} \leqslant \nu_{\max}$ where $\hat{N} = \sum_n n|n\rangle\langle n|$ is the photon-number operator. Then, using Eq. (C1), we have that

$$\nu_{\max} \geqslant \langle E_{j,\boldsymbol{g}}|\hat{N}|E_{j,\boldsymbol{g}}\rangle_{E_k} = \epsilon^{(k)}\langle e_{j,\boldsymbol{g}}|\hat{N}|e_{j,\boldsymbol{g}}\rangle_{E_k} \geqslant \epsilon^{(k)}, \tag{C7}$$

with equality if and only if $|e_{j,\boldsymbol{g}}\rangle_{E_k}$ is a single-photon state. We note that one could obtain slightly tighter bounds if one had more information on the back-reflected light, such as it being a coherent state [35].

### 2. Mixed output light

In this case, we assume that the back-reflected light system $\tilde{E}_k$ is purified by an ancillary system $E'_k$ that is in Eve's hands [79]. We can then express the joint state of the back-reflected light and this ancillary system as

$$|E_{j,\boldsymbol{g}}\rangle_{E_k} = \sum_c \sqrt{p_c}|c\rangle_{E'_k}\Big(\sqrt{1-\epsilon^{(k)}_c}|v\rangle_{\tilde{E}_k} + \sqrt{\epsilon^{(k)}_c}|e_{j,\boldsymbol{g},c}\rangle_{\tilde{E}_k}\Big), \tag{C8}$$

where $E_k := E'_k, \tilde{E}_k$ and $\{|c\rangle\}$ forms an orthonormal basis. If $\epsilon^{(k)}_c = \epsilon^{(k)}$ for all $c$, it is straightforward to write the emitted state on round $k$ in the form of Eq. (C2), with $|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} = |\omega_{j,\boldsymbol{g}}\rangle_{B_k}|\lambda\rangle_{E_k}$, where $|\lambda\rangle_{E_k} := \sum_c \sqrt{p_c}|c\rangle_{E'_k}|v\rangle_{\tilde{E}_k}$, and $|\phi^{\perp}_{j,\boldsymbol{g}}\rangle_{B_k,E_k} := |\omega_{j,\boldsymbol{g}}\rangle_{B_k}\sum_c \sqrt{p_c}|c\rangle_{E'_k}|e_{j,\boldsymbol{g},c}\rangle_{\tilde{E}_k}$. If $\epsilon^{(k)}_c \leqslant \epsilon^{(k)}$ for all $c$, then the worst case scenario is such that the bound is saturated for all $c$, and one can assume that $\epsilon^{(k)}_c = \epsilon^{(k)}$. In both of these cases, one can directly use the experimental bound $\nu_{\max} \geqslant \epsilon^{(k)}$ in Eq. (C7) and apply our security analysis.

However, even if the conditions above do not hold, our security analysis can still be applied as long as we have an upper bound on $\sum_c p_c\epsilon^{(k)}_c$, which, as we will show below, can be related to $\nu_{\max}$. To use the RT in this case, one can select the reference states as follows:

$$\begin{aligned}|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} &= |\omega_{j,\boldsymbol{g}}\rangle_{B_k}|\lambda\rangle_{E_k}\\ &:= \frac{|\omega_{j,\boldsymbol{g}}\rangle_{B_k}\sum_c \sqrt{p_c}|c\rangle_{E'_k}\sqrt{1-\epsilon^{(k)}_c}|v\rangle_{\tilde{E}_k}}{\sqrt{\sum_c p_c\big(1-\epsilon^{(k)}_c\big)}},\end{aligned} \tag{C9}$$

for $j \in \{0_Z, 1_Z, 0_X, 1_X\}$. Then, the required inner products $|\langle\psi_{j,\boldsymbol{g}}|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k}|$ can be calculated as

$$|\langle\psi_{j,\boldsymbol{g}}|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k}| = \sqrt{1-\sum_c p_c\epsilon^{(k)}_c}, \tag{C10}$$

where we have used the fact that $\langle\phi^{\perp}_{j,\boldsymbol{g}}|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} = 0$ for any $j$.

*Experimental method to upper bound $\sum_c p_c\epsilon^{(k)}_c$.* Similarly to the pure state case, we can obtain a bound on $\sum_c p_c\epsilon^{(k)}_c$ using only information about the output light intensity. In particular, from Eq. (C8), we have that

$$\begin{aligned}\nu_{\max} &\geqslant \langle E_{j,\boldsymbol{g}}|\hat{\mathbb{1}}_{E'_k} \otimes \hat{N}_{\tilde{E}_k}|E_{j,\boldsymbol{g}}\rangle_{E_k}\\ &= \sum_c p_c\epsilon^{(k)}_c\langle e_{j,\boldsymbol{g},c}|\hat{N}_{\tilde{E}_k}|e_{j,\boldsymbol{g},c}\rangle_{\tilde{E}_k} \geqslant \sum_c p_c\epsilon^{(k)}_c,\end{aligned} \tag{C11}$$

with equality if $|e_{j,\boldsymbol{g},c}\rangle_{\tilde{E}_k}$ is a single-photon state.

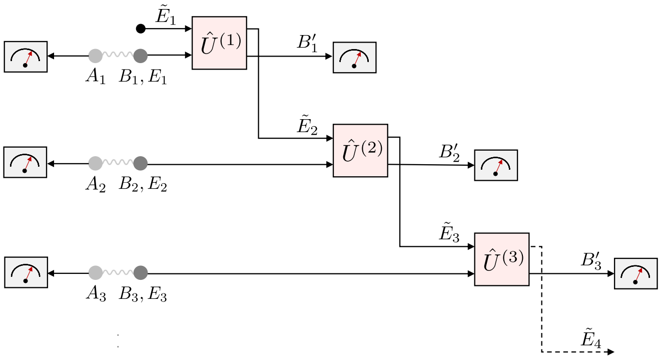

FIG. 3. Diagram of the entanglement-based scenario and Eve's most general coherent attack when the QKD protocol is run sequentially. Eve must perform her attack on the $k$th pulse before she learns information about systems $B_{k+1}, E_{k+1}$, and therefore $\hat{U}^{(k)}$ cannot take systems $B_{k+1}, E_{k+1}$ as an input. Conversely, Eve's attack on the $k$th pulse can depend on information that she has learned about the systems $B_{k-1}, E_{k-1}$. This is why $\hat{U}^{(k)}$ takes Eve's updated ancilla $\tilde{E}_k$ as an input.

## APPENDIX D: PROOF OF Eq. (6)

Here, we prove Eq. (6) and show that when the QKD protocol is run sequentially [see Assumption (A4) in Sec. II A], the operator $\hat{M}^{(k)}_{\gamma_X}$ satisfies $0 \leqslant \hat{M}^{(k)}_{\gamma_X} \leqslant \hat{\mathbb{1}}$. To do so, we consider an entanglement-based scenario that is equivalent to the actual protocol. For a fixed $\boldsymbol{g}$ (see Appendix B), the transmission of $N$ pulses can then be described by Alice first preparing the following entangled state:

$$\begin{aligned}|\Psi_{\boldsymbol{g}}\rangle_{\mathbf{A},\mathbf{B},\mathbf{E}} &= \bigotimes_{k=1}^{N} \sum_{j_k} \sqrt{p_{j_k}} |j_k\rangle_{A_k} |\psi_{j_k,\boldsymbol{g}}\rangle_{B_k,E_k} \\ &=: \bigotimes_{k=1}^{N} |\Psi'_{\boldsymbol{g}}\rangle_{A_k,B_k,E_k}, \end{aligned} \tag{D1}$$

then keeping system $\mathbf{A}$ in her laboratory and sending systems $\mathbf{B}, \mathbf{E}$ through the quantum channel. Here, $\mathbf{A} := A_1, \dots, A_N$ and $\mathbf{B}, \mathbf{E} := B_1, E_1, \dots, B_N, E_N$ refers to the composite systems of Alice's ancillae and to the pulses sent to Bob, respectively, where $A_k$ and $B_k$ for $k \in \{1, \dots, N\}$ denote Alice's and Bob's $k$th systems and $E_k$ denotes any other systems that are emitted by Alice on round $k$, such as the back-reflected light from a THA (see Appendix C for more details). Also, in Eq. (D1), $\{|j_k\rangle_{A_k}\}_{j_k \in \{0_Z, 1_Z, 0_X, 1_X\}}$ is a set of orthonormal states and $|\psi_{j_k,\boldsymbol{g}}\rangle_{B_k,E_k}$ is defined in Eq. (1). Note that, for ease of discussion, in this Appendix we explicitly write the setting choice on the $k$th round as $j_k$, rather than $j$.

As already discussed, in our analysis we assume that the QKD protocol is run sequentially such that Alice only generates the $k$th pulse once Bob has obtained his $k-1$th measurement result. Under this restriction, Eve's most general coherent attack can be described as the application of a sequence of $N$ unitary operators $\hat{U}^{(N)}_{B_N,E_N,\tilde{E}_N} \dots \hat{U}^{(1)}_{B_1,E_1,\tilde{E}_1}$, where $\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}$ acts on the $k$th photonic system $B_k, E_k$ and on Eve's updated ancilla $\tilde{E}_k$, resulting in systems $B'_k$ and $\tilde{E}_k$, respectively, as shown in Fig. 3.

After Eve's attack on each round, Alice and Bob measure their local systems. First, Alice measures system $A_k$ to know if $j_k \in \{0_Z, 1_Z\}$ or if $j_k \in \{0_X, 1_X\}$, whose respective outcomes are denoted by $a = Z$ and $a = X$, and Bob selects the measurement basis $\beta \in \{Z, X\}$ with probability $p_{\beta_B}$. In the modified BB84 protocol, the secret key is generated from the rounds in which $(a, \beta) = (Z, Z)$. As explained in Sec. III A, to prove the security of these rounds, one needs to consider the number of phase errors that Alice and Bob would have observed if they had performed their local measurements in the phase basis instead, i.e., $\{|\tau_X^{(\mathrm{vir})}\rangle_{A_k} = (|0_Z\rangle_{A_k} + (-1)^{\tau}|1_Z\rangle_{A_k})/\sqrt{2}\}_{\tau \in \{0,1\}}$ and $\{\hat{m}_{0_X}, \hat{m}_{1_X}, \hat{m}_f\}$, respectively. We can then define a virtual protocol in which, in the key generation rounds, Alice and Bob perform these measurements, and in all the other rounds they perform measurements in their selected basis $(a, \beta)$.

In this virtual protocol, let us denote all possible outcomes for round $k$ as $o_k \in \{(a, \beta, \tau, \gamma), (f)\}$, where $\tau, \gamma \in \{0, 1\}$ are Alice's and Bob's observed bit values, respectively, and $f$ is associated with an inconclusive outcome. Now, let us define the POVM element associated with obtaining the outcome $o_k$ by $\hat{F}^{o_k}_{A_k,B'_k}$. Then, we can summarize the possible POVM elements as

$$\begin{aligned}\hat{F}^{(Z,Z,\tau,\gamma)}_{A_k,B'_k} &= \big|\tau_X^{(\mathrm{vir})}\big\rangle\!\big\langle\tau_X^{(\mathrm{vir})}\big|_{A_k} \otimes p_{Z_B}\hat{m}_{\gamma_X}, \\ \hat{F}^{(Z,X,\tau,\gamma)}_{A_k,B'_k} &= |\tau_Z\rangle\langle\tau_Z|_{A_k} \otimes p_{X_B}\hat{m}_{\gamma_X}, \\ \hat{F}^{(X,\beta,\tau,\gamma)}_{A_k,B'_k} &= |\tau_X\rangle\langle\tau_X|_{A_k} \otimes p_{\beta_B}\hat{m}_{\gamma_\beta}, \\ \hat{F}^{(f)}_{A_k,B'_k} &= \hat{\mathbb{1}}_{A_k} \otimes \hat{m}_f. \end{aligned} \tag{D2}$$

Note that $\hat{F}^{(Z,Z,\tau,\gamma)}_{A_k,B'_k}$ is the POVM element associated with the estimation of the number of phase errors. A phase error on the $k$th round can then be defined as obtaining the outcome $o_k = (Z, Z, 0, 1)$ or $o_k = (Z, Z, 1, 0)$. Therefore, the probability of a phase error on the $k$th round conditional on all the previous outcomes $o_{\overrightarrow{k-1}}$ can be expressed as

$$\begin{aligned}P^{(k)}_{\boldsymbol{g}}(\mathrm{ph}|\mathrm{Act}) &= P(o_k = (Z, Z, 0, 1) \text{ or } (Z, Z, 1, 0)|o_{\overrightarrow{k-1}}) \\ &= \frac{P((o_k = (Z, Z, 0, 1) \text{ or } (Z, Z, 1, 0)), o_{\overrightarrow{k-1}})}{P(o_{\overrightarrow{k-1}})}, \end{aligned} \tag{D3}$$

where $P(o_{\overrightarrow{k-1}})$ is the probability of obtaining Alice's and Bob's previous $k-1$ outcomes. Note that this conditional probability on all the previous outcomes is required to take into account any correlations between the measurement outcomes of different rounds of the protocol, which could arise due to Eve's coherent attack. Because of this dependence, Azuma's or Kato's inequality can be employed in order to estimate the total number of phase errors in the protocol.

