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 Tamaki4

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

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 work, 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 work 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.

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 minimise 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 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.

∗ mpereira@com.uvigo.es
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 work, 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.

II. DESCRIPTION OF THE Emitted STATES AND ASSUMPTIONS

In this section, we first describe in detail the state of the emitted pulses in the modified BB84 protocol (see Appendix A for a full protocol description), and then list the assumptions imposed by our security proof on Alice’s and Bob’s devices. The security analysis is presented afterwards, in Section III. For ease of discussion, we consider a BB84 protocol with an imperfect single-photon source. Nevertheless, our analysis could also be combined with the decoy-state method [49–51] to deal with phase-randomised coherent sources.

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 $\mathcal{G} := \mathcal{G}_1, \ldots, \mathcal{G}_N$, where $\mathcal{G}_k$ represents all the setting-independent factors that affect the state emitted in the $k\textsuperscript{th}$ 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 $j_1, \ldots, j_N$, the global state emitted is mixed over the probability distribution of $\mathcal{G}$. However, as shown in Appendix B, as long as one demonstrates that the protocol is secure for any particular outcome $g := g_1, \ldots, g_N$ of $\mathcal{G}$, 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, $g$ 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 $k\textsuperscript{th}$ 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 work, 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 $k\textsuperscript{th}$ setting choice, but also on Alice’s previous $l_e$ setting choices, for some known correlation length $l_e$. 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 $j_1, \ldots, j_N$, 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 [52].

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

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

where $B_k$ is a two-dimensional system and $E_k$ includes any other systems that carry information about the $k\textsuperscript{th}$
pulse, such as the back-reflected light from a THA (see Appendix C for more details). Note that $|\psi_{j,g}\rangle_{B_k, E_k}$ in Eq. (1) is a uniquely determined pure state once $j$ and $g_k$ are fixed. However, here we write $g$ rather than $g_k$, because as explained above, in our security proof the parameter $g$, 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 $|\psi_{j,g}\rangle_{B_k, E_k}$ in the basis $\{ |\phi^\perp_{j,g}\rangle_{B_k, E_k}, |\phi^\perp_{g}\rangle_{B_k, E_k} \}$ [31, 53]. In Eq. (1) the parameter $c_{j,g}^{(k)} \in [0, 1]$ quantifies the deviation of $|\psi_{j,g}\rangle_{B_k, E_k}$ from the qubit state $|\phi_{j,g}\rangle_{B_k, E_k} \equiv |\omega_{j,g}\rangle_{B_k} |\lambda_{g}\rangle_{E_k}$, where $|\omega_{j,g}\rangle_{B_k}$ is the state that Alice would send to Bob in the absence of side channels and $|\lambda_{g}\rangle_{E_k}$ is a setting-independent state for the current round. Note that the state $|\omega_{j,g}\rangle_{B_k}$ incorporates any imperfections in a qubit space, such as SPFs and phase fluctuations. The side channels are represented in Eq. (1) by the state $|\phi^\perp_{j,g}\rangle_{B_k, E_k}$, which can live in a Hilbert space of arbitrary dimension and is orthogonal to $|\phi_{j,g}\rangle_{B_k, E_k}$. In other words, the state $|\phi^\perp_{j,g}\rangle_{B_k, E_k}$ 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,

\[
|\psi_{j,g}\rangle_{j_{k-1},...j_{k-l_c}} \rangle_{B_k, E_k} = \sqrt{1 - c_{j,g}^{(k)}} |\phi_{j,g}\rangle_{B_k, E_k} + \sqrt{c_{j,g}^{(k)}} |\phi^\perp_{j,g}\rangle_{j_{k-1},...j_{k-l_c}} \rangle_{B_k, E_k},
\]

where $j_{k-1},...,j_{k-l_c}$ 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 $|\psi_{j,g}\rangle_{j_{k-1},...j_{k-l_c}} \rangle_{B_k, E_k}$ in the basis $\{ |\phi_{j,g}\rangle_{B_k, E_k}, |\phi^\perp_{j,g}\rangle_{j_{k-1},...j_{k-l_c}} \rangle_{B_k, E_k} \}$, and within the framework of classical pulse correlations, this is the most general description of the transmitted states. Note that the state $|\phi^\perp_{j,g}\rangle_{j_{k-1},...j_{k-l_c}} \rangle_{B_k, E_k}$ in Eq. (2), besides incorporating all the side channels in $|\phi^\perp_{j,g}\rangle_{B_k, E_k}$, 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 $p_0 = p_{1z}$.

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:

a. The emitted states do not depend on Alice’s previous setting choices – Eq. (1).

We assume that an upper bound $\epsilon^U \geq c_{j,g}^{(k)}$ is known for all $k$, $j$ and $g$. 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 $g_1, \ldots, g_N$ that represent the setting-independent factors may be correlated between consecutive rounds. We show the security analysis under this scenario in Section III A.

b. The emitted states depend on Alice’s previous $l_c$ setting choices – Eq. (2).

We assume that an upper bound $\epsilon^U \geq c_{j,g}^{(k)}$ is known for all $k$, $j$, $g$, and $j_{k-1},...,j_{k-l_c}$. Moreover, we assume that the state of the $k$th pulse is affected by $g$ and Alice’s previous $l_c$ setting choices, and that $l_c$ is a known parameter. The analysis under this scenario is given in Section III B. As we shall see, the data post-processing 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).

We emphasise that, while knowing the upper bound $\epsilon^U (\epsilon^U)$ is a requirement to apply the RT, the characterisation of the side-channel states $|\phi^\perp_{j,g}\rangle_{B_k, E_k}$, $|\phi^\perp_{j,g}\rangle_{j_{k-1},...j_{k-l_c}} \rangle_{B_k, E_k}$ is not needed. In other words, the inner products $\langle \phi^\perp_{j,g}| \phi^\perp_{j',g}\rangle_{B_k, E_k}$, $\langle \phi^\perp_{j,g}| \phi^\perp_{j',g}\rangle_{B_k, E_k}$, $\langle \phi^\perp_{j,g}| \phi^\perp_{j',g}\rangle_{B_k, E_k}$, $\langle \phi^\perp_{j,g}| \phi^\perp_{j',g}\rangle_{B_k, E_k}$, $\langle \phi^\perp_{j,g}| \phi^\perp_{j',g}\rangle_{B_k, E_k}$, $\langle \phi^\perp_{j,g}| \phi^\perp_{j',g}\rangle_{B_k, E_k}$ 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 characterisation of the side-channel states is very challenging in practice, previous theoretical works [31, 53], as well as this work, have exploited this advantage to consider device models that require minimal experimental characterisation. Nonetheless, it is important to emphasise 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 characterisation 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].
A partial characterisation of the qubit state $|\omega_{j,g}\rangle_{B_k}$ in Eqs. (1) and (2) can be obtained.

In the analysis presented in [31], for simplicity, the qubit state $|\omega_{j,g}\rangle_{B_k}$ in $|\phi_{j,g}\rangle_{B_k,E_k} := |\omega_{j,g}\rangle_{B_k}$ $|g\rangle_{E_k}$ is assumed to be perfectly characterised and stable in time. Here, we go a step further and allow $|\omega_{j,g}\rangle_{B_k}$ to vary slightly round by round, hence its dependence on $g$. However, we assume that one can at least partially characterise this state, such that the upper bounds $c_{g_j}^U$ and $p_{\alpha_X}^{(\text{vir})U}$ 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 Section IV A we show for illustration purposes that, for a typical phase-encoding setup in which the qubit component $|\omega_{j,g}\rangle_{B_k}$ of all the emitted states is in a standard basis plane (such as the XZ plane) and the exact encoded phase $\theta_{j,g}^{(k)}$ fluctuates over time, this requirement translates to being able to determine the range of these fluctuations, i.e. guaranteeing that $\theta_{j,g}^{(k)} \in [\theta_j^L, \theta_j^U]$ for all $k, j$ and $g$, where $\{\theta_j^L, \theta_j^U\} \subset \mathbb{R}$ is known.

Alice only emits her $k^{\text{th}}$ pulse after Bob has performed his $k - 1^{\text{th}}$ measurement [54].

This guarantees that the measurement operators $\hat{M}_Z^{(k)}$ in Eq. (6) ($\hat{D}_X^{(k)}$ in Eq. (27)) satisfies $0 \leq \hat{M}_Z^{(k)} \leq \mathbb{1}$ ($0 \leq \hat{D}_X^{(k)} \leq \mathbb{1}$) (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 generalised entropy accumulation theorem [48] to prove the security of prepare-and-measure protocols against coherent attacks.

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

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

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

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 [55, 56].

(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 [53]).

### 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 Section III A, the emitted states only depend on setting-independent factors and in Section III B, in addition to these factors, the emitted states also depend on Alice’s previous $l_e$ setting choices. Their corresponding assumptions are (A2.a) and (A2.b) in Section 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 $|\psi_{j,g}\rangle_{B_k,E_k}$ 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 [57, 58]. 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

$$|\Psi_g^Z\rangle_{A_k,B_k,E_k} = \frac{1}{\sqrt{2}} \sum_{a \in \{0,1\}} |aZ\rangle_{A_k} |\psi_{aZ,g}\rangle_{B_k,E_k} ,$$

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

\[ |\psi_{\alpha X, g}^{(vir)}\rangle_{B_k, E_k} = \frac{|\psi_{0 Z, g}\rangle_{B_k, E_k} + (1-\alpha)|\psi_{1 Z, g}\rangle_{B_k, E_k}}{2 \sqrt{p_{Z_A}^k}}, \]

where \( \alpha \in \{0, 1\} \), with probabilities

\[ \tilde{p}_{\alpha X, g}^{(k,vir)} = \frac{1}{2} p_{Z_A} \left[ 1 + (1-\alpha) \text{Re} \left( \langle \psi_{0 Z, g} | \psi_{1 Z, g} \rangle_{B_k, E_k} \right) \right], \]

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

A phase error occurs when Alice selects the virtual state associated to \( 1_X \) (0X) 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

\[ P_g^{(k)}(\text{ph|Act}) := \sum_{\alpha, \gamma \in \{0, 1\}} \tilde{p}_{\alpha X, g}^{(k,vir)} p_{Z_B} \text{Tr} \left[ \tilde{\sigma}_{\alpha X, g}^{(k,vir)} M_{\gamma X}^{(k)} \right], \]

where \( \tilde{\sigma}_{\alpha X, g}^{(k,vir)} := |\psi_{\alpha X, g}^{(vir)}\rangle \langle \psi_{\alpha X, g}^{(vir)}|_{B_k, E_k} \) and \( M_{\gamma X}^{(k)} \) with \( \gamma \in \{0, 1\} \) is Bob’s effective POVM element for the \( k \)th pulse after a coherent attack (see Eq. (D7) in Appendix D for more details).

