Complete physical simulation of the entangling-probe attack on the Bennett-Brassard 1984 protocol

Taehyun Kim,[*] Ingo Stork genannt Wersborg, Franco N. C. Wong, and Jeffrey H. Shapiro
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

(Received 22 November 2006; published 25 April 2007)

We have used deterministic single-photon two-qubit quantum logic to implement the most powerful individual-photon attack against the Bennett-Brassard 1984 (BB84) quantum key distribution protocol. Our measurement results, including physical source and gate errors, are in good agreement with theoretical predictions for the Rényi information obtained by Eve as a function of the errors she imparts to Alice and Bob's sifted key bits. The current experiment is a physical simulation of a true attack, because Eve has access to Bob's physical receiver module. Nevertheless, the physical simulation allows investigation of the fundamental security limit of the BB84 protocol against eavesdropping in the presence of realistic physical errors, and it affords the opportunity to study the effectiveness of error correction and privacy amplification when the BB84 protocol is attacked.

DOI: 10.1103/PhysRevA.75.042327 PACS number(s): 03.67.Dd, 03.67.−a, 42.65.Lm

I. INTRODUCTION

In 1984 Bennett and Brassard [1] proposed a protocol [Bennett-Brassard 1984 (BB84)] for quantum key distribution in which the sender (Alice) transmits single-photon pulses to the receiver (Bob) in such a way that security is vouchsafed by physical laws. Since then, the BB84 protocol has been implemented in free space [2] as well as in fibers [3], and also its security has been the subject of many analyses [4,5], particularly for configurations that involve nonideal operating conditions [6], such as the use of weak laser pulses instead of single photons. A more fundamental question is how much information the eavesdropper (Eve) can gain under ideal BB84 operating conditions. Papers by Fuchs and Peres [7], Slutsky *et al.* [8], and Brandt [9] show that the most powerful individual-photon attack can be accomplished with a controlled-NOT (CNOT) gate. In this scheme, Eve supplies the target qubit to the CNOT gate, which entangles it with the BB84 qubit that Alice is sending to Bob. Eve then makes her measurement of the target qubit to obtain information on the shared key bit at the expense of imposing detectable errors between Alice and Bob [9,10].

We have recently shown [10] that this Fuchs-Peres-Brandt (FPB) entangling probe can be implemented using single-photon two-qubit (SPTQ) quantum logic in a proof-of-principle experiment. In SPTQ logic a single photon carries two independent qubits: the polarization and the momentum (or spatial path) states of the photon. Compared to standard two-photon quantum gates, SPTQ gates are deterministic and can be efficiently implemented using only linear optical elements [11,12].

In this work we use SPTQ logic to implement the FPB probe as a complete physical simulation of the most powerful individual-photon attack on BB84 key distribution, including physical errors. This is to our knowledge the first experiment on attacking the BB84 protocol, and the results are in good agreement with theoretical predictions. It is only a physical simulation because the two qubits of a single photon carrier must be measured jointly, so that Eve needs access to Bob's receiver, but *not* his measurement. The SPTQ probe could become a true attack if quantum nondemolition measurements were available to Eve [10].

II. THEORETICAL BACKGROUND

In the BB84 protocol, Alice sends Bob a single photon randomly chosen from the four polarization states of the horizontal-vertical (H-V) and ±45° diagonal-antidiagonal (D-A) bases. In the FPB attack, Eve sets up her CNOT gate with its control-qubit computational basis $\{|0\rangle_C, |1\rangle_C\}$ given by a $\pi/8$ rotation from the BB84 H-V basis, as shown in Fig. 1(a),

$$|0\rangle_C = \cos(\pi/8)|H\rangle + \sin(\pi/8)|V\rangle,$$

$$|1\rangle_C = -\sin(\pi/8)|H\rangle + \cos(\pi/8)|V\rangle. \tag{1}$$

Having selected the error probability P_E that she is willing to create, Eve prepares her probe qubit (CNOT's target) in the initial state

$$|T_{\text{in}}\rangle = \{(C+S)|0\rangle_T + (C-S)|1\rangle_T\}/\sqrt{2}$$
$$\equiv \cos\,\theta_{\text{in}}|0\rangle_T + \sin\,\theta_{\text{in}}|1\rangle_T, \tag{2}$$

where $C=\sqrt{1-2P_E}$, $S=\sqrt{2P_E}$, and $\{|0\rangle_T, |1\rangle_T\}$ is the target