Now, using the form of the full emitted state in Eq. (D1), the description of Eve's coherent attack when the QKD protocol is run sequentially, and the POVM elements in Eq. (D2), we can mathematically express the time evolution illustrated in Fig. 3 and calculate the numerator of Eq. (D3) as follows:

$$
\begin{aligned}
&P((o_k=(\mathrm{Z},\mathrm{Z},0,1)\text{ or }(\mathrm{Z},\mathrm{Z},1,0)),o_{\overrightarrow{k-1}})\\
&=\sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}}\mathrm{Tr}\left[\hat{F}^{(\mathrm{Z},\mathrm{Z},\tau,\gamma)}_{A_k,B'_k}\hat{P}\left(\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}\prod_{n=1}^{k-1}\sqrt{\hat{F}^{o_{k-n}}_{A_{k-n},B'_{k-n}}}\hat{U}^{(k-n)}_{B_{k-n},E_{k-n},\tilde{E}_{k-n}}|0\rangle_{\tilde{E}_1}\bigotimes_{n=1}^{k-1}|\Psi'_{\boldsymbol{g}}\rangle_{A_n,B_n,E_n}|\Psi'_{\boldsymbol{g}}\rangle_{A_k,B_k,E_k}\bigotimes_{n=k+1}^{N}|\Psi'_{\boldsymbol{g}}\rangle_{A_n,B_n,E_n}\right)\right]\\
&=\sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}}\mathrm{Tr}\left[\hat{F}^{(\mathrm{Z},\mathrm{Z},\tau,\gamma)}_{A_k,B'_k}\hat{P}\left(\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}\prod_{n=1}^{k-1}\sqrt{\hat{F}^{o_{k-n}}_{A_{k-n},B'_{k-n}}}\hat{U}^{(k-n)}_{B_{k-n},E_{k-n},\tilde{E}_{k-n}}|0\rangle_{\tilde{E}_1}\bigotimes_{n=1}^{k-1}|\Psi'_{\boldsymbol{g}}\rangle_{A_n,B_n,E_n}|\Psi'_{\boldsymbol{g}}\rangle_{A_k,B_k,E_k}\right)\right]\\
&=\sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}}\mathrm{Tr}\Big[{}_{\mathbf{A}_{\leqslant k-1},\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1},\tilde{E}_k}\big\langle\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big|\hat{U}^{(k)\dagger}_{B_k,E_k,\tilde{E}_k}\hat{F}^{(\mathrm{Z},\mathrm{Z},\tau,\gamma)}_{A_k,B'_k}\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}\big|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big\rangle_{\mathbf{A}_{\leqslant k-1},\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1},\tilde{E}_k}|\Psi'_{\boldsymbol{g}}\rangle\langle\Psi'_{\boldsymbol{g}}|_{A_k,B_k,E_k}\Big]\\
&=\sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}}\mathrm{Tr}\Big[{}_{\mathbf{A}_{\leqslant k-1},\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1},\tilde{E}_k}\big\langle\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big|\hat{U}^{(k)\dagger}_{B_k,E_k,\tilde{E}_k}\hat{m}_{\gamma_X}\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}\big|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big\rangle_{\mathbf{A}_{\leqslant k-1},\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1},\tilde{E}_k}\tilde{p}^{(k,\mathrm{vir})}_{\tau_X,\boldsymbol{g}}p_{Z_B}\big|\psi^{(\mathrm{vir})}_{\tau_X,\boldsymbol{g}}\big\rangle\big\langle\psi^{(\mathrm{vir})}_{\tau_X,\boldsymbol{g}}\big|_{B_k,E_k}\Big],
\end{aligned}
\tag{D4}
$$

where

$$
\begin{aligned}
&\big|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big\rangle_{\mathbf{A}_{\leqslant k-1},\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1},\tilde{E}_k}\\
&\quad:=\prod_{n=1}^{k-1}\sqrt{\hat{F}^{o_{k-n}}_{A_{k-n},B'_{k-n}}}\hat{U}^{(k-n)}_{B_{k-n},E_{k-n},\tilde{E}_{k-n}}|0\rangle_{\tilde{E}_1}\bigotimes_{n=1}^{k-1}|\Psi'_{\boldsymbol{g}}\rangle_{A_n,B_n,E_n},
\end{aligned}
\tag{D5}
$$

with $\mathbf{A}_{\leqslant k-1}:=A_1,\dots,A_{k-1}$ and $\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1}:=B'_1,E_1,\dots,B'_{k-1},E_{k-1}$, and where $|\psi^{(\mathrm{vir})}_{\tau_X,\boldsymbol{g}}\rangle_{B_k,E_k}$ and $\tilde{p}^{(k,\mathrm{vir})}_{\tau_X,\boldsymbol{g}}$ are defined in Eqs. (4) and (5), respectively. In the second equality of Eq. (D4), we took the partial trace over systems $\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+1},\mathbf{E}_{\geqslant k+1}$. In the third equality of Eq. (D4), we defined the state $|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\rangle_{\mathbf{A}_{\leqslant k-1},\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1},\tilde{E}_k}$ using Eq. (D5) and used the cyclic property of the trace operation. Finally, in the last equality of Eq. (D4), we used Eq. (3) and the definition of $\hat{F}^{(\mathrm{Z},\mathrm{Z},\tau,\gamma)}_{A_k,B'_k}$ in Eq. (D2) for $\tau,\gamma\in\{0,1\}$ such that $\tau\neq\gamma$.

By substituting Eq. (D4) in Eq. (D3), we can now express $P^{(k)}_{\boldsymbol{g}}(\mathrm{ph}|\mathrm{Act})$ as

$$
P^{(k)}_{\boldsymbol{g}}(\mathrm{ph}|\mathrm{Act})=\sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}}\tilde{p}^{(k,\mathrm{vir})}_{\tau_X,\boldsymbol{g}}p_{Z_B}\mathrm{Tr}\big[\tilde{\sigma}^{(k,\mathrm{vir})}_{\tau_X,\boldsymbol{g}}\hat{M}^{(k)}_{\gamma_X}\big],
\tag{D6}
$$

with $\tilde{\sigma}^{(k,\mathrm{vir})}_{\tau_X,\boldsymbol{g}}:=|\psi^{(\mathrm{vir})}_{\tau_X,\boldsymbol{g}}\rangle\langle\psi^{(\mathrm{vir})}_{\tau_X,\boldsymbol{g}}|_{B_k,E_k}$ and

$$
\hat{M}^{(k)}_{\gamma_X}:=\frac{\big\langle\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big|\hat{U}^{(k)\dagger}_{B_k,E_k,\tilde{E}_k}\hat{m}_{\gamma_X}\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}\big|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big\rangle}{P(o_{\overrightarrow{k-1}})},
\tag{D7}
$$

where we have omitted the mode subscripts in the quantum states for simplicity of notation. This concludes the derivation of Eq. (6).

Next, we prove that in this case, the operator $\hat{M}^{(k)}_{\gamma_X}$ in Eq. (D7) satisfies $0\leqslant\hat{M}^{(k)}_{\gamma_X}\leqslant\hat{\mathbb{1}}$. For this, first note that $\hat{M}^{(k)}_{\gamma_X}\geqslant 0$ holds because for any state $|\varphi\rangle$ we have that $\langle\varphi|\hat{M}^{(k)}_{\gamma_X}|\varphi\rangle=||\sqrt{\hat{m}_{\gamma_X}}\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\rangle|\varphi\rangle P(o_{\overrightarrow{k-1}})^{-1/2}||^2\geqslant 0$. Then, since $\hat{m}_{\gamma_X}$ is a POVM element, we have that $\hat{m}_{\gamma_X}\leqslant\hat{\mathbb{1}}$, and thus we can upper bound $\hat{M}^{(k)}_{\gamma_X}$ as

$$
\begin{aligned}
\hat{M}^{(k)}_{\gamma_X}&\leqslant\frac{\big\langle\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big|\hat{U}^{(k)\dagger}_{B_k,E_k,\tilde{E}_k}\hat{\mathbb{1}}\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}\big|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big\rangle}{P(o_{\overrightarrow{k-1}})}\\
&=\frac{\big\|\big|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big\rangle\big\|^2\hat{\mathbb{1}}_{B_k,E_k}}{P(o_{\overrightarrow{k-1}})}=\frac{P(o_{\overrightarrow{k-1}})\hat{\mathbb{1}}_{B_k,E_k}}{P(o_{\overrightarrow{k-1}})}=\hat{\mathbb{1}}_{B_k,E_k}.
\end{aligned}
\tag{D8}
$$

Therefore, by combining these two arguments, we have that, when the QKD protocol is run sequentially, $0\leqslant\hat{M}^{(k)}_{\gamma_X}\leqslant\hat{\mathbb{1}}$, as required.

## APPENDIX E: RESULTS USED IN Sec. III B

### 1. Derivation of $|\tilde{\psi}_{j_k,\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}\rangle_{B_k,E'_k}$ and Eq. (26)

Here, we show how to derive $|\tilde{\psi}_{j_k,\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}\rangle_{B_k,\boldsymbol{E}'_k}$ and Eq. (26), which are required to prove the security of the modified BB84 protocol in the presence of setting-dependent pulse correlations. To do so, we consider an entanglement-based picture of the protocol. In this case, for a fixed $\boldsymbol{g}$ (see Appendix B), the transmission of $N$ pulses is described by Alice first preparing $N$ ancilla systems and $N$ pulses in the state

$$
\begin{aligned}
|\Psi_{\boldsymbol{g}}\rangle_{\mathbf{A},\mathbf{B},\mathbf{E}}=&\sum_{j_1}\sqrt{p_{j_1}}|j_1\rangle_{A_1}\big|\psi_{j_1,\boldsymbol{g}}\big\rangle_{B_1,E_1}\\
&\otimes\sum_{j_2}\sqrt{p_{j_2}}|j_2\rangle_{A_2}\big|\psi_{j_2,\boldsymbol{g}|j_1}\big\rangle_{B_2,E_2}\otimes\dots\\
&\otimes\sum_{j_N}\sqrt{p_{j_N}}|j_N\rangle_{A_N}\big|\psi_{j_N,\boldsymbol{g}|j_{N-1},\dots,j_{N-l_c}}\big\rangle_{B_N,E_N},
\end{aligned}
\tag{E1}
$$

and then sending systems $\mathbf{B},\mathbf{E}$ through the quantum channel. Here, $\mathbf{A}:=A_1,\dots,A_N$ and $\mathbf{B},\mathbf{E}:=B_1,E_1,\dots,B_N,E_N$ refers to the composite systems of Alice's ancillae and to the pulses sent to Bob, respectively, where $A_k$ and $B_k$ for $k\in\{1,2,\dots,N\}$ denote Alice's and Bob's $k$th systems, and $E_k$ denotes any other systems sent by Alice through the quantum channel. In Eq. (E1), $\{|j_k\rangle_{A_k}\}_{j_k\in\{0_Z,1_Z,0_X,1_X\}}$ is a set of orthonormal states and $|\psi_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,E_k}$ is defined in Eq. (2). Note that, for ease of discussion, in this Appendix we explicitly write the setting choice on the $k$th round as $j_k$, rather than $j$.