The detection probabilities \( \text{Tr} \left[ \tilde{\sigma}_{\alpha X, g}^{(k,vir)} M_{\gamma X}^{(k)} \right] \) 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 \( P_g^{(k)}(\text{ph|Act}) \) 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: \( P_g^{(k)}(\text{ph|Ref}) \). Then, by evaluating the deviation between the reference and actual states, we obtain \( P_g^{(k)}(\text{ph|Act}) \) from \( P_g^{(k)}(\text{ph|Ref}) \). 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 \( \{ \phi_{j g} \}_{j \in \{0_Z, 1_Z, 0_X, 1_X\}} \), 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

\[ |\Phi_g^Z\rangle_{A_k, B_k, E_k} = \frac{1}{\sqrt{2}} \sum_{\alpha \in \{0, 1\}} |\alpha Z\rangle_{A_k} |\phi_{\alpha Z g}\rangle_{B_k, E_k}, \]

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

\[ P_g^{(k)}(\text{ph|Ref}) := \sum_{\alpha, \gamma \in \{0, 1\}} \tilde{p}_{\alpha X, g}^{(k,vir)} p_{Z_B} \text{Tr} \left[ \tilde{\sigma}_{\alpha X, g}^{(k,vir)} M_{\gamma X}^{(k)} \right], \]

where \( \tilde{\sigma}_{\alpha X, g}^{(k,vir)} := |\psi_{\alpha X, g}^{(vir)}\rangle \langle \psi_{\alpha X, g}^{(vir)}|_{B_k, E_k} \). Here, \( P_g^{(k)}(\text{ph|Ref}) \) could be interpreted as the probability of a phase-error on the \( k \)th round when using the reference states. We emphasise 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 \( \tilde{\sigma}_{\alpha X, g}^{(k,vir)} \) live in the same qubit space as the reference states \( \sigma_{\alpha X, g}^{(k)} := |\phi_{j g}\rangle \langle \phi_{j g}|_{B_k, E_k} \), 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 [59] for a more general treatment. Then, we have that

\[ \tilde{\sigma}_{\alpha X, g}^{(k,vir)} = \sum_j c_j^{(k)} \sigma_{\alpha j, g}^{(k)}, \]

for \( j \in \{0_Z, 1_Z, 0_X, 1_X\} \), where \( c_j^{(k)} \) 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 Section IV C). We note that, in the case of the three-state protocol, the \( 1_X \) state is not emitted and thus \( c_{11 X, g} = c_{01 X, g} = 0 \). 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

\[ P_g^{(k)}(\text{ph|Ref}) = \sum_{\alpha, \gamma \in \{0, 1\}} \tilde{p}_{\alpha X, g}^{(k,vir)} p_{Z_B} \sum_j c_j^{(k)} \text{Tr} \left( \sigma_{j g}^{(k)} M_{\gamma X}^{(k)} \right), \]

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

$$P_g^{(k)}(\text{ph|Ref}) ≤ \sum_{α, γ \in \{0, 1\}} p_{α, x}^{(k, \text{vir})} p_{Z_B} \sum_j c_{α, j, γ}^{(k)} \sum_{j'} G_{j, j'}^{(k)} M_{γ, x}^{(k)}.$$  

(11)

The fictitious probabilities $\operatorname{Tr}[\sigma_j^{(k)} \hat{M}_{γ, x}^{(k)}]$ 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 $P_g^{(k)}(\text{ph|Ref})$. In doing so, we may adopt a number of different inequalities; but here, we choose the following inequality [31]

$$G_-(\operatorname{Tr}[|A⟩⟨A| \hat{M}], ⟨A|R⟩) \leq \operatorname{Tr}[|R⟩⟨R| \hat{M}] \leq G_+(\operatorname{Tr}[|A⟩⟨A| \hat{M}], ⟨A|R⟩),$$  

(12)

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

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

(13)

and

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

(14)

with

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

(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 $P_g^{(k)}(\text{ph|Ref})$. We can then express $P_g^{(k)}(\text{ph|Ref})$ as a function of the detection probabilities of the actual states:

$$P_g^{(k)}(\text{ph|Ref}) ≤ \sum_{α, γ \in \{0, 1\}} p_{α, x}^{(k, \text{vir})} p_{Z_B} \left[ \sum_j c_{α, j, γ}^{(k)} \sum_{j'} G_{j, j'}^{(k)} M_{γ, x}^{(k)} \right]$$  

(16)

where $\hat{M}_{γ, x}^{(k)} = |ψ_j g⟩⟨ψ_j g|_{B_x, E_k}$.

Using the definitions of the actual and reference states, we find that the inner products in Eq. (16) have the form $⟨|ψ_j g⟩⟨ψ_j g|_{B_x, E_k}| = \sqrt{1 - ε_{j, g}^{(k)}}$. For simplicity, we now use the upper bound $ε_{j, g}^{(k)} ≤ ε_{j, g}$ for all $k$, $j$ and $g$ (see assumption (A2.a) in Section II A). Using the fact that the functions $G_+(y, z)$ and $-G_-(y, z)$ are decreasing with respect to $z$ and that $\sqrt{1 - ε_{j, g}^{(k)}} ≤ \sqrt{1 - ε_{j, g}}$, Eq. (16) can then be upper bounded by

$$P_g^{(k)}(\text{ph|Ref}) ≤ \sum_{α, γ \in \{0, 1\}} p_{α, x}^{(k, \text{vir})} p_{Z_B} \left[ \sum_j c_{α, j, γ}^{(k)} G_+(\sqrt{1 - ε_{j, g}^{(k)}}), \sqrt{1 - ε_{j, g}} \right] +$$

$$\sum_j c_{α, j}^{(k)} G_-(\sqrt{1 - ε_{j, g}^{(k)}}, \sqrt{1 - ε_{j, g}}),$$  

(17)

where $P_g^{(k)}(j, γX|\text{Act}) := p_j p_{X_B} \operatorname{Tr}[\hat{M}_{γ, x}^{(k)}]$, $p_j p_{X_B}$ is the joint probability that Alice prepares the actual state $|ψ_j g⟩_{B_x, E_k}$, Bob chooses the $X$ basis and his measurement outcome is $γ$. 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 $P_g^{(k)}(\text{ph|Ref})$, the next step is to obtain an upper bound on the probability of a phase error in the actual protocol: $P_g^{(k)}(\text{ph|Act})$. For this, first note that $P_g^{(k)}(\text{ph|Ref})$ in Eq. (8) can be written as

$$P_g^{(k)}(\text{ph|Ref}) = p_{X_A} p_{Z_B} \operatorname{Tr}[|\hat{Φ}_g^Z⟩⟨\hat{Φ}_g^Z|_{B_x, E_k} \hat{M}_{ph}^{(k)}],$$  

(18)

where

$$\hat{M}_{ph}^{(k)} = \hat{p} \left( |0⟩⟨0|_{A_x} - |1⟩⟨1|_{A_x} \right) \otimes \hat{M}_{0, x}^{(k)} + \hat{p} \left( |0⟩⟨0|_{A_x} + |1⟩⟨1|_{A_x} \right) \otimes \hat{M}_{1, x}^{(k)},$$  

(19)
with \( \hat{P}(|\cdot|) = |\cdot| \). Similarly, \( P_g^{(k)}(\text{ph}|\text{Act}) \) in Eq. (6) can be written as
\[
P_g^{(k)}(\text{ph}|\text{Act}) = p_{Z_A}p_{Z_B}\text{Tr}\left[|\Psi_g^{Z}|^2|\Psi_g^{Z}|_{B_k,E_k} M_g^{(k)}\right].
\] (20)

Hence, one can simply employ the bound in Eq. (12) again to obtain the following expression
\[
P_g^{(k)}(\text{ph}|\text{Act}) \\
\leq p_{Z_A}p_{Z_B}G_+(\frac{p_g^{(k)}(\text{ph}|\text{Ref})}{p_{Z_A}p_{Z_B}}, \sqrt{1 - \epsilon^U}).
\] (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 \( |\langle \Psi_g^Z|\phi_g^Z|_{B_k,E_k} \rangle| = (\sqrt{1 - \epsilon_{yz}^{(k)} + \sqrt{1 - \epsilon_{1z,y}^{(k)}}})/2 \geq \sqrt{1 - \epsilon^U} \).

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 [60] to the right-hand side:
\[
\frac{1}{N} \sum_{k=1}^{N} P_g^{(k)}(\text{ph}|\text{Act}) \\
\leq \frac{1}{N} \sum_{k=1}^{N} p_{Z_A}p_{Z_B}G_+(\frac{p_g^{(k)}(\text{ph}|\text{Ref})}{p_{Z_A}p_{Z_B}}, \sqrt{1 - \epsilon^U}).
\] (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
\[
\frac{1}{N} \sum_{k=1}^{N} P_g^{(k)}(\text{ph}|\text{Ref}) \leq \sum_{\alpha,\gamma \in \{0,1\}} \sum_{\alpha \neq \gamma} \frac{p^{(\text{vir})\text{U}}_{\alpha,\gamma}}{p_{Z_A}} p_{Z_B} \\
\left[ \sum_{c_{\alpha,j}^{(\text{U})} > 0} \sum_{c_{\gamma,j}^{(\text{U})} > 0} \frac{1}{N} \sum_{k=1}^{N} P_g^{(k)}(j, \gamma|\text{Act}) \frac{1}{p_{j}\text{p}_{X_B}}, \sqrt{1 - \epsilon^U} \right] + \\
\sum_{c_{\alpha,j}^{(\text{U})} < 0} \sum_{c_{\gamma,j}^{(\text{U})} > 0} \frac{1}{N} \sum_{k=1}^{N} P_g^{(k)}(j, \gamma|\text{Act}) \frac{1}{p_{j}\text{p}_{X_B}}, \sqrt{1 - \epsilon^U} \right],
\] (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 [61] or Kato’s inequality [62] to substitute each sum of probabilities into its corresponding observable. In particular, for the asymptotic case where \( N \to \infty \), we have that
\[
\sum_{k=1}^{N} P_g^{(k)}(\text{ph}|\text{Act}) \approx N_{\text{ph}}, \text{ where } N_{\text{ph}} \text{ is the number of phase errors. Similarly, for } N \to \infty, \text{ we find that}
\sum_{k=1}^{N} P_g^{(k)}(j, \gamma|\text{Act}) \approx N_{j,\gamma X}, \text{ where } N_{j,\gamma X} \text{ is the number of observed events in which Alice’s setting choice is } j, \text{ Bob selects the } X \text{ basis and his measurement outcome is } \gamma.
\]

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
\[
N_{\text{ph}} \leq N_{\text{ph}}p_{Z_A}p_{Z_B}G_+ \left( \sum_{\alpha,\gamma \in \{0,1\}} \sum_{\alpha \neq \gamma} \frac{p^{(\text{vir})\text{U}}_{\alpha,\gamma}}{p_{Z_A}} \left[ \sum_{c_{\alpha,j}^{(\text{U})} > 0} \sum_{c_{\gamma,j}^{(\text{U})} > 0} \frac{1}{N} \sum_{k=1}^{N} P_g^{(k)}(j, \gamma|\text{Act}) \frac{1}{p_{j}\text{p}_{X_B}}, \sqrt{1 - \epsilon^U} \right] + \sum_{c_{\alpha,j}^{(\text{U})} < 0} \sum_{c_{\gamma,j}^{(\text{U})} > 0} \frac{1}{N} \sum_{k=1}^{N} P_g^{(k)}(j, \gamma|\text{Act}) \frac{1}{p_{j}\text{p}_{X_B}}, \sqrt{1 - \epsilon^U} \right] \right) =: N_{\text{ph}}^{U}.
\] (24)

Finally, we can calculate an upper bound on the phase-error rate of the actual protocol by using
\[
\epsilon_{\text{ph}} := \frac{N_{\text{ph}}^{U}}{N_{\text{det}}^{U}},
\] (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 \( p_g^{(k)}(\text{ph}|\text{Act}) \) and \( P_g^{(k)}(j, \gamma|\text{Act}) \) 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 \( N \) [55, 59, 63, 64], and therefore the performance of the protocol should not be significantly affected in the finite-key regime.

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 \( |\psi_j, g|_{j_k, l_{c-1}, \ldots, j_k, l_{c-1}} \rangle_{B_k, E_k} \), 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
of the $k$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 E1, we show that one can consider the states emitted in round $k$ to be $|\tilde{\psi}_{j,g}|_{j_{k-1},...,j_{k-l_{c}}} B_k E_k$ instead, where $E_k$ includes not only system $E_k$ but also the systems of all rounds after the $k$th

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

$$|\Psi^Z_{g}|_{j_{k-1},...,j_{k-l_{c}}}) A_k B_k E_k = \frac{1}{\sqrt{Z}} \sum_{\alpha \in \{0,1\}} |\alpha Z\rangle A_k |\psi_{\alpha g}|_{j_{k-1},...,j_{k-l_{c}}}) B_k E_k , \quad (26)$$

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 Section IIIA. The reason is the following. To derive the state $|\tilde{\psi}_{j,g}|_{j_{k-1},...,j_{k-l_{c}}}) B_k E_k$ 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 Section IIIA, 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 [65], 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 $w = k \mod (l_{c} + 1)$, 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 keys with round tag $w$, while for the rest of the keys 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, $e_{w}^{\phi}$, 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 $e_{w}^{\phi}$, 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 non-commutativity 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 [65].

