SECURITY OF QUANTUM KEY DISTRIBUTION USING
WEAK COHERENT STATES WITH NONRANDOM PHASES

HOI-KWONG LO

Center for Quantum Information and Quantum Control
Department of Electrical and Computer Engineering and Department of Physics
University of Toronto, Toronto, Canada M5G 3G4

JOHN PRESKILL

Institute for Quantum Information
California Institute of Technology, Pasadena, CA 91125, USA

We prove the security of the Bennett-Brassard (BB84) quantum key distribution protocol in the case where the key information is encoded in the relative phase of a coherent-state reference pulse and a weak coherent-state signal pulse, as in some practical implementations of the protocol. In contrast to previous work, our proof applies even if the eavesdropper knows the phase of the reference pulse, provided that this phase is not modulated by the source, and even if the reference pulse is bright. The proof also applies to the case where the key is encoded in the photon polarization of a weak coherent-state pulse with a known phase, but only if the phases of the four BB84 signal states are judiciously chosen. The achievable key generation rate scales quadratically with the transmission in the channel, just as for BB84 with phase-randomized weak coherent-state signals (when decoy states are not used). For the case where the phase of the reference pulse *is* strongly modulated by the source, we exhibit an explicit attack that allows the eavesdropper to learn every key bit in a parameter regime where a protocol using phase-randomized signals is provably secure.

Keywords: Quantum cryptography, quantum key distribution

Communicated by: S Braunstein & R Hughes

1 Introduction

In quantum key distribution (QKD) [1], two parties (Alice and Bob) use quantum signals to establish a shared key that can be used to encrypt and decrypt classical messages. An eavesdropper (Eve) who collects information about the key by interacting with the signals produces a detectable disturbance; therefore Alice and Bob can detect the eavesdropper's activity, and they can reject the key if they fear that the eavesdropper knows too much about it. But if the detected disturbance is weak enough, then Alice and Bob can use classical error correction and privacy amplification protocols to extract a shared key that is very nearly uniformly distributed and almost certainly private [2, 3, 4, 5, 6, 7, 8]. The security of the QKD protocol is said to be *unconditional*, because the security can be proven for any attack consistent with the laws of quantum physics, and without any assumptions about computational hardness.

Experiments have recently demonstrated the feasibility of QKD over 150 km telecom fibers [9, 10], and at least two firms are now marketing commercial QKD systems [11]. But how secure are these systems, really? To assess the security of practical implementations of QKD, it is important to consider how well the actual systems match the performance assumed in the security proofs. In particular, the signals used in typical practical realizations of QKD are dim laser pulses, which occasionally contain more than one photon. Multi-photon signals together with loss in the optical fiber can threaten security, but proofs of security for QKD using weak coherent states have been found [12, 13]. (We note that for QKD protocols that use decoy states [14, 15, 16], security can be proven even for rather strong coherent-state signals. In this paper, however, we will focus on QKD with weak coherent states.)

A key assumption in the security proofs in [12, 13] (and also in [15]) is that the phase of the quantum signal is uniformly random. A coherent state of one mode of the electromagnetic field can be expressed as

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle , \tag{1}$$

where $|n\rangle$ denotes the state with photon number n. We may write $\alpha = \sqrt{\nu} e^{i\theta}$, where $\nu = |\alpha|^2$ denotes the mean photon number and $e^{i\theta}$ is the *phase* of the coherent state. To an eavesdropper with no *a priori* knowledge of the phase, a signal whose phase is selected uniformly at random is indistinguishable from the state

$$\rho_\nu = \int_0^{2\pi} \frac{d\theta}{2\pi} |\sqrt{\nu} e^{i\theta}\rangle\langle\sqrt{\nu} e^{i\theta}| = e^{-\nu} \sum_{n=0}^{\infty} \frac{\nu^n}{n!} |n\rangle\langle n| , \tag{2}$$

a Poisson distributed mixture of photon number eigenstates. Therefore, for a security analysis, we may suppose that a source emitting weak coherent state signals is actually emitting signals in the state ρ_ν.

With probability $p_0 = e^{-\nu}$, which is close to one for small ν, the source emits no photon; exactly one photon is emitted with probability $p_1 = \nu e^{-\nu}$. The probability that two or more photons are emitted is

$$p_M = 1 - e^{-\nu} (1 + \nu)) \leq \frac{1}{2}\nu^2 . \tag{3}$$

Multi-photons can pose a security risk, but if each signal has a random phase, p_M is sufficiently small, and the loss in the channel is not too high, then it is possible to prove security of the Bennett-Brassard (BB84) protocol [1] against arbitrary eavesdropping attacks [12, 13].

However, no previously known security proof applies if the eavesdropper has some *a priori* knowledge about the phase of the signal states. Conceivably, such phase information might be accessible in realistic implementations of QKD. For example, in a "plug-and-play" scheme [17], a strong signal is sent from Bob to Alice, who attenuates and modulates the signal before returning it to Bob; in unidirectional schemes as well, strong ancillary pulses are sometimes used to monitor the channel. The phase of a strong pulse is accurately measurable in principle, and could be correlated with the phase of the key-generating pulse. Furthermore, the key information itself might be encoded in the relative phase of a bright reference pulse and a weak signal pulse, in which case the eavesdropper could plausibly learn the phase of the reference pulse. Even if strong pulses are not used, the phase coherence of a realistic source might be

maintained during the emission of many weak signals, allowing the phase to be determined accurately.

For all these reasons, it is worthwhile to investigate the security of QKD under the assumption that the eavesdropper knows something about the phase of the signals. Our main result in this paper is a proof of security for the BB84 protocol in the case where the key information is encoded in the relative phase of a reference pulse and a signal pulse. Our proof works even if the reference pulse is bright, with a phase known by the eavesdropper. It also applies if the key is encoded in the photon polarization of weak coherent states with nonrandom phases, provided the phases of the signal states are chosen appropriately. The proof is founded on the observation that, if the signal pulse is weak, Alice's source reveals relatively little information to Eve about the basis that Alice uses to encode her key bits. Privacy of the final key is demonstrated using an argument due to Koashi [18, 19] that invokes the uncertainty principle. We also point out that this argument establishes that the key distribution protocol is *universally composable* — the key can be used in any subsequent application without compromising security.

We describe our quasi-realistic model for sources and detectors in Sec. 2, and present the security analysis (which relies heavily on Koashi's ideas [18, 19]) in Sec. 3, Sec. 4, and Appendix A. In Sec. 5 we note that the key generation rate when the signals have nonrandom phases is comparable to the rate for phase randomized signals (when decoy states are not used). In Sec. 6 and Appendix B we point out that if the key is encoded in the photon polarization, and the (nonrandom) phases of the signals are not chosen judiciously, then BB84 is vulnerable to an *unambiguous key discrimination* attack, a measurement that, when conclusive, determines the key bit with certainty. Sec. 7 contains our conclusions.

2   Modeling sources and detectors

2.1   The source

In the ideal BB84 protocol [1], each signal is carried by a single qubit sent by Alice and received by Bob. The qubit encodes a key bit in one of two conjugate orthonormal bases, which we will call the x basis and the y basis. When Alice uses the x basis, her signal states are

$$
\begin{aligned}
|0_X\rangle &= \left(|0_Z\rangle + |1_Z\rangle\right)/\sqrt{2}\,, \\
|1_X\rangle &= \left(|0_Z\rangle - |1_Z\rangle\right)/\sqrt{2}\,,
\end{aligned}
\tag{4}
$$

the eigenstates with eigenvalues ± 1 of the Pauli operator

$$
X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\,;
\tag{5}
$$

when Alice uses the y basis, her signal states are

$$
\begin{aligned}
|0_Y\rangle &= \left(|0_Z\rangle + i|1_Z\rangle\right)/\sqrt{2}\,, \\
|1_Y\rangle &= \left(|0_Z\rangle - i|1_Z\rangle\right)/\sqrt{2}\,,
\end{aligned}
\tag{6}
$$

the eigenstates with eigenvalues ± 1 of the Pauli operator

$$
Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\,.
\tag{7}
$$

Alice's source emits one of these four states, chosen equiprobably. Bob measures the qubit that he receives in either the x basis or the y basis, chosen equiprobably, to determine his key bit. Later, through public discussion, Alice and Bob "sift" their key by retaining only the key bits for which Bob measured in the same basis that Alice used. The final key is extracted from the sifted key via a classical protocol that reconciles Alice's key with Bob's and amplifies the privacy of the key.

In practice, the key information is carried by dim laser pulses transmitted through an optical fiber. In the protocol we will analyze, Alice's source emits one of four states of a pair of photon modes (described by annihilation operators a_R and a_S):

$$
\begin{aligned}
|\tilde{0}_X\rangle &= e^{-|\alpha|^2} e^{\alpha\left(a_R^\dagger + a_S^\dagger\right)} |\mathrm{vac}\rangle_\mathrm{R} \otimes |\mathrm{vac}\rangle_\mathrm{S} = |\alpha\rangle_\mathrm{R} \otimes |\alpha\rangle_\mathrm{S} \ , \\
|\tilde{1}_X\rangle &= e^{-|\alpha|^2} e^{\alpha\left(a_R^\dagger - a_S^\dagger\right)} |\mathrm{vac}\rangle_\mathrm{R} \otimes |\mathrm{vac}\rangle_\mathrm{S} = |\alpha\rangle_\mathrm{R} \otimes |-\alpha\rangle_\mathrm{S} \ , \\
|\tilde{0}_Y\rangle &= e^{-|\alpha|^2} e^{\alpha\left(a_R^\dagger + ia_S^\dagger\right)} |\mathrm{vac}\rangle_\mathrm{R} \otimes |\mathrm{vac}\rangle_\mathrm{S} = |\alpha\rangle_\mathrm{R} \otimes |i\alpha\rangle_\mathrm{S} \ , \\
|\tilde{0}_Y\rangle &= e^{-|\alpha|^2} e^{\alpha\left(a_R^\dagger - ia_S^\dagger\right)} |\mathrm{vac}\rangle_\mathrm{R} \otimes |\mathrm{vac}\rangle_\mathrm{S} = |\alpha\rangle_\mathrm{R} \otimes |-i\alpha\rangle_\mathrm{S} \ .
\end{aligned}
\tag{8}
$$

Here the phase of α is defined relative to a fixed classical phase reference frame that Eve can access. We will refer to the state of mode S as the "signal" pulse and to the state of mode R as the "reference" pulse. The key is encoded in the relative phase of the two pulses, and it is actually quite important for our analysis that the phase of the reference pulse is the same in all four of the signal states — *only* the phase of the signal pulse is modulated by the source. (We will discuss in Sec. 6 and Appendix B how security can be compromised when the phase of the reference pulse is also modulated.) In some implementations, the two pulses are spatially separated in the optical fiber by a distance large compared to the pulse width; this scheme is called "phase encoding." In some implementations, the two modes represent the polarization states of the same spatial mode; this scheme is called polarization encoding. Phase encoding has been used in most fiber-based QKD experiments. Our security analysis will also apply to phase encoding in the case where the reference pulse is brighter than the signal pulse. We will return to this generalization in Sec. 4, but for now we will stick with the signals eq. (8), which could arise in either a phase-encoding or polarization-encoding scheme.