We are interested in the post-measurement state after Alice has measured her ancillae $\mathbf{A}_{\leqslant k-1} := A_1, \ldots, A_{k-1}$. Once we trace out the systems $\mathbf{A}_{\leqslant k-1}, \mathbf{B}_{\leqslant k-1}, \mathbf{E}_{\leqslant k-1}$, this post-measurement state can be expressed as

$$\begin{aligned}
&\left|\Psi_{\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\right\rangle_{A_k,B_k,\mathbf{E}'_k}\\
&:= \sum_{j_k}\sqrt{p_{j_k}}|j_k\rangle_{A_k}\left|\psi_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\right\rangle_{B_k,E_k}\bigotimes_{n=1}^{l_c}\sum_{j_{k+n}}\sqrt{p_{j_{k+n}}}|j_{k+n}\rangle_{A_{k+n}}\left|\psi_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}\right\rangle_{B_{k+n},E_{k+n}}\\
&\quad\otimes\sum_{j_{k+l_c+1}}\sqrt{p_{j_{k+l_c+1}}}\left|j_{k+l_c+1}\right\rangle_{A_{k+l_c+1}}\left|\psi_{j_{k+l_c+1},\boldsymbol{g}|j_{k+l_c},\ldots,j_{k+1}}\right\rangle_{B_{k+l_c+1},E_{k+l_c+1}}\cdots\sum_{j_N}\sqrt{p_{j_N}}|j_N\rangle_{A_N}\left|\psi_{j_N,\boldsymbol{g}|j_{N-1},\ldots,j_{N-l_c}}\right\rangle_{B_N,E_N},
\end{aligned}\tag{E2}$$

where $j_{k-1}, \ldots, j_{k-l_c}$ denotes Alice's measurement outcomes on systems $A_{k-1}, \ldots, A_{k-l_c}$, and $\mathbf{E}'_k := \mathbf{A}_{\geqslant k+1}, \mathbf{B}_{\geqslant k+1}, \mathbf{E}_{\geqslant k}$ with $\mathbf{A}_{\geqslant k+1}$ and $\mathbf{B}_{\geqslant k+1}, \mathbf{E}_{\geqslant k}$ defined as $A_{k+1}, \ldots, A_N$ and $E_k, B_{k+1}, E_{k+1}, \ldots, B_N, E_N$, respectively. Note that, to derive Eq. (E2), we need to assume that the outcomes $j_{k-1}, \ldots, j_{k-l_c} \in \{0_Z, 1_Z, 0_X, 1_X\}$, i.e., that Alice has measured her ancillae $A_{k-1}, \ldots, A_{k-l_c}$ in the basis $\{|0_Z\rangle, |1_Z\rangle, |0_X\rangle, |1_X\rangle\}$. This has important consequences when using the complementary approach to prove the security of the protocol; see discussion after Eq. (26) in the main text.

In Eq. (E2), one can explicitly see that some information about the $k$th setting choice is leaked to the subsequent $l_c$ pulses. However, the pulses $k + l_c + 1$ to $N$ are independent of $j_k$. To simplify the description in Eq. (E2), we now introduce two definitions. First, we define a state that is independent of $j_k$ as

$$\begin{aligned}
\left|\psi_{j_{k+1},\ldots,j_{k+l_c},\boldsymbol{g}}\right\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+l_c+1},\mathbf{E}_{\geqslant k+l_c+1}} &:= |j_{k+1}\rangle_{A_{k+1}}\cdots|j_{k+l_c}\rangle_{A_{k+l_c}}\sum_{j_{k+l_c+1}}\sqrt{p_{j_{k+l_c+1}}}|j_{k+l_c+1}\rangle_{A_{k+l_c+1}}\left|\psi_{j_{k+l_c+1},\boldsymbol{g}|j_{k+l_c},\ldots,j_{k+1}}\right\rangle_{B_{k+l_c+1},E_{k+l_c+1}}\\
&\quad\cdots\sum_{j_N}\sqrt{p_{j_N}}|j_N\rangle_{A_N}\left|\psi_{j_N,\boldsymbol{g}|j_{N-1},\ldots,j_{N-l_c}}\right\rangle_{B_N,E_N},
\end{aligned}\tag{E3}$$

which forms a set of orthogonal states as $\{|\psi_{j_{k+1},\ldots,j_{k+l_c},\boldsymbol{g}}\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+l_c+1},\mathbf{E}_{\geqslant k+l_c+1}}\}_{j_{k+1},\ldots,j_{k+l_c}\in\{0_Z,1_Z,0_X,1_X\}}$. Next, using Eq. (E3), we define the following state:

$$\left|\lambda_{j_k,\ldots,j_{k+1-l_c},\boldsymbol{g}}\right\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+1},\mathbf{E}_{\geqslant k+1}} := \sum_{j_{k+1}}\sqrt{p_{j_{k+1}}}\cdots\sum_{j_{k+l_c}}\sqrt{p_{j_{k+l_c}}}\left|\psi_{j_{k+1},\ldots,j_{k+l_c},\boldsymbol{g}}\right\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+l_c+1},\mathbf{E}_{\geqslant k+l_c+1}}\bigotimes_{n=1}^{l_c}\left|\psi_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}\right\rangle_{B_{k+n},E_{k+n}}.\tag{E4}$$

Note that $|\lambda_{j_k,\ldots,j_{k+1-l_c},\boldsymbol{g}}\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+1},\mathbf{E}_{\geqslant k+1}}$ depends on $j_k$ even though it does not include system $B_k$. This means that it is effectively a side channel to the $k$th pulse. Finally, using Eq. (E4), we can rewrite Eq. (E2) as

$$\left|\Psi_{\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\right\rangle_{A_k,B_k,\mathbf{E}'_k} = \sum_{j_k}\sqrt{p_{j_k}}|j_k\rangle_{A_k}\left|\psi_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\right\rangle_{B_k,E_k}\left|\lambda_{j_k,\ldots,j_{k+1-l_c},\boldsymbol{g}}\right\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+1},\mathbf{E}_{\geqslant k+1}}.\tag{E5}$$

Now, it is useful to decompose $|\psi_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,E_k}$ $|\lambda_{j_k,\ldots,j_{k+1-l_c},\boldsymbol{g}}\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+1},\mathbf{E}_{\geqslant k+1}}$ in Eq. (E5). For this, we first combine Eqs. (E5) and (E4), and then use Eq. (2) such that

$$\begin{aligned}
&\left|\psi_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\right\rangle_{B_k,E_k}\sum_{j_{k+1}}\sqrt{p_{j_{k+1}}}\cdots\sum_{j_{k+l_c}}\sqrt{p_{j_{k+l_c}}}\left|\psi_{j_{k+1},\ldots,j_{k+l_c},\boldsymbol{g}}\right\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+l_c+1},\mathbf{E}_{\geqslant k+l_c+1}}\bigotimes_{n=1}^{l_c}\left|\psi_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}\right\rangle_{B_{k+n},E_{k+n}}\\
&= \left(\sqrt{1-\epsilon^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}\left|\phi_{j_k,\boldsymbol{g}}\right\rangle_{B_k,E_k} + \sqrt{\epsilon^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}\left|\phi^{\perp}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\right\rangle_{B_k,E_k}\right)\\
&\quad\otimes\sum_{j_{k+1}}\sqrt{p_{j_{k+1}}}\cdots\sum_{j_{k+l_c}}\sqrt{p_{j_{k+l_c}}}\left|\psi_{j_{k+1},\ldots,j_{k+l_c},\boldsymbol{g}}\right\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+l_c+1},\mathbf{E}_{\geqslant k+l_c+1}}\\
&\quad\otimes\bigotimes_{n=1}^{l_c}\left(\sqrt{1-\epsilon^{(k)}_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}}\left|\phi_{j_{k+n},\boldsymbol{g}}\right\rangle_{B_{k+n},E_{k+n}} + \sqrt{\epsilon^{(k)}_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}}\left|\phi^{\perp}_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}\right\rangle_{B_{k+n},E_{k+n}}\right)\\
&=: \sqrt{1-\tilde{\epsilon}^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}\left|\tilde{\phi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k+1-l_c}}\right\rangle_{B_k,\mathbf{E}'_k} + \sqrt{\tilde{\epsilon}^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}\left|\tilde{\phi}^{\perp}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\right\rangle_{B_k\mathbf{E}'_k}\\
&=: \left|\tilde{\psi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\right\rangle_{B_k\mathbf{E}'_k}.
\end{aligned}\tag{E6}$$

In this equation, we have defined

$$\tilde{\epsilon}^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}} := 1 - \big(1 - \epsilon^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\big)\prod_{n=1}^{l_c}\sum_{j_{k+n}} p_{j_{k+n}}\big(1 - \epsilon^{(k)}_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}\big), \tag{E7}$$

and have used the fact that $\langle j_k|j'_k\rangle_{A_k} = \delta_{j_k,j'_k}$. Moreover, the normalized state $|\tilde{\phi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k+1-l_c}}\rangle_{B_k,\mathbf{E}'_k}$ is expressed as

$$\frac{\big|\phi_{j_k,\boldsymbol{g}}\big\rangle_{B_k,E_k}\sum_{j_{k+1}}\sqrt{p_{j_{k+1}}}\ldots\sum_{j_{k+l_c}}\sqrt{p_{j_{k+l_c}}}\big|\psi_{j_{k+1},\ldots,j_{k+l_c},\boldsymbol{g}}\big\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+l_c+1},\mathbf{E}_{\geqslant k+l_c+1}}\bigotimes_{n=1}^{l_c}\sqrt{1-\epsilon^{(k)}_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}}\big|\phi_{j_{k+n},\boldsymbol{g}}\big\rangle_{B_{k+n},E_{k+n}}}{\prod_{n=1}^{l_c}\sqrt{\sum_{j_{k+n}}p_{j_{k+n}}\big(1-\epsilon^{(k)}_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}\big)}}$$

$$=: \big|\phi_{j_k,\boldsymbol{g}}\big\rangle_{B_k,E_k}\big|\Lambda_{j_k,\ldots,j_{k+1-l_c},\boldsymbol{g}}\big\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+1},\mathbf{E}_{\geqslant k+1}} =: \big|\tilde{\phi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k+1-l_c}}\big\rangle_{B_k,\mathbf{E}'_k}, \tag{E8}$$

and $|\tilde{\phi}^{\perp}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,\mathbf{E}'_k}$ is a state orthogonal to $|\tilde{\phi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k+1-l_c}}\rangle_{B_k,\mathbf{E}'_k}$, living in a Hilbert space of any dimension, and whose explicit form is omitted here for simplicity. Importantly, substituting Eq. (E6) into Eq. (E5) we obtain

$$\big|\Psi_{\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\big\rangle_{A_k,B_k,\mathbf{E}'_k} = \sum_{j_k}\sqrt{p_{j_k}}|j_k\rangle_{A_k}\big|\tilde{\psi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\big\rangle_{B_k,\mathbf{E}'_k}, \tag{E9}$$

which implies that we can regard the emitted states on round $k$ to be $|\tilde{\psi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,\mathbf{E}'_k}$.

Now, suppose that Alice performs a measurement on system $A_k$ such that she learns if $j_k \in \{0_Z, 1_Z\}$ or if $j_k \in \{0_X, 1_X\}$, i.e., her choice of basis for the $k$th round. If she obtains the former outcome, her post-measurement state can be expressed as

$$\big|\Psi^{Z}_{\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\big\rangle_{A_k,B_k,\mathbf{E}'_k} = \frac{1}{\sqrt{2}}\sum_{\tau\in\{0,1\}}|\tau_Z\rangle_{A_k}\big|\tilde{\psi}_{\tau_Z,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\big\rangle_{B_k,\mathbf{E}'_k}, \tag{E10}$$

where we have used $p_{0_Z} = p_{1_Z}$ [see Assumption (A1) in Sec. II A]. This concludes the derivation of Eq. (26).

### 2. Proof of Eq. (27)

Here, we prove Eq. (27) and show that when the QKD protocol is run sequentially [see Assumption (A4) Sec. II A], the operator $\hat{D}^{(k)}_{\gamma_X}$ satisfies $0 \leqslant \hat{D}^{(k)}_{\gamma_X} \leqslant \hat{\mathbb{1}}$. To do so, we consider again an entanglement-based scenario that is equivalent to the actual protocol. For a fixed $\boldsymbol{g}$, the transmission of $N$ pulses can then be described by Alice first preparing the entangled state $|\Psi_{\boldsymbol{g}}\rangle_{\mathbf{A},\mathbf{B},\mathbf{E}}$ in Eq. (E1), keeping system $\mathbf{A}$ in her laboratory and sending systems $\mathbf{B}$, $\mathbf{E}$ through the quantum channel. In our analysis we assume that the QKD protocol is run sequentially. As explained in Appendix D, under this restriction, Eve's most general attack is described by $\hat{U}^{(N)}_{B_N,E_N,\tilde{E}_N}\ldots\hat{U}^{(1)}_{B_1,E_1,\tilde{E}_1}$, where $\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}$ acts on the $k$th photonic system $B_k$, $E_k$ and on Eve's updated ancilla $\tilde{E}_k$, resulting in systems $B'_k$ and $\tilde{E}_k$ (see Fig. 3 for more details). After Eve's attack on each round, Alice and Bob measure their local systems to obtain the experimental data. The secret key is generated from the rounds in which both have selected the $Z$ basis, i.e., $(a, \beta) = (Z, Z)$, where $a$ ($\beta$) denotes Alice's (Bob's) basis selection.