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 E2, 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

$$P_{g}^{(k)}(\text{ph}|\text{Act}) \equiv \sum_{\alpha,\gamma \in \{0,1\}} P_{\alpha X,g}^{(k,\text{vir})} |\psi_{\alpha g}|_{j_{k-1},...,j_{k-l_{c}}} B_k E_k \| P_{Z\alpha}^{(k)}.$$  

where

$$P_{\alpha X,g}^{(k,\text{vir})} = \frac{1}{2} P_{\alpha X} A_k \left[ 1 + (-1)^{\alpha} \right].$$

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, $D_{\chi}^{(k)}$ with $\chi \in \{0,1\}$ is Bob's effective POVM element for the $k$th pulse after Eve's coherent attack (see Eq. (E15) in Appendix E2 for more details), $P_{\alpha X,g}^{(k,\text{vir})}$ is defined as

$$P_{\alpha X,g}^{(k,\text{vir})} = \frac{1}{2} P_{\alpha X} |\psi_{\alpha g}|_{j_{k-1},...,j_{k-l_{c}}} B_k E_k$$

where

$$|\psi_{\alpha g}|_{j_{k-1},...,j_{k-l_{c}}} B_k E_k = \frac{1}{\sqrt{Z}} \sum_{\alpha \in \{0,1\}} |\alpha Z\rangle A_k |\psi_{\alpha g}|_{j_{k-1},...,j_{k-l_{c}}} B_k E_k . \quad (27)$$

(28)

Importantly, note that Eq. (27) is very similar to Eq. (6), and thus one can apply the RT in the same way as in Section IIIA 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 $\{ |\phi_{\alpha g}|_{B_k E_k} \}_{\alpha \in \{0,1\}}$, 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:

$$|\psi_{j,g}|_{B_k E_k} \rightarrow |\psi_{j,g}|_{j_{k-1},...,j_{k-l_{c}}} B_k E_k , \quad |\Psi^Z_{g}|_{A_k B_k E_k} \rightarrow |\Psi^Z_{g}|_{j_{k-1},...,j_{k-l_{c}}} A_k B_k E_k , \quad |\phi_{\alpha g}|_{B_k E_k} \rightarrow |\phi_{\alpha g}|_{j_{k-1},...,j_{k-l_{c}}} B_k E_k$$

and $\Phi_{g}^{Z}$ to $\Phi_{g}^{Z}$, where $|\phi_{\alpha g}|_{B_k E_k}$ is defined in Eq. (E17). In the derivations, the inner products $|\langle \psi_{j,g}|_{j_{k-1},...,j_{k-l_{c}}} |\phi_{g_{\alpha}}|_{B_k E_k} \rangle$ and $|\langle \psi^{Z}_{g}|_{j_{k-1},...,j_{k-l_{c}}} |\phi_{g_{\alpha}}|_{A_k B_k E_k} \rangle$ will now appear. In Appendix E3, we show that these inner products are both upper bounded by $\sqrt{1 - e_{U}}$, where $e_{U} = 1 - (1 - e^{U})^{l_{c} + 1}$. Using this result, we obtain Eqs. (17) and (21), which now apply to all rounds with tag $w$ in the $w$th virtual protocol.

By taking the average over all rounds with tag $w$ on both sides of Eq. (21) and using the concavity of $G_{+}(y, z)$
with respect to \( y \), we obtain

\[
\frac{1}{N_w} \sum_{k \in N_w} P^{(k)}_g(\phi|\text{Act}) \leq \frac{1}{N_w} \sum_{k \in N_w} p_{Z_A} p_{Z_B} G_+ \left( \sum \frac{P^{(k)}_g(\phi|\text{Ref})}{p_{Z_A} p_{Z_B}}, \sqrt{1 - \epsilon^U} \right) \leq p_{Z_A} p_{Z_B} G_+ \left( \frac{1}{N_w} \sum_{k \in N_w} P^{(k)}_g(\phi|\text{Ref}), \sqrt{1 - \epsilon^U} \right),
\]

where \( N_w \) is the set of all rounds with tag \( w \) and \( N_w = |N_w| = 1/N \). As in Eq. (23), we now apply Jensen’s inequality to \( \frac{1}{N_w} \sum_{k \in N_w} P^{(k)}_g(\phi|\text{Ref}) \), using the concavity of \( G_+(y, z) \) and \( -G_-(y, z) \) with respect to \( y \). Then, by substituting this expression back in Eq. (30), and applying Azuma’s or Kato’s inequality [66] under the assumption \( N_w \rightarrow \infty \), we have that

\[
N_{\phi, w} \leq N_w p_{Z_A} p_{Z_B} G_+ \left( \sum_{a, \gamma \in \{0, 1\}} \sum_{\alpha \neq \gamma} c_{\alpha,j}^{(k)} p^{(k)}_{a X} \left( \sum_{\alpha \neq \gamma} c_{\alpha,j}^{(k)} \right) \right),
\]

\[
G_+ \left( \frac{N_{j, \gamma X, w}}{N_w p_{j} p_{X_B}}, \sqrt{1 - \epsilon^U} \right) + \sum_{\alpha \neq \gamma} c_{\alpha,j}^{(k)} \left( \sum_{\alpha \neq \gamma} c_{\alpha,j}^{(k)} \right)^{-1} \left( \sum_{\alpha \neq \gamma} c_{\alpha,j}^{(k)} \right)^{-1} \right) = N_{\phi, w}^{(U)},
\]

with \( j \in \{0_Z, 1_Z, 0_X, 1_X\} \), where \( N_{j, \gamma X, w} \) 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,

\[
\epsilon_{\phi, w} \left( \frac{N_{\phi, w}}{N_{(Z)\det}} \right),
\]

where \( N_{(Z)\det} \) 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 \( h(\epsilon_{\phi, w}^{(U)}) \) 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

\[
\sum_{w=0}^{l_v} q_w h(\epsilon_{\phi, w}^{(U)}) \leq h \left( \sum_{w=0}^{l_v} q_w \epsilon_{\phi, w}^{(U)} \right) \leq h(\epsilon_{\phi, w}^{(U)}),
\]

where \( q_w := N_{(Z)\det}^{i w}/N_{(Z)\det} \) and \( c_{\phi}^{(k)} \) is the bound obtained using Eq. (25) in Section III A, which in this case can be regarded as an upper bound on the average phase-error rate. The first inequality in Eq. (33) 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 \( c_{\phi}^{(k)} \) and then apply privacy amplification to the total sifted key at once. One needs to compute each \( \epsilon_{\phi, w}^{(U)} \), and apply privacy amplification separately to each of the \((l_v + 1)\) 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 \( c_{\alpha,j}^{(k)} \) and on the probabilities \( P_{a X, g}^{(k,v)} \). 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 Section 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 \( |\omega_{j,g}^{(k)}\rangle_{B_k} \) in the qubit part of these equations: \( |\phi_{j,g}^{(k)}\rangle_{B_k,E_k} = |\omega_{j,g}^{(k)}\rangle_{B_k} \), satisfies

\[
|\omega_{j,g}^{(k)}\rangle_{B_k} = \cos \left( \frac{\theta_{j,g}^{(k)}}{2} \right) |0_Z\rangle_{B_k} + \sin \left( \frac{\theta_{j,g}^{(k)}}{2} \right) |1_Z\rangle_{B_k} \text{, (34)}
\]

where \( \theta_{j,g}^{(k)} \in [0, 2\pi) \) is the encoding phase, which depends on the round \( k \). Recall that the state in Eq. (34) 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 \( \sigma_{\alpha X, g}^{(k,v)} \) and the reference states \( \sigma_{\alpha}^{(k)} \) through Eq. (9). Then, for both virtual states, we choose the unique solution of Eq. (9) such that \( c_{\phi}^{(k)} = c_{\alpha}^{(k)} = 0 \), which provides the best numerical results for our device model. We elaborate on this point in Section IV C. Using the definition of \( |\phi_{j,g}^{(k)}\rangle_{B_k,E_k} \) and Eq. (34), we can solve the resulting systems of linear equations to find analytical expressions for the other
coefficients $c_{\alpha, j}^{(k)}$ as a function of the encoding phases, i.e. $c_{\alpha, j}^{(k)}(\theta_{1x}^{(k)}, \theta_{1z}^{(k)}, \theta_{0x}^{(k)}, \theta_{0z}^{(k)})$ for $j \in \{0, 1, 2\}$ and $c_{\alpha, j}^{(k)}(\theta_{1x}^{(k)}, \theta_{1z}^{(k)}, \theta_{0x}^{(k)}, \theta_{0z}^{(k)})$ for $j \in \{0, 1, 2\}$ (see Appendix G for full expressions). Similarly, we can express $p_{x, vir}^{(k)}$ as a function of $\theta_{0z}^{(k)}$ and $\theta_{1z}^{(k)}$. 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. $\theta_{j, k}^{(k)} \in [\theta_{j, l}^{(k)}, \theta_{j, l}^{(k)}]$ for $j \in \{0, 1, 2\}$ and $k \in \{0, 1\}$. Then, we can obtain upper bounds $c_{\alpha, j}^{(k)}$ and $p_{x, vir}^{(k)}$ on each individual coefficient $c_{\alpha, j}^{(k)}$ and probability $p_{x, vir}^{(k)}$ 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 $\epsilon_{ph}^{U}$. A lower bound on the asymptotic secret-key rate can then be expressed as

$$R \geq Y_{Z}[1 - h(\epsilon_{ph}^{U}) - fh(\epsilon_{bin})],$$

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 $\epsilon_{bin}$ is the bit error rate. Note that $Y_{Z}$ and $\epsilon_{bin}$ would be directly observed in a practical implementation of the protocol. The lower bound in Eq. (35) 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 $L$ setting choices; the latter is due to Eq. (33). However, we emphasise that this lower bound depends on the value of $\epsilon^{U}$, whose definition differs in the two cases. Namely, in the former, $\epsilon^{U}$ is an upper bound on $\epsilon_{j, g}^{(k)}$ in Eq. (1); while in the latter, $\epsilon^{U} = 1 - (1 - \epsilon^{U})^{L+1}$, where $\epsilon^{U}$ is an upper bound on $\epsilon_{j, g}^{(k)}$ in Eq. (2).

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

To apply our analysis to this particular device model, one needs to experimentally quantify the parameter $\epsilon^{U}$, and $\{\theta_{j}^{(k)}, \epsilon_{j, g}^{(k)}\}$ for all $j$, which represents the uncertainty on the phase of the qubit component. In our simulations, we consider a range of values for $\epsilon^{U}$ and, as an example, we take the uncertainty on the phase to be the same for all $j$, i.e. $\theta_{j}^{(k)} = \theta_{j} - \Delta$ and $\epsilon_{j, g}^{(k)} = \theta_{j} + \Delta$ for some known $\theta_{j}$ and $\Delta$. Moreover, we assume that $\theta_{j}$ deviates from the ideal encoding angles due to SFPs; in particular, we assume that $\theta_{0x} = 0$, $\theta_{1x} = \kappa \pi$, $\theta_{0z} = \kappa \pi / 2$ and $\theta_{1z} = \kappa \pi / 2$, where $\kappa = 1 + \delta / \pi$ and $\delta \geq 0$ 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 $\theta_{j}$ when she selects the setting $j$. This is because $\Delta$ and $\epsilon^{U}$ 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 [67], where $p_{X} = p_{X} \rightarrow 0$ in the asymptotic scenario.

The results for the modified BB84 protocol are shown in Fig. 1. This figure shows that, as $\epsilon^{U}$ increases, the secret-key rate $R$ decreases. This is expected because $\epsilon^{U}$ takes into account multiple side channels, and therefore, an increase in $\epsilon^{U}$ corresponds to a bigger leakage of secret-key information. 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.!
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 Section IV A). There is, however, a difference in the security proof: for the three-state protocol, Eq. (9) can only be expressed as

$$\sigma_{\alpha X g}^{(k, \text{vir})} = \sum_j c_{\alpha j g}^{(k)} \sigma_{j g}^{(k)}$$

(36)

for \(j \in \{0, 1, 0, 1\}\). In this case, the coefficients \(c_{\alpha j g}^{(k)}\) for \(\alpha \in \{0, 1\}\) are both functions of the same encoding phases: \(c_{\alpha j g}^{(k)}(g_{0z g}, g_{1g}, g_{0X g})\). The form of these coefficients as well as their upper bounds are defined in Appendix H. For the simulations, we take \(c^U \in \{10^{-6}, 10^{-3}\}\) 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 \(c^U \in \{10^{-6}, 10^{-3}\}\). This indicates that using four states rather than three, while redundant in an idealised scenario [19], is advantageous in the presence of multiple source imperfections.

![Figure 2](image)

Figure 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 \(\delta = 0.126\).

As pointed out in Section III A, this improvement comes from the extra choice in the coefficients of Eq. (9) when adding a fourth state. Loosely speaking, for each \(\alpha \in \{0, 1\}\), the estimation should be tightest when \(\sum_j |c_{\alpha j g}^{(k)}|\) is minimised, 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 \(\sigma_{1X g}^{(k)}\) and \(\sigma_{1X g}^{(k, \text{vir})}\) (\(\sigma_{0X g}^{(k)}\) and \(\sigma_{0X g}^{(k, \text{vir})}\)) are close in the Bloch sphere, if we take \(c_{1,0X g}^{(k)} = 0\) \((c_{0,1X g}^{(k)} = 0)\), then \(c_{1,1X g}^{(k)}\) \((c_{0,0X g}^{(k)}\) is close to one, while \(c_{1,0X g}^{(k)}\) and \(c_{1,1X g}^{(k)}\) \((c_{0,0X g}^{(k)}\) and \(c_{0,1X g}^{(k)}\) are close to zero, minimising the sum of absolute values. We have numerically tested all possible combinations in which \(\sigma_{\alpha X g}^{(k, \text{vir})}\) can be expressed as a function of \(\sigma_{j g}^{(k)}\) for \(j \in \{0, 1, 0, 1\}\), and confirmed that the choices \(c_{1,0X g}^{(k)} = c_{0,1X g}^{(k)}\) 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 \(\sigma_{1X g}^{(k, \text{vir})}\) as a function of \(\sigma_{1X g}^{(k)}\) rather than \(\sigma_{0X g}^{(k)}\).