To summarize, the crucial features of our source model are: (1) the *single-mode signal pulse assumption* (the signal pulse is carried by a single bosonic mode, the same mode for all four of the signal states), and (2) the *unmodulated reference pulse assumption* (the state of the reference pulse is the same for all four signal states). These assumptions are reasonably well fulfilled by realistic sources used for the BB84 protocol with phase-encoded signals. Previous proofs of the security of BB84 have used (2′) the *phase-randomization assumption* (the phase of the reference pulse, but not of course the relative phase of signal and reference pulse that encodes the key information, is chosen uniformly at random and is not known by the eavesdropper). Our main new contribution in this paper is to prove security of BB84 using assumption (2) instead of assumption (2′).

2.2   The detectors

When Bob wants to measure the signal he receives in the x basis, he (in effect) combines the two modes in an interferometer, and directs modes $a_{0,X}$, $a_{1,X}$ to two different "threshold"

photon detectors, where

$$\begin{aligned} a_{0,X} &= (a_R + a_S)/\sqrt{2} \ , \\ a_{1,X} &= (a_R - a_S)/\sqrt{2} \ . \end{aligned} \tag{9}$$

Therefore, if Alice sends $|\tilde{0}_X\rangle$ and the ideal (unattenuated) signal enters Bob's measurement device (which is configured for x measurement), then detector 0 receives a coherent state with mean photon number

$$\mu = 2|\alpha|^2 \tag{10}$$

and detector 1 receives the vacuum state. Likewise, if Alice sends $|\tilde{1}_X\rangle$, then detector 0 receives the vacuum, and detector 1 receives the coherent state with mean photon number μ. When Bob wants to measure in the y basis, the modes directed to the photon detectors are

$$\begin{aligned} a_{0,Y} &= (a_R + ia_S)/\sqrt{2} \ , \\ a_{1,Y} &= (a_R - ia_S)/\sqrt{2} \ ; \end{aligned} \tag{11}$$

again, depending on the key bit, one mode receives mean photon number μ while the other receives vacuum.

We assume that each detector used by Bob to measure the signal he receives is a "threshold" photon detector. This means that the detector cannot distinguish a single photon from many photons. An *ideal* threshold detector "clicks" (registers a count) if it collects one or more photons, and does not click if it receives no photon; thus it performs a POVM with the two outcomes

$$\begin{aligned} E_{\text{no click}} &= |\text{vac}\rangle\langle\text{vac}| \ , \\ E_{\text{click}} &= I - |\text{vac}\rangle\langle\text{vac}| \ . \end{aligned} \tag{12}$$

But in this paper we consider a threshold detector that is not ideal: we allow the detector to have imperfect efficiency (it sometimes fails to click even when it receives one or more photons) and to record occasional "dark counts" (it sometimes clicks even when it receives the vacuum state).

Because Bob uses two threshold detectors, there are four possible outcomes when he measures a signal. If neither detector clicks, then the measurement is inconclusive and the signal is rejected. If detector $b \in \{0,1\}$ clicks and the other detector does not, then Bob records the key bit b. If *both* detectors click, then Bob records a key bit that he chooses uniformly at random. As noted in [12] and as we will explain below, it is important for the security analysis that double clicks are interpreted as key bits rather than regarded as inconclusive measurements. Actually, the security proof works the same way no matter how we map the double click events to key bits, but recording a random key bit is a particularly symmetric and natural choice. Given any fixed prescription for mapping double clicks to key bits, Bob's measurement realizes a POVM with three outcomes: 0, 1, and (when neither detector clicks) inconclusive.

Thus the protocol prescribes that Bob perform one of two possible POVMs, depending on whether Bob declares the x basis or the y basis, where each POVM has three outcomes. In fact, for our security proof we will not need to specify Bob's measurement in detail. All

we require is that the measurement satisfy one assumption: Bob's three-outcome POVM is equivalent to a (basis-dependent) two-outcome POVM that determines the key bit, preceded by a *basis-independent* "filter" that allows the signal to pass if and only if Bob's measurement conclusively determines the key bit.

Following Koashi [19], we note that this property, that the inconclusive measurement outcome can be attributed to a basis-independent filter, follows if we adopt a particular well-motivated model of dark counts and detector inefficiency, and if we suppose that both of Bob's detectors have the same efficiency. First we note that the inconclusive outcome can occur only if neither detector has a dark count. Let us model the dark counts as a random background process, where the rate of background counts is an intrinsic property of the detector, not dependent on the state of the mode that is being measured. That is, a background click occurs in detector 0 with probability d_0, whether mode $a_{0,X}$ or mode $a_{0,Y}$ is being measured, and in detector 1 with probability d_1, whether mode $a_{1,X}$ or mode $a_{1,Y}$ is being measured. Thus, each time a signal is received, the probability $(1-d_0)(1-d_1)$ that there is no background click in either detector does not depend on whether Bob declares the x basis or the y basis.

Furthermore, let us also assume [20] that when mode $a_{0,X}$ (or $a_{0,Y}$) contains n photons the probability that detector 0 fails to click is

$$\text{Prob}(n \text{ photons}, \text{no click}) = (1 - \xi)^n \ , \tag{13}$$

where ξ is the detector's efficiency, and similarly for detector 1. We may imagine that the detector induces decoherence in the photon number measurement before measuring — the real detector might not actually destroy the coherence of a superposition of states with two different nonzero photon numbers, but we are entitled to suppose that it does, because if so Eve's information and the measurement outcome would not be affected. Thus, each signal Bob receives can be regarded as a mixture of eigenstates of the total photon number n, where $m \leq n$ of the photons are directed to detector 0 and the remaining $n - m$ photons are directed to mode 1. But if both detectors have the same efficiency ξ, and if the probability of no click is given by eq. (13), then the probability that *neither* detector clicks depends only on n, not on m, and therefore it makes no difference whether Bob's detectors receive modes $a_{0,X}$ and $a_{1,X}$ or $a_{0,Y}$ and $a_{1,Y}$; in other words the probability of an inconclusive outcome does not depend on whether Bob measures in the x basis or the y basis.

We conclude, then, that when Bob receives an n photon signal, the probability that no click occurs in either detector, so that the measurement is inconclusive, is

$$\text{Prob}(n, \text{inconclusive}) = (1 - d_0)(1 - d_1)(1 - \xi)^n \ , \tag{14}$$

which does not depend on whether Bob declares the x basis or the y basis. Thus, we can imagine that a basis-independent filter is applied that measures the total photon number n and blocks the signal with probability $\text{Prob}(n, \text{inconclusive})$; if the signal passes the filter, Bob then performs a two-outcome POVM, either M_x or M_y depending on whether Bob declares the x basis or the y basis.

In order to describe Bob's measurement this way, we need a prescription for mapping the double click events to key bits. This prescription is needed because the probability of a double click event might depend on Bob's declared basis; therefore if the double clicks were

regarded as inconclusive we would not be able to attribute the inconclusive outcome to a basis-independent filter. The prescription is incorporated into the definition of the measurements M_x and M_y, and for the security proof, the details of this prescription do not matter.

We note that if Bob receives a coherent state with mean photon number μ, then the inconclusive outcome occurs with probability

$$
\begin{aligned}
\text{Prob}(\mu, \text{inconclusive}) &= e^{-\mu} \sum_{n=0}^{\infty} \frac{\mu^n}{n!} \text{Prob}(n, \text{inconclusive}) \\
&= (1 - d_0)(1 - d_1)e^{-\xi\mu} .
\end{aligned}
\tag{15}
$$

In particular, eq. (15) applies if Bob receives one of the signal states in eq. (8) and performs the measurement specified in eq. (9) or eq. (11), where $\mu = 2|\alpha|^2$. In that case, when Alice and Bob use the same basis, double clicks occur only because of dark counts, and when the measurement is conclusive and there are no dark counts Bob's key bit always agrees with Alice's. Attenuation in the channel reduces the probability of a conclusive outcome, but does not cause bit errors if the signals are otherwise unmodified and if Bob's measurement is ideal.

To summarize, the crucial feature of our detector model is the *basis-independent filter assumption*, which says that Bob's measurement can be modeled by a basis-independent filter that removes all the signals for which the measurement of the key bit is inconclusive, followed by a (basis-dependent) two-outcome POVM that always produces a conclusive result. This assumption is reasonably fulfilled by realistic measurement devices used in BB84 provided that (1) Bob uses two threshold detectors with the same efficiency, (2) the probability that the detector fails to click decays exponentially as a function of the number of photons received (as in the standard theory of photodetection), and (3) a prescription (either deterministic or probabilistic) is adopted for mapping double click events to key bits. We note that attacks have been proposed and analyzed [21, 22] that exploit a mismatch in the efficiency of Bob's two photon detectors, highlighting the importance of assumption (1).

The basis-independent filter assumption limits Eve's ability to enhance her information about the signals by taking advantage of detector inefficiency. But we emphasize that our model still incorporates quite general flaws in the detection system (including for example a reduction in visibility due to an imperfect alignment of Bob's interferometer), because the POVMs M_x and M_y are arbitrary. Security is not compromised even if the detectors are very noisy, because in that case the bit error rate will be high and the key will almost certainly be rejected.

3 Security analysis

3.1 The unbalanced quantum coin

We can analyze the security of this protocol by applying the methods of [13, 18], where security was proven for the case where Alice's source emits signals that, averaged over the key bits, have a small dependence on the basis. To apply this method, we must quantify the basis-dependence of the signal set eq. (8).

It is convenient to imagine that Alice launches each signal by performing a perfect measurement on a qubit that is entangled with the signal mode (the reference mode is always prepared in the state $|\alpha\rangle$ no matter which signal is being sent). When Alice wants to send

one of the two signals $\{|\tilde{0}_X\rangle, |\tilde{1}_X\rangle\}$, we choose the entangled state of her qubit and the signal mode to be

$$|\Psi_x\rangle = \left(|0_X\rangle \otimes |\alpha\rangle + |1_X\rangle \otimes |-\alpha\rangle\right)/\sqrt{2} \,, \tag{16}$$

and we instruct Alice to determine her key bit by measuring her qubit in the basis $\{|0_X\rangle, |1_X\rangle\}$. When Alice wants to send one of the two signals $\{|\tilde{0}_Y\rangle, |\tilde{1}_Y\rangle\}$, we choose the entangled state of her qubit and the mode to be

$$|\Psi_y\rangle = \left(|1_Y\rangle \otimes |i\alpha\rangle + |0_Y\rangle \otimes |-i\alpha\rangle\right)/\sqrt{2} \,, \tag{17}$$

and we instruct Alice to determine her key bit by measuring her qubit in the basis $\{|1_Y\rangle, |0_Y\rangle\}$. See Fig. 1.