As before, to prove the security of these rounds, one needs to consider a virtual protocol to estimate the number of phase errors that Alice and Bob would have obtained if they had performed their local measurements in the phase basis instead, i.e., $\{|\tau^{(\mathrm{vir})}_X\rangle_{A_k} = (|0_Z\rangle_{A_k} + (-1)^{\tau}|1_Z\rangle_{A_k})/\sqrt{2}\}_{\tau\in\{0,1\}}$ and $\{\hat{m}_{0_X}, \hat{m}_{1_X}, \hat{m}_f\}$, respectively. Unlike in Appendix D, however, each emitted pulse $k$ now depends on the previous $l_c$ setting choices $\{j_{k-1},\ldots,j_{k-l_c}\}$. As explained in Sec. III B, this dependence does not allow us to naively consider a virtual protocol in which Alice and Bob perform phase basis measurements on all key generation rounds. Instead, we consider that Alice and Bob assign a tag $w \in \{0,\ldots,l_c\}$ to each round $k$ according to the value $w = k \bmod (l_c + 1)$ and construct $(l_c + 1)$ virtual protocols whose respective $w$th sifted keys are subsets of the total sifted key originating from the rounds with a tag $w$. In each $w$th virtual protocol, Alice's and Bob's measurements on the $w$-tagged rounds are described by the POVM $\{\hat{F}^{o_k}_{A_k,B'_k}\}_{o_k}$, whose elements are defined in Eq. (D2), and on the other rounds, Alice's and Bob's measurements are described by the POVM $\{\hat{J}^{o_k}_{A_k,B'_k}\}_{o_k}$, whose elements are defined as $\hat{J}^{(a,\beta,\tau,\gamma)}_{A_k,B'_k} = |\tau_a\rangle\langle\tau_a|_{A_k}\otimes p_{\beta_B}\hat{m}_{\gamma_\beta}$ and $\hat{J}^{(f)}_{A_k,B'_k} = \hat{\mathbb{1}}_{A_k}\otimes\hat{m}_f$, where $\tau, \gamma \in \{0, 1\}$ are Alice's and Bob's observed bit value respectively, and $f$ is associated with an inconclusive outcome. Note that each of the these virtual protocols is indistinguishable from the actual protocol.

The probability that, in the $w$th virtual protocol, a phase error is obtained on some round $k$ with a tag $w$, conditioned on all the previous outcomes of the $w$th virtual protocol, can then be expressed as Eq. (D3), rewritten here for convenience:

$$P^{(k)}_{\boldsymbol{g}}(\mathrm{ph}|\mathrm{Act}) = \frac{P((o_k = (Z, Z, 0, 1) \text{ or } (Z, Z, 1, 0)), o_{\overrightarrow{k-1}})}{P(o_{\overrightarrow{k-1}})}, \tag{E11}$$

where $P(o_{\overrightarrow{k-1}})$ is the probability of obtaining Alice's and Bob's previous outcomes. Let us now introduce the POVM $\{\hat{L}^{o_k}_{A_k,B'_k}\}_{o_k}$, which is defined as $\{\hat{F}^{o_k}_{A_k,B'_k}\}_{o_k}$ when a pulse $k$ has the tag $w$ and defined as $\{\hat{J}^{o_k}_{A_k,B'_k}\}_{o_k}$ when it has another tag. Therefore, by using this POVM, the form of the emitted state in Eq. (E1) as well as the description of Eve's coherent attack when the QKD protocol is run sequentially, we have that

$$\begin{aligned}
&P((o_k = (Z, Z, 0, 1) \text{ or } (Z, Z, 1, 0)), o_{\overrightarrow{k-1}})\\
&\quad = \sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}} \mathrm{Tr}\left[\hat{F}^{(Z,Z,\tau,\gamma)}_{A_k,B'_k}\hat{P}\left(\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}\prod_{n=1}^{k-1}\sqrt{\hat{L}^{o_{k-n}}_{A_{k-n},B'_{k-n}}}\hat{U}^{(k-n)}_{B_{k-n},E_{k-n},\tilde{E}_{k-n}}|0\rangle_{\tilde{E}_1}\big|\Psi_{\boldsymbol{g}}\big\rangle_{\mathbf{A},\mathbf{B},\mathbf{E}}\right)\right]\\
&\quad = \sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}} \mathrm{Tr}\left[\hat{F}^{(Z,Z,\tau,\gamma)}_{A_k,B'_k}\hat{P}\left(\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}\left[\prod_{n=1}^{k-1}\sqrt{\hat{L}^{o^B_{k-n}}_{B'_{k-n}}}\hat{U}^{(k-n)}_{B_{k-n},E_{k-n},\tilde{E}_{k-n}}\right]|0\rangle_{\tilde{E}_1}\left[\prod_{n=1}^{k-1}\sqrt{\hat{L}^{o^A_{k-n}}_{A_{k-n}}}\big|\Psi_{\boldsymbol{g}}\big\rangle_{\mathbf{A},\mathbf{B},\mathbf{E}}\right]\right)\right]\\
&\quad = \sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}} \mathrm{Tr}\left[\hat{F}^{(Z,Z,\tau,\gamma)}_{A_k,B'_k}\hat{P}\left(\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}\left[\prod_{n=1}^{k-1}\sqrt{\hat{L}^{o^B_{k-n}}_{B'_{k-n}}}\hat{U}^{(k-n)}_{B_{k-n},E_{k-n},\tilde{E}_{k-n}}\right]|0\rangle_{\tilde{E}_1}\right.\right.\\
&\qquad\left.\left.\otimes\left[\bigotimes_{n=1}^{k-1}\sqrt{p_{j_n}}|j_n\rangle_{A_n}\big|\psi_{j_n,\boldsymbol{g}|j_{n-1},\dots,j_{n-l_c}}\big\rangle_{B_n}\right]\big|\Psi_{\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}\big\rangle_{A_k,B_k,\mathbf{E}'_k}\right)\right]\\
&\quad = \sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}} \mathrm{Tr}\big[_{\mathbf{A}_{\leqslant k-1},\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1},\tilde{E}_k}\big\langle\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big|\hat{U}^{(k)\dagger}_{B_k,E_k,\tilde{E}_k}\hat{F}^{(Z,Z,\tau,\gamma)}_{A_k,B'_k}\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}\big|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big\rangle_{\mathbf{A}_{\leqslant k-1},\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1},\tilde{E}_k}\big|\Psi_{\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}\big\rangle\big\langle\Psi_{\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}\big|_{A_k,B_k,\mathbf{E}'_k}\big]\\
&\quad = \sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}} \mathrm{Tr}\big[_{\mathbf{A}_{\leqslant k-1},\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1},\tilde{E}_k}\big\langle\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big|\hat{U}^{(k)\dagger}_{B_k,E_k,\tilde{E}_k}\hat{m}_{\gamma_X}\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}\big|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big\rangle_{\mathbf{A}_{\leqslant k-1},\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1},\tilde{E}_k}\\
&\qquad\otimes\tilde{p}^{(k,\mathrm{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}p_{Z_B}\big|\psi^{(\mathrm{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}\big\rangle\big\langle\psi^{(\mathrm{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}\big|_{B_k,\boldsymbol{E}'_k}\big],
\end{aligned}\tag{E12}$$

where

$$\big|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big\rangle_{\mathbf{A}_{\leqslant k-1},\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1},\tilde{E}_k} := \left[\prod_{n=1}^{k-1}\sqrt{\hat{L}^{o^B_{k-n}}_{B'_{k-n}}}\hat{U}^{(k-n)}_{B_{k-n},E_{k-n},\tilde{E}_{k-n}}\right]|0\rangle_{\tilde{E}_1}\left[\bigotimes_{n=1}^{k-1}\sqrt{p_{j_n}}|j_n\rangle_{A_n}\big|\psi_{j_n,\boldsymbol{g}|j_{n-1},\dots,j_{n-l_c}}\big\rangle_{B_k}\right],\tag{E13}$$

with $\mathbf{A}_{\leqslant k-1} := A_1, \dots, A_{k-1}$ and $\mathbf{B}'_{\leqslant k-1}, \mathbf{E}_{\leqslant k-1} := B'_1, E_1, \dots, B'_{k-1}, E_{k-1}$, and where $|\psi^{(\mathrm{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}\rangle_{B_k,\boldsymbol{E}'_k}$ and $\tilde{p}^{(k,\mathrm{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}$ are defined in Eqs. (29) and (28), respectively. In the second equality of Eq. (E12), we have used the fact that Alice's measurements on her ancillae $\mathbf{A}_{\leqslant k-1}$ commute with any measurements or operations on systems $B_1, E_1, \tilde{E}_1, \dots, B_{k-1}, E_{k-1}, \tilde{E}_{k-1}$, and the fact that $\hat{L}^{o_k}_{A_k,B'_k}$ can be decomposed as $\hat{L}^{o^A_k}_{A_k} \otimes \hat{L}^{o^B_k}_{B'_k}$, where Alice's outcome $o^A_k$ and Bob's outcome $o^B_k$ on round $k$ are defined as $(a, \tau)$ and $(\beta, \gamma)$, respectively. In the third equality of Eq. (E12), we have used the definition of the post-measurement state $|\Psi_{\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}\rangle_{A_k,B_k,\mathbf{E}'_k}$ in Eq. (E2), and in the fourth equality of Eq. (E12), we have defined the state $|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\rangle_{\mathbf{A}_{\leqslant k-1},\mathbf{B}'_{\leqslant k-1},\mathbf{E}_{\leqslant k-1},\tilde{E}_k}$ using Eq. (E13) and we have taken advantage of the cyclic property of the trace operation. Finally, in the last equality of Eq. (E12), we have used Eq. (E9), Eq. (26) and the definition of $\hat{F}^{(Z,Z,\tau,\gamma)}_{A_k,B'_k}$ in Eq. (D2) for $\tau, \gamma \in \{0, 1\}$ such that $\tau \neq \gamma$.

By substituting Eq. (E12) in Eq. (E11), we can now express $P^{(k)}_{\boldsymbol{g}}(\mathrm{ph}|\mathrm{Act})$ as

$$P^{(k)}_{\boldsymbol{g}}(\mathrm{ph}|\mathrm{Act}) := \sum_{\substack{\tau,\gamma\in\{0,1\}\\ \tau\neq\gamma}}\tilde{p}^{(k,\mathrm{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}p_{Z_B}\mathrm{Tr}\big[\tilde{\sigma}^{(k,\mathrm{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}\hat{D}^{(k)}_{\gamma_X}\big],\tag{E14}$$

with $\tilde{\sigma}^{(k,\mathrm{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}} := \hat{P}(|\psi^{(\mathrm{vir})}_{\tau_X,\boldsymbol{g}|j_{k-1},\dots,j_{k-l_c}}\rangle_{B_k,\mathbf{E}'_k})$ and

$$\hat{D}^{(k)}_{\gamma_X} := \frac{\big\langle\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big|\hat{U}^{(k)\dagger}_{B_k,E_k,\tilde{E}_k}\hat{m}_{\gamma_X}\hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k}\big|\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\big\rangle}{P(o_{\overrightarrow{k-1}})},\tag{E15}$$

where $P(o_{\overrightarrow{k-1}}) = |||\Psi^{\overrightarrow{k-1}}_{\boldsymbol{g}}\rangle||^2$, and where we have omitted the mode subscripts in the quantum states for simplicity of notation. This concludes the derivation of Eq. (27).

Finally, note that the operator $\hat{D}^{(k)}_{\gamma_X}$ in Eq. (E15) has a very similar form to the operator $\hat{M}^{(k)}_{\gamma_X}$ in Eq. (D7). Therefore, we can use an analogous argument to that in Appendix D to show that when the QKD protocol is run sequentially, the operator $\hat{D}^{(k)}_{\gamma_X}$ satisfies $0 \leqslant \hat{D}^{(k)}_{\gamma_X} \leqslant \mathbb{1}$.