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 work, 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 \(c^U\). In other words, as expected, the RT is robust to SPFs independently of the number of states used.

V. CONCLUSION

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, 58, 72–74], but the security of the BB84 protocol with imperfect sources has not yet been fully established. In this work, 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 flawed, leaky and correlated sources. More precisely, our security proof accommodates multiple source imperfections, including state preparation flaws and side channels, such as Trojan-horse attacks, mode dependencies and classical pulse correlations.

Additionally, in this paper we have compared the achievable secret-key rates of the modified BB84 protocol with those obtained when using the three-state loss-tolerant protocol. We have shown that, in the presence of source imperfections, the addition of a fourth state
results in a significant performance improvement, indicating that the modified BB84 protocol offers a clear advantage in guaranteeing the practical security of QKD sources.

In short, this work 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.

VI. ACKNOWLEDGEMENTS

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 Number PID2020-118178RB-C21 and MICIN with funding from the European Union NextGenerationEU (PRTR-C17.I1) and the Galician Regional Government. G.K. acknowledges support from JSPS Postdoctoral Fellowships for Research Communication. M.P. and G.C.-L. acknowledge support from I+D+I con las Comunidades Autónomas in Quantum (PRTR-C17.I1) and the Galician Regional Government funding from the European Union NextGenerationEU Grant Number PID2020-118178RB-C21 and MICIN with Europeo de Desarrollo Regional (FEDER) through Economy and Competitiveness (MINECO), the Fondo search Units: AtlantTIC), the Spanish Ministry of the Galician Regional Government (consolidation of Res-

Appendix A: Description of the modified BB84 QKD protocol

1. **State preparation:** For each round $k \in \{1, \ldots, N\}$, 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 $p_{Z\beta}$ and $p_{X\beta}$, respectively.

3. **Sifting:** Bob announces which rounds were detected, and Alice and Bob reveal their basis choices in these rounds. Let $K$ be the set of detected rounds in which both users employed the $Z$ basis and $T$ 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 $K$, respectively. As for rounds $T$, Alice and Bob announce their respective bit values.

4. **Parameter estimation:** Alice quantifies the number of events $N_{j, \gamma, X}$ on rounds $T$ 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 $N_{\text{ph}, w}^\text{U}$ on $N_{\text{ph}, w}$, the number of phase errors in her sifted key, using Eq. (24).

5. **Data post-processing:** For error correction, Alice sends Bob encrypted syndrome information [75] 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:

- **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 $w = k \mod (l_c + 1)$, where $l_c$ is the correlation length.

Then, Alice and Bob perform the steps 1-3, with $K_w (T_w)$ 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 $K_w$. After that, the users perform parameter estimation and employ data post-processing to obtain a final secure key:

- **Parameter estimation:** Alice quantifies the number of events $N_{j, \gamma, X, w}$ on rounds $T_w$ 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 $N_{\text{ph}, w}^\text{U}$ on $N_{\text{ph}, w}$, the number of phase errors in her $w$th sifted key, using Eq. (31).

- **Data post-processing:** 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.

6. **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 \( G = G_1, \ldots, G_N \) denote the random variables that represent the setting-independent factors affecting the form of Alice’s emitted states, and let \( g = g_1, \ldots, g_N \) denote a particular outcome 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 \( A := A_1, \ldots, A_N \) that she can measure to learn her setting choices. Since in the actual protocol Alice’s emitted states depend on \( g \), this entangled state can be expressed as

\[
\int_{g \in \text{dom}(f_G)} f_G(g) |\Psi_g\rangle \otimes |\psi_{A,B,E}\rangle dg.
\]  

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

\[
|\Psi_g\rangle_{A,B,E} = \bigotimes_{k=1}^{N} \sum_{j_k} \sqrt{p_{j_k}} |j_k\rangle_{A_k} |\psi_{jk,g}\rangle_{B_k,E_k},
\]  

where \( |\psi_{jk,g}\rangle_{B_k,E_k} \) 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 \( |\Psi_g\rangle_{A,B,E} \) in Eq. (B1) takes the form

\[
|\Psi_g\rangle_{A,B,E} = \bigotimes_{k=1}^{N} \sum_{j_k} \sqrt{p_{j_k}} |j_k\rangle_{A_k} |\psi_{jk,g}\rangle_{B_{k-1},j_k-1,B_k,E_k},
\]  

where \( |\psi_{jk,g}\rangle_{B_{k-1},j_k-1,B_k,E_k} \) 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,

\[
\int_{g \in \text{dom}(f_G)} \sqrt{f_G(g)} |g\rangle \langle g|_{G} |\Psi_g\rangle_{A,B,E} dg,
\]  

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

\[
\Pr[N_{\text{ph}} > N^U_{\text{ph}} | G = g] \leq \epsilon,
\]  

where \( \epsilon \) is the failure probability. Importantly, however, this bound is valid for all possible outcomes \( g \), as neither \( N^U_{\text{ph}} \) nor \( \epsilon \) depend on \( g \). 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

\[
\Pr[N_{\text{ph}} > N^U_{\text{ph}}] = \int_{g \in \text{dom}(f_G)} f_G(g) \Pr[N_{\text{ph}} > N^U_{\text{ph}} | G = g]
\]

\[
\leq \int_{g \in \text{dom}(f_G)} f_G(g) \epsilon = \epsilon.
\]  

By a similar argument, we conclude that the phase-error rate bounds derived in Section 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 work, as shown below.

1. Pure output light

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

\[
|E_{jk,g}\rangle_{E_k} = \sqrt{1 - \epsilon(k)} |v\rangle_{E_k} + \sqrt{\epsilon(k)} |e_{jk,g}\rangle_{E_k},
\]  

is the back-reflected light from the THA. Here, \( \epsilon(k) \in [0, 1] \) quantifies the deviation of \( |E_{jk,g}\rangle_{E_k} \) from the vacuum state \( |v\rangle_{E_k} \), and \( |e_{jk,g}\rangle_{E_k} \) is a setting-dependent state orthogonal to \( |v\rangle_{E_k} \), i.e. it belongs to the Fock subspace \( \{1\}, \{2\}, \ldots \). Note that \( |E_{jk,g}\rangle_{E_k} \) 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):

\[
|\psi_{jk,g}\rangle_{B_k,E_k} = \sqrt{1 - \epsilon(k)} |\phi_{jk,g}\rangle_{B_k,E_k} + \sqrt{\epsilon(k)} |\phi^+_g\rangle_{B_k,E_k},
\]  

(C2)
where the qubit state $|\phi_{j,g}\rangle_{B_k,E_k} := |\omega_{j,g}\rangle_{B_k} |v\rangle_{E_k}$ and the side-channel state $|\phi^+_{j,g}\rangle_{B_k,E_k} := |\omega_{j,g}\rangle_{B_k} |e_{j,g}\rangle_{E_k}$; 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:

$$\Omega_{j,g} := \sqrt{1 - c_k} |\omega_{j,g}\rangle_{B_k} + \sqrt{c_k} |\phi_{j,g}\rangle_{B_k},$$

(C3)

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

$$\psi_{j,g} := \sqrt{1 - \epsilon^{(k)}_{j,g}} |\omega_{j,g}\rangle_{B_k} + \sqrt{\epsilon^{(k)}_{j,g}} |\phi_{j,g}\rangle_{B_k},$$

(C4)

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

$$\phi_{j,g} := \sqrt{1 - \epsilon^{(k)}_{j,g}} |\omega_{j,g}\rangle_{B_k} + \sqrt{\epsilon^{(k)}_{j,g}} |\phi_{j,g}\rangle_{B_k}.$$  

(C5)

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

$$\Omega_{j,g|\mathcal{X}_{j-1},\ldots,j_{k-1}} := |\omega_{j,g}\rangle_{B_k}$$

and the side-channel state would become $|\phi_{j,g|\mathcal{X}_{j-1},\ldots,j_{k-1}}\rangle_{B_k,E_k} := |\omega_{j,g}\rangle_{B_k} \otimes |E_{j,g}\rangle_{E_k}$; 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 characterising the value of the attenuation in Alice’s setup [23], one can bound the maximum intensity of the output light $\nu_{\text{max}}$. Mathematically, this bound can be expressed as $|E_{j,g}|N|E_{j,g}\rangle_{E_k} \leq \nu_{\text{max}}$ where $N = \sum_n n|n\rangle\langle n|$ is the photon-number operator. Then, using Eq. (C1), we have that

$$\nu_{\text{max}} \geq \langle E_{j,g}|N|E_{j,g}\rangle_{E_k} = \epsilon^{(k)} \langle e_{j,g}|N|e_{j,g}\rangle_{E_k} \geq \epsilon^{(k)},$$

(C7)

with equality if and only if $|e_{j,g}\rangle_{E_k}$ 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 $\hat{E}_k$ is purified by an ancillary system $\hat{E}_k'$ that is in Eve’s hands [76]. We can then express the joint state of the back-reflected light and this ancillary system as

$$|E_{j,g}\rangle_{E_k} = \sum_c \sqrt{p_c} |c\rangle \langle c|_{\hat{E}_k} \left(\sqrt{1 - \epsilon^{(k)}_c} |v\rangle_{\hat{E}_k}ight.$$ 

$$+ \sqrt{\epsilon^{(k)}_c} |e_{j,g,c}\rangle_{\hat{E}_k}\right),$$

(C8)

where $E_k := E_k' \otimes \hat{E}_k$ 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 $|\phi_{j,g}\rangle_{B_k,E_k} = |\omega_{j,g}\rangle_{B_k} |\lambda\rangle_{E_k}$, where $|\lambda\rangle_{E_k} := \sum_c \sqrt{p_c} |c\rangle \langle c|_{\hat{E}_k} |v\rangle_{\hat{E}_k}$, and $|\phi^+_{j,g}\rangle_{B_k,E_k} := |\omega_{j,g}\rangle_{B_k} \sum_c \sqrt{p_c} |c\rangle \langle c|_{\hat{E}_k} |e_{j,g,c}\rangle_{\hat{E}_k}$. If $\epsilon^{(k)}_c \leq \epsilon^{(k)}$ 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 $\nu_{\text{max}} \geq \epsilon^{(k)}$ 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 $\nu_{\text{max}}$. To use the RT in this case, one can select the reference states as follows

$$|\phi_{j,g}\rangle_{B_k,E_k} := |\omega_{j,g}\rangle_{B_k} |\lambda\rangle_{E_k},$$

$$|\phi^+_{j,g}\rangle_{B_k,E_k} := |\omega_{j,g}\rangle_{B_k} \sum_c \sqrt{p_c} |c\rangle \langle c|_{\hat{E}_k} \left(\sqrt{1 - \epsilon^{(k)}_c} |v\rangle_{\hat{E}_k}ight.$$ 

$$+ \sqrt{\epsilon^{(k)}_c} |e_{j,g,c}\rangle_{\hat{E}_k}\right),$$

(C9)

where $\nu_{\text{max}} \geq \epsilon^{(k)}$ in Eq. (C7) and apply our security analysis.

Experimental method to upper bound $\sum_c p_c \epsilon^{(k)}$

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

$$\nu_{\text{max}} \geq \langle E_{j,g}|\hat{N} |E_{j,g}\rangle_{E_k} \otimes \hat{N}_E |E_{j,g}\rangle_{E_k}$$

where $\hat{N}_E$ is the photon-number operator.
\[
\sum_{c} p_{c} e^{(k)}_{c} (e_{j,c} | \hat{N}_{E_k} | e_{j,c}) E_k \geq \sum_{c} p_{c} e^{(k)}_{c},
\]
(C11)

with equality if \( |e_{j,c} \rangle \) is a single-photon state.