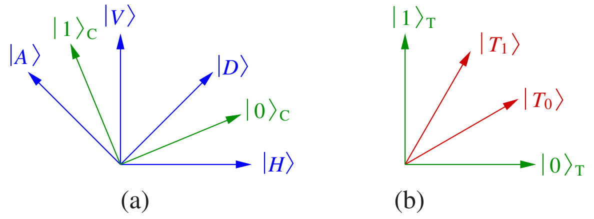

FIG. 1. (Color online) Relations between different bases. (a) Control qubit basis for Eve's CNOT gate referenced to the BB84 polarization states. (b) $|T_0\rangle$ and $|T_1\rangle$ relative to the target qubit basis.

[*] Electronic address: thkim@mit.edu

qubit's computational basis. After the CNOT operation—with inputs from Alice's photon and Eve's probe—the two qubits become entangled. For each of Alice's four possible inputs $|H\rangle$, $|V\rangle$, $|D\rangle$, and $|A\rangle$, the output of the CNOT gate is

$$|H\rangle|T_{\text{in}}\rangle \rightarrow |H_{\text{out}}\rangle \equiv |H\rangle|T_0\rangle + |V\rangle|T_E\rangle, \tag{3}$$

$$|V\rangle|T_{\text{in}}\rangle \rightarrow |V_{\text{out}}\rangle \equiv |V\rangle|T_1\rangle + |H\rangle|T_E\rangle, \tag{4}$$

$$|D\rangle|T_{\text{in}}\rangle \rightarrow |D_{\text{out}}\rangle \equiv |D\rangle|T_0\rangle - |A\rangle|T_E\rangle, \tag{5}$$

$$|A\rangle|T_{\text{in}}\rangle \rightarrow |A_{\text{out}}\rangle \equiv |A\rangle|T_1\rangle - |D\rangle|T_E\rangle, \tag{6}$$

where (un-normalized) $|T_0\rangle$, $|T_1\rangle$, and $|T_E\rangle$ are defined in the target qubit's computational basis [see Fig. 1(b)] as

$$|T_0\rangle \equiv \left(\frac{C}{\sqrt{2}} + \frac{S}{2}\right)|0\rangle_T + \left(\frac{C}{\sqrt{2}} - \frac{S}{2}\right)|1\rangle_T, \tag{7}$$

$$|T_1\rangle \equiv \left(\frac{C}{\sqrt{2}} - \frac{S}{2}\right)|0\rangle_T + \left(\frac{C}{\sqrt{2}} + \frac{S}{2}\right)|1\rangle_T, \tag{8}$$