### 3. Upper bounds on $|\langle\tilde{\boldsymbol{\psi}}_{j_k,g|j_{k-1},\ldots,j_{k-l_c}}|\boldsymbol{\phi}_{j_k,g}\rangle_{B_k,\mathbf{E}'_k}|$ and $|\langle\boldsymbol{\Psi}_{g|j_{k-1},\ldots,j_{k-l_c}}|\boldsymbol{\Phi}_g\rangle_{A_k,B_k,\mathbf{E}'_k}|$

When employing the RT to prove the security of the protocol, we define the reference states

$$\begin{aligned}\left|\phi_{j_k,\boldsymbol{g}}\right\rangle_{B_k,\mathbf{E}'_k} &= \left|\phi_{j_k,\boldsymbol{g}}\right\rangle_{B_k,E_k}\sum_{j_{k+1}}\sqrt{p_{j_{k+1}}}\cdots\sum_{j_{k+l_c}}\sqrt{p_{j_{k+l_c}}}\left|\psi_{j_{k+1},\ldots,j_{k+l_c},\boldsymbol{g}}\right\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+l_c+1},\mathbf{E}_{\geqslant k+l_c+1}}\bigotimes_{n=1}^{l_c}\left|\phi_{j_{k+n},\boldsymbol{g}}\right\rangle_{B_{k+n},E_{k+n}}\\ &=: \left|\phi_{j_k,\boldsymbol{g}}\right\rangle_{B_k,E_k}|\Lambda_{\boldsymbol{g}}\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+1},\mathbf{E}_{\geqslant k+1}},\end{aligned}\tag{E16}$$

and

$$\left|\Phi^Z_{\boldsymbol{g}}\right\rangle_{A_k,B_k,\mathbf{E}'_k} = \frac{1}{\sqrt{2}}\sum_{\tau\in\{0,1\}}|\tau_Z\rangle_{A_k}\left|\phi_{\tau_Z,\boldsymbol{g}}\right\rangle_{B_k,\mathbf{E}'_k},\tag{E17}$$

which are analogous to $|\tilde{\psi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{B_k,\mathbf{E}'_k}$ and $|\Psi^Z_{\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\rangle_{A_k,B_k,\mathbf{E}'_k}$, respectively. Note that, unlike $|\tilde{\phi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k+1-l_c}}\rangle_{B_k,\mathbf{E}'_k}$ in Eq. (E8), $|\phi_{j_k,\boldsymbol{g}}\rangle_{B_k,\mathbf{E}'_k}$ in Eq. (E16) is a qubit state, i.e., $\{|\phi_{j_k,\boldsymbol{g}}\rangle_{B_k,\mathbf{E}'_k}\}_{j_k=0_Z,1_Z,0_X,1_X}$ spans a qubit space, because $|\Lambda_{\boldsymbol{g}}\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+1},\mathbf{E}_{\geqslant k+1}}$ is a state independent of $j_k$.

Using Eqs. (E6) and (E16), we can then calculate an upper bound on the inner product $|\langle\tilde{\psi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}|\phi_{j_k,\boldsymbol{g}}\rangle_{B_k,\mathbf{E}'_k}|$ such that

$$\begin{aligned}&\left|\left\langle\tilde{\psi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\middle|\phi_{j_k,\boldsymbol{g}}\right\rangle_{B_k\mathbf{E}'_k}\right|\\ &= \left|\sqrt{1-\tilde{\epsilon}^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}\left\langle\tilde{\phi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k+1-l_c}}\middle|\phi_{j_k,\boldsymbol{g}}\right\rangle_{B_k\mathbf{E}'_k}+\sqrt{\tilde{\epsilon}^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}\left\langle\tilde{\phi}^{\perp}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\middle|\phi_{j_k,\boldsymbol{g}}\right\rangle_{B_k\mathbf{E}'_k}\right|\\ &= \left|\sqrt{1-\tilde{\epsilon}^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}\left\langle\phi_{j_k,\boldsymbol{g}}\middle|\phi_{j_k,\boldsymbol{g}}\right\rangle_{B_kE_k}\left\langle\Lambda_{j_k,\ldots,j_{k+1-l_c},\boldsymbol{g}}\middle|\Lambda_{\boldsymbol{g}}\right\rangle_{\mathbf{A}_{\geqslant k+1},\mathbf{B}_{\geqslant k+1},\mathbf{E}_{\geqslant k+1}}\right|\\ &= \sqrt{1-\tilde{\epsilon}^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}\frac{\prod_{n=1}^{l_c}\sum_{j_{k+n}}p_{j_{k+n}}\sqrt{1-\epsilon^{(k)}_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}}}{\prod_{n=1}^{l_c}\sqrt{\sum_{j_{k+n}}p_{j_{k+n}}\left(1-\epsilon^{(k)}_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}\right)}}\\ &= \sqrt{1-\epsilon^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}\prod_{n=1}^{l_c}\sqrt{\sum_{j_{k+n}}p_{j_{k+n}}\left(1-\epsilon^{(k)}_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}\right)}\frac{\prod_{n=1}^{l_c}\sum_{j_{k+n}}p_{j_{k+n}}\sqrt{1-\epsilon^{(k)}_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}}}{\prod_{n=1}^{l_c}\sqrt{\sum_{j_{k+n}}p_{j_{k+n}}\left(1-\epsilon^{(k)}_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}\right)}}\\ &= \sqrt{1-\epsilon^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}}\prod_{n=1}^{l_c}\sum_{j_{k+n}}p_{j_{k+n}}\sqrt{1-\epsilon^{(k)}_{j_{k+n},\boldsymbol{g}|j_{k+n-1},\ldots,j_{k+n-l_c}}}\geqslant\prod_{n=1}^{l_c+1}\sqrt{1-\epsilon'^{\mathrm{U}}}=(\sqrt{1-\epsilon'^{\mathrm{U}}})^{l_c+1}=:\sqrt{1-\epsilon^{\mathrm{U}}},\end{aligned}\tag{E18}$$

where $\epsilon^{\mathrm{U}} := 1-(1-\epsilon'^{\mathrm{U}})^{l_c+1}$. In the second equality of Eq. (E18), we have used the fact that $\langle\tilde{\phi}^{\perp}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}|\phi_{j_k,\boldsymbol{g}}\rangle_{B_k,\mathbf{E}'_k}=0$ for any $j_k$, in the third equality we have used the definition in Eq. (E8), and in the forth equality, we have used the definition in Eq. (E7). Finally, in the inequality of Eq. (E18), we have used that $\sum_{j_k}p_{j_k}=1$ for any round $k$ and that $\epsilon^{(k)}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\leqslant\epsilon'^{\mathrm{U}}$ for all $k$, $j_k$, $\boldsymbol{g}$, and $j_{k-1},\ldots,j_{k-l_c}$ [see assumption (A2.b) in Sec. II A].

Similarly, using Eqs. (E10) and (E17), we can calculate an upper bound on the inner product $|\langle\Psi^Z_{\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}|\Phi^Z_{\boldsymbol{g}}\rangle_{A_k,B_k,\mathbf{E}'_k}|$ such that

$$\begin{aligned}\left|\left\langle\Psi^Z_{\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}\middle|\Phi^Z_{\boldsymbol{g}}\right\rangle_{A_k,B_k,\mathbf{E}'_k}\right| &= \tfrac{1}{2}\left|\left\langle\tilde{\psi}_{0_Z,\boldsymbol{g}|j_{k-1},\ldots,j_{k-1}}\middle|\phi_{0_Z,\boldsymbol{g}}\right\rangle_{B_k,\mathbf{E}'_k}+\left\langle\tilde{\psi}_{1_Z,\boldsymbol{g}|j_{k-1},\ldots,j_{k-1}}\middle|\phi_{1_Z,\boldsymbol{g}}\right\rangle_{B_k,\mathbf{E}'_k}\right|\\ &\leqslant \tfrac{1}{2}(\sqrt{1-\epsilon}+\sqrt{1-\epsilon})=\sqrt{1-\epsilon},\end{aligned}\tag{E19}$$

where we have used the fact that $\langle j_k|j'_k\rangle_{A_k}=\delta_{j_k,j'_k}$. Also, in the inequality of Eq. (E19) we have used the upper bound in Eq. (E18). This concludes the calculation of the upper bounds on the inner products $|\langle\tilde{\psi}_{j_k,\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}|\phi_{j_k,\boldsymbol{g}}\rangle_{B_k,\mathbf{E}'_k}|$ and $|\langle\Psi_{\boldsymbol{g}|j_{k-1},\ldots,j_{k-l_c}}|\Phi_{\boldsymbol{g}}\rangle_{A_k,B_k,\mathbf{E}'_k}|$.

## APPENDIX F: PROOF OF Eq. (32)

Here, we provide a proof of the second inequality in Eq. (32). From Eq. (25), we can express $e^{\mathrm{U}}_{\mathrm{ph}}$ as

$$e^{\mathrm{U}}_{\mathrm{ph}} = \frac{N^{\mathrm{U}}_{\mathrm{ph}}}{N^{(Z)}_{\mathrm{det}}} = \frac{N}{N^{(Z)}_{\mathrm{det}}}f\left(\frac{\mathbf{x}}{N}\right),\tag{F1}$$

where $\mathbf{x}$ is a tuple whose elements are $N_{j,\gamma_X}$ for all $j$, $\gamma$, and $f$ is a multivariate function that is concave with respect to all of its arguments due to the concavity of $G_+(y,z)$ and $-G_-(y,z)$ with respect to $y$. Similarly, from Eq. (31), we can express $e^{\mathrm{U}}_{\mathrm{ph},w}$ as

$$e^{\mathrm{U}}_{\mathrm{ph},w}=\frac{N^{\mathrm{U}}_{\mathrm{ph},w}}{N^{(Z)}_{\mathrm{det},w}}=\frac{N_w}{N^{(Z)}_{\mathrm{det},w}}f\left(\frac{\mathbf{x}_w}{N_w}\right), \tag{F2}$$

where $\mathbf{x}_w$ is a tuple whose components are $N_{j,\gamma_X,w}$ for all $j$, $\gamma$, and $f$ is the same function as in Eq. (F1). We have that

$$\begin{aligned}\sum_{w=0}^{l_c}q_w e^{\mathrm{U}}_{\mathrm{ph},w}&=\sum_{w=0}^{l_c}\frac{N^{(Z)}_{\mathrm{det},w}}{N^{(Z)}_{\mathrm{det}}}\frac{N_w}{N^{(Z)}_{\mathrm{det},w}}f\left(\frac{\mathbf{x}_w}{N_w}\right)\\ &=\sum_{w=0}^{l_c}\frac{N_w}{N^{(Z)}_{\mathrm{det}}}f\left(\frac{\mathbf{x}_w}{N_w}\right)\\ &=\frac{N}{N^{(Z)}_{\mathrm{det}}}\sum_{w=0}^{l_c}\frac{N_w}{N}f\left(\frac{\mathbf{x}_w}{N_w}\right)\\ &\leqslant\frac{N}{N^{(Z)}_{\mathrm{det}}}f\left(\sum_{w=0}^{l_c}\frac{N_w}{N}\frac{\mathbf{x}_w}{N_w}\right)\\ &=\frac{N}{N^{(Z)}_{\mathrm{det}}}f\left(\frac{\mathbf{x}}{N}\right)=e^{\mathrm{U}}_{\mathrm{ph}},\end{aligned} \tag{F3}$$

where the inequality is due to the concavity of $f$, and the second to last equality is due to $\sum_w\mathbf{x}_w=\mathbf{x}$, since $\sum_w N_{j,\gamma_X,w}=N_{j,\gamma_X}$.