**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 Section II A), the operator \( M^{(k)}_{\gamma \alpha} \) satisfies \( 0 \leq M^{(k)}_{\gamma \alpha} \leq \hat{1} \). To do so, we consider an entanglement-based scenario that is equivalent to the actual protocol. For a fixed \( g \) (see Appendix B), the transmission of \( N \) pulses can then be described by Alice first preparing the following entangled state

\[
|\Psi^g_{A,B,E} \rangle = \bigotimes_{k=1}^{N} \sum_{j_k} \sqrt{p_{j_k}} |j_k \rangle_{A_k} |\psi_{j_k,g} \rangle_{B_k,E_k}
\]

then keeping system \( A \) in her laboratory and sending systems \( B,E \) through the quantum channel. Here, \( A := A_1, \ldots, A_N \) and \( B,E := B_1, E_1, \ldots, B_N, E_N \) refers to the composite systems of Alice’s ancillae and to the pulses sent to Bob, respectively, where \( A_k \) and \( B_k \) for \( k \in \{1, \ldots, N\} \) denote Alice’s and Bob’s kth 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), \( \{|j_k \rangle_{A_k}\} \in \{0_Z, 1_Z, 0_X, 1_X\} \) is a set of orthonormal states and \( |\psi_{j_k,g} \rangle_{B_k,E_k} \) is defined in Eq. (1). Note that, for ease of discussion, in this appendix we explicitly write the setting choice on the kth 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 kth 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 \( \hat{U}^{(N)}_{B_N,E_N,\tilde{E}_N} \cdots \hat{U}^{(1)}_{B_{12},E_{12},\tilde{E}_{12}} \), where \( \hat{U}^{(k)}_{B_k,E_k,\tilde{E}_k} \) acts on the kth 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 \( j_k \in \{0_X, 1_X\} \), whose respective outcomes are denoted by \( a = Z \) and \( a = X \), and Bob selects the measurement basis \( \beta \in \{Z,X\} \) with probability \( p_{\beta \gamma} \). In the modified BB84 protocol, the secret key is generated from the rounds in which \( (a, \beta) = (Z,Z) \). As explained in Section 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. \( \{|\alpha_{X,0}^{(\text{vir})} \rangle_{A_k} = (|0_Z \rangle_{A_k} + (-1)^{\alpha} |1_Z \rangle_{A_k})/\sqrt{2} \}_{\alpha \in \{0,1\}} \) and \( \{\hat{m}_{0X}, \hat{m}_{1X}, \hat{m}_{0Y}, \hat{m}_{1Y}\} \). 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 \( o_k \in \{(a, \beta, \alpha, \gamma), (f)\} \), where \( \alpha, \gamma \in \{0,1\} \) 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 summarise the possible POVM elements as

\[
\begin{align*}
\hat{F}^{(Z,Z,\alpha,\gamma)}_{A_k, B'_k} &= |\alpha_X^{(\text{vir})} \rangle \langle \alpha_X^{(\text{vir})} |_{A_k} \otimes p_{ZB} \hat{m}_{\gamma_X}, \\
\hat{F}^{(Z,X,\alpha,\gamma)}_{A_k, B'_k} &= |\alpha_Z \rangle \langle \alpha_Z |_{A_k} \otimes p_{XB} \hat{m}_{\gamma_X}, \\
\hat{F}^{(X,\beta,\alpha,\gamma)}_{A_k, B'_k} &= |\alpha_X \rangle \langle \alpha_X |_{A_k} \otimes p_{\beta\gamma} \hat{m}_{\gamma_B}, \\
\hat{F}^{(f)}_{A_k, B'_k} &= \hat{1}_{A_k} \otimes \hat{1}_{f}.
\end{align*}
\]

(D2)

Note that \( \hat{F}^{(Z,Z,\alpha,\gamma)}_{A_k, B'_k} \) is the POVM element associated with the estimation of the number of phase errors. A phase error on the kth 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 kth round conditional on all the previous outcomes \( o_{k-1} \) can be expressed as

\[
P^{(k)}_{g}(\text{ph}\text{Act} | o_{k-1}) = \frac{P(o_k = (Z, Z, 0, 1) \text{ or } (Z, Z, 1, 0) | o_{k-1})}{P(o_{k-1})},
\]

(D3)
where $P(\omega_{k-1})$ 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

\[
\begin{align*}
\hat{P} \left( \sum_{\alpha, \gamma \in \{0, 1\} \atop \alpha \neq \gamma} \text{Tr} \left[ F_{A_{\alpha, \gamma}}^{(k)} \hat{U}_{B_{\alpha, \gamma}}^{(k-n)} \hat{U}_{B_{\alpha, \gamma}}^{(k-n)} |0 \rangle \prod_{n=1}^{k-1} |\Psi_{A_{\alpha, \gamma}}^{(n)} \rangle_{A_{\alpha, \gamma}, B_{\alpha, \gamma}, E_{\alpha, \gamma}} \right] \right] \\
&= \sum_{\alpha, \gamma \in \{0, 1\} \atop \alpha \neq \gamma} \text{Tr} \left[ F_{A_{\alpha, \gamma}}^{(k)} \hat{U}_{B_{\alpha, \gamma}}^{(k-n)} \hat{U}_{B_{\alpha, \gamma}}^{(k-n)} |0 \rangle \prod_{n=1}^{k-1} |\Psi_{A_{\alpha, \gamma}}^{(n)} \rangle_{A_{\alpha, \gamma}, B_{\alpha, \gamma}, E_{\alpha, \gamma}} \right] \\
&= \sum_{\alpha, \gamma \in \{0, 1\} \atop \alpha \neq \gamma} \text{Tr} \left[ \hat{U}_{B_{\alpha, \gamma}}^{(k-n)} \hat{U}_{B_{\alpha, \gamma}}^{(k-n)} |0 \rangle \prod_{n=1}^{k-1} |\Psi_{A_{\alpha, \gamma}}^{(n)} \rangle_{A_{\alpha, \gamma}, B_{\alpha, \gamma}, E_{\alpha, \gamma}} \right]
\end{align*}
\]

(D4)

where

\[
\begin{align*}
|\Psi_{A_{\alpha, \gamma}}^{(k-n)} \rangle_{A_{\alpha, \gamma}, B_{\alpha, \gamma}, E_{\alpha, \gamma}} := & \prod_{n=1}^{k-1} \hat{U}_{B_{\alpha, \gamma}}^{(k-n)} |0 \rangle \\
& \prod_{n=1}^{k-1} |\Psi_{A_{\alpha, \gamma}}^{(n)} \rangle_{A_{\alpha, \gamma}, B_{\alpha, \gamma}, E_{\alpha, \gamma}}
\end{align*}
\]

(D5)

with $A_{\leq k-1} := A_{1}, \ldots, A_{k-1}$ and $B_{\leq k-1}, E_{\leq k-1} := B_{1}, E_{1}, \ldots, B_{k-1}, E_{k-1}$, and where $|\psi_{\alpha, \gamma}^{(k)} \rangle_{B_{k}, E_{k}}$ and $p_{\alpha, \gamma}^{(k)}$ are defined in Eqs. (4) and (5), respectively.

In the second equality of Eq. (D4), we took the partial trace over systems $A_{\geq k+1}, B_{\geq k+1}, E_{\geq k+1}$. In the third equality of Eq. (D4), we defined the state $|\Psi_{A_{\alpha, \gamma}}^{(k-n)} \rangle_{A_{\alpha, \gamma}, B_{\alpha, \gamma}, E_{\alpha, \gamma}}$ 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 $\hat{F}_{A_{\alpha, \gamma}}^{(k)}$ in Eq. (D2) for $\alpha, \gamma \in \{0, 1\}$ such that $\alpha \neq \gamma$.

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

\[
P_{g}^{(k)}(\text{ph|Act}) = \sum_{\alpha, \gamma \in \{0, 1\} \atop \alpha \neq \gamma} p_{\alpha, \gamma}^{(k)} |\psi_{\alpha, \gamma}^{(k)} \rangle_{B_{k}, E_{k}} \prod_{\gamma} \hat{U}_{B_{\gamma}}^{(k)} |\Psi_{A_{\alpha, \gamma}}^{(k)} \rangle_{A_{\alpha, \gamma}, B_{\alpha, \gamma}, E_{\alpha, \gamma}}
\]

(D6)

with $\hat{M}_{\alpha, \gamma}^{(k)} := |\psi_{\alpha, \gamma}^{(k)} \rangle_{B_{k}, E_{k}}$ and

\[
\hat{M}_{\gamma}^{(k)} := \frac{\langle \Psi_{\alpha, \gamma}^{(k)} | \hat{U}_{B_{k}}^{(k)} | \Psi_{\alpha, \gamma}^{(k)} \rangle}{P(o_{k-1})},
\]

(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 $\hat{M}_{\gamma}^{(k)}$ in Eq. (D7) satisfies $0 \leq \hat{M}_{\gamma}^{(k)} \leq \hat{1}$. For this, first note that $\hat{M}_{\gamma}^{(k)} \geq 0$ holds because for any state $|\varphi \rangle$ we have that $\langle \varphi | \hat{M}_{\gamma}^{(k)} | \varphi \rangle = \langle \sqrt{\hat{m}}_{\gamma}^{|k\rangle} | \hat{U}_{B_{k}}^{(k)} | \hat{m}_{\gamma}^{|k\rangle} | \varphi \rangle P(o_{k-1})^{-1/2} \rangle \geq 0$. Then, since $\hat{m}_{\gamma}^{|k\rangle}$ is a POVM element, we have that $\hat{m}_{\gamma}^{|k\rangle} \leq \hat{1}$, and thus we can upper bound $\hat{M}_{\gamma}^{(k)}$ as

\[
\hat{M}_{\gamma}^{(k)} \leq \frac{\langle \Psi_{\alpha, \gamma}^{(k)} | \hat{U}_{B_{k}}^{(k)} | \Psi_{\alpha, \gamma}^{(k)} \rangle}{P(o_{k-1})}
\]

(D8)

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

**Appendix E: Results used in Section III B**

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

Here, we show how to derive $|\tilde{\psi}_{jk,g} j_{k-1} \ldots j_{k-l_c} \rangle_{B_k, E'_k}$ 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 $g$ (see Appendix B), the transmission of $N$ pulses is described by Alice first preparing $N$ ancilla systems and $N$ pulses in the state

$$|\Psi g\rangle_{A,B,E} = \sum_{j_1} \sqrt{p_{j_1}} |j_1\rangle_{A_1} |\psi_{j_1.g} \rangle_{B_1, E_1}$$

$$|\Psi g j_{k-1} \ldots j_{k-l_c} \rangle_{A_k, B_k, E'_k} := \sum_{j_k} \sqrt{p_{j_k}} |j_k\rangle_{A_k} |\psi_{jk.g} j_{k-1} \ldots j_{k-l_c} \rangle_{B_k, E'_k} \sum_{n=1}^{l_c} \sum_{j_{k+n}} \sqrt{p_{j_{k+n}}} |j_{k+n}\rangle_{A_{k+n}} |\psi_{jk+n,g} j_{k+n-1} \ldots j_{k-l_c} \rangle_{B_{k+n}, E_{k+n}}$$

$$\otimes \sum_{j_N} \sqrt{p_{j_N}} |j_N\rangle_{A_N} |\psi_{jN,g} j_{N-1} \ldots j_{k-l_c} \rangle_{B_N, E_N},$$

and then sending systems $B, E$ through the quantum channel. Here, $A := A_1, \ldots, A_N$ and $B, E := B_1, E_1, \ldots, B_N, E_N$ refers to the composite systems of Alice’s ancillae and to the pulses sent to Bob, respectively, where $A_k$ and $B_k$ for $k \in \{1, 2, \ldots, N\}$ 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), $\{ |j_k\rangle_{A_k} \}_{j_k \in \{0,1\}}$ is a set of orthonormal states and $|\psi_{jk,g} j_{k-1} \ldots j_{k-l_c} \rangle_{B_k, E'_k}$ 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 ancilla $A_{k-1} := A_1, \ldots, A_{k-1}$. Once we trace out the systems $A_{k-1} := A_1, \ldots, A_{k-1}, B_{k-1}, E_{k-1}$, this post-measurement state can be expressed as

$$|\tilde{\psi}_{jk+lc+1} j_{k+lc+1} \rangle_{A_{k+lc+1}} |\psi_{jk+lc+1,g} j_{k+lc+1} \ldots j_{k+lc} \rangle_{B_{k+lc+1}, E_{k+lc+1}} \otimes \sum_{j_N} \sqrt{p_{j_N}} |j_N\rangle_{A_N} |\psi_{jN,g} j_{N-1} \ldots j_{k-l_c} \rangle_{B_N, E_N},$$

which forms a set of orthogonal states as $
\{ |\psi_{jk+lc+1,g} \rangle_{A_{k+lc+1}, B_{k+lc+1}, E_{k+lc+1}} \}_{j_k \in \{0,1\}, j_{k+lc+1} \in \{0,1\}}$. Next, using Eq. (E3), we define the following state

$$|\tilde{\lambda}_{jk\ldots j_{k+lc}} \rangle_{A_{k+lc+1}, B_{k+lc+1}, E_{k+lc+1}} := \sum_{j_{k+1}} \sqrt{p_{j_{k+1}}} \otimes \sum_{j_{k+lc}} |\psi_{jk+lc+1,g} j_{k+lc+1} \rangle_{A_{k+lc+1}, B_{k+lc+1}, E_{k+lc+1}}$$

$$|\tilde{\psi}_{jk+lc+1} j_{k+lc+1} \rangle_{A_{k+lc+1}} |\psi_{jk+lc+1,g} j_{k+lc+1} \ldots j_{k+lc} \rangle_{B_{k+lc+1}, E_{k+lc+1}} \otimes \sum_{j_N} \sqrt{p_{j_N}} |j_N\rangle_{A_N} |\psi_{jN,g} j_{N-1} \ldots j_{k-l_c} \rangle_{B_N, E_N},$$