$$|T_E\rangle \equiv \frac{S}{2}(|0\rangle_T - |1\rangle_T). \tag{9}$$

Consider the case in which Bob measures in the same basis that Alice employed and his outcome matches what Alice sent. Then, according to Eqs. (3)–(6), the target qubit is projected into either $|T_0\rangle$ or $|T_1\rangle$. After Alice and Bob compare their basis selections over the classical channel, Eve can learn about their shared bit value by distinguishing between the $|T_0\rangle$ and $|T_1\rangle$ output states of her target qubit. To do so, she employs the minimum error probability receiver for distinguishing between $|T_0\rangle$ and $|T_1\rangle$ by performing a projective measurement along $|0\rangle_T$ and $|1\rangle_T$. Eve can then correlate the measurement of $|0\rangle_T$ ($|1\rangle_T$) with $|T_0\rangle$ ($|T_1\rangle$). Note that this projective measurement is not perfect unless $|T_0\rangle$ and $|T_1\rangle$ are orthogonal and hence coincide with the target's computational basis, $|0\rangle_T$ and $|1\rangle_T$. Also note that regardless of the basis that Alice and Bob choose (*H*-*V* or *D*-*A*), Eve needs only to distinguish between $|0\rangle_T$ and $|1\rangle_T$. Therefore she can measure her probe qubit immediately, obviating the need for any quantum memory in the FPB probe attack.

Of course, Eve's information gain comes at a cost: Eve has caused an error event whenever Alice and Bob choose a common basis and Eve's probe output state is $|T_E\rangle$. When Alice sent $|H\rangle$ and Bob measured in the *H*-*V* basis, Eq. (3) then shows that Alice and Bob will have an error event if the measured output state is $|V\rangle|T_E\rangle$. The probability that this will occur is $\langle T_E|T_E\rangle = S^2/2 = P_E$. For the other three cases in Eqs. (4)–(6), the error event corresponds to the last term in each expression. Therefore the conditional error probabilities are identical, and hence P_E is the unconditional error probability.

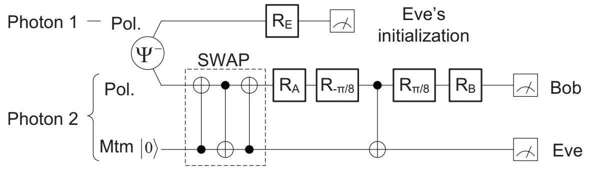

FIG. 2. Quantum circuit diagram for the FPB-probe attack. Photon 1 of a polarization-entangled singlet photon pair heralds photon 2 and sets Eve's probe qubit to its initial state. The SWAP gate allows Alice's qubit to be set in the polarization mode of photon 2, whose momentum mode is Eve's probe qubit. The CNOT gate entangles Alice's qubit with Eve's qubit. R_E, rotation by Eve; R_A, rotation by Alice; R_B, rotation by Bob; $R_{\pm\pi/8}$, rotation by angle $\pm\pi/8$.

We use Rényi information to quantify Eve's information gain about the sift events in which Bob measures because privacy amplification [5] requires an estimated upper bound for Eve's Rényi information about the corrected data [8]. Let $B = \{0,1\}$ and $E = \{0,1\}$ denote the ensembles of possible bit values that Bob and Eve receive on an error-free sift event. The Rényi information (in bits) that Eve learns about each error-free sift event is

$$I_R \equiv -\log_2\left(\sum_{b=0}^{1} P^2(b)\right) + \sum_{e=0}^{1} P(e)\log_2\left(\sum_{b=0}^{1} P^2(b|e)\right), \tag{10}$$

where $\{P(b), P(e)\}$ are the *a priori* probabilities for Bob's and Eve's bit values, and $P(b|e)$ is the conditional probability for Bob's bit value to be b given that Eve's is e. With a perfect channel and perfect equipment, this leads to the theoretical prediction [10],

$$I_R = \log_2\left(1 + \frac{4P_E(1 - 2P_E)}{(1 - P_E)^2}\right). \tag{11}$$

Under these ideal conditions, Eve's Rényi information is the same for both bases, but in actual experiments it may differ, owing to differing equipment errors in each basis.