## APPENDIX G: PARAMETERS $c^{(k)}_{\tau,j,\boldsymbol{g}}$, $c^{\mathrm{U}}_{\tau,j}$, AND $p^{(\mathrm{vir})\mathrm{U}}_{\tau_X}$ FOR THE MODIFIED BB84 PROTOCOL

In this Appendix, we provide the full expressions of the coefficients $c^{(k)}_{\tau,j,\boldsymbol{g}}$, with $\tau\in\{0,1\}$ and $j\in\{0_Z,1_Z,0_X,1_X\}$, associated with the modified BB84 protocol, for the particular phase-encoding scheme considered in Sec. IV A. Direct calculations show that these coefficients are given by

$$\begin{aligned}
c^{(k)}_{1,0_Z,\boldsymbol{g}}&=\frac{\sin\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{1_X,\boldsymbol{g}}}{2}\right)-\sin\left(\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{1_X,\boldsymbol{g}}}{2}\right)}{\sin\left(\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}-\theta^{(k)}_{0_Z,\boldsymbol{g}}+\frac{\theta^{(k)}_{1_X,\boldsymbol{g}}}{2}\right)+2\sin\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{1_X,\boldsymbol{g}}}{2}\right)-\sin\left(\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{1_X,\boldsymbol{g}}}{2}\right)},\\
c^{(k)}_{1,1_Z,\boldsymbol{g}}&=\frac{-\sin\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{1_X,\boldsymbol{g}}}{2}\right)+\sin\left(\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{1_X,\boldsymbol{g}}}{2}\right)}{\sin\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}-\theta^{(k)}_{1_Z,\boldsymbol{g}}+\frac{\theta^{(k)}_{1_X,\boldsymbol{g}}}{2}\right)-\sin\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{1_X,\boldsymbol{g}}}{2}\right)+2\sin\left(\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{1_X,\boldsymbol{g}}}{2}\right)},\\
c^{(k)}_{1,1_X,\boldsymbol{g}}&=\frac{\cos\left(\theta^{(k)}_{0_Z,\boldsymbol{g}}-\theta^{(k)}_{1_Z,\boldsymbol{g}}\right)-1}{\left[\begin{array}{r}\cos\left(\theta^{(k)}_{0_Z,\boldsymbol{g}}-\theta^{(k)}_{1_Z,\boldsymbol{g}}\right)-\cos\left(\theta^{(k)}_{0_Z,\boldsymbol{g}}-\theta^{(k)}_{1_X,\boldsymbol{g}}\right)-\cos\left(\theta^{(k)}_{1_Z,\boldsymbol{g}}-\theta^{(k)}_{1_X,\boldsymbol{g}}\right)\\ +2\cos\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}+\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}-\theta^{(k)}_{1_X,\boldsymbol{g}}\right)-2\cos\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}\right)+1\end{array}\right]},\\
c^{(k)}_{0,0_Z,\boldsymbol{g}}&=\frac{\sin\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{0_X,\boldsymbol{g}}}{2}\right)+\sin\left(\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{0_X,\boldsymbol{g}}}{2}\right)}{2\sin\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{0_X,\boldsymbol{g}}}{2}\right)-\sin\left(\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}-\theta^{(k)}_{0_Z,\boldsymbol{g}}+\frac{\theta^{(k)}_{0_X,\boldsymbol{g}}}{2}\right)+\sin\left(\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{0_X,\boldsymbol{g}}}{2}\right)},\\
c^{(k)}_{0,1_Z,\boldsymbol{g}}&=\frac{\sin\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{0_X,\boldsymbol{g}}}{2}\right)+\sin\left(\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{0_X,\boldsymbol{g}}}{2}\right)}{\sin\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{0_X,\boldsymbol{g}}}{2}\right)-\sin\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}-\theta^{(k)}_{1_Z,\boldsymbol{g}}+\frac{\theta^{(k)}_{0_X,\boldsymbol{g}}}{2}\right)+2\sin\left(\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{0_X,\boldsymbol{g}}}{2}\right)},\\
c^{(k)}_{0,0_X,\boldsymbol{g}}&=\frac{\cos\left(\theta^{(k)}_{0_Z,\boldsymbol{g}}-\theta^{(k)}_{1_Z,\boldsymbol{g}}\right)-1}{\left[\begin{array}{r}\cos\left(\theta^{(k)}_{0_Z,\boldsymbol{g}}-\theta^{(k)}_{1_Z,\boldsymbol{g}}\right)-\cos\left(\theta^{(k)}_{0_Z,\boldsymbol{g}}-\theta^{(k)}_{0_X,\boldsymbol{g}}\right)-\cos\left(\theta^{(k)}_{1_Z,\boldsymbol{g}}-\theta^{(k)}_{0_X,\boldsymbol{g}}\right)\\ -2\cos\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}+\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}-\theta^{(k)}_{0_X,\boldsymbol{g}}\right)+2\cos\left(\frac{\theta^{(k)}_{0_Z,\boldsymbol{g}}}{2}-\frac{\theta^{(k)}_{1_Z,\boldsymbol{g}}}{2}\right)+1\end{array}\right]}.
\end{aligned} \tag{G1}$$

These coefficients are constrained by the fact that $\theta_{j,\boldsymbol{g}}^{(k)} \in [\theta_j^{\mathrm{L}}, \theta_j^{\mathrm{U}}]$ for some known $\theta_j^{\mathrm{L}}$ and $\theta_j^{\mathrm{U}}$. While one can in principle find numerical upper bounds on these coefficients regardless of the value of $\theta_j^{\mathrm{L}}$ and $\theta_j^{\mathrm{U}}$, here we obtain analytical bounds under the following assumption: $-\pi/6 \leqslant \theta_{0_Z}^{\mathrm{L}} \leqslant \theta_{0_Z}^{\mathrm{U}} \leqslant \pi/6$, $5\pi/6 \leqslant \theta_{1_Z}^{\mathrm{L}} \leqslant \theta_{1_Z}^{\mathrm{U}} \leqslant 7\pi/6$, $\pi/3 \leqslant \theta_{0_X}^{\mathrm{L}} \leqslant \theta_{0_X}^{\mathrm{U}} \leqslant 2\pi/3$, and $4\pi/3 \leqslant \theta_{1_X}^{\mathrm{L}} \leqslant \theta_{1_X}^{\mathrm{U}} \leqslant 5\pi/3$. Note that this assumption is reasonable because a deviation of $\pm\pi/6$ from the ideal phase value is much larger than the modulation errors characterized in recent experiments [16].

To derive the analytical upper bounds, we consider the partial differential equations of $c_{\tau,j,\boldsymbol{g}}^{(k)}$ with respect to each $\theta_{j,\boldsymbol{g}}^{(k)}$ and then select the values of $\theta_{j,\boldsymbol{g}}^{(k)} \in [\theta_j^{\mathrm{L}}, \theta_j^{\mathrm{U}}]$ that maximise them. For example, since $\partial_{\theta_{0_Z,\boldsymbol{g}}} c_{1,0_Z,\boldsymbol{g}}^{(k)} > 0$ when $\theta_{0_Z,\boldsymbol{g}}^{(k)} \in [-\pi/6, \pi/6]$, $\theta_{1_Z,\boldsymbol{g}}^{(k)} \in [5\pi/6, 7\pi/6]$, and $\theta_{0_X,\boldsymbol{g}}^{(k)} \in [\pi/3, 2\pi/3]$, $c_{1,0_Z,\boldsymbol{g}}^{(k)}$ is maximized when $\theta_{0_Z,\boldsymbol{g}}^{(k)} = \theta_{0_Z}^{\mathrm{U}}$. Alternatively, if the function is not always increasing or decreasing with respect to a particular argument, but given the ranges stated above it is convex with respect to that argument, e.g. $\partial^2_{\theta_{0_Z,\boldsymbol{g}}} c_{1,1_X,\boldsymbol{g}}^{(k)} > 0$, then we conclude that $c_{1,1_X,\boldsymbol{g}}$ is maximized either when $\theta_{0_Z,\boldsymbol{g}}^{(k)} = \theta_{0_Z}^{\mathrm{L}}$ or when $\theta_{0_Z,\boldsymbol{g}}^{(k)} = \theta_{0_Z}^{\mathrm{U}}$, and we select the appropriate value by inspection. The overall solution is the following:

$$
\begin{aligned}
c_{1,0_Z}^{\mathrm{U}} &\rightarrow c_{1,0_Z,\boldsymbol{g}}^{(k)}\big(\theta_{0_Z}^{\mathrm{U}}, \theta_{1_Z}^{\mathrm{U}}, \theta_{1_X}^{\mathrm{L}}\big),\\
c_{1,1_Z}^{\mathrm{U}} &\rightarrow c_{1,1_Z,\boldsymbol{g}}^{(k)}\big(\theta_{0_Z}^{\mathrm{L}}, \theta_{1_Z}^{\mathrm{L}}, \theta_{1_X}^{\mathrm{U}}\big),\\
c_{1,1_X}^{\mathrm{U}} &\rightarrow \max_{x_1,y_1,z_1\in\{\mathrm{L},\mathrm{U}\}} c_{1,1_X,\boldsymbol{g}}^{(k)}\big(\theta_{0_Z}^{x_1}, \theta_{1_Z}^{y_1}, \theta_{1_X}^{z_1}\big),\\
c_{0,0_Z}^{\mathrm{U}} &\rightarrow c_{0,0_Z,\boldsymbol{g}}^{(k)}\big(\theta_{0_Z}^{\mathrm{L}}, \theta_{1_Z}^{\mathrm{L}}, \theta_{1_X}^{\mathrm{U}}\big),\\
c_{0,1_Z}^{\mathrm{U}} &\rightarrow c_{0,1_Z,\boldsymbol{g}}^{(k)}\big(\theta_{0_Z}^{\mathrm{U}}, \theta_{1_Z}^{\mathrm{U}}, \theta_{1_X}^{\mathrm{L}}\big),\\
c_{0,0_X}^{\mathrm{U}} &\rightarrow \max_{x_0,y_0,z_0\in\{\mathrm{L},\mathrm{U}\}} c_{0,0_X,\boldsymbol{g}}^{(k)}\big(\theta_{0_Z}^{x_0}, \theta_{1_Z}^{y_0}, \theta_{1_X}^{z_0}\big).
\end{aligned}
\tag{G2}
$$

Similarly, we can obtain upper bounds on the probabilities $p_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$. For this, first recall that

$$
p_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})} = \tfrac{1}{2} p_{Z_A}\big[1 + (-1)^{\tau}\Re\big(\langle\phi_{0_Z,\boldsymbol{g}}|\phi_{1_Z,\boldsymbol{g}}\rangle_{B_k,E_k}\big)\big]. \tag{G3}
$$

Then, by using the definition of $|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k}$ and Eq. (33), and by considering the partial differential equations of $p_{\tau_X,\boldsymbol{g}}^{(k,\mathrm{vir})}$ with respect to each phase $\theta_{j,\boldsymbol{g}}^{(k)}$ we have that

$$
\begin{aligned}
p_{1_X}^{(\mathrm{vir})\mathrm{U}} &\rightarrow \frac{1}{2} p_{Z_A}\left[1 - \cos\left(\frac{\theta_{0_Z}^{\mathrm{L}} - \theta_{1_Z}^{\mathrm{U}}}{2}\right)\right],\\
p_{0_X}^{(\mathrm{vir})\mathrm{U}} &\rightarrow \frac{1}{2} p_{Z_A}\left[1 + \cos\left(\frac{\theta_{0_Z}^{\mathrm{U}} - \theta_{1_Z}^{\mathrm{L}}}{2}\right)\right].
\end{aligned}
\tag{G4}
$$

## APPENDIX H: PARAMETERS $c_{\tau,j,\boldsymbol{g}}^{(k)}$, $c_{\tau,j}^{\mathrm{U}}$, AND $p_{\tau_X}^{(\mathrm{vir})\mathrm{U}}$ FOR THE THREE-STATE PROTOCOL

Here, we provide the full expressions of the coefficients $c_{1,j,\boldsymbol{g}}^{(k)}$ with $j \in \{0_Z, 1_Z, 0_X\}$ associated with the three-state protocol for the particular phase-encoding scheme considered in Sec. IV A, as well as their upper bounds. Direct calculations show that $c_{1,j,\boldsymbol{g}}^{(k)}$ are given by

$$
\begin{aligned}
c_{1,0_Z,\boldsymbol{g}}^{(k)} &= \frac{\sin\left(\frac{\theta_{0_Z,\boldsymbol{g}}^{(k)}}{2} - \frac{\theta_{0_X,\boldsymbol{g}}^{(k)}}{2}\right) - \sin\left(\frac{\theta_{1_Z,\boldsymbol{g}}^{(k)}}{2} - \frac{\theta_{0_X,\boldsymbol{g}}^{(k)}}{2}\right)}{\sin\left(\frac{\theta_{1_Z,\boldsymbol{g}}^{(k)}}{2} - \theta_{0_Z,\boldsymbol{g}}^{(k)} + \frac{\theta_{0_X,\boldsymbol{g}}^{(k)}}{2}\right) + 2\sin\left(\frac{\theta_{0_Z,\boldsymbol{g}}^{(k)}}{2} - \frac{\theta_{0_X,\boldsymbol{g}}^{(k)}}{2}\right) - \sin\left(\frac{\theta_{1_Z,\boldsymbol{g}}^{(k)}}{2} - \frac{\theta_{0_X,\boldsymbol{g}}^{(k)}}{2}\right)},\\
c_{1,1_Z,\boldsymbol{g}}^{(k)} &= \frac{-\sin\left(\frac{\theta_{0_Z,\boldsymbol{g}}^{(k)}}{2} - \frac{\theta_{0_X,\boldsymbol{g}}^{(k)}}{2}\right) + \sin\left(\frac{\theta_{1_Z,\boldsymbol{g}}^{(k)}}{2} - \frac{\theta_{0_X,\boldsymbol{g}}^{(k)}}{2}\right)}{\sin\left(\frac{\theta_{0_Z,\boldsymbol{g}}^{(k)}}{2} - \theta_{1_Z,\boldsymbol{g}}^{(k)} + \frac{\theta_{0_X,\boldsymbol{g}}^{(k)}}{2}\right) - \sin\left(\frac{\theta_{0_Z,\boldsymbol{g}}^{(k)}}{2} - \frac{\theta_{0_X,\boldsymbol{g}}^{(k)}}{2}\right) + 2\sin\left(\frac{\theta_{1_Z,\boldsymbol{g}}^{(k)}}{2} - \frac{\theta_{0_X,\boldsymbol{g}}^{(k)}}{2}\right)},\\
c_{1,0_X,\boldsymbol{g}}^{(k)} &= \frac{\cos\left(\theta_{0_Z,\boldsymbol{g}}^{(k)} - \theta_{1_Z,\boldsymbol{g}}^{(k)}\right) - 1}{\left[\begin{array}{r} \cos\big(\theta_{0_Z,\boldsymbol{g}}^{(k)} - \theta_{1_Z,\boldsymbol{g}}^{(k)}\big) - \cos\big(\theta_{0_Z,\boldsymbol{g}}^{(k)} - \theta_{0_X,\boldsymbol{g}}^{(k)}\big) - \cos\big(\theta_{1_Z,\boldsymbol{g}}^{(k)} - \theta_{0_X,\boldsymbol{g}}^{(k)}\big) \\ + 2\cos\big(\frac{\theta_{0_Z,\boldsymbol{g}}^{(k)}}{2} + \frac{\theta_{1_Z,\boldsymbol{g}}^{(k)}}{2} - \theta_{0_X,\boldsymbol{g}}^{(k)}\big) - 2\cos\big(\frac{\theta_{0_Z,\boldsymbol{g}}^{(k)}}{2} - \frac{\theta_{1_Z,\boldsymbol{g}}^{(k)}}{2}\big) + 1 \end{array}\right]}.
\end{aligned}
\tag{H1}
$$