Now we notice that the states $|\Psi_x\rangle$ and $|\Psi_y\rangle$ are hard to distinguish when α is small. Using

$$\langle i\alpha|\alpha\rangle = e^{-|\alpha|^2}e^{-i|\alpha|^2} = \langle -i\alpha|-\alpha\rangle \,,$$
$$\langle -i\alpha|\alpha\rangle = e^{-|\alpha|^2}e^{i|\alpha|^2} = \langle i\alpha|-\alpha\rangle \,, \tag{18}$$

we find

$$
\begin{aligned}
\langle \Psi_y|\Psi_x\rangle &= e^{-|\alpha|^2}\left(\cos|\alpha|^2 + \sin|\alpha|^2\right) \\
&= e^{-\mu/2}\left(\cos(\mu/2) + \sin(\mu/2)\right) \\
&= 1 - \mu^2/4 + O(\mu^3) \,.
\end{aligned}
\tag{19}
$$

Note that there is no term linear in $\mu \equiv 2|\alpha|^2$ – that is, if we ignore the multi-photon contribution, the two states are indistinguishable. This happens because, to linear order in α, the two-dimensional subspace spanned by the two x-basis signals and the two-dimensional subspace spanned by the y-basis signals coincide. The large overlap of $|\Psi_x\rangle$ and $|\Psi_y\rangle$ indicates that, averaged over the value of the key bit, the states emitted by the source when the x basis is chosen are hard to distinguish from the states emitted when the y basis is chosen. (We have chosen the purifications $|\Psi_x\rangle$ and $|\Psi_y\rangle$ that have the maximal overlap compatible with the key-averaged signal states [23]; that is why we have paired $|1_Y\rangle$ with $|i\alpha\rangle$ and $|0_Y\rangle$ with $|-i\alpha\rangle$ in eq. (17).)

A protocol such that the signals emitted by the source, averaged over the key bit, have a small dependence on Alice's basis choice can be analyzed by considering an equivalent protocol in which the basis-dependence of the signals can be related to the "balance" of a "quantum coin" [13, 18]. In this equivalent protocol, Alice measures the coin in the basis $\{|0_Z\rangle, |1_Z\rangle\}$ to determine whether her signal is encoded in the x basis or the y basis, and we may take the joint state of the coin and Alice's source states to be

$$|\Phi\rangle = \left(|0_Z\rangle \otimes |\Psi_x\rangle + |1_Z\rangle \otimes |\Psi_y\rangle\right)/\sqrt{2} \,; \tag{20}$$

furthermore we may imagine that the measurement of the coin is delayed until after Eve is finished interacting with the signals.

If $|\Psi_x\rangle$ and $|\Psi_y\rangle$ were equal, then the signals, averaged over the two possible values of the key bit, would be independent of whether Alice chose the x basis or the y basis. If $|\Psi_x\rangle$ and $|\Psi_y\rangle$ are nearly equal, then the source leaks a small amount of information about Alice's
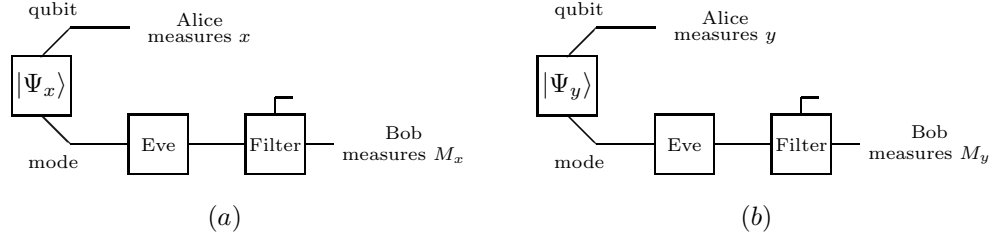
Fig. 1. Roles of Alice, Bob, and Eve in the key distribution protocol. Alice determines her key bit by performing an ideal measurement on a qubit that is entangled with the signal mode. (*a*) If Alice declares the *x* basis, then Alice measures in the *x* basis and the entangled state of her qubit and the mode is $|\Psi_x\rangle$. (*b*) If she declares the *y* basis, she measures in the *y* basis and the entangled state is $|\Psi_y\rangle$. Bob's detector applies a basis-independent filter that determines whether his measurement has a conclusive outcome, followed by a two-outcome POVM, either M_x or M_y depending on Bob's declared basis. If Alice and Bob both declare the *x* basis, the probability that their measurement outcomes disagree is δ_x, and if both declare the *y* basis, the probability that they disagree is δ_y.

basis choice that Eve might exploit. A useful way to quantify the leaked basis information is to consider what would happen if Alice were to measure each of her coins in the basis of X eigenstates rather than the basis of Z eigenstates; then the outcome $X = -1$ would occur with probability Δ, where

$$\langle\Psi_y|\Psi_x\rangle = 1 - 2\Delta \ . \tag{21}$$

Thus we say that the coin is "Δ-balanced," where Δ quantifies the basis dependence of Alice's signals. For the signal states eq. (8), we have

$$
\begin{aligned}
\Delta &= \frac{1}{2}\left(1 - e^{-\mu/2}\left(\cos(\mu/2) + \sin(\mu/2)\right)\right) \\
&= \mu^2/8 + O(\mu^3) \ .
\end{aligned}
\tag{22}
$$

However, not all of the signals emitted by the source are detected; Eve, by carefully choosing which signals to block, might be able to enhance the basis dependence of the detected signals. Let us pessimistically assume that the detected signals are chosen to maximize the imbalance of the coin. For a perfect channel and for perfect detection efficiency, the fraction of signals detected would be $1 - e^{-\mu} \approx \mu$, where μ is the mean photon number of the signals emitted by the source. The actual fraction of detected signals is $\eta\mu$, where η is an effective rate of loss due to absorption in the channel and imperfect detector performance. In the worst case, for all of the signals that are removed, the coin is an $X = 1$ eigenstate, which would enhance the probability of the outcome $X = -1$ by

$$\Delta \rightarrow \Delta' = \Delta/(\eta\mu) \approx \mu/(8\eta) \ . \tag{23}$$

Thus we will use Δ' to quantify the basis dependence in the signals that determine Alice's sifted key.

We will argue that for a fixed Δ' that is sufficiently small, secure key can be extracted from sifted key at a fixed nonzero asymptotic rate. Therefore, the number of private key bits

that can be generated per transmitted signal is proportional to $\eta\mu \approx 8\Delta'\eta^2$ and hence scales like η^2 — the same scaling as for a source with random phases [12, 13] (assuming that decoy states are not used).

3.2   The uncertainty principle and privacy amplification

Koashi [18], generalizing the techniques in [13], explained how to relate the imbalance of the quantum coin to the privacy of a key that is generated by performing ideal measurements on qubits; here we wish to apply his reasoning to Alice's key bits, which we regard as determined by outcomes of measurements of (fictitious) qubits entangled with the signal states. For our argument, it is important that the imbalance of the coin arises solely from the basis dependence of Alice's signals. That is why we required that Bob's measurement be modeled by a basis-independent filter followed by a two-outcome measurement; if the probability of an inconclusive outcome actually depended on the basis, then the basis-dependence of the filter might have further enhanced the imbalance of the coins. As it is, after the inconclusive measurement results are discarded, we may imagine that Bob performs his two-outcome POVM on signals that are entangled with Alice's qubits and the quantum coins, where the imbalance Δ' of the coins is determined only by the basis dependence of the source and the loss in the channel.

We recall that the BB84 protocol includes a verification test conducted by Alice and Bob. In the test, Alice and Bob publicly compare a randomly selected subset of their key bits to estimate a bit error rate, the fraction of the key bits for which Alice and Bob disagree. Signals sent in the x basis and the y basis can be tested separately, to estimate both a bit error rate δ_x for signals sent in the x basis, and a bit error rate δ_y for signals sent in the y basis. Let us consider the key bits that are generated by measuring in the x basis. Using Koashi's method [18], we may show that, in the asymptotic limit of a very long key, private key can be extracted from Alice's sifted key at a rate

$$R = 1 - H(\delta_x) - H(\delta'_y) \ , \tag{24}$$

where δ_x is the bit error rate, $\delta'_y \geq \delta_y$ is a function of δ_y and Δ', and $H(\delta) = -\delta \log_2 \delta - (1-\delta)\log_2(1-\delta)$ is the binary Shannon entropy. A similar formula applies for the key bits generated by measuring in the y basis, but with δ_x and δ_y interchanged. The term $H(\delta_x)$ is the rate at which sifted key must be sacrificed to perform error correction that reconciles Bob's key with Alice's (according to standard Shannon theory), and $H(\delta'_y)$ is the cost of performing privacy amplification to ensure that Eve has negligible information about Alice's final key.

What is the quantity δ'_y? Recall that when the x basis is declared, Alice measures her qubit in the x basis and Bob performs measurement M_x, while if the y basis is declared, Alice measures her qubit in the y basis and Bob performs measurement M_y. Furthermore, when the x basis is declared, the entangled state of Alice's qubit and the signal mode is $|\Psi_x\rangle$, while when the y basis is declared, the entangled state is $|\Psi_y\rangle$. The quantity δ'_y is an upper bound on what the bit error rate would have been if Alice measured in the y basis and Bob performed measurement M_y, but where the entangled state of the qubit and mode is $|\Psi_x\rangle$ instead of $|\Psi_y\rangle$. This hypothetical error rate is not known directly from the test, but because the state $|\Psi_x\rangle$ is close to $|\Psi_y\rangle$ when Δ is small, the hypothetical "phase error rate" δ'_y is close to the actual observed bit error rate δ_y. See Fig. 2.
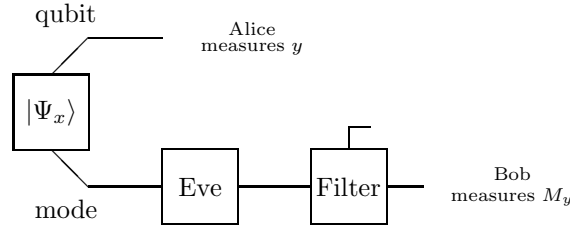
Fig. 2. Setting that defines the "phase error rate" δ_y'. The entangled state of Alice's qubit and the signal mode is $|\Psi_x\rangle$ as though Alice had declared the x basis, but Alice and Bob both measure in the "wrong" basis — Alice measures y instead of x, and Bob measures M_y instead of M_x. The probability that Bob's outcome disagrees with Alice's is δ_y'. The quantity δ_y' cannot be directly measured in the protocol, but an upper bound can be inferred from the observed bit error rate δ_y and the known basis dependence of Alice's signal states.

The estimate of δ_y' is explained in Appendix 1. The result is that for any $\varepsilon > 0$, with high probability we may express δ_y' as

$$\begin{aligned} \delta_y' &= \delta_y + 4\Delta'(1-\Delta')(1-2\delta_y) + 4(1-2\Delta')\sqrt{\Delta'(1-\Delta')\delta_y(1-\delta_y)} + \varepsilon \\ &\leq \delta_y + 4\Delta' + 4\sqrt{\Delta'\delta_y} + \varepsilon \ . \end{aligned} \tag{25}$$

The upper bound on δ_y' in the last line is obtained by retaining only the terms of lowest order in δ_y and Δ', and gives a reasonable approximation to the tighter bound on the previous line. From the tighter bound and eq. (24), we see that the key generation rate R remains positive, in the limit of negligible bit error rates δ_x, δ_y, for $\Delta' < 0.146$, which corresponds to μ smaller than about $(1.16)\eta$ when η is small.

Why does δ_y' characterize the cost of amplifying privacy? Following Koashi [18] (and generalizing an idea used in [2] and [6]) we observe that when the x basis is declared, Eve's knowledge about Alice's key would not have been affected if Bob had chosen to measure in the wrong basis (to measure M_y rather than M_x). Furthermore, neither the key nor Eve's knowledge of the key would have been affected if Alice had chosen to delay measuring her qubits until after Bob measured the signals that he received. Therefore, we may imagine that Alice generated her key by measuring in the x basis a string of qubits that, conditioned on the outcome of Bob's M_y measurement, are close to y-basis eigenstates (if δ_y' is small).