Note that $|\tilde{\lambda}_{jk\ldots j_{k+lc}} \rangle_{A_{k+lc+1}, B_{k+lc+1}, E_{k+lc+1}}$ 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

\[ |\Psi_{jk} \rangle_{j_k \ldots j_{k-1}} = \sum_{j_k} \sqrt{P_{j_k}} |j_k \rangle A_k \]
\[ |\hat{\phi}_{jk} \rangle_{j_k \ldots j_{k-1}} = \sum_{j_k} \sqrt{P_{j_k}} |j_k \rangle A_k \]

Now, it is useful to decompose \(|\psi_{jk} \rangle_{j_k \ldots j_{k-1}} \rangle_{A_{k+1}, E_k} \) in Eq. (E5). For this, we first combine Eq. (E5) and Eq. (E4), and then use Eq. (2) such that

\[ |\psi_{jk} \rangle_{j_k \ldots j_{k-1}} \rangle_{A_{k+1}, \cdots, A_{k+1}, \cdots, A_{k+1}, E_{k+1}, \cdots, E_{k+1}} = |\lambda_{jk} \rangle_{j_k \ldots j_{k+1}, \cdots, j_{k+1}, \cdots, j_{k+1}, E_{k+1}, \cdots, E_{k+1}} \]

In this equation, we have defined

\[ \epsilon^{(k)}_{j_k,j_{k-1}, \ldots, j_{k-1}} := 1 - \left( 1 - \epsilon^{(k)}_{j_k,j_{k-1}, \ldots, j_{k-1}} \right) \]
\[ \prod_{n=1}^{j_k} \sqrt{P_{j_k}} \sum_{j_{k+1}} \sqrt{P_{j_{k+1}}} |j_{k+1} \rangle \cdots |\lambda_{j_k, \ldots, j_{k+1}, \cdots, j_{k+1}, \cdots, j_{k+1}, E_{k+1}, \cdots, E_{k+1}} \rangle = |\hat{\phi}_{jk} \rangle_{B_k, E_k} \]

and \(|\hat{\phi}_{jk} \rangle_{B_k, E_k} \) is a state orthogonal to \(|\lambda_{jk} \rangle_{j_k \ldots j_{k+1}, \cdots, j_{k+1}, \cdots, j_{k+1}, E_{k+1}, \cdots, E_{k+1}} \rangle_{A_{k+1}, E_k} \), 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

\[ |\Psi_{jk} \rangle_{j_k \ldots j_{k-1}} \rangle_{A_k, B_k, E_k} \]
\[ = \sum_{j_k} \sqrt{P_{j_k}} |\lambda_{jk} \rangle_{j_k \ldots j_{k-1}} \rangle_{B_k, E_k} \]

which implies that we can regard the emitted states on round \( k \) to be \(|\hat{\phi}_{jk} \rangle_{B_k, E_k} \).

Now, suppose that Alice performs a measurement on system \( A_k \) such that she learns if \( j_k \in \{0Z, 1Z\} \) or if \( j_k \in \{0X, 1X\} \), 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

\[ |\Psi^Z_{jk} \rangle_{j_k \ldots j_{k-1}} \rangle_{A_k, B_k, E_k} \]
\[ = \frac{1}{\sqrt{2}} \sum_{\alpha \in \{0, 1\}} |\alpha Z \rangle_{A_k} |\hat{\phi}_{\alpha Z, j_k \ldots j_{k-1}} \rangle_{B_k, E_k} \]

where we have used \( p_{0Z} = p_{1Z} \) (see assumption (A1) in Section 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) Section II A), the operator $\hat{D}_{x}^{(k)}$ satisfies $0 \leq \hat{D}_{x}^{(k)} \leq \hat{1}$. To do so, we consider again an entanglement-based scenario that is equivalent to the actual protocol. For a fixed $g$, the transmission of $N$ pulses can then be described by Alice first preparing the entangled state $|\Psi_{g}\rangle_{A,B,E}$ in Eq. (E1), keeping system $A$ in her laboratory and sending systems $B,E$ 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 $\hat{U}_{B_{1},E_{1}}^{(N)} \ldots \hat{U}_{A_{1}}^{(N)}$, where $\hat{U}_{A_{k},B_{k},E_{k}}^{(k)}$ acts on the $k^{th}$ photonic system $B_{k}$, $E_{k}$ and on Eve’s updated ancilla $\tilde{E}_{k}$, resulting in systems $B_{k}^{o}$ 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. $(\alpha, \beta) = (Z, Z)$, where $\alpha$ ($\beta$) 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. $\{|a_{\lambda}^{(\text{vir})}\rangle_{A_{k}} = (|0\rangle_{A_{k}} + (-1)^{\lambda} |1\rangle_{A_{k}})/\sqrt{2}\}_{\lambda \in \{0, 1\}}$ and $\{\tilde{m}_{0}, \tilde{m}_{1}, \tilde{m}_{f}\}$, 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 Section 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 $w = k \mod (l_{c}+1)$ 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 $\{F_{A_{k},B_{k}^{o}}^{(w)}\}_{o_{k}}$, whose elements are defined in Eq. (D2), and on the other rounds, Alice’s and Bob’s measurements are described by the POVM $\{J_{A_{k},B_{k}^{o}}^{(w)}\}_{o_{k}}$, whose elements are defined as $J_{A_{k},B_{k}^{o}}^{(w)} = |a_{\alpha}\rangle_{A_{k}} \otimes p_{\beta} \tilde{m}_{\gamma}$ and $J_{A_{k},B_{k}^{o}}^{(f)} = \tilde{1}_{A_{k}} \otimes \tilde{m}_{f}$, where $\alpha, \gamma \in \{0, 1\}$ 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:

$$P^{(k)}(\text{ph|Act}) = \frac{P(o_{k} = (Z, Z, 0, 1) \text{ or } (Z, Z, 1, 0), o_{k-1})}{P(o_{k-1})},$$

(E11)

where $P(o_{k-1})$ is the probability of obtaining Alice’s and Bob’s previous outcomes. Let us now introduce the POVM $\{\hat{L}_{A_{k},B_{k}^{o}}^{(w)}\}_{o_{k}}$, which is defined as $\{F_{A_{k},B_{k}^{o}}^{(w)}\}_{o_{k}}$ when a pulse $k$ has the tag $w$ and defined as $\{J_{A_{k},B_{k}^{o}}^{(w)}\}_{o_{k}}$ when it has another tag. Therefore, by using this POVM, the form of the emitted state in Eq. (E1) as well as the de-
\[
\begin{align*}
&= \sum_{\alpha, \gamma \in \{0, 1\}} \alpha \neq \gamma \quad \text{Tr} \left[ A_{\leq k-1, B'_{\leq k-1}, E_{\leq k-1}, \tilde{E}_k} \left( \Psi_g^{k-1} \right) | \tilde{U}^{(k)}_{B_k, E_k, \tilde{E}_k} \tilde{F}^{(Z, Z, \alpha, \gamma)}_{A_k, B'_k} \tilde{U}^{(k)}_{B_k, E_k, \tilde{E}_k} \Psi_g^{k-1} \right] A_{\leq k-1, B'_{\leq k-1}, E_{\leq k-1}, \tilde{E}_k} \\
&= \sum_{\alpha, \gamma \in \{0, 1\}} \alpha \neq \gamma \quad \text{Tr} \left[ A_{\leq k-1, B'_{\leq k-1}, E_{\leq k-1}, \tilde{E}_k} \left( \Psi_g^{k-1} \right) | \tilde{U}^{(k)}_{B_k, E_k, \tilde{E}_k} \tilde{m}_{\gamma \times} \tilde{U}^{(k)}_{B_k, E_k, \tilde{E}_k} \Psi_g^{k-1} \right] A_{\leq k-1, B'_{\leq k-1}, E_{\leq k-1}, \tilde{E}_k} \\
&= P^{(k, \text{vir})}_{\alpha X, g, j_{k-1}, \ldots, j_{k-\ell}} P_{ZB} \left| \psi^{(\text{vir})}_{\alpha X, g, j_{k-1}, \ldots, j_{k-\ell}} \right\rangle \left\langle \psi^{(\text{vir})}_{\alpha X, g, j_{k-1}, \ldots, j_{k-\ell}} \right| B_k, E_k', \right]
\end{align*}
\]

where

\[
|\Psi_g^{k-1}\rangle_{A_{\leq k-1}, B'_{\leq k-1}, E_{\leq k-1}, \tilde{E}_k} := \left[ \prod_{n=1}^{k-1} \sqrt{P_{j_n}} | j_n \rangle_{A_n} \left| \psi_{\alpha X, g, j_{k-1}, \ldots, j_{k-\ell}} \right\rangle_{B_k, E_k'} \right],
\]

with \(A_{\leq k-1} := A_1, \ldots, A_{k-1}\) and \(B'_{\leq k-1}, E_{\leq k-1} := B'_1, E_1, \ldots, B'_{k-1}, E_{k-1}\), and where \(|\psi^{(\text{vir})}_{\alpha X, g, j_{k-1}, \ldots, j_{k-\ell}}\rangle_{B_k, E_k'}\) and \(P^{(k, \text{vir})}_{\alpha X, g, j_{k-1}, \ldots, j_{k-\ell}}\) 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 \(A_{\leq k-1}\) commute with any measurements or operations on systems \(B_1, E_1, \ldots, B_{k-1}, E_{k-1}, \tilde{E}_k\), and the fact that \(\hat{L}_{A_k, B'_k}\) can be decomposed as \(\hat{L}_{A_k} \otimes \hat{L}_{B'_k}\), where Alice’s outcome \(\alpha_k\) and Bob’s outcome \(\beta_k\) on round \(k\) are defined as \((d, \alpha, \beta)\), respectively. In the third equality of Eq. (E12), we have used the definition of the post-measurement state \(|\Psi_g^{(\text{j}, \ldots, \text{j}_{k-\ell})}, A_k, B_k, E_k\rangle\) in Eq. (E2), and in the fourth equality of Eq. (E12), we have defined the state \(|\Psi_g^{k-1}\rangle_{A_{\leq k-1}, B'_{\leq k-1}, E_{\leq k-1}, \tilde{E}_k}\) 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. (E26) and the definition of \(\hat{F}^{(Z, Z, \alpha, \gamma)}\) in Eq. (D2) for \(\alpha, \gamma \in \{0, 1\}\) such that \(\alpha \neq \gamma\).

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

\[
P^{(k, \text{ph})}_{g}(\text{ph}|\text{Act}) := \sum_{\alpha, \gamma \in \{0, 1\}} \alpha \neq \gamma \quad \text{Tr} \left[ \sigma^{(k, \text{vir})}_{\alpha X, g, j_{k-1}, \ldots, j_{k-\ell}} \hat{D}^{(k)}_{\gamma X} \right],
\]

with \(\sigma^{(k, \text{vir})}_{\alpha X, g, j_{k-1}, \ldots, j_{k-\ell}} := \hat{P} \left( |\psi^{(\text{vir})}_{\alpha X, g, j_{k-1}, \ldots, j_{k-\ell}} \rangle_{B_k, E_k'} \right)\)

and

\[
\hat{D}^{(k)}_{\gamma X} := \frac{\langle \Psi_g^{k-1}| \tilde{U}^{(k)}| \tilde{U}^{(k)*} \rangle_{B_k, E_k, \tilde{E}_k} \tilde{m}_{\gamma \times} \tilde{U}^{(k)}| \Psi_g^{k-1} \rangle_{B_k, E_k, \tilde{E}_k}}{P(a_{k-1}^{-1})},
\]

where \(P(a_{k-1}^{-1}) = ||\Psi_g^{k-1}||^2\), 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 \(\tilde{M}_{\gamma X}\) 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 \(0 \leq \hat{D}^{(k)}_{\gamma X} \leq \hat{1}\).