Their upper bounds can be derived by following a similar approach as in Appendix G and are summarized below [29]:

$$
\begin{aligned}
c_{1,0_Z}^{\mathrm{U}} &\rightarrow c_{1,0_Z,\boldsymbol{g}}^{(k)}\big(\theta_{0_Z}^{\mathrm{U}}, \theta_{1_Z}^{\mathrm{L}}, \theta_{0_X}^{\mathrm{L}}\big),\\
c_{1,1_Z}^{\mathrm{U}} &\rightarrow c_{1,1_Z,\boldsymbol{g}}^{(k)}\big(\theta_{0_Z}^{\mathrm{U}}, \theta_{1_Z}^{\mathrm{L}}, \theta_{0_X}^{\mathrm{U}}\big),\\
c_{1,0_X}^{\mathrm{U}} &\rightarrow \begin{cases} c_{1,0_X,\boldsymbol{g}}^{(k)}\big(\theta_{0_Z}^{\mathrm{L}}, \theta_{1_Z}^{\mathrm{U}}, \theta_{0_X}^{\mathrm{U}}\big) & \text{if } \theta_{0_X}^{\mathrm{U}} < \frac{\theta_{0_Z}^{\mathrm{L}} + \theta_{1_Z}^{\mathrm{U}}}{2},\\ c_{1,0_X,\boldsymbol{g}}^{(k)}\left(\theta_{0_Z}^{\mathrm{L}}, \theta_{1_Z}^{\mathrm{U}}, \frac{\theta_{0_Z}^{\mathrm{L}} + \theta_{1_Z}^{\mathrm{U}}}{2}\right) & \text{if } \frac{\theta_{0_Z}^{\mathrm{L}} + \theta_{1_Z}^{\mathrm{U}}}{2} \in \big[\theta_{0_X}^{\mathrm{L}}, \theta_{0_X}^{\mathrm{U}}\big],\\ c_{1,0_X,\boldsymbol{g}}^{(k)}\big(\theta_{0_Z}^{\mathrm{L}}, \theta_{1_Z}^{\mathrm{U}}, \theta_{0_X}^{\mathrm{L}}\big) & \text{if } \frac{\theta_{0_Z}^{\mathrm{L}} + \theta_{1_Z}^{\mathrm{U}}}{2} < \theta_{0_X}^{L}. \end{cases}
\end{aligned}
\tag{H2}
$$

The full expressions for the coefficients $c_{0,j,\boldsymbol{g}}^{(k)}$ with $j \in \{0_Z, 1_Z, 0_X\}$ as well as their upper bounds, and the probabilities $p_{\tau_X}^{(\mathrm{vir})\mathrm{U}}$ are the same as in Appendix G.

[1] H.-K. Lo, M. Curty, and K. Tamaki, Secure quantum key distribution, Nat. Photonics **8**, 595 (2014).

[2] F. Xu, X. Ma, Q. Zhang, H.-K. Lo, and J.-W. Pan, Secure quantum key distribution with realistic devices, Rev. Mod. Phys. **92**, 025002 (2020).

[3] S. Pirandola, U. L. Andersen, L. Banchi, M. Berta, D. Bunandar, R. Colbeck, D. Englund, T. Gehring, C. Lupo, C. Ottaviani *et al.* Advances in quantum cryptography, Adv. Opt. Photonics **12**, 1012 (2020).

[4] C. H. Bennett and G. Brassard, Quantum cryptography: Public key distribution and coin tossing, Theore. Comput. Sci. **560**, 7 (2014).

[5] H.-K. Lo, M. Curty, and B. Qi, Measurement-Device-Independent Quantum Key Distribution, Phys. Rev. Lett. **108**, 130503 (2012).

[6] A. Rubenok, J. A. Slater, P. Chan, I. Lucio-Martinez, and W. Tittel, Real-World Two-Photon Interference and Proof-of-Principle Quantum Key Distribution Immune to Detector Attacks, Phys. Rev. Lett. **111**, 130501 (2013).

[7] T. Ferreira da Silva, D. Vitoreti, G. B. Xavier, G. C. do Amaral, G. P. Temporão, and J. P. von der Weid, Proof-of-principle demonstration of measurement-device-independent quantum key distribution using polarization qubits, Phys. Rev. A **88**, 052303 (2013).

[8] Y. Liu, T.-Y. Chen, L.-J. Wang, H. Liang, G.-L. Shentu, J. Wang, K. Cui, H.-L. Yin, N.-L. Liu, L. Li *et al.*, Experimental Measurement-Device-Independent Quantum Key Distribution, Phys. Rev. Lett. **111**, 130502 (2013).

[9] Z. Tang, Z. Liao, F. Xu, B. Qi, L. Qian, and H.-K. Lo, Experimental Demonstration of Polarization Encoding Measurement-Device-Independent Quantum Key Distribution, Phys. Rev. Lett. **112**, 190503 (2014).

[10] H.-L. Yin, T.-Y. Chen, Z.-W. Yu, H. Liu, L.-X. You, Y.-H. Zhou, S.-J. Chen, Y. Mao, M.-Q. Huang, W.-J. Zhang *et al.*, Measurement-Device-Independent Quantum Key Distribution Over a 404 km Optical Fiber, Phys. Rev. Lett. **117**, 190501 (2016).

[11] M. Lucamarini, Z. L. Yuan, J. F. Dynes, and A. J. Shields, Overcoming the rate-distance limit of quantum key distribution without quantum repeaters, Nature (London) **557**, 400 (2018).

[12] M. Pittaluga, M. Minder, M. Lucamarini, M. Sanzaro, R. I. Woodward, M.-J. Li, Z. Yuan, and A. J. Shields, 600-km repeater-like quantum communications with dual-band stabilization, Nat. Photonics **15**, 530 (2021).

[13] J.-P. Chen, C. Zhang, Y. Liu, C. Jiang, W.-J. Zhang, Z.-Y. Han, S.-Z. Ma, X.-L. Hu, Y.-H. Li, H. Liu *et al.* Twin-field quantum key distribution over a 511 km optical fibre linking two distant metropolitan areas, Nat. Photonics **15**, 570 (2021).

[14] S. Wang, Z.-Q. Yin, D.-Y. He, W. Chen, R.-Q. Wang, P. Ye, Y. Zhou, G.-J. Fan-Yuan, F.-X. Wang, W. Chen *et al.* Twin-field quantum key distribution over 830-km fibre, Nat. Photonics **16**, 154 (2022).

[15] T. Honjo, K. Inoue, and H. Takahashi, Differential-phase-shift quantum key distribution experiment with a planar lightwave circuit Mach-Zehnder interferometer, Opt. Lett. **29**, 2797 (2004).

[16] F. Xu, K. Wei, S. Sajeed, S. Kaiser, S. Sun, Z. Tang, L. Qian, V. Makarov, and H.-K. Lo, Experimental quantum key distribution with source flaws, Phys. Rev. A **92**, 032305 (2015).

[17] Z. Tang, K. Wei, O. Bedroya, L. Qian, and H.-K. Lo, Experimental measurement-device-independent quantum key distribution with imperfect sources, Phys. Rev. A **93**, 042308 (2016).

[18] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, Security of quantum key distribution with imperfect devices, Quantum Inf. Comput. **4**, 325 (2004).

[19] K. Tamaki, M. Curty, G. Kato, H.-K. Lo, and K. Azuma, Loss-tolerant quantum cryptography with imperfect sources, Phys. Rev. A **90**, 052314 (2014).

[20] M. Pereira, M. Curty, and K. Tamaki, Quantum key distribution with flawed and leaky sources, npj Quantum Inf. **5**, 62 (2019).

[21] N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, Trojan-horse attacks on quantum-key-distribution systems, Phys. Rev. A **73**, 022320 (2006).

[22] A. Vakhitov, V. Makarov, and D. R. Hjelme, Large pulse attack as a method of conventional optical eavesdropping in quantum cryptography, J. Mod. Opt. **48**, 2023 (2001).

[23] M. Lucamarini I. Choi, M. B. Ward, J. F. Dynes, Z. L. Yuan, and A. J. Shields, Practical Security Bounds Against the Trojan-Horse Attack in Quantum Key Distribution, Phys. Rev. X **5**, 031030 (2015).

[24] K. Tamaki, M. Curty, and M. Lucamarini, Decoy-state quantum key distribution with a leaky source, New J. Phys. **18**, 065008 (2016).

[25] W. Wang, K. Tamaki, and M. Curty, Finite-key security analysis for quantum key distribution with leaky sources, New J. Phys. **20**, 083027 (2018).

[26] S. Nauerth, M. Fürst, T. Schmitt-Manderbach, H. Weier, and H. Weinfurter, Information leakage via side channels in freespace BB84 quantum cryptography, New J. Phys. **11**, 065001 (2009).

[27] A. Duplinskiy, and D. Sych, Bounding passive light-source side channels in quantum key distribution via Hong-Ou-Mandel interference, Phys. Rev. A **104**, 012601 (2021).

[28] J. E. Bourassa, A. Gnanapandithan, L. Qian, and H.-K. Lo, Measurement-device-independent quantum key distribution with passive time-dependent source side channels, Phys. Rev. A **106**, 062618 (2022).

[29] A. Mizutani, G. Kato, K. Azuma, M. Curty, R. Ikuta, T. Yamamoto, N. Imoto, H.-K. Lo, and K. Tamaki, Quantum key distribution with setting-choice-independently correlated light sources, npj Quantum Inf. **5**, 8 (2019).

[30] K. Yoshino, M. Fujiwara, K. Nakata, T. Sumiya, T. Sasaki, M. Takeoka, M. Sasaki, A. Tajima, M. Koashi, and A. Tomita, Quantum key distribution with an efficient countermeasure against correlated intensity fluctuations in optical pulses, npj Quantum Inf. **4**, 8 (2018).

[31] M. Pereira, G. Kato, A. Mizutani, M. Curty, and K. Tamaki, Quantum key distribution with correlated sources, Sci. Adv. **6**, eaaz4487 (2020).

[32] F. Grünenfelder, A. Boaron, D. Rusca, A. Martin, and H. Zbinden, Performance and security of 5 GHz repetition rate polarization-based quantum key distribution, Appl. Phys. Lett. **117**, 144003 (2020).

[33] V. Zapatero, Á. Navarrete, K. Tamaki, and M. Curty, Security of quantum key distribution with intensity correlations, Quantum **5**, 602 (2021).

[34] X. Sixto, V. Zapatero, and M. Curty, Security of Decoy-State Quantum Key Distribution with Correlated Intensity Fluctuations, Phys. Rev. Appl. **18**, 044069 (2022).

[35] Á. Navarrete, and M. Curty, Improved finite-key security analysis of quantum key distribution against Trojan-horse attacks, Quantum Sci. Technol. **7**, 035021 (2022).

[36] M. Lucamarini, K. A. Patel, J. F. Dynes, B. Fröhlich, A. W. Sharpe, A. R. Dixon, Z. L. Yuan, R. V. Penty, and A. J. Shields, Efficient decoy-state quantum key distribution with quantified security, Opt. Express **21**, 24550 (2013).