If, after Bob's measurement, each of Alice's qubits were *exactly* a y-basis eigenstate, then Eve would be powerless to predict the key bits that Alice obtains by measuring the qubits in the x basis. In that case we could say that the privacy of the key is founded on the uncertainty principle: a qubit that produces a deterministic outcome when measured in the y basis will produce a uniformly random outcome when measured in the x basis. From this perspective, we may say that the purpose of the privacy amplification is to ensure that Alice's final key is very nearly equivalent to a key determined by measuring y-basis eigenstates in the x basis.

Let the n-component binary vector v denote Alice's sifted key for the signals in which the x basis was declared by Alice and Bob, and Bob reported successfully detecting the signal. Alice extracts her k-bit final key $\kappa = vG$ by applying the random rank-k binary $n \times k$ matrix G to the sifted key. Equivalently, we may say that Alice's final key is obtained by measuring

the k binary observables

$$\bar{X}_i = \bigotimes_{j:G_{ji}=1} X_j \;, \quad i = 1, 2, 3, \ldots k \;, \tag{26}$$

where X_j denotes the Pauli matrix X acting on Alice's jth qubit.

Before selecting the matrix G, we might have first selected a random rank-$(n-k)$ binary $n \times (n-k)$ matrix H, and then chosen G subject to the constraint $G^T H = 0$ (so that each column of G is orthogonal in the binary inner product to each column of H). In that case, each of the $n-k$ binary observables

$$\bar{Y}_m = \bigotimes_{j:H_{jm}=1} Y_j \;, \quad m = 1, 2, 3, \ldots n - k \tag{27}$$

(where Y_j denotes the Pauli matrix Y acting on Alice's jth qubit) would commute with each of the $\bar{X}_i$'s. We may therefore imagine (since it would have no effect on Alice's final key or on Eve's information about the key) that Alice measured each of the $\bar{Y}_m$'s before measuring the $\bar{X}_i$'s to determine her final key bits.

Now, if Alice were to measure all of her qubits in the y basis, then with probability exponentially close to 1 for n large, her outcome would be one of $N_{\text{typ}} = 2^{n(H(\delta'_y)+\varepsilon)}$ typical n-bit strings. Note that if u and v are two distinct n-bit strings, and s is a randomly chosen n-bit string, then the bits us^T and vs^T are distinct with probability $1/2$. Therefore if we choose the number of randomly chosen binary observables in the set $\{\bar{Y}_m\}$ (the number of columns in the matrix H) to be

$$n - k = n\big(H(\delta'_y) + 2\varepsilon\big) \;, \tag{28}$$

then the probability that more than one typical binary string is compatible with the outcomes of the $\{\bar{Y}_m\}$ measurements is at most $N_{\text{typ}} \cdot 2^{k-n} = 2^{-\varepsilon n}$. This means that, conditioned on the results of measuring all the observables in the set $\{\bar{Y}_m\}$, Alice's n qubits have been projected to a state that is exponentially close to a product of y-basis eigenstates. Thus Alice's final key is essentially the same as it would have been if the sifted key had been generated by measuring y-basis eigenstates in the x basis, and therefore is guaranteed to be private. Since $k = n\big(1 - H(\delta'_y) - 2\varepsilon\big)$, and $n\big(H(\delta_x) + \varepsilon'\big)$ key bits are sacrificed to reconcile Bob's sifted key with Alice's, we obtain the asymptotic key rate eq. (24).

3.3   Universal composability

This argument shows not just that Eve's mutual information with the final key is negligible, but also that a stronger definition of security is satisfied: the protocol is *universally composable* [24, 25, 26]. This means that the final key bits can be safely used in any subsequent application without compromising security. To demonstrate composability, we consider in more depth (following [26]) the joint state shared by Alice and Eve before Alice performs her x-basis measurements to generate her key. We have already seen that, after Alice's fictitious measurement of the observables $\{\bar{Y}_m\}$, her density operator ρ_A has high fidelity with the pure state $|y_0\rangle$, where $|y_0\rangle$ denotes a particular product of y-basis eigenstates:

$$F(\rho_A, |y_0\rangle\langle y_0|) \equiv \langle y_0|\rho_A|y_0\rangle = 1 - \zeta^2 \;, \tag{29}$$

where ζ is exponentially small. Therefore we may introduce an auxiliary system E, which we pessimistically assume to be under the eavesdropper's control, such that ρ_A has a purification $|\Upsilon\rangle_{AE}$ where $|\Upsilon\rangle_{AE}$ has a large overlap with $|y_0\rangle_A \otimes |\lambda\rangle_E$ and $|\lambda\rangle_E$ is a pure state of E:

$$F\big((|\Upsilon\rangle\langle\Upsilon|)_{AE}, (|y_0\rangle\langle y_0|)_A \otimes (|\lambda\rangle\langle\lambda|)_E\big) = 1 - \zeta^2 . \tag{30}$$

We can relate the large overlap to proximity in the trace norm, using the inequality

$$\frac{1}{2}\|\rho - \sigma\|_{\mathrm{tr}} \leq \sqrt{1 - F(\rho,\sigma)} , \tag{31}$$

obtaining

$$\frac{1}{2}\| (|\Upsilon\rangle\langle\Upsilon|)_{AE} - (|y_0\rangle\langle y_0|)_A \otimes (|\lambda\rangle\langle\lambda|)_E \|_{\mathrm{tr}} \leq \zeta . \tag{32}$$

Now when Alice measures the observables $\{\bar{X}_i\}$ to determine her final key, the state $|y_0\rangle\langle y_0|$ yields a uniformly random key described by the maximally mixed density operator

$$\rho_A^{\mathrm{uni}} = \frac{1}{2^k} \sum_{\kappa=0}^{2^k-1} |\kappa\rangle\langle\kappa| , \tag{33}$$

and the resulting state shared by Alice and Eve is a product state $\rho_A^{\mathrm{uni}} \otimes \sigma_E$, where $\sigma_E = (|\lambda\rangle\langle\lambda|)_E$. Quantum operations cannot increase the trace distance; therefore, if ρ_{AE} is the state shared by Alice and Eve after Alice's measurement, we find that

$$\frac{1}{2}\|\rho_{AE} - \rho_A^{\mathrm{uni}} \otimes \sigma_E\|_{\mathrm{tr}} \leq \zeta . \tag{34}$$

Eq. (34) is a security criterion shown to be universally composable in [25], which we have now seen is satisfied by a protocol where the signal states have nonrandom phases. This demonstration of universal composability for the BB84 quantum key distribution protocol also applies to the other cases addressed by Koashi in [18, 19].

4   Bright reference pulse

Our security proof also applies if the reference pulse is bright. Suppose that, for each of the BB84 signals, the state of the reference pulse emitted by the source is $|\beta\rangle_R$, while the emitted state of the signal pulse is as in eq. (8), where $|\beta| \geq |\alpha|$. Then we may calculate the imbalance of the quantum coin as before, except in eq. (22) we replace $\mu/2$ by $\mu_S = |\alpha|^2$, the mean photon number of the signal pulse. When Bob receives the signals, he first attenuates the reference pulse so that its strength matches the strength of the signal pulse, and then performs the interferometric measurement described earlier. Thus we replace $\eta\mu$ in eq. (23) by $2\eta\mu_S$, the mean total photon number of the two modes after the attenuation of the reference pulse (and taking loss into account). Therefore, in both formulas we replace μ by $2\mu_S$, finding

$$\Delta' \approx \mu_S/(4\eta) , \tag{35}$$

when the signal pulse is weak. It is obviously essential for this argument that the (possibly bright) reference pulse has exactly the same state (or at least very nearly the same state) for each of the BB84 signals, so that it conveys no information (or very little information) to Eve

concerning which signal is emitted. Eve might tamper with Bob's measurement of the signal by attacking the reference pulse, but because she knows little about the basis in which the key bit is encoded, she is still incapable of collecting much information about the key without creating a detectable disturbance.

Why is it that Eve seems to be no better off when the reference pulse is bright than when it is as dim as the signal pulse? The reason is that even when the reference pulse is dim we have pessimistically assumed that Eve knows the phase of the reference pulse perfectly. When we say that Eve "knows" that the state of the reference pulse is the coherent state $|\beta\rangle_R$, we are taking it for granted that the phase of β has a fixed value relative to a classical phase reference that Eve controls. If the phase of β were not already known, then increasing $|\beta|$ would make it easier for Eve to determine the phase relative to her reference frame, but if the phase is already known then increasing $|\beta|$ does not give Eve any additional advantage.

The observation that the reference pulse can be bright without compromising security is relevant to, for example, the "double Mach-Zehnder" scheme for implementing phase encoding [27, 28]. In this set-up, Alice uses an unbalanced interferometer to split a single pulse into two spatially separated pulses that travel to Bob through the same optical fiber; one of these pulses, whose phase is unmodulated, is the reference pulse, and the other, whose phase is modulated, is the signal pulse. Bob uses another unbalanced interferometer to combine the two pulses, with a modulated relative phase that determines his measurement basis.

For a given strength of the signal pulse, using a brighter reference pulse increases the key rate by reducing the probability that Bob fails to detect the signal. In particular, if the reference pulse is much brighter than the signal pulse, then nearly twice as many signals will be detected compared to the case where the reference and signal pulses have equal strength. Our proof shows that the key rate can be doubled by using a bright reference pulse, without reducing the privacy of the final key.

We can understand this doubling of the key rate by considering how Bob's detection system operates in the double Mach-Zehnder scheme. Let us suppose that the reference pulse follows the longer path in Alice's unbalanced interferometer and therefore lags behind the signal pulse during transmission through the fiber; we therefore use a_L (L for "long") to denote the reference mode, and a_S (S for "short" as well as "signal") to denote the signal mode. The two paths in Bob's interferometer also have unequal lengths, to compensate for the spatial separation in the fiber of the reference and signal pulses: the portion of the reference pulse that follows the short path in Bob's interferometer (let us call it the LS mode, since it follows the long path in Alice's interferometer and the short path in Bob's) interferes with the portion of the signal pulse that follows the long path in Bob's interferometer (the SL mode). The first beam splitter in Bob's interferometer is asymmetric, so that the reference and signal pulses are split according to

$$|\tilde{\beta}\rangle_L \mapsto |c\tilde{\beta}\rangle_{LL} \otimes |s\tilde{\beta}\rangle_{LS} \ ,$$
$$|\tilde{\alpha}\rangle_S \mapsto |c\tilde{\alpha}\rangle_{SL} \otimes |s\tilde{\alpha}\rangle_{SS} \ , \tag{36}$$

where $c^2 + s^2 = 1$. Here LL refers to the part of the reference pulse that enters the long arm of Bob's interferometer, and SS refers to the part of the signal pulse that enters the short arm; these LL and SS modes are ignored by Bob. We denote the strengths of the reference and signal pulses by $\tilde{\beta}$ and $\tilde{\alpha}$ rather than β and α to take into account the attenuation of the

pulses during transmission through the fiber; thus, if Eve does not intervene, $|\tilde{\alpha}|/|\tilde{\beta}| = |\alpha|/|\beta|$.