III. EXPERIMENTAL SETUP AND RESULTS

Figure 2 shows the quantum circuit diagram of our SPTQ implementation of the FPB probe. We start with a pair of polarization-entangled photons in the singlet state. Photon 1 is used as a trigger to herald photon 2 as a single-photon pulse for the BB84 protocol. A SWAP operation applied to photon 2 exchanges its polarization and momentum qubits so that the polarization of photon 1 and the momentum of photon 2 are now entangled in a singlet state. Eve encodes her probe qubit in the momentum state of photon 2 by projecting photon 1 along an appropriate polarization state set by a polarization rotation R_E. The polarization state of photon 2 after the SWAP gate is Alice's qubit, which is set by rotation R_A. Similarly, Bob's polarization analysis of Alice's qubit is set by R_B. In this configuration, Eve is heralding the photon on which Alice is encoding polarization. However, an equivalent experiment could be performed if Alice polarization-encoded a single-photon source, after which Eve imposed her momentum qubit by polarization control applied in between a pair of SWAP gates.

The CNOT gate in Fig. 2 is preceded by a $-\pi/8$ rotation and followed by a $+\pi/8$ rotation because the basis for the CNOT's control qubit is rotated by $\pi/8$ from the BB84 bases,

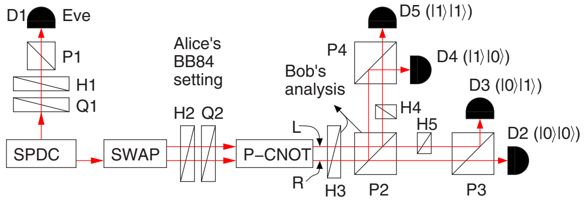

FIG. 3. (Color online) Experimental configuration for a complete physical simulation of the FPB attack on BB84. SPDC, spontaneous parametric down-conversion source; H, half-wave plate; Q, quarter-wave plate; P, polarizing beam splitter; D, single-photon detector. R, L refer to spatial paths.

as noted in Fig. 1. The CNOT gate that Eve employs is a polarization-controlled NOT (P-CNOT) gate that uses the polarization qubit as the control and the momentum qubit of the same photon as the target. We have previously demonstrated such a gate in a polarization Sagnac interferometer with an embedded dove prism [11]. We have also demonstrated the SWAP operation [12] by cascading three CNOT gates: a momentum-controlled NOT (M-CNOT), a P-CNOT, and another M-CNOT.

Figure 3 shows our experimental setup for implementing the quantum circuit. We used a bidirectionally pumped Sagnac interferometric down-conversion source [13] with a periodically poled $\mathrm{KTiOPO_4}$ crystal to generate polarization-entangled photons at 810 nm in the singlet state. The measured flux was ~700 pairs/s per mW of pump in a 1 nm bandwidth at ~99.45% quantum-interference visibility. The collimated output beam had a beam waist of w_0=0.53 mm. In a collimated configuration, the momentum state of a photon is the same as the spatial orientation of the beam, for which we use the right-left (R-L) basis, and we aligned the output path of photon 2 to match the R input path of the SWAP gate as shown in Fig. 3. This is equivalent to setting the momentum qubit to $|0\rangle$ in Fig. 2 under the mapping of R (L) to $|0\rangle$ ($|1\rangle$). The L and R beams were separated by ~2 mm.

For each photon pair, photon 1 is used to herald the arrival of photon 2 and also to remotely control the momentum qubit of photon 2 by postselection. R_E polarization rotation by Eve was implemented using a quarter-wave plate (QWP) Q1 and a half-wave plate (HWP) H1, followed by single-photon detection (D1) through a polarizing beam splitter (PBS) P1 along H. Q1 was used to compensate an intrinsic phase shift ξ imposed by the SWAP gate on the target-qubit basis $\{|0\rangle_T, |1\rangle_T\}$. We have independently measured $\xi \simeq 88°$. Therefore, the R_E operation prepared the momentum qubit in $|T'_{\text{in}}\rangle$ with a Q1-imposed phase shift ξ:

$$|T'_{\text{in}}\rangle \equiv \cos\theta_{\text{in}}|0\rangle_T + e^{i\xi}\sin\theta_{\text{in}}|1\rangle_T. \tag{12}$$

The extra phase shift of the SWAP gate would bring Eve's probe qubit to be in $|T_{\text{in}}\rangle$ of Eq. (2).