3. Upper bounds on \(|\langle \psi_{(\text{j}, \ldots, \text{j}_{k-\ell})}| \phi_{(\text{j}, \ldots, \text{j}_{k-\ell})} \rangle_{B_k, E_k'}|\)

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

\[
|\phi_{(\text{j}, \ldots, \text{j}_{k-\ell})} \rangle_{B_k, E_k'} = |\phi_{(\text{j}, \ldots, \text{j}_{k-\ell})} \rangle_{B_k, E_k} \sum_{j_{k+1}} \sqrt{P_{j_{k+1}}} \cdots \sum_{j_{k+\ell}} \sqrt{P_{j_{k+\ell}}} \langle \psi_{j_{k+1}, \ldots, j_{k+\ell}}|_{A_{\geq k+1}, B_{\geq k+1}, E_{\geq k+1}} \rangle_{B_{k+n}, E_{k+n}} \quad (\text{E16})
\]

and

\[
|\psi_{(g)}^{Z}_{(\text{f})} \rangle_{A_k, B_k, E_k'} = \frac{1}{\sqrt{2}} \sum_{\alpha \in \{0, 1\}} |\alpha Z\rangle_{A_k} |\phi_{\alpha Z} \rangle_{B_k, E_k'},
\]

which are analogous to \(|\tilde{U}^{(k)}| \tilde{U}^{(k)*} \rangle \tilde{m}_{\gamma \times} \tilde{U}^{(k)}| \Psi_g^{k-1} \rangle_{B_k, E_k, \tilde{E}_k}\) and \(|\psi_{(\text{j}, \ldots, \text{j}_{k-\ell})} \rangle_{A_k, B_k, E_k'}\), respectively. Note that, unlike \(|\phi_{(\text{j}, \ldots, \text{j}_{k-\ell})} \rangle_{B_k, E_k'}\) in Eq. (E8), \(|\phi_{(\text{j}, \ldots, \text{j}_{k-\ell})} \rangle_{B_k, E_k'}\) in Eq. (E16) is a qubit state,
i.e., \( \{|\phi_{j,k,g}\rangle_{A_k,B_k,E_k}\}_{j,k=0,1,x,0,1,\ldots} \) spans a qubit space, because \( |A_g\rangle_{A_{\geq k+1},B_{\geq k+1},E_{\geq k+1}} \) is a state independent of \( j_k \).

Using Eqs. (E6) and (E16), we can then calculate an upper bound on the inner product \( \langle \tilde{\psi}_{j,k,g}|j_{k-1},\ldots,j_{k-1}\rangle \phi_{j,k,g}\rangle_{B_k,E_k}^* \) such that

\[
\langle \tilde{\psi}_{j,k,g}|j_{k-1},\ldots,j_{k-1}\rangle \phi_{j,k,g}\rangle_{B_k,E_k}^* \geq \langle \tilde{\psi}_{j,k,g}|j_{k-1},\ldots,j_{k-1}\rangle \phi_{j,k,g}\rangle_{B_k,E_k}^* \geq \sqrt{1 - \epsilon_U}^{l+1} =: \sqrt{1 - \epsilon_U}; 
\]

where \( \epsilon_U := 1 - (1 - \epsilon_{U})^{l+1} \). In the second equality of Eq. (18), we have used the fact that \( \langle \tilde{\psi}_{j,k,g}|j_{k-1},\ldots,j_{k-1}\rangle \phi_{j,k,g}\rangle_{B_k,E_k}^* = 0 \) 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 \( \epsilon_{(k)}_{j_k,g}j_{k-1},\ldots,j_{k-1} \leq \epsilon_U \) for all \( k, j_k, g \), and \( j_{k-1},\ldots,j_{k-1} \) (see assumption (A2.b) in Section II A).

Similarly, using Eqs. (E10) and (E17), we can calculate an upper bound on the inner product \( \langle \tilde{\psi}_{j,k,g}|j_{k-1},\ldots,j_{k-1}\rangle \phi_{j,k,g}\rangle_{A_k,B_k,E_k}^* \) such that

\[
\langle \tilde{\psi}_{j,k,g}|j_{k-1},\ldots,j_{k-1}\rangle \phi_{j,k,g}\rangle_{A_k,B_k,E_k}^* \geq \langle \tilde{\psi}_{j,k,g}|j_{k-1},\ldots,j_{k-1}\rangle \phi_{j,k,g}\rangle_{A_k,B_k,E_k}^* \geq \sqrt{1 - \epsilon_U}^{l+1} =: \sqrt{1 - \epsilon_U}; 
\]

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 \( \langle \tilde{\psi}_{j,k,g}|j_{k-1},\ldots,j_{k-1}\rangle \phi_{j,k,g}\rangle_{B_k,E_k}^* \) and \( \langle \tilde{\psi}_{j,k,g}|j_{k-1},\ldots,j_{k-1}\rangle \phi_{j,k,g}\rangle_{A_k,B_k,E_k}^* \).

### Appendix F: Proof of Eq. (33)

Here, we provide a proof of the second inequality in Eq. (33). From Eq. (25), we can express \( \epsilon_{U}^{\parallel} \) as

\[
\epsilon_{U}^{\parallel} = \frac{N_{ph}^{U}}{N_{det}^{2(Z)}} = \frac{N}{N_{det}^{2(Z)}} f \left( \frac{X}{N} \right),
\]

where \( x \) 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. (32), we can express \( \epsilon_{U}^{\parallel, w} \) as

\[
\epsilon_{U}^{\parallel, w} = \frac{N_{w}^{U}}{N_{det}^{2(Z)}} = \frac{N_{w}}{N_{det}^{2(Z)}} f \left( \frac{X_{w}}{N_{w}} \right),
\]

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

\[
\sum_{w=0}^{l_c} q_{w} \epsilon_{U}^{\parallel, w} = \sum_{w=0}^{l_c} N_{det}^{2(Z)} N_{w} f \left( \frac{X_{w}}{N_{w}} \right) 
\]

Using Eqs. (F6) and (F16), we can then calculate the upper bound on the inner product \( \langle \tilde{\psi}_{j,k,g}|j_{k-1},\ldots,j_{k-1}\rangle \phi_{j,k,g}\rangle_{B_k,E_k}^* \) such that
\[
\begin{align*}
\leq \frac{N}{N_{\text{det}}} f \left( \sum_{w=0}^{L} N_{w} x_{w} \right) \\
= \frac{N}{N_{\text{det}}} f \left( \frac{X}{N} \right) = r_{\text{ph}},
\end{align*}
\]

where the inequality is due to the concavity of \( f \), and the second to last equality is due to \( \sum_{w=0}^{L} N_{w} x_{w} = x \), since \( \sum_{w=0}^{L} N_{j,\gamma X,w} = N_{j,\gamma X} \).

### Appendix G: Parameters \( c_{\alpha,j,g}^{(k)} \), \( c_{\alpha,j}^{(1)} \) and \( p_{\alpha X}^{(n)iU} \) for the modified BB84 protocol

In this appendix, we provide the full expressions of the coefficients \( c_{\alpha,j,g}^{(k)} \), with \( \alpha \in \{0,1\} \) 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 Section IV A. Direct calculations show that these coefficients are given by

\[
\begin{align*}
c_{1,0_Z,g}^{(k)} &= \sin \left( \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{0_Z,1}^{(k)}}{2} \right) - \sin \left( \frac{\theta_{0_Z,1}^{(k)}}{2} - \frac{\theta_{1_Z,1}^{(k)}}{2} \right), \\
c_{1,1_Z,g}^{(k)} &= \sin \left( \frac{\theta_{1_Z,0}^{(k)}}{2} - \frac{\theta_{1_Z,1}^{(k)}}{2} \right) + 2 \sin \left( \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{1_Z,1}^{(k)}}{2} \right) - \sin \left( \frac{\theta_{0_Z,1}^{(k)}}{2} - \frac{\theta_{1_Z,1}^{(k)}}{2} \right), \\
c_{1,1_X,g}^{(k)} &= \cos \left( \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{1_Z,1}^{(k)}}{2} \right) - 1 \left[ \cos \left( \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{1_Z,1}^{(k)}}{2} \right) - \cos \left( \frac{\theta_{0_Z,1}^{(k)}}{2} - \frac{\theta_{1_Z,1}^{(k)}}{2} \right) \right. \\
&\quad \left. - \cos \left( \frac{\theta_{0_Z,1}^{(k)}}{2} - \frac{\theta_{1_Z,1}^{(k)}}{2} \right) - \cos \left( \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{1_Z,1}^{(k)}}{2} \right) \right], \\
c_{0,0_Z,g}^{(k)} &= 2 \sin \left( \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{0_Z,1}^{(k)}}{2} \right) - \sin \left( \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{0_Z,1}^{(k)}}{2} + \frac{\theta_{0_X,1}^{(k)}}{2} \right) + \sin \left( \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{0_X,1}^{(k)}}{2} \right), \\
c_{0,1_Z,g}^{(k)} &= \sin \left( \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{0_Z,1}^{(k)}}{2} \right) + \sin \left( \frac{\theta_{0_Z,1}^{(k)}}{2} - \frac{\theta_{0_X,1}^{(k)}}{2} \right), \\
c_{0,1_X,g}^{(k)} &= \cos \left( \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{1_Z,1}^{(k)}}{2} \right) - 1 \left[ \cos \left( \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{1_Z,1}^{(k)}}{2} \right) - \cos \left( \frac{\theta_{0_Z,1}^{(k)}}{2} - \frac{\theta_{0_X,1}^{(k)}}{2} \right) \right. \\
&\quad \left. - \cos \left( \frac{\theta_{0_Z,1}^{(k)}}{2} - \frac{\theta_{0_X,1}^{(k)}}{2} \right) - \cos \left( \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{0_X,1}^{(k)}}{2} \right) \right] = \cos \left( \frac{\theta_{0_Z,0}^{(k)}}{2} + \frac{\theta_{1_Z,1}^{(k)}}{2} - \frac{\theta_{0_Z,1}^{(k)}}{2} - \frac{\theta_{0_X,1}^{(k)}}{2} + \frac{\theta_{0_Z,1}^{(k)}}{2} - \frac{\theta_{1_Z,1}^{(k)}}{2} + \frac{\theta_{0_Z,0}^{(k)}}{2} - \frac{\theta_{0_X,1}^{(k)}}{2} \right). \\
\end{align*}
\]
iments [16].

To derive the analytical upper bounds, we consider the partial differential equations of $c_{\alpha_{\beta_1},g}^{(k)}$ with respect to each $\theta_{\beta_1}^{(k)}$ and then select the values of $\theta_{\beta_1}^{(k)} \in [\theta_{\beta_1}^L, \theta_{\beta_1}^U]$ that maximise them. For example, since $\partial_{\theta_{0z}^{(k)}g}c_{1,0z,g}^{(k)} > 0$ when $\theta_{0z}^{(k)}g \in [-\pi/6, \pi/6]$, $\theta_{1z}^{(k)}g \in [5\pi/6, 7\pi/6]$ and $\theta_{0x}^{(k)}g \in [\pi/3, 2\pi/3]$, $c_{1,0z,g}^{(k)}$ is maximised when $\theta_{0z}^{(k)}g = \theta_{0z}^L$. 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. $\partial_{\theta_{0z}^{(k)}g} c_{1,1x,g}^{(k)}$ is maximised either when $\theta_{0z}^{(k)}g = \theta_{0z}^L$ or when $\theta_{0z}^{(k)}g = \theta_{0z}^U$, and we select the appropriate value by inspection. The overall solution is the following:

\[
c_{1,0z}^U \rightarrow c_{1,0z,g}^{(k)}(\theta_{0z}^U, \theta_{1z}^L, \theta_{1x}^L), \quad c_{1,1z}^U \rightarrow c_{1,1z,g}^{(k)}(\theta_{0z}^L, \theta_{1z}^L, \theta_{1x}^L), \quad c_{1,1x}^U \rightarrow \max_{\theta_1^0, \theta_1^L, \theta_1^L \in [0,1,0]} c_{1,1x,g}^{(k)}(\theta_{1z}^0, \theta_{1z}^L, \theta_{1x}^L), \quad c_{0,0z}^U \rightarrow c_{0,0z,g}^{(k)}(\theta_{0z}^U, \theta_{1z}^L, \theta_{1x}^L), \quad c_{0,1z}^U \rightarrow c_{0,1z,g}^{(k)}(\theta_{0z}^L, \theta_{1z}^L, \theta_{1x}^L), \quad c_{0,0x}^U \rightarrow \max_{\theta_0^0, \theta_0^L, \theta_0^L \in [0,1,0]} c_{0,0x,g}^{(k)}(\theta_{0z}^0, \theta_{1z}^L, \theta_{1x}^L). \quad (G2)
\]

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

\[
p_{\alpha_{X},g}^{(k,vir)} = \frac{1}{2} p_{Z_A} \left[ 1 + (-1)\alpha \Re \left( \langle \phi_{0z,g} | \phi_{1z,g} \rangle B_z, E_z \right) \right]. \quad (G3)
\]

Then, by using the definition of $| \phi_{\beta_1,g} \rangle B_z, E_z$ and Eq. (34), and by considering the partial differential equations of $p_{\alpha_{X},g}^{(k,vir)}$ with respect to each phase $\theta_{\beta_1}^{(k)}$, we have that

\[
p_{1X}^{(vir)U} \rightarrow \frac{1}{2} p_{Z_A} \left[ 1 - \cos \left( \frac{\theta_{0z}^L - \theta_{1z}^L}{2} \right) \right], \quad p_{0X}^{(vir)U} \rightarrow \frac{1}{2} p_{Z_A} \left[ 1 + \cos \left( \frac{\theta_{0z}^U - \theta_{1z}^L}{2} \right) \right]. \quad (G4)
\]

Appendix H: Parameters $c_{\alpha_{\beta_1},g}^{(k)}$, $c_{\alpha_{\beta_1},g}^{(k)}$, and $p_{\alpha_{X},g}^{(k,vir)}$ for the three-state protocol

Here, we provide the full expressions of the coefficients $c_{\alpha_{\beta_1},g}^{(k)}$ with $j \in \{0z, 1z, 0x\}$ associated with the three-state protocol for the particular phase-encoding scheme considered in Section IV A, as well as their upper bounds.