Bob's first beam splitter is chosen so that $s|\tilde{\beta}| = c|\tilde{\alpha}|$, or $s/c = |\alpha|/|\beta|$; therefore the incoming LS and SL modes at Bob's second beam splitter have equal strength. This second beam splitter is symmetric, and Bob reads out his device by detecting the photons in the two output ports of this symmetric beam splitter. The probability of detection is proportional to the total mean photon number in the LS and SL modes as they arrive at Bob's second beam splitter, which is

$$s^2|\tilde{\beta}|^2 + c^2|\tilde{\alpha}|^2 = 2c^2|\tilde{\alpha}|^2 \ , \tag{37}$$

where

$$c^2 = \frac{1}{1 + s^2/c^2} = \frac{|\tilde{\beta}|^2}{|\tilde{\alpha}|^2 + |\tilde{\beta}|^2} = \frac{|\beta|^2}{|\alpha|^2 + |\beta|^2} \ . \tag{38}$$

The factor c^2 suppressing the detection rate arises because the strength of the signal pulse following the long path in Bob's interferometer is attenuated by the factor c at the first beam splitter; the rest of the signal pulse follows the short path and is wasted. This factor c^2 is $1/2$ for $|\alpha| = |\beta|$ but close to 1 for $|\beta|^2 \gg |\alpha|^2$. Thus when the reference pulse and signal pulse have the same strength half of the signal is wasted, but when the reference pulse is bright hardly any of the signal is wasted.

5   Comparison to phase-randomized signals

It is instructive to compare our estimate of the key generation rate with the rate that would be achievable if the phases of the signals were random. In that case, the quantum coin is undisturbed if the source emits a single photon, but it is strongly affected if two or more photons are emitted. Thus the fraction of the coins that are damaged is at worse $\Delta' = p_M/p_D$, where p_M is the probability for emission of a multi-photon, and p_D is the fraction of the signals that are detected. Therefore, for small μ we have [13]

$$\Delta' = p_M/p_D \approx \frac{\frac{1}{2}\mu^2}{\eta\mu} = \mu/2\eta \ . \tag{39}$$

The key generation rate again has the form eq. (24), but for coins in this particular type of state (a fraction Δ' badly damaged and a fraction $1 - \Delta'$ undisturbed), the estimate of δ'_y can be improved to

$$\delta'_y - \delta_y = \Delta'/2 \approx \mu/4\eta \ . \tag{40}$$

(Eq. (40) is an improvement found in [29] of the result $\delta'_y - \delta_y = \Delta'$ found in [13].) For comparison, by combining eq. (23) with eq. (25) we find

$$\delta'_y - \delta_y \approx \mu/2\eta + \sqrt{2\mu\delta_y/\eta} \tag{41}$$

for the case where Eve knows the phase of the signals; thus we see that phase randomization improves the rate, but only by a relatively modest amount when the bit error rate is small.

In fact, as shown in [13], for a phase-randomized source a higher key generation rate can be achieved than implied by eq. (24) and eq. (40), furthermore as shown in [19] this higher rate applies for the detector model assumed in Sec. 2. The achievable number G of key bits *per pulse* can be expressed as

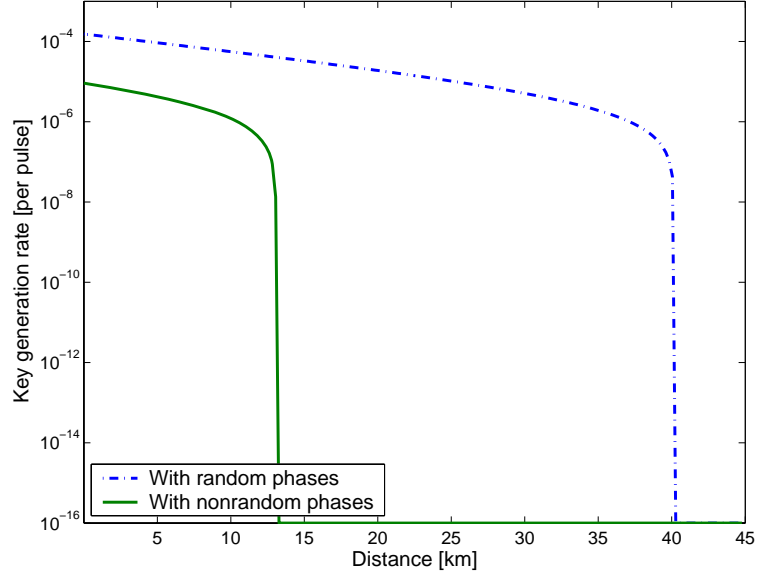$$G = \frac{1}{2}\Big(Q_1\big(1 - H(e_1)\big) - Qf(e)H(e)\Big) \ . \tag{42}$$

Fig. 3. The key generation rate in bits per pulse as a function of distance for sources with random and nonrandom phases, using experimental parameters from [10], and assuming an error-correction inefficiency $f(e) = 1.22$. Here for the phase-randomized case we assume that decoy states are not used. For each value of the distance, the signal strength has been chosen to optimize the rate.

Here Q is the fraction of pulses for which Bob detects the signal, e is the bit error rate observed in the verification test, and $f(e) \geq 1$ parametrizes the inefficiency of the error correction used to perform key reconciliation; furthermore e_1 is the error rate for the single-photon signals emitted by the source and Q_1 is the fraction of the single-photon signals that result in a detection event. (The factor of $\frac{1}{2}$ in front of the expression for G arises because half of the detected signals, those for which Alice and Bob used different bases, are discarded during sifting.) In a protocol that uses decoy states, Q_1 and e_1 can be estimated directly, but if decoy states are not used then we may pessimistically assume all the bit errors are due to single-photon signals, so that

$$e_1 = e/(1 - \Delta') \tag{43}$$

(with $\Delta' = p_M/Q$), and that all the multi-photon signals are detected, so that

$$Q_1 = Q(1 - \Delta') . \tag{44}$$

For the case of a source with nonrandom phases, the achievable key rate per pulse is

$$G = \frac{1}{2}Q\big(1 - f(e)H(e) - H(e_{\mathrm{ph}})\big) ; \tag{45}$$

here e_{ph} as a function of e is given as in eq. (25) by

$$e_{\mathrm{ph}} = e + 4\Delta'(1 - \Delta')(1 - 2e) + 4(1 - 2\Delta')\sqrt{\Delta'(1 - \Delta')e(1 - e)} , \tag{46}$$

where

$$\Delta' = \Delta/Q = \frac{1}{2Q}\left(1 - e^{-\mu/2}\left(\cos(\mu/2) + \sin(\mu/2)\right)\right) . \tag{47}$$

We have plotted the key rates per pulse given by eq. (42) and eq. (45) as a function of distance in Fig. 3. For this comparison, we have used experimental parameters (channel loss, dark count rate, and detector efficiency) from [10], and assumed an error-correction inefficiency $f(e) = 1.22$; for each value of the distance, the signal strength μ has been chosen to optimize the rate. The maximal distance for which secure key exchange is possible is about three times longer in the phase-randomized case.

6   Phase modulated reference pulse

The security proof works because the overlap of $|\Psi_x\rangle$ and $|\Psi_y\rangle$ is $1 - O(\mu_S^2)$. This property holds because, if we assume that the phase of the reference pulse is unmodulated, then the four BB84 signal states span a two-dimensional space when the multi-photons in the signal pulse are neglected. But, even when the reference pulse is weak, the story can change if the source modulates the phase of the reference pulse as well as the signal pulse. In [30] we studied such a source, and we concluded that the eavesdropper can exploit the modulation of the reference pulse to improve the effectiveness of her attack.

In particular, in Appendix B (see also [30]), we study the BB84 signal states

$$
\begin{aligned}
|\tilde{0}_X\rangle &= |\alpha\rangle_R \otimes |\alpha\rangle_S \ , \\
|\tilde{1}_X\rangle &= |-i\alpha\rangle_R \otimes |i\alpha\rangle_S \ , \\
|\tilde{0}_Y\rangle &= |e^{-i\pi/4}\alpha\rangle_R \otimes |e^{i\pi/4}\alpha\rangle_S \ , \\
|\tilde{1}_Y\rangle &= |e^{i\pi/4}\alpha\rangle_R \otimes |e^{-i\pi/4}\alpha\rangle_S \ ,
\end{aligned}
\tag{48}
$$

in which the reference pulse for each signal has a distinct phase relative to Eve's classical phase reference. Thus key information is encoded not just in the relative phase of the signal and reference pulses, but also in the phase of the (dim) reference pulse relative to Eve's reference frame. The signal set eq. (48) seems more natural if we recognize that, when re-expressed in a different basis, it becomes the signal set eq. (B.1); thus it arises in a polarization encoding scheme where the single-photon component of each signal is the standard BB84 single-qubit signal, and where the relative phase of the vacuum and single-photon component is the same for all four signals.

We note that, even if we ignore the (small) multi-photon component of the weak coherent state signals, these four BB84 signals belong to a three-dimensional Hilbert space spanned by the vacuum (no-photon) state and two distinguishable one-photon states. On the other hand, the two signals that convey a particular key bit value (0 or 1) span a two-dimensional space. Therefore, an eavesdropper who knows the relative phase of the vacuum and single photon component can launch an attack on the protocol that we call *unambiguous key discrimination*. She performs a POVM with three outcomes: 0, 1, and inconclusive; if either of the conclusive outcomes occurs, then Eve knows with certainty the key bit (0 or 1) encoded in the BB84 signal emitted by the source, though she does not gain any information about which of the two possible BB84 states compatible with that key bit was emitted.

For these signal states, we show in Appendix B that Eve's key discrimination has a conclusive outcome with probability $p_C \approx .146\mu$ when the mean photon number μ is small. If the signals had random phases, then only the multi-photon signals would be insecure, a fraction $O(\mu^2)$ of all the signals. But when the phase of α is known and the signals eq. (48) are used, this fraction increases to $O(\mu)$.

As van Enk [31] has astutely pointed out to us, the unambiguous key discrimination attack becomes much less effective if the encoding of the signals is chosen judiciously. And for the signal set eq. (8) in which the reference pulse has the same phase in all four signals, the signals span a two-dimensional space when the multi-photons are neglected, so that the conclusive key discrimination succeeds with probability $p_C = O(\mu^2)$. (For polarization encoding, this corresponds to particular choices for the phases of the coherent states and for the orientation of the BB84 polarization states in the Bloch sphere.) Thus as we have seen, unconditional security is provable for this particular encoding, but the proof does not apply to the signal set eq. (48). Therefore, the protocol is vulnerable to a "Trojan horse" attack, in which a malicious manufacturer in cahoots with Eve provides Alice with a source that modulates the phase of the reference pulse [32].

There are other possible Trojan horse attacks that could be even more devastating — for example the source might encode the four BB84 states in four distinct bosonic modes so that Eve can distinguish the signals perfectly. It is useful, however to differentiate such multi-mode Trojan horse attacks from the more subtle attacks where the protocol is compromised because the source modulates the phase of a single mode. The single-mode Trojan horse might be harder for Alice and Bob to detect, and might be less susceptible to countermeasures [33] such as filtering out the unwanted modes.

7   Conclusion

In summary, we have shown that the BB84 quantum key distribution protocol with phase-encoded signals is secure if the signal pulse is weak. Our proof applies even if the reference pulse is bright and has a phase known by the adversary, provided that the phase of the reference pulse is the same for all four of the BB84 signals, and the measurement device satisfies the basis-independent filter assumption formulated in Sec. 2. Furthermore, our proof shows that the BB84 protocol is universally composable under these conditions. Our security proof also applies to polarization encoding, if the signal states are chosen appropriately. And we have discussed a new type of *single-mode* Trojan horse attack, in which security can be compromised because the source modulates the phase of the reference pulse depending on which signal is being emitted.

