Security of Quantum Key Distribution Against All Collective Attacks

Eli Biham[(1)], Michel Boyer[(2)], Gilles Brassard[(2)], Jeroen van de Graaf[(2)] and Tal Mor[(2)]
(1) Computer Science Department, Technion, Haifa 32000, Israel; (2) DIRO, Université de Montréal, Montréal, Canada;
(12th January 1998)

Security of quantum key distribution against sophisticated attacks is among the most important issues in quantum information theory. In this work we prove security against a very important class of attacks called *collective attacks* (under a compatible noise model) which use quantum memories and gates, and which are directed against the final key. Although attacks stronger than the collective attacks can exist in principle, no explicit example was found and it is conjectured that security against collective attacks implies also security against any attack.

a. Introduction Quantum cryptography is one of the most surprising consequences of processing information using quantum two-state systems (qubits) instead of classical bits. Quantum key distribution was invented 14 years ago [1], to provide a new type of solution to one of the most important cryptographic problems, the transmission of secret messages.

For many years physicists and computer scientists have been trying to prove the security of various quantum key distribution schemes. Many particular "simple" cases where analyzed, such as the *intercept-resend attacks* and the *individual particle attacks*, for which there is a clear intuition that classical privacy amplification provides the desired security, but no explicit bound on the information available to an eavesdropper has been proven. The security in case of the general *joint attacks* which are using quantum gates and quantum memory and are directed against the *final key* was also considered in several works (see [2, 3] and references there in). In this paper we complete the work started in [4, 5, 2] to conclude that the four-state scheme [1] for quantum key distribution is secure against any *collective attack* (an important subclass of the joint attacks) under a compatible error model.

In the four-state scheme Alice sends to Bob a classical string of length n'' using a quantum channel, by sending qubits; she sends either $|0\rangle_z$ or $|0\rangle_x = (|0\rangle_z + |1\rangle_z)/\sqrt{2}$ to encode a bit value 0, or she sends either $|1\rangle_z$ or $|1\rangle_x = (|0\rangle_z - |1\rangle_z)/\sqrt{2}$ to encode a bit value 1. Alice and Bob are also connected by a classical channel which is insecure but unjammable. At a later stage Alice tells Bob (classically), regarding each qubit, whether she used the z basis or the x basis. If Bob has used the same basis for his measurement, they keep the bit (which is supposed to be the same as Alice's), so they are left with n' similar bits. Alice and Bob now estimate the error rate using some (n_{test}) test bits. If the estimated error-rate p_{test} is less than some pre-agreed threshold $p_{allowed}$, then the test succeeds and Alice and Bob obtain a final key from the remaining n-bit string (where $n = n' - n_{test}$), by performing error correction and privacy amplification. They choose parities of k substrings for error-correction and parities of m substrings for privacy amplification. The parity of each of the k substrings is announced in order to correct the string, and the parities of the m substrings are kept secret, and used as the final key. We consider $m = 1$ in the following and leave the general case to a review paper.

In the most general (so called "joint") attack, Eve can do whatever she likes (the most general unitary transformation using an ancila) to the qubits, and delay all her measurements till receiving all classical data. We restrict ourself to "collective" attacks [5] where each qubit is attached to a separate probe (unentangled to the other probes), and the measurement is delayed, and is performed collectively on all probes, after all classical data is obtained. There are good reasons [5, 2] to believe that collective attacks are the strongest joint attacks (when n is large). Furthermore, no particular joint attack was shown to be stronger than collective attacks. In a collective attack, after Alice sends n'' qubits to Bob, each is attached to a separate probe by Eve. Then, the global state of the Eve-Bob system is $\rho_1 \otimes \ldots \otimes \rho_{n''}$ where each ρ_i is a density operator on the space $\mathcal{H}^{E_i} \otimes \mathcal{H}^{B_i}$ where the spaces $\mathcal{H}^{E_i}$ and $\mathcal{H}^{B_i}$ belong respectively to Eve and Bob.

b. Bounds on information We shall first fix some notations from information theory. Let B and X be random variables (describing the input and output of a channel). When the context is clear we write $p(b)$ for $p(B{=}b)$ and $p(x)$ for $p(X{=}x)$. The joint probability $p(x, b)$ satisfies $p(x) = \sum_{b\in B} p(x, b)$ and $p(b) = \sum_{x\in X} p(x, b)$. The conditional probability is denoted by $p_b(x) \equiv p(X = x|B{=}b)$ and $p_x(b) \equiv p(B = b|X = x)$. It satisfies the Bayes formula $p_b(x)p(b) = p(x, b) = p_x(b)p(x)$. The mutual information between the input and the output probability distributions, $I(X; B) = -\sum_{b\in B} p(b)log_2p(b) + \sum_{x\in X} p(x)\sum_{b\in B} p_x(b)log_2p_x(b)$, tells us the increase of knowledge about the input, if the output becomes known to us.

For a binary input B with equal input probabilities $I(B; X) = \sum_{x\in X} p(x)I_2(p_x(0))$, where $I_2(p) = 1 + p\log_2 p + (1-p)\log_2(1-p)$. Distinguishing the input when the output is given, is then equivalent to distinguishing the two probability distributions $