After the SWAP gate, R_A and $R_{-\pi/8}$ were combined in a single operation. The P-CNOT gate had the same phase shift problem as the SWAP gate, so we used a HWP (H2) and a QWP (Q2) to compensate for this phase shift and to impose the required rotation. After H2 and Q2, Alice's qubit becomes

$$|\Psi_A\rangle \equiv \cos\theta_A|0\rangle_C + e^{i\chi}\sin\theta_A|1\rangle_C, \tag{13}$$

where χ ($\simeq 98°$) is the compensating phase shift and θ_A, which is the sum of Alice's angle and −22.5°, is −22.5°, 22.5°, 67.5°, or 112.5° for $|H\rangle$, $|D\rangle$, $|V\rangle$, or $|A\rangle$, respectively, as shown in Fig. 1(a). Similarly we combined $R_{\pi/8}$ and R_B into a single HWP (H3) in Fig. 3 and a PBS (P2) was used by Bob to analyze the polarization of Alice's qubit.

Eve measured her qubit by a projective measurement along the $|0\rangle_T - |1\rangle_T$ (spatially, R-L) basis. A HWP (H4/H5) was placed in the R or L beam path, as indicated in Fig. 3, so that the R and L beams would be distinguished by their orthogonal polarizations. This polarization tagging simplified their measurements by a PBS (P3, P4) and single-photon detectors. The four detectors uniquely identified the two qubits of photon 2. D2, D3, D4, and D5 correspond to $|H\rangle|R\rangle$, $|H\rangle|L\rangle$, $|V\rangle|R\rangle$, $|V\rangle|L\rangle$ ($|D\rangle|R\rangle$, $|D\rangle|L\rangle$, $|A\rangle|R\rangle$, $|A\rangle|L\rangle$), respectively, when the H-V (D-A) basis is chosen. Therefore, in our physical simulation, these joint measurements yield Bob's polarization information and Eve's momentum information.

TABLE I. Data samples, estimated probabilities, and theoretical values for D and A inputs with Bob using the same basis as Alice, and for predicted error probabilities P_E=0, 0.1, and 0.33. $|0\rangle|1\rangle$ corresponds to Bob's measuring $|D\rangle$ and Eve's measuring $|1\rangle_T$. Column 1 shows the state Alice sent and column 2 shows the predicted error probability P_E. "Coincidence" columns show coincidence counts over a 40 s interval. "Estimated" columns show the measured coincidence counts normalized by the total counts of all four detectors, and "Expected" columns show the theoretical values under ideal operating conditions.

		Coincidence				Estimated				Expected			
Alice	P_E	$\|1\rangle\|0\rangle$	$\|1\rangle\|1\rangle$	$\|0\rangle\|1\rangle$	$\|0\rangle\|0\rangle$	$\|1\rangle\|0\rangle$	$\|1\rangle\|1\rangle$	$\|0\rangle\|1\rangle$	$\|0\rangle\|0\rangle$	$\|1\rangle\|0\rangle$	$\|1\rangle\|1\rangle$	$\|0\rangle\|1\rangle$	$\|0\rangle\|0\rangle$
$\|D\rangle=\|0\rangle$	0	1356	1836	23408	23356	0.027	0.037	0.469	0.468	0	0	0.500	0.500
	0.1	2840	4220	9664	32592	0.058	0.086	0.196	0.661	0.050	0.050	0.167	0.733
	0.33	7512	9496	1512	30916	0.152	0.192	0.031	0.625	0.167	0.167	0	0.667
$\|A\rangle=\|1\rangle$	0	22664	23388	1140	1112	0.469	0.484	0.024	0.023	0.500	0.500	0	0
	0.1	8480	34492	4088	2052	0.173	0.702	0.083	0.042	0.167	0.733	0.050	0.050
	0.33	1096	32360	9384	6564	0.022	0.655	0.19	0.133	0	0.667	0.167	0.167