The achievable key rate established by our proof scales quadratically with the transmission η in the channel, as for BB84 with phase-randomized weak coherent-state signals. In the phase-randomized case, a key rate linear in η can be achieved with the decoy state method. An interesting open question is whether employing decoy states can also extend the range over which secure BB84 quantum key distribution is possible in the case where the eavesdropper knows the phase of the (possibly bright) reference pulse and the detector is modeled as in Sec. 2. We note that a key rate linear in η has been established for the B92 [27] protocol assuming that the reference pulse is bright and the detector has suitable properties (one proof requires Bob to have a local oscillator phase-locked with the reference pulse [34], another requires that Bob's detector can distinguish single-photon signals from multi-photons [35]), but these proofs do not apply to the threshold detectors used in current experiments. It may also be fruitful to investigate the impact on security of using imperfect sources and detectors for other quantum key distribution protocols, such as the six-state protocol [36] and SARG04 [37].

Finally we remark that, since a higher key generation rate has been established for the case where the reference pulse has a random phase than for the case where the phase is known, it may be advantageous to deliberately randomize the phases of the signals in order to reduce the eavesdropper's power. An experimental demonstration of quantum key distribution using active phase randomization was recently reported in [38].

Acknowledgments

References

1. C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," in *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing* (IEEE, New York, 1984), pp. 175-179.
2. D. Mayers, "Quantum key distribution and string oblivious transfer in noisy channels," in *Advances in Cryptology–Proc. Crypto '96*, Lecture Notes in Computer Science vol. 1109 (Springer-Verlag, New York, 1996), pp. 343-357; D. Mayers, "Unconditional security in quantum cryptography," J. of ACM 48, 351 (2001), arXiv:quant-ph/9802025.
3. H.-K. Lo and H. F. Chau, "Unconditional security of quantum key distribution over arbitrarily long distances," Science 283, 2050–2056 (1999), arXiv:quant-ph/9803006.
4. E. Biham, M. Boyer, P. O. Boykin, T. Mor, and V. Roychowdhury, "A proof of security of quantum key distribution," in *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing (STOC'00)* (ACM Press, New York, 2000), pp. 715-724, arXiv:quant-ph/9912053.
5. P. W. Shor and J. Preskill, "Simple proof of security of the BB84 quantum key distribution protocol", Phys. Rev. Lett. 85, 441, (2000), arXiv:quant-ph/0003004.
6. M. Koashi and J. Preskill, "Secure quantum key distribution with an uncharacterized source," Phys. Rev. Lett. 90, 057902 (2003), arXiv:quant-ph/0208155.
7. M. Christandl, R. Renner, and A. Ekert, "A Generic Security Proof for Quantum Key Distribution," arXiv:quant-ph/0402131 (2004).
8. R. Renner, N. Gisin, and B. Kraus, "Lower and upper bounds on the secret key rate for QKD protocols using one–way classical communication," Phys. Rev. Lett. 95, 080501 (2005), arXiv:quant-ph/0410215; "An information-theoretic security proof for QKD protocols," Phys. Rev. A 72, 012332 (2005), arXiv:quant-ph/0502064.
9. Kimura, T. *et al.*, "Single-photon interference over 150-km transmission using silica-based integrated-optic interferometers for quantum cryptography," arXiv:quant-ph/0403104 (2004).
10. C. Gobby, Z. L. Yuan, and A. J. Shields, "Quantum key distribution over 122 km of standard

telecom fiber," Applied Phys. Lett. 84, 3762-3764, (2004), arXiv:quant-ph/0412171.

11. MagiQ Technologies, Inc. (website: `http://www.magiqtech.com/`) and id Quantique (website: `http://www.idquantique.com/`).

12. H. Inamori, N. Lütkenhaus,and D. Mayers, "Unconditional security of practical quantum key distribution," arXiv:quant-ph/0107017 (2001).

13. D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, "Security of quantum key distribution with imperfect devices," Quantum Inf. Comp. 4, 325-360 (2004), arXiv:quant-ph/0212066.

14. W.-Y. Hwang, "Quantum key distribution with high loss: toward global secure communication," Phys. Rev. Lett. 91, 057901 (2003), arXiv:quant-ph/0211153.

15. H.-K. Lo, X. Ma and K. Chen "Decoy state quantum key distribution," Phys. Rev. Lett. 94, 230504 (2005), arXiv:quant-ph/0411004 (2004); X. Ma, "Security of quantum key distribution with realistic devices," arXiv:quant-ph/0503057 (2005).

16. X.-B. Wang, "Beating the PNS attack in practical quantum cryptography," Phys. Rev. Lett. 94, 230503 (2005), arXiv:quant-ph/0410075; "A decoy-state protocol for quantum cryptography with 4 intensities of coherent states," Phys. Rev. A 72, 012322 (2005), arXiv:quant-ph/0411047.

17. A. Muller, T. Herzog, B. Huttner, W. Tittel, H. Zbinden, and N. Gisin, "Plug and play systems for quantum cryptography," Appl. Phys. Lett. 70, 793 (1997), arXiv:quant-ph/9611042.

18. M. Koashi, "Simple security proof of quantum key distribution via uncertainty principle," arXiv:quant-ph/05005108 (2005).

19. M. Koashi, "Efficient quantum key distribution with practical sources and detectors," arXiv:quant-ph/0609180 (2006).

20. D. F. Walls and G. J. Milburn, *Quantum Optics* (Springer, Berlin, 1994).

21. V. Makarov, A. Anisimov, and J. Skaar, "Effects of detector efficiency mismatch on security of quantum cryptosystems," Phys. Rev. A 74, 022313 (2006), arXiv:quant-ph/0511032.

22. B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma, "Time-shift attack in practical quantum cryptosystems," arXiv:quant-ph/0512080 (2005).

23. P. Raynal and N. Lütkenhaus "Optimal unambiguous state discrimination of two density matrices and its link with the fidelity," arXiv:quant-ph/0502165 (2005).

24. M. Ben-Or, M. Horodecki, D. W. Leung, D. Mayers, and J. Oppenheim, "The universal composable security of quantum key distribution," Lecture Notes in Computer Science, vol. 3378, pp. 386-406 (2005), arXiv:quant-ph/0409078.

25. R. Renner and R. König, "Universally composable privacy amplification against quantum adversaries," Proc. of TCC 2005, Lecture Notes in Computer Science, vol. 3378, pp. 407-425 (2005), arXiv:quant-ph/0403133.

26. R. König, R. Renner, A. Bariska, and U. Maurer, "Locking of accessible information and implications for the security of quantum cryptography," arXiv:quant-ph/0512021 (2005).

27. C. H. Bennett, "Quantum cryptography using any two nonorthogonal states" Phys. Rev. Lett. 68, 3121-3124 (1992).

28. N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," Rev. Mod. Phys. 74, 145195 (2002), arXiv:quant-ph/0101098.

29. J. C. Boileau, private communication.

30. H.-K. Lo and J. Preskill, "Phase randomization improves the security of quantum key distribution," arXiv:quant-ph/0504209 (2005).

31. S. van Enk, private communication (2005).

32. N. Lütkenhaus, private communication (2005).

33. N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, "Trojan horse attacks on quantum key distribution systems," arXiv:quant-ph/0507063 (2005).

34. M. Koashi, "Unconditional security of coherent-state quantum key distribution with strong phase-reference pulse," Phys. Rev. Lett 93, 120501 (2004), arXiv:quant-ph/0403131.

35. K. Tamaki, N. Lütkenhaus, M. Koashi, and J. Batuwantudawe, Unconditional security of the Bennett 1992 quantum key-distribution scheme with strong reference pulse, arXiv:quant-ph/0607082 (2006).

36. H.-K. Lo, "Proof of unconditional security of six-state quantum key distribution scheme," Quantum Inf. Comp. 1, 81-94 (2001), arXiv:quant-ph/0102138.
37. V. Scarani, A. Acin, G. Ribordy, N. Gisin, "Quantum cryptography protocols robust against photon number splitting attacks for weak laser pulses implementations," Phys. Rev. Lett. 92, 057901 (2004), arXiv:quant-ph/0211131.
38. Y. Zhao, B. Qi, and H.-K. Lo, Experimental quantum key distribution with active phase randomization, arXiv:quant-ph/0611059 (2006).
39. K. Tamaki, M. Koashi, and N. Imoto, "Unconditionally secure key distribution based on two nonorthogonal states" Phys. Rev. Lett. 90, 167904 (2003), arXiv:quant-ph/0212162.
40. R. Koenig and R. Renner, "A de Finetti representation for finite symmetric quantum states," J. Math. Phys. 46, 122108 (2005), arXiv:quant-ph/0410229.
41. M. Christandl, R. Koenig, G. Mitchison, and R. Renner, "One-and-a-half quantum de Finetti theorems," arXiv:quant-ph/0602130 (2006).
42. M. Dusek, M. Jahma, and N. Lütkenhaus, "Unambiguous state discrimination in quantum cryptography with weak coherent states," Phys. Rev. A 62, 022306 (2000); arXiv:quant-ph/9910106.
43. D. Gottesman, and H.-K. Lo, "Proof of security of quantum key distribution with two-way classical communications," IEEE Trans. Inf. Theory 49, 457 (2003), arXiv:quant-ph/0105121.
44. J. Sheriff, Masters Report, University of Toronto, September 2005.

Appendix A Bounding the phase error rate

The argument in Sec. 3 invokes the uncertainty principle to establish that private key can be extracted from sifted key at the asymptotic rate R shown in eq. (24); this rate depends on an observed bit error rate δ_x and an unobserved "phase error rate" δ_y' for the signals sent in the x basis. In this Appendix, we derive eq. (25), which expresses δ_y' in terms of the observed y-basis bit error rate δ_y and the imbalance Δ' of the quantum coin that determines the basis choice. The derivation is nearly identical to an argument used by Koashi in [18]; we include it in this paper so that our security analysis will be self-contained.

The derivation uses the "Bloch sphere bound" proved in [39]. Note that for an arbitrary state of a single qubit, expectation values of the Pauli operators X and Z obey the inequality

$$\langle X \rangle^2 + \langle Z \rangle^2 \leq 1 ; \tag{A.1}$$

the polarization vector in the qubit lies within the Bloch sphere of unit radius. The Bloch sphere bound asserts that a similar inequality applies with high probability if $\langle X \rangle$ and $\langle Z \rangle$ are estimated by randomly sampling from a correlated state of many qubits. That is, consider an arbitrary state of $2n$ qubits, and randomly select a subset of n qubits. Measure in the x basis each of the n qubits in the subset, and let $n\gamma_x$ denote the number of those qubits for which the outcome of the measurement is $X = -1$. Then measure the rest of the qubits in the z basis, and let $n\gamma_z$ denote the number of qubits for which the outcome is $Z = -1$. Then for any $\varepsilon > 0$ and with probability exponentially (in n) close to 1,

$$(1 - 2\gamma_x)^2 + (1 - 2\gamma_z)^2 \leq 1 + \varepsilon . \tag{A.2}$$

Thus, despite the possible correlations among the qubits, the results are constrained by the naive Bloch sphere bound in the limit of large n. It is convenient to rewrite this inequality (suppressing the ε) as

$$1 - 2\gamma_x \leq 2\sqrt{\gamma_z(1 - \gamma_z)} . \tag{A.3}$$

We could apply the Bloch sphere bound to the quantum coin; the coin is measured in the z basis to determine whether Alice will encode her key bit in the x basis or the y basis. Thus γ_z is the probability that Alice declares the y basis, $1 - \gamma_z$ is the probability that Alice declares the x basis, and the "imbalance" of the coin is $\Delta' = \gamma_x$.