Direct calculations show that $c_{1,1,g}^{(k)}$ are given by

\[
c_{1,0z,g}^{(k)} = \cos \left( \frac{\theta_{0z}^L - \theta_{1z}^L}{2} \right) - 1 + 2 \cos \left( \frac{\theta_{0z}^L - \theta_{1z}^L}{2} \right) + \cos \left( \frac{\theta_{0z}^L - \theta_{1z}^L}{2} \right) + \cos \left( \frac{\theta_{0z}^L - \theta_{1z}^L}{2} \right) + \cos \left( \frac{\theta_{0z}^L - \theta_{1z}^L}{2} \right) + \cos \left( \frac{\theta_{0z}^L - \theta_{1z}^L}{2} \right) + 1.
\]

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

\[
c_{1,0z}^U \rightarrow c_{1,0z,g}^{(k)}(\theta_{0z}^L, \theta_{1z}^L, \theta_{0x}^L), \quad c_{1,1g}^U \rightarrow c_{1,1g}^{(k)}(\theta_{0z}^U, \theta_{1z}^L, \theta_{0x}^L),
\]
The full expressions for the coefficients $c_{i,j,g}^{(k)}$ with $j \in \{0, 1, 2\}$ as well as their upper bounds, and the probabilities $p_{i,j,g}^{(vir)}U$ are the same as in Appendix G.
A. Duplinskiy and D. Sych, Bounding passive light-source
A. Mizutani, G. Kato, K. Azuma, M. Curty, J. E. Bourassa, A. Gnanapandithan, L. Qian, and H.-K. Yoshino, M. Fujiwara, K. Nakata, T. Sumiya, M. Peev, C. Pacher, R. Alléaume, C. Barreiro, J. Bouda, M. Pereira, G. Kato, A. Mizutani, M. Curty, and K. Tamaki, Quantum key distribution with setting-choice-independently correlated light sources, npj Quantum Information 5, 8 (2019).
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 Information 5, 8 (2019).
K. Yoshino, M. Fujiwara, N. 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 Information 4, 8 (2018).
M. Pereira, G. Kato, A. Mizutani, M. Curty, and K. Tamaki, Quantum key distribution with correlated sources, Science Advances 6 (2020).
F. Grünfenfelder, A. Boaron, D. Rusca, A. Martin, and H. Zbinden, Performance and security of 5 GHz repetition rate polarization-based quantum key distribution, Applied Physics Letters 117 (2020).
V. Zapatero, Á. Navarrete, K. Tamaki, and M. Curty, Security of quantum key distribution with intensity correlations, Quantum 5, 602 (2021).
X. Sixto, V. Zapatero, and M. Curty, Security of decoy-state quantum key distribution with correlated intensity fluctuations, arXiv:2206.06700 (2022).
Á. Navarrete and M. Curty, Improved finite-key security analysis of quantum key distribution against Trojan-horse attacks, Quantum Sci. Technol. 7, 035021 (2022).
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).
K. Yoshino, T. Ochi, M. Fujiwara, M. Sasaki, and A. Tajima, Maintenance-free operation of WDM quantum key distribution system through a fiber field over 30 days, Opt. Express 21, 31395 (2013).
Z. Yuan, A. Plews, R. Takahashi, K. Doi, W. Tam, A. Sharpe, A. Dixon, E. Lavelle, J. Dynes, A. Murakami, M. Kujiraoka, M. Lucamarini, Y. Tanizawa, H. Sato, and A. J. Shields, 10-Mb/s quantum key distribution, J. Lightwave Technol. 36, 3427 (2018).
C. Elliott, A. Colvin, D. Pearson, O. Pikalo, J. Schlafer, and H. Yeh, Current status of the DARPA quantum network, in Quantum Information and Computation III, Vol. 5815, edited by E. J. Donkor, A. R. Pirich, and H. E. Brandt, International Society for Optics and Photonics (SPIE, 2005) pp. 138 – 149.
M. Peev, C. Pacher, R. Alléaume, C. Barreiro, J. Bouda, W. Boxleitner, T. Debuisschert, E. Diamanti, M. Dianati, J. F. Dynes, S. Fasel, S. Fossier, M. Fürst, J.-D. Gautier, O. Gay, N. Gisin, P. Grangier, A. Happe, Y. Hasani, M. Hentschel, H. Hübel, G. Humer, T. Länger, M. Legré, R. Lieger, J. Lodewyck, T. Lorinser, N. Lütkenhaus, A. Marhold, T. Matyus, O. Maurhart, L. Monat, S. Nauerth, J.-B. Page, A. Poppe, E. Querasser, G. Ribordy, S. Robyr, L. Salvail, A. W. Sharpe, A. J. Shields, D. Stucki, M. Suda, C. Tamas, T. Themel, R. T. Thew, Y. Thoma, A. Treiber, P. Trinkler, R. Tualle-Brouri, F. Vannel, N. Walenta, H. Weier, H. Weinfurter, I. Wimberger, Z. L. Yuan, H. Zbinden, and A. Zeilinger, The SECOQC quantum key distribution network in Vienna, New Journal of Physics 11, 056001 (2009).
T.-Y. Chen, J. Wang, H. Liang, W.-Y. Liu, Y. Liu, X. Jiang, Y. Wang, X. Wan, W.-Q. Cai, L. Ju, L.-K. Chen, L.-J. Wang, Y. Gao, K. Chen, C.-Z. Peng, Z.-B. Chen, and J.-W. Pan, Metropolitan all-pass and intercity quantum communication network, Opt. Express 18, 27217 (2010).
D. Stucki, M. Legré, F. Buntschu, B. Clausen, N. Felber, N. Gisin, L. Henzen, P. Junod, G. Litzister, P. Monbaron, L. Monat, J. B. Page, D. Perroud, G. Ribordy, A. Rochas, S. Robyr, J. Tavares, R. Thew, P. Trinkler, S. Ventura, R. Voirol, N. Walenta, and H. Zbinden, Long-term performance of the SwissQuantum quantum key distribution network in a field environment, New Journal of Physics 13, 123001 (2011).
M. Sasaki, M. Fujiwara, H. Ishizuka, W. Kraus, K. Wakui, M. Takeoka, S. Miki, T. Yamashita, Z. Wang, A. Tanaka, K. Yoshino, Y. Nambu, S. Takahashi, A. Tajima, A. Tomita, T. Domeki, T. Hasegawa, Y. Sakai, H. Kobayashi, T. Asai, K. Shimizu, T. Tokura, T. Tsurumaru, M. Matsui, T. Houjo, K. Tamaki, H. Takesue, Y. Tokura, J. F. Dynes, A. R. Dixon, A. W. Sharpe, Z. L. Yuan, A. J. Shields, S. Uchikoga, M. Legré, S. Robyr, P. Trinkler, L. Monat, J.-B. Page, G. Ribordy, A. Poppe, A. Allacher, O. Maurhart, T. Länger, M. Peev, and A. Zeilinger, Field test of quantum key distribution in the Tokyo QKD Network, Opt. Express 19, 10387 (2011).
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, Y. Tanizawa, J. P. Elbers, H. Greißer, I. H. White, R. V. Penty, and A. J. Shields, Cambridge quantum network, npj Quantum Information 5, 101 (2019).
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, S.-L. Han, Q. Yu, K. Liang, F. Zhou, X. Yuan, M.-S. Zhao, T.-Y. Wang, X. Jiang, L. Zhang, W.-Y. Liu, Y. Li, Q. Shen, Y. Cao, C.-Y. Lu, R. Shu, J.-Y. Wang, L. Li, N.-L. Liu, F. Xu, X.-B. Wang, C.-Z. Peng, and J.-W. Pan, An integrated space-to-ground quantum communication network over 4,600 kilometres, Nature 589, 214 (2021).
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, X.-W. Chen, L.-H. Sun, J.-J. Jia, J.-C. Wu, X.-J. Jiang, J.-F. Wang, Y.-M. Huang, Q. Wang, Y.-L. Zhou, L. Deng, T. Xi, L. Ma, T. Hu, Q. Zhang, Y.-A. Chen, N.-L. Liu, X.-B. Wang, Z.-C. Zhu, C.-Y. Lu, R. Shu, C.-Z. Peng, J.-Y. Wang, and J.-W. Pan, Satellite-to-ground quantum key distribution, Nature 549, 43 (2017).
L. Zhang, D. Rauch, M. Fink, J.-G. Ren, W.-Y. Liu, Y. Li, Q. Shen, Y. Cao, F.-Z. Li, J.-F. Wang, Y.-M. Huang, L. Deng, T. Xi, L. Ma, T. Hu, L. Li, N.-L. Liu, F. Koidl, P. Wang, Y.-A. Chen, X.-B. Wang, M. Steindorfer, G. Kirchner, C.-Y. Lu, R. Shu, R. Ursin, T. Scheidl, C.-Z. Peng, J.-Y. Wang, A. Zeilinger, and J.-W. Pan, 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 (2022).

[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] Even though a QKD source is not likely to spontaneously produce entangled pulses, in principle, a powerful Eve could inject entangled light into Alice’s device. Given the fragile nature of entangled states, one could expect this entanglement to almost completely vanish from the output light, especially if the source employs optical isolators. Nonetheless, one cannot be certain that all entanglement has been destroyed. In our analysis, we do not consider such attacks, although we believe that the security of QKD in the presence of these quantum correlations is an interesting problem to consider in future works.

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

[54] We believe that this assumption could be relaxed to the requirement that Alice’s setting choices on rounds \( \{k, \ldots , N\} \) do not affect Bob’s measurement result on round \( k - 1 \). This would allow us to take into account the maximum speed at which information about Alice’s \( k^{th} \) setting choice could travel to Bob’s measurement unit, given the distance separating the two, when determining the constraint on the maximum repetition rate of Alice’s source. However, here we make the stronger assumption for conceptual simplicity.

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

[56] 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, Nature Communications 2, 349 (2011).

[57] M. Koashi, Simple security proof of quantum key distribution via uncertainty principle, arXiv:quant-ph/0505108 (2005).

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

[59] 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).

[60] J. L. W. V. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, Acta Mathematica 30 (1906).

[61] H. Zhou, T. Sasaki, and M. Koashi, Efficient quantum key distribution scheme and a proof of its unconditional security, Journal of Cryptology 18, 133 (2005).

[62] 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 Information 5, 17 (2019).

[63] 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).

[64] P. J. Coles, E. M. Metodiev, and N. Lütkenhaus, Numerical approach for unstructured quantum key distribution, Nature Communications 7, 11712 (2016).

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

[66] 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, Berlin, Heidelberg, 1996) pp. 343–357.

[67] P. W. Shor and J. Preskill, Simple proof of security of the BB84 quantum key distribution protocol, Phys. Rev. Lett. 85 (2000).

[68] 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 (Association for Computing Machinery, 2000) pp. 715–724.

[69] 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 [57, 58] for more information).

[70] 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.