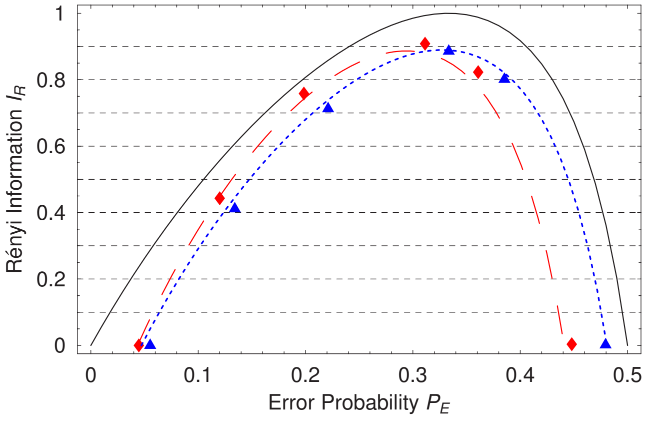

FIG. 4. (Color online) Eve's Rényi information I_R about Bob's error-free sifted bits as a function of the error probability P_E that her eavesdropping creates. Solid curve: theoretical result from Eq. (11). Diamonds (triangles): measured values for *H*-*V* (*D*-*A*) basis. Dashed (dotted) curves are fits to the data with error model for the *H*-*V* (*D*-*A*) basis.

In data collection, we measured coincidences between $D1$ and one of the detectors for photon 2. Table I shows two data sets for Alice's input of D and A polarizations and compares them with the expected values for the ideal case. From the raw data, we calculate the Rényi information I_R based on Eq. (10), and Fig. 4 plots I_R as a function of the error probability P_E. The solid curve shows the ideal case [Eq. (11)] and diamonds (triangles) represent I_R for the measured values with inputs in the *H*-*V* (*D*-*A*) basis. We note that accidental coincidences were negligible and the coincidence window was ~3 ns.

In the ideal case with $P_E{=}0$, Eve gets no information, $I_R{=}0$, and Alice and Bob have no error bits. However, due to experimental errors such as imperfect gate fidelities, we found that ~5% of the sifted bits had errors. For $P_E{=}1/3$, Eve obtains perfect information, $I_R{=}1$ under ideal conditions, but in our experiment, Eve gained a maximum I_R $=0.9$, corresponding to her having 95% probability of correctly receiving one of Alice and Bob's error-free sifted bits.

IV. ERROR ANALYSIS

To understand the errors involved in the experiment, we model our experimental setup with some nonideal parameters. We assume that the phases ξ in Eq. (12) and χ in Eq. (13) could be inaccurate, and similarly for the setting of θ_A in Eq. (13) that might be caused by the wave plates. We also model the unitary P-CNOT gate as

$$\begin{pmatrix} \cos\alpha & ie^{-i\delta}\sin\alpha & 0 & 0 \\ ie^{i\delta}\sin\alpha & \cos\alpha & 0 & 0 \\ 0 & 0 & -ie^{i\delta}\sin\alpha & \cos\alpha \\ 0 & 0 & \cos\alpha & -ie^{-i\delta}\sin\alpha \end{pmatrix}, \tag{14}$$

where $\alpha{=}0$ and $\delta{=}0$ for an ideal P-CNOT gate. Finally we assume that Bob's HWP H3 setting of θ_B was imperfect, so that

$$|H\rangle \rightarrow \cos\theta_B|0\rangle_C - \sin\theta_B|1\rangle_C, \tag{15}$$

$$|V\rangle \rightarrow \sin\theta_B|0\rangle_C + \cos\theta_B|1\rangle_C, \tag{16}$$

where θ_B should equal 22.5° (−22.5°) in the *H*-*V* (*D*-*A*) basis.