However, our goal is to relate the bit error rate δ_y that is observed when the y basis is declared to the bit error rate δ'_y that would have been observed when the x basis is declared, if both Alice and Bob had measured in the "wrong" basis (the y basis rather than the x basis). Therefore, we will imagine that Alice and Bob always measure in the y basis, and we will divide the signals into two sets — the set for which Alice and Bob find key bits that agree (the set "n" for "no error") and the set for which Alice and Bob disagree (the set "e" for "error").

Let X_{coin} denote the Pauli operator X acting on the coin, let Z_{coin} denote the Pauli operator Z acting on the coin, and define

$$
\begin{aligned}
\gamma_x^{(e)} &= \mathrm{Prob}(X_{\mathrm{coin}} = -1 \mid Y \text{ error}) , \\
\gamma_z^{(e)} &= \mathrm{Prob}(Z_{\mathrm{coin}} = -1 \mid Y \text{ error}) , \\
\gamma_x^{(n)} &= \mathrm{Prob}(X_{\mathrm{coin}} = -1 \mid \text{no } Y \text{ error}) , \\
\gamma_z^{(n)} &= \mathrm{Prob}(Z_{\mathrm{coin}} = -1 \mid \text{no } Y \text{ error}) .
\end{aligned}
\tag{A.4}
$$

Applying the Bloch sphere bound to the signals with a y-basis error we obtain

$$
1 - 2\gamma_x^{(e)} \le 2\sqrt{\gamma_z^{(e)}(1 - \gamma_z^{(e)})} ,
\tag{A.5}
$$

and applying it to the signals with no y-basis error we obtain

$$
1 - 2\gamma_x^{(n)} \le 2\sqrt{\gamma_z^{(n)}(1 - \gamma_z^{(n)})} .
\tag{A.6}
$$

Using Bayes's rule, we have

$$
\begin{aligned}
&\mathrm{Prob}(Z_{\mathrm{coin}} = -1, Y \text{ error}) \\
&= \mathrm{Prob}(Z_{\mathrm{coin}} = -1) \cdot \mathrm{Prob}(Y \text{ error} \mid Z_{\mathrm{coin}} = -1) \\
&= \gamma_z \delta_y \\
&= \mathrm{Prob}(Z_{\mathrm{coin}} = -1 \mid Y \text{ error}) \cdot \mathrm{Prob}(Y \text{ error}) \\
&= \gamma_z^{(e)} \cdot \mathrm{Prob}(Y \text{ error}) ,
\end{aligned}
\tag{A.7}
$$

and likewise

$$
\begin{aligned}
&\mathrm{Prob}(Z_{\mathrm{coin}} = 1, Y \text{ error}) \\
&= \mathrm{Prob}(Z_{\mathrm{coin}} = 1) \cdot \mathrm{Prob}(Y \text{ error} \mid Z_{\mathrm{coin}} = 1) \\
&= (1 - \gamma_z)\delta'_y \\
&= \mathrm{Prob}(Z_{\mathrm{coin}} = 1 \mid Y \text{ error}) \cdot \mathrm{Prob}(Y \text{ error}) \\
&= (1 - \gamma_z^{(e)}) \cdot \mathrm{Prob}(Y \text{ error}) .
\end{aligned}
\tag{A.8}
$$

Furthermore,

$$
\begin{aligned}
& \text{Prob}(X_{\text{coin}} = -1, Y \text{ error}) \\
& = \text{Prob}(X_{\text{coin}} = -1 \mid Y \text{ error}) \cdot \text{Prob}(Y \text{ error}) \\
& = \gamma_x^{(e)} \cdot \text{Prob}(Y \text{ error}) \ ;
\end{aligned}
\tag{A.9}
$$

therefore, multiplying both sides of eq. (A.5) by $\text{Prob}(Y \text{ error})$, we find

$$
\begin{aligned}
& \text{Prob}(Y \text{ error}) - 2\,\text{Prob}(X_{\text{coin}} = -1, Y \text{ error}) \\
& \leq 2\sqrt{\gamma_z(1 - \gamma_z)\delta_y \delta_y'} \ .
\end{aligned}
\tag{A.10}
$$

By applying the same reasoning to the signals with no y-basis error, we find from eq. (A.6) that

$$
\begin{aligned}
& \text{Prob}(\text{no } Y \text{ error}) - 2\,\text{Prob}(X_{\text{coin}} = -1, \text{ no } Y \text{ error}) \\
& \leq 2\sqrt{\gamma_z(1 - \gamma_z)(1 - \delta_y)(1 - \delta_y')} \ .
\end{aligned}
\tag{A.11}
$$

Now adding together eq. (A.10) and eq. (A.11), we have

$$
\begin{aligned}
1 - 2\Delta' &= 1 - 2\,\text{Prob}(X_{\text{coin}} = -1) \\
&\leq\ 2\sqrt{\gamma_z(1 - \gamma_z)}\left(\sqrt{\delta_y \delta_y'} + \sqrt{(1 - \delta_y)(1 - \delta_y')}\right) \\
&\leq\ \sqrt{\delta_y \delta_y'} + \sqrt{(1 - \delta_y)(1 - \delta_y')} \ .
\end{aligned}
\tag{A.12}
$$

This inequality says that δ_y' must be close to δ_y if Δ' is small. After some algebra (and after reinstating ε), we may re-express eq. (A.12) in the form eq. (25).

In our analysis, we have allowed a basis asymmetry in the detected signals ($\gamma_z \neq 1/2$). This asymmetry might occur because the source leaks a small amount of information about the basis choice, and Eve could use that basis information to (say) block more x-basis signals than y-basis signals. However, using the available basis information to enhance the basis asymmetry is not a good strategy for Eve. Rather, her attack is most effective when $\delta_y' - \delta_y$ is largest, and eq. (A.12) indicates that the maximal possible value of $\delta_y' - \delta_y$ occurs when x-basis and y-basis signals are detected with equal likelihood ($\gamma_z = 1/2$).

In the end, the inequality eq. (A.12) has a simple form and in fact it admits a simple interpretation. Recall that if ρ_x and ρ_y are two density operators and the operators $\{E_a\}$ are the elements of a POVM, then there is an inequality relating the fidelity of the two density operators to the statistical overlap of the measurement outcomes:

$$
\sqrt{F(\rho_x, \rho_y)} \equiv \|\sqrt{\rho_x}\sqrt{\rho_y}\|_{\text{tr}} \leq \sum_a \sqrt{\text{tr}(\rho_x E_a)\text{tr}(\rho_y E_a)} \ .
\tag{A.13}
$$

Ignore for the moment the correlations among the signal states that might be induced by Eve's attack, and suppose that, for each signal, the joint state of the qubit measured by Alice and of the bosonic modes measured by Bob is ρ_x when Alice declares the x basis and ρ_y when Alice declares the y basis. Recall that for the detector model formulated in Sec. 2, the probability that Bob's measurement yields a conclusive outcome does not depend on

whether Bob measures in the x basis or the y basis; let us denote this probability that Bob's measurement is conclusive as Q_x in the state ρ_x and Q_y in the state ρ_y. Let $\{E_a\}$ be the three-outcome POVM that outputs "inconclusive" when Bob's measurement is inconclusive, outputs "agree" when Bob's y measurement is conclusive and agrees with Alice's y measurement, and outputs "disagree" when Bob's y measurement is conclusive and disagrees with Alice's y measurement. Then eq. (A.13) becomes

$$\sqrt{F(\rho_x, \rho_y)} \;\leq\; \sqrt{(1 - Q_x)(1 - Q_y)} + \sqrt{Q_x Q_y} \left(\sqrt{\delta_y \delta_y'} + \sqrt{(1 - \delta_y)(1 - \delta_y')} \right) \;. \quad \text{(A.14)}$$

For a fixed value of $Q_x + Q_y$, the right-hand side of eq. (A.14) is maximized for $Q_x = Q_y$; therefore, defining $Q = (Q_x + Q_y)/2$, we find

$$\sqrt{F(\rho_x, \rho_y)} \;\leq\; 1 - Q + Q \left(\sqrt{\delta_y \delta_y'} + \sqrt{(1 - \delta_y)(1 - \delta_y')} \right) \;. \quad \text{(A.15)}$$

But Eve's attack cannot improve the distinguishability of ρ_x and ρ_y; thus

$$1 - 2\Delta = |\langle \Psi_y | \Psi_x \rangle| \leq \sqrt{F(\rho_x, \rho_y)} \;, \quad \text{(A.16)}$$

and so we obtain

$$1 - 2\Delta' \leq \sqrt{\delta_y \delta_y'} + \sqrt{(1 - \delta_y)(1 - \delta_y')} \quad \text{(A.17)}$$

where $\Delta' = \Delta/Q$, in agreement with eq. (A.12).

This fidelity argument establishes security against individual attacks, but we have not seen how to turn it into a rigorous proof of security against collective attacks. Recall that we need eq. (A.12) to hold with high probability even if (after Eve's attack) the signals are highly correlated with one another. For this purpose it would be natural to use the quantum de Finetti theorem [40, 41] to show that the state of all the signals can be well approximated by a convex combination of product states. But unfortunately, because the Hilbert space of the signal and reference modes is infinite dimensional, the de Finetti theorem does not provide a useful approximation. Fortunately, though, the argument based on the imbalance of a quantum coin applies for any two-outcome POVM that Bob might apply (after the filtering), irrespective of the dimension of Bob's Hilbert space.

Appendix B Unambiguous key discrimination

In this Appendix, we elaborate on the unambiguous key discrimination attack against the signal set eq. (48). It is convenient to express the signals in an alternative basis, which is natural from the perspective of polarization encoding:

$$\begin{aligned}
|\tilde{0}_Z\rangle &= e^{-\mu/2} \left(|\text{vac}\rangle + \sqrt{2}\,\alpha|0_\text{Z}\rangle + \ldots \right) \;, \\
|\tilde{1}_Z\rangle &= e^{-\mu/2} \left(|\text{vac}\rangle + \sqrt{2}\,\alpha|1_\text{Z}\rangle + \ldots \right) \;, \\
|\tilde{0}_X\rangle &= e^{-\mu/2} \left(|\text{vac}\rangle + \sqrt{2}\,\alpha|0_\text{X}\rangle + \ldots \right) \;, \\
|\tilde{1}_X\rangle &= e^{-\mu/2} \left(|\text{vac}\rangle + \sqrt{2}\,\alpha|1_\text{X}\rangle + \ldots \right) \;;
\end{aligned} \quad \text{(B.1)}$$

here $\{|0_Z\rangle, |1_Z\rangle, |0_X\rangle, |1_X\rangle\}$ are the ideal BB84 states encoded in the polarization of a single photon, and the ellipsis indicates the multi-photon contribution. In this basis, the phase of

each signal is $e^{i\theta} = 1$, so that $\alpha = \sqrt{\mu/2}$ is real and positive. We assume that α is small, so that multi-photons are unlikely. Ignoring the small multi-photon component, the signals reside in a qutrit Hilbert space with basis $\{|\text{vac}\rangle, |0_Z\rangle, |1_Z\rangle\}$; expanded in this basis, they may be re-expressed as

$$
\begin{aligned}
|\tilde{0}_Z\rangle &\approx e^{-\mu/2}\left(1, \sqrt{2}\,\alpha, 0\right), \\
|\tilde{0}_X\rangle &\approx e^{-\mu/2}\left(1,\ \alpha,\ \alpha\right), \\
|\tilde{1}_Z\rangle &\approx e^{-\mu/2}\left(1, 0, \sqrt{2}\,\alpha\right), \\
|\tilde{1}_X\rangle &\approx e^{-\mu/2}\left(1,\ \alpha, -\alpha\right).
\end{aligned}
\tag{B.2}
$$

We will describe an intercept/resend attack on BB84 using these signals, based on *unambiguous key discrimination*. By exploiting her knowledge of the phase of the signals, Eve performs a POVM with three outcomes: 0, 1, and DK (don't know). The DK outcome is inconclusive, but if either of the other outcomes occurs, then Eve knows with certainty the key bit (0 or 1) encoded in the BB84 signal emitted by the source, though she does not gain any information about which of the two possible BB84 states compatible with that key bit was emitted. Eve blocks the signals when her outcome is DK, but if the outcome is conclusive she sends on to Bob a uniform coherent superposition of the two compatible BB84 states. This procedure generates a bit error rate $\delta = \frac{1}{2} - \frac{1}{2\sqrt{2}} \approx .146$. Evidently, Eve has the same key information as Alice and Bob, so that, if she also knows their protocol for error correction and privacy amplification, she will have perfect knowledge of every bit of the final key.