[37] K. Yoshino, T. Ochi, M. Fujiwara, M. Sasaki, and A. Tajima, Maintenance-free operation of WDM quantum key distribution system through a field fiber over 30 days, Opt. Express **21**, 31395 (2013).

[38] Z. Yuan, A. Plews, R. Takahashi, K. Doi, W. Tam, A. W. Sharpe, A. R. Dixon, E. Lavelle, J. F. Dynes, A. Murakami, M. Kujiraoka *et al.* 10-Mb/s quantum key distribution, J. Lightwave Technol. **36**, 3427 (2018).

[39] C. Elliott, A. Colvin, D. Pearson, O. Pikalo, J. Schlafer, and H. Yeh, Current status of the DARPA quantum network, *Quantum Information and Computation III*, edited by E. J. Donkor, A. R. Pirich and H. E. Brandt, Vol. 5815 (SPIE, Philadelphia, 2005), pp. 138–149.

[40] M. Peev, C. Pacher, R. Alléaume, C. Barreiro, J. Bouda, W. Boxleitner, T. Debuisschert, E. Diamanti, M. Dianati, J. F. Dynes *et al.* The SECOQC quantum key distribution network in Vienna, New J. Phys. **11**, 075001 (2009).

[41] T.-Y. Chen, J. Wang, H. Liang, W.-Y. Liu, Y. Liu, X. Jiang, Y. Wang, X. Wan, W.-Q. Cai, L. Ju *et al.* Metropolitan all-pass and inter-city quantum communication network, Opt. Express **18**, 27217 (2010).

[42] D. Stucki, M. Legré, F. Buntschu, B. Clausen, N. Felber, N. Gisin, L. Henzen, P. Junod, G. Litzistorf, P. Monbaron, L. Monat *et al.* Long-term performance of the SwissQuantum quantum key distribution network in a field environment, New J. Phys. **13**, 123001 (2011).

[43] M. Sasaki, M. Fujiwara, H. Ishizuka, W. Klaus, K. Wakui, M. Takeoka, S. Miki, T. Yamashita, Z. Wang, A. Tanaka *et al.* Field test of quantum key distribution in the Tokyo QKD Network, Opt. Express **19**, 10387 (2011).

[44] J. F. Dynes, A. Wonfor, W. W. -S. Tam, A. W. Sharpe, R. Takahashi, M. Lucamarini, A. Plews, Z. L. Yuan, A. R. Dixon, J. Cho *et al.* Cambridge quantum network, npj Quantum Inf. **5**, 101 (2019).

[45] Y.-A. Chen, Q. Zhang, T.-Y. Chen, W.-Q. Cai, S.-K. Liao, J. Zhang, K. Chen, J. Yin, J.-G. Ren, Z. Chen *et al.* An integrated space-to-ground quantum communication network over 4,600 kilometres, Nature (London) **589**, 214 (2021).

[46] S.-K. Liao, W.-Q. Cai, W.-Y. Liu, L. Zhang, Y. Li, J.-G. Ren, J. Yin, Q. Shen, Y. Cao, Z.-P. Li, F.-Z. Li *et al.* Satellite-to-ground quantum key distribution, Nature (London) **549**, 43 (2017).

[47] S.-K. Liao, W.-Q. Cai1, J. Handsteiner, B. Liu, J. Yin, L. Zhang, D. Rauch, M. Fink, J.-G. Ren, W.-Y. Liu *et al.*, Satellite-Relayed Intercontinental Quantum Network, Phys. Rev. Lett. **120**, 030501 (2018).

[48] T. Metger, and R. Renner, Security of quantum key distribution from generalised entropy accumulation, arXiv:2203.04993.

[49] W.-Y. Hwang, Quantum Key Distribution with High Loss: Toward Global Secure Communication, Phys. Rev. Lett. **91**, 057901 (2003).

[50] H.-K. Lo, X. Ma, and K. Chen, Decoy State Quantum Key Distribution, Phys. Rev. Lett. **94**, 230504 (2005).

[51] X.-B. Wang, Beating the Photon-Number-Splitting Attack in Practical Quantum Cryptography, Phys. Rev. Lett. **94**, 230503 (2005).

[52] Our security proof could be applied to the case in which the users employ phase-randomized weak coherent sources. Essentially, the set of single-photon emissions would constitute a "subprotocol" to which our results are directly applicable. However, the outcomes corresponding to this "subprotocol" would not be directly observable, because the users would not know which specific rounds correspond to single-photon emissions. More specifically, Eq. (24) would still hold as an upper bound on the number of phase errors within the single-photon events, but the quantities $\{N_{j,\gamma_X}\}$, which now also refer to single-photon events, would not be directly observable. However, we could use the decoy-state method to find appropriate upper or lower bounds on these quantities [depending on whether their coefficient in Eq. (24) is positive or negative] via linear programming. In a similar way, we could obtain a lower bound on the number of bits in the sifted key that originated from single-photon emissions. By combining both results, we could finally derive a bound on the information leakage to Eve, which is enough to ensure the security of the final key.

[53] A. Navarrete, M. Pereira, M. Curty, and K. Tamaki, Practical quantum key distribution that is secure against side channels, Phys. Rev. Appl. **15**, 034072 (2021).

[54] We believe that this assumption could be relaxed to the requirement that Alice's setting choice on round $k$ does not affect Bob's measurement results on rounds $\{1, \ldots, k-1\}$. This would allow us to take into account the fact that the information about Alice's setting choice cannot travel faster than the speed of light to Bob's measurement unit when determining the constraint on the maximum repetition rate of Alice's source. For example, if Alice and Bob are, say, 100 km apart, then this information must take at least slightly over 0.3 ms to travel between them, and thus it may be enough if Alice emits the $k$th pulse at time $t - 0.3$ ms, where $t$ is the time at which Bob performed his $k-1$th measurement. We note that a similar argument has been made in a recent security proof [55] based on the generalized entropy accumulation theorem [48], which requires a similar condition as our proof. While this argument could be used to achieve a higher repetition rate, here we make the stronger assumption for conceptual simplicity. Another way in which Assumption (A4) could perhaps be relaxed would be to use an approach similar to that in Sec. III B. For example, by assigning rounds a tag $w \in \{0, \ldots, 9\}$ according to the value $w = k \bmod 10$, we could form 10 groups of rounds and then prove the security of each of these groups separately. By doing so, the sequential assumption would perhaps only need to hold for the rounds in each group. Effectively, this would mean that Alice could send her $k$th pulse after Bob had performed his $k-10$th measurement, which would speed up the repetition rate by a factor of 10. However, one would need to make this intuition rigorous before it could be used to replace Assumption (A4), and thus we leave it for future works.

[55] M. Sandfuchs, M. Haberland, V. Vilasini and R. Wolf, Security of differential phase shift QKD from relativistic principles, arXiv:2301.11340.

[56] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Hacking commercial quantum cryptography systems by tailored bright illumination, Nat. Photonics **4**, 686 (2010).

[57] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov, Full-field implementation of a perfect eavesdropper on a quantum cryptography system, Nat. Commun. **2**, 349 (2011).

[58] In fact, the RT has already been applied to an MDI-type protocol [53], and the extensions to the RT introduced in this paper, such as the emission of four states and the consideration of setting-independent factors, do not affect its applicability. As discussed in [53], in this case one would need to consider the joint emitted state by Alice and Bob, which combines the imperfections present in both sources. For example, if Alice and Bob each emit a state in the form of Eq. (1), then their joint emitted state on round $k$ can be expressed as $|\psi_{j,\boldsymbol{g},\tilde{j},\tilde{\boldsymbol{g}}}\rangle_{B_k,E_k,\tilde{B}_k,\tilde{E}_k} = \sqrt{1-\epsilon^{(k)}_{j,\boldsymbol{g},\tilde{j},\tilde{\boldsymbol{g}}}}|\phi_{j,\boldsymbol{g}}\rangle_{B_k,E_k} \otimes |\phi_{\tilde{j},\tilde{\boldsymbol{g}}}\rangle_{\tilde{B}_k,\tilde{E}_k} + \sqrt{\epsilon^{(k)}_{j,\boldsymbol{g},\tilde{j},\tilde{\boldsymbol{g}}}}|\phi^{\perp}_{j,\boldsymbol{g},\tilde{j},\tilde{\boldsymbol{g}}}\rangle_{B_k,E_k,\tilde{B}_k,\tilde{E}_k}$, where $|\phi^{\perp}_{j,\boldsymbol{g},\tilde{j},\tilde{\boldsymbol{g}}}\rangle_{B_k,E_k,\tilde{B}_k,\tilde{E}_k}$ is the joint side-channel state and $\epsilon^{(k)}_{j,\boldsymbol{g},\tilde{j},\tilde{\boldsymbol{g}}} := 1-(1-\epsilon^{(k)}_{j,\boldsymbol{g}})(1-\epsilon^{(k)}_{\tilde{j},\tilde{\boldsymbol{g}}})$.

[59] M. Koashi, Simple security proof of quantum key distribution via uncertainty principle, arXiv:quant-ph/0505108.

[60] M. Koashi, Simple security proof of quantum key distribution based on complementarity, New J. Phys. **11**, 045018 (2009).

[61] G. Currás-Lorenzo, A. Navarrete, M. Pereira, and K. Tamaki, Finite-key analysis of loss-tolerant quantum key distribution based on random sampling theory, Phys. Rev. A **104**, 012406 (2021).

[62] J. L. W. V. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, Acta Math. **30**, 175 (1906).

[63] K. Azuma, Weighted sums of certain dependent random variables, Tohoku Math. J. **19**, 357 (1967).

[64] G. Kato, Concentration inequality using unconfirmed knowledge, arXiv:2002.04357.

[65] G. Currás-Lorenzo, Á. Navarrete, K. Azuma, G. Kato, M. Curty, and M. Razavi, Tight finite-key security for twin-field quantum key distribution, npj Quantum Inf. **7**, 22 (2021).

[66] H. Zhou, T. Sasaki, and M. Koashi, Numerical method for finite-size security analysis of quantum key distribution, Phys. Rev. Res. **4**, 033126 (2022).

[67] A. Mizutani, and G. Kato, Security of round-robin differential-phase-shift quantum-key-distribution protocol with correlated light sources, Phys. Rev. A **104**, 062611 (2021).

[68] H.-K. Lo, H. Chau, and M. Ardehali, Efficient quantum key distribution scheme and a proof of its unconditional security, J. Cryptology **18**, 133 (2005).

[69] A. Chefles, and S. M. Barnett, Optimum unambiguous discrimination between linearly independent symmetric states, Phys. Lett. A **250**, 223 (1998).

[70] M. Dušek, M. Jahma, and N. Lütkenhaus, Unambiguous state discrimination in quantum cryptography with weak coherent states, Phys. Rev. A **62**, 022306 (2000).

[71] Y. Wang, I. W. Primaatmaja, E. Lavie, A. Varvitsiotis, and C. C. W. Lim, Characterising the correlations of prepare-and-measure quantum networks, npj Quantum Inf. **5**, 17 (2019).

[72] I. W. Primaatmaja, E. Lavie, K. T. Goh, C. Wang, and C. C. W. Lim, Versatile security analysis of measurement-device-independent quantum key distribution, Phys. Rev. A **99**, 062332 (2019).

[73] P. J. Coles, E. M. Metodiev, and N. Lütkenhaus, Numerical approach for unstructured quantum key distribution, Nat. Commun. **7**, 11712 (2016).

[74] A. Winick, N. Lütkenhaus, and P. J. Coles, Reliable numerical key rates for quantum key distribution, Quantum **2**, 77 (2018).

[75] D. Mayers, Quantum key distribution and string oblivious transfer in noisy channels, in *Advances in Cryptology — CRYPTO '96*, edited by N. Koblitz (Springer, Berlin, Heidelberg, 1996), pp. 343–357.

[76] P. W. Shor, and J. Preskill, Simple Proof of Security of the BB84 Quantum Key Distribution Protocol, Phys. Rev. Lett. **85**, 441 (2000).

[77] E. Biham, M. Boyer, P. O. Boykin, T. Mor, and V. Roychowdhury, A proof of the security of quantum key distribution, in *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, Portland Oregon, USA* (Association for Computing Machinery, New York, 2000), pp. 715–724.

[78] Encrypting the syndrome information ensures that the actual protocol and the virtual protocol employed in the security proof are equivalent from Eve's perspective (see [59,60] for more information).

[79] In practice, even if Eve injects half of a maximally entangled state into the source, the joint state of the back-reflected light and Eve's ancilla will become mixed as the light is attenuated in Alice's setup. Mathematically, this mixedness could be purified by another ancillary system that Eve has no access to. However, for simplicity, here we consider the worst case scenario in which she has full access to an ancillary system that purifies the back-reflected light.