We fit the data by minimizing the differences between 96 measurements and the calculated numbers based on this error model. The fitting results were $\Delta\xi \simeq 3°$, $\Delta\chi \simeq -11°$, $\Delta\theta_A(H,D,V,A)=\{3.2°, 0.9°, -0.7°, -2.3°\}$, $\alpha=12.3°$, δ $=3.6°$, $\Delta\theta_B(H/V,D/A)=\{-1.8°, 0°\}$. As expected, the phase errors are relatively small and those associated with θ_A and θ_B are within the resolution of the rotating mounts housing the wave plates. The nonzero α also agrees with the measured classical visibility of 94% for the P-CNOT gate. Figure 4 shows the fitted I_R based on this model for the *H*-*V* basis (dashed curve) and the *D*-*A* basis (dotted curve).

V. CONCLUSION

In summary, we have demonstrated experimentally a complete physical simulation of the entangling-probe attack, showing that Eve can gain Rényi information of up to 0.9 under realistic operating conditions, including a CNOT gate that does not have an ultrahigh fidelity. Our results suggest the possible amount of information gain by Eve with current technology and the need to evaluate the required level of privacy amplification.

ACKNOWLEDGMENT

This work was supported by BBN Technologies under the DARPA QuIST program.

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

[2] C. H. Bennett *et al.*, Lect. Notes Comput. Sci. **473**, 253 (1991); B. C. Jacobs and J. D. Franson, Opt. Lett. **21**, 1854 (1996); W. T. Buttler, R. J. Hughes, S. K. Lamoreaux, G. L. Morgan, J. E. Nordholt, and C. Peterson, Phys. Rev. Lett. **84**, 5652 (2000); R. J. Hughes *et al.*, New J. Phys. **4**, 43 (2002); C. Kurtsiefer *et al.*, Nature (London) **419**, 450 (2002).

[3] J. D. Franson and H. Ilves, Appl. Opt. **33**, 2949 (1994); C. Marand and P. D. Townsend, Opt. Lett. **20**, 1695 (1995); R. J. Hughes *et al.*, J. Mod. Opt. **47**, 533 (2000).

[4] P. W. Shor and J. Preskill, Phys. Rev. Lett. **85**, 441 (2000); D. Mayers, J. ACM **48**, 351 (2001); H.-K. Lo, J. Phys. A **34**, 6957 (2001).

[5] C. H. Bennett *et al.*, J. Cryptology **5**, 3 (1992); C. H. Bennett *et al.*, IEEE Trans. Inf. Theory **41**, 1915 (1995).

[6] B. Slutsky *et al.*, J. Mod. Opt. **44**, 953 (1997); G. Brassard, N. Lutkenhaus, T. Mor, and B. C. Sanders, Phys. Rev. Lett. **85**, 1330 (2000); G. Gilbert and M. Hamrick, e-print quant-ph/0009027; V. Makarov and D. R. Hjelme, J. Mod. Opt. **52**, 691 (2005).

[7] C. A. Fuchs and A. Peres, Phys. Rev. A **53**, 2038 (1996).

[8] B. A. Slutsky, R. Rao, P. C. Sun, and Y. Fainman, Phys. Rev. A **57**, 2383 (1998).

[9] H. E. Brandt, Phys. Rev. A **71**, 042312 (2005).

[10] J. H. Shapiro and F. N. C. Wong, Phys. Rev. A **73**, 012315 (2006).

[11] M. Fiorentino and F. N. C. Wong, Phys. Rev. Lett. **93**, 070502 (2004).

[12] M. Fiorentino, T. Kim, and F. N. C. Wong, Phys. Rev. A **72**, 012318 (2005).

[13] T. Kim, M. Fiorentino, and F. N. C. Wong, Phys. Rev. A **73**, 012316 (2006); F. N. C. Wong, J. H. Shapiro, and T. Kim, Laser Phys. **16**, 1517 (2006).