The multi-photon component of the state can help Eve, but to keep our analysis simple, we will consider an attack on this source that makes no use of the multi-photons. Eve performs an orthogonal measurement that distinguishes photon number less than two from photon number greater than or equal to two, and she discards the state if the latter outcome is found. (We will be interested in values of μ that are sufficiently small that Eve would not benefit very much from taking advantage of the multi-photons.) Thus the states she retains are qutrits. She then performs a three-outcome POVM (unambiguous key distribution) to identify the key bit. The two conclusive outcomes of the POVM are projections onto the states:

$$
\begin{aligned}
|0^\perp\rangle &= N_{\mu,0}\left(-\sqrt{2}\,\alpha - \alpha,\ 1 + \frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}\right), \\
|1^\perp\rangle &= N_{\mu,1}\left(-\alpha,\ 1 + \frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}\right),
\end{aligned}
\tag{B.3}
$$

where $N_{\mu,0}, N_{\mu,1}$ are normalization factors such that

$$
\begin{aligned}
N_{\mu,0}^{-2} &= (2 + \sqrt{2})\left[1 + \left(\frac{1}{2} + \frac{1}{2\sqrt{2}}\right)\mu\right], \\
N_{\mu,1}^{-2} &= (2 + \sqrt{2})\left[1 + \left(\frac{1}{2} - \frac{1}{2\sqrt{2}}\right)\mu\right].
\end{aligned}
\tag{B.4}
$$

The vector $|0^\perp\rangle$ is orthogonal to both of the two states $|\tilde{0}_Z\rangle$ and $|\tilde{0}_X\rangle$ that indicate the key bit 0. Hence, if this outcome is found, Eve knows for sure that the key bit could not be 0

and so must be 1. Similarly, the vector $|1^\perp\rangle$ is orthogonal to both of the states $|\tilde{1}_Z\rangle$ and $|\tilde{1}_X\rangle$ that indicate the key bit 1.

The vectors $|0^\perp\rangle$ and $|1^\perp\rangle$ are nearly parallel for small α. To ensure that all three POVM elements are positive, we may choose

$$E_0 = \frac{1}{2}|1^\perp\rangle\langle 1^\perp| , \quad E_1 = \frac{1}{2}|0^\perp\rangle\langle 0^\perp| ,$$
$$E_{\mathrm{DK}} = I - E_0 - E_1 . \tag{B.5}$$

(For small positive μ, the strength of the conclusive POVM elements can be pushed up slightly, but this is a small effect that we ignore.) Thus we find the probability p_C of a conclusive outcome (taking into account that Eve might detect multi-photons and reject the state)

$$\langle \tilde{0}_Z|E_0|\tilde{0}_Z\rangle = \langle \tilde{0}_X|E_0|\tilde{0}_X\rangle$$
$$= \left(\frac{1}{2} - \frac{1}{2\sqrt{2}}\right)\mu e^{-\mu}\left[1 + \left(\frac{1}{2} + \frac{1}{2\sqrt{2}}\right)\mu\right]^{-1} \tag{B.6}$$

if the key bit is 0, and

$$\langle \tilde{1}_Z|E_1|\tilde{1}_Z\rangle = \langle \tilde{1}_X|E_1|\tilde{1}_X\rangle$$
$$= \left(\frac{1}{2} - \frac{1}{2\sqrt{2}}\right)\mu e^{-\mu}\left[1 + \left(\frac{1}{2} - \frac{1}{2\sqrt{2}}\right)\mu\right]^{-1} \tag{B.7}$$

if the key bit is 1. We note that the conclusive outcome is slightly more likely when the key bit is 1. This asymmetry can be traced to the property that the overlap $|\langle \tilde{0}_X|\tilde{0}_Z\rangle|$ of the two signals that indicate the key bit 0 is slightly larger than the overlap $|\langle \tilde{1}_X|\tilde{1}_Z\rangle|$ of the two signals that indicate the key bit 1. (For other choices of the phase $e^{i\varphi}$ that determines the plane in the Bloch sphere occupied by the BB84 signals, the asymmetry is substantially larger.) In any case, for either value of the key bit, the probability of a conclusive outcome obeys

$$p_C \geq (.146)\mu e^{-\mu}\left[1 + (.854)\mu\right]^{-1} . \tag{B.8}$$

Now p_C is the probability that Eve resends the signal to Bob, and therefore it is also the probability p_D that Bob detects a signal, if his detector is perfectly efficient. If there were no interference by the eavesdropper (and no loss in the quantum channel connecting Alice and Bob), all non-vacuum signals would be detected, and then $p_D = 1 - e^{-\mu} = \mu + O(\mu^2)$. If Eve uses the unambiguous key discrimination POVM, then of course p_C vanishes in the limit $\mu \to 0$, but what is noteworthy is that p_C vanishes *linearly* with μ. Thus for μ small, a fraction $\eta \approx .146$ of order one of all the non-vacuum signals sent by Alice are received by Bob. (Since multi-photon signals span a space of even higher dimension, there is a POVM — unambiguous state discrimination — that, when conclusive, identifies not just the key bit but also the basis [42]; however for that POVM the probability of a conclusive outcome is higher order in μ.)

Having characterized Eve's attack against a source that emits the signal set eq. (B.1), let us now consider the security analysis of a phase-randomized source. We will refer to the source that emits weak coherent states with a definite phase known by the eavesdropper as source P (for phase), and to a source that emits phase-randomized weak coherent states, with the same

mean photon number μ, as source R (for random). If source P emits the signals eq. (B.1), and Eve launches the intercept/resend attack using unambiguous key discrimination, then Eve has perfect knowledge of every key bit, and Alice and Bob detect a bit error rate of $\delta = .146$. But if source R is used instead, we will see that Alice and Bob can extract private key at a nonzero asymptotic rate for a bit error rate up to $\delta = .189$. For the security analysis of source R, we will adopt a more restricted (and less realistic) model of the detector than the model described in Sec. 2: the detector applies a basis-independent "squash" that maps the incoming signal to a qubit, and then an ideal BB84 measurement is performed on that qubit in the appropriate basis.

We suppose that the source emits a multi-photon signal with probability p_M, that Bob detects a fraction p_D of all the signals sent by Alice, and that Eve's attack is unrestricted. Of the signals that are received, the fraction that were emitted as multi-photons is no more than $\Delta = p_M/p_D$; the rest are single photon signals.

To establish a nonzero key generation rate for a relatively high bit error rate, we will need to consider schemes for key reconciliation and privacy amplification that involve two-way communication between Alice and Bob, for which the argument in [18] based on the uncertainty principle does not apply. Instead we can prove security following [5], by relating the BB84 protocol to a protocol in which the key is generated by measuring noisy entangled pairs shared by Alice and Bob. Private key can be extracted at a positive asymptotic rate if it is possible to distill high fidelity entanglement from the noisy entanglement. Entanglement distillation will succeed if the noisy entangled pairs have a bit error rate and a phase error rate that are both sufficiently small. The bit error rate δ is inferred directly from the verification test in the BB84 protocol; the phase error rate δ_p is also inferred, but by a less direct argument.

If the source and detector used in the protocol were perfect, then a symmetry argument would suffice to show $\delta = \delta_p$. This symmetry is broken if the equipment is imperfect, but it is still possible to bound the difference between the two error rates using an appropriate characterization of the imperfections. For the case where Bob has a perfect detector, but Alice's source sometimes emits multi-photons, it can be shown as in eq. (40) that

$$|\delta_p - \delta| < \Delta/2 \; , \tag{B.9}$$

where Δ is the fraction of all the detected signals that were emitted as multi-photons. Eq. (40) is an improvement found by Boileau [29] of the result $|\delta_p - \delta| = \Delta$ found in [13].

The security proof in [5], which relates one-way privacy amplification to one-way entanglement distillation using quantum error-correcting codes, does not apply for a bit error rate above $\delta = .110$. But security of BB84 was established in [43] for a bit error rate as high as .189, by relating two-way privacy amplification to entanglement distillation with two-way communication between Alice and Bob. The original argument in [43] assumed a perfect source and detector. But the two-way entanglement distillation succeeds if both δ and δ_p are below .189; therefore the argument can be applied to a protocol with imperfect equipment if there is a strong enough bound on $|\delta - \delta_p|$.

If the bit error rate δ is .146, then the two-way BB84 protocol is secure for $|\delta - \delta_p| < .189 - .146 = .043$. And for a source that emits phase-randomized coherent states, it suffices

that $\Delta < .086$, where

$$\Delta = \frac{p_M}{p_D} \quad \leq \quad \left(\frac{\frac{1}{2}\mu^2}{(.146)\mu e^{-\mu}\left[1+(.854)\mu\right]^{-1}} \right)$$
$$= \quad (3.42)\mu e^{\mu}\left[1+(.854)\mu\right] \ . \tag{B.10}$$

Thus $\Delta < .086$, and the protocol is provably secure, for $\mu < .0240$. The security proof still applies if Bob's detector, rather than being perfectly efficient, has an efficiency that is independent of the basis in which the detector measures, where whether the detector fires is decided randomly, uninfluenced by the eavesdropper [44].

We have shown, therefore, that the BB84 QKD protocol is less secure using the phase-coherent source P than using the phase-randomized source R. Eve can exploit her knowledge of the phase of the signals emitted by the source P to implement a POVM that, when its outcome is conclusive, unambiguously identifies the key bit. But for the same bit error rate $\delta \approx .146$, signal strength μ $(< .0240)$, and signal detection rate $p_D \approx .146\mu$, if the signals have random phases then Alice and Bob can generate a final key about which Eve has negligible knowledge.

In fact, for source P using the signal set eq. (B.1) we do not have a security proof that establishes *any* positive bit error rate δ such that provably secure key can be generated at a positive asymptotic rate. However, as we have emphasized, it is a different story for the signal set eq. (8). In this paper, we have proven the security of BB84 using these signals, even when Eve knows the phase of the reference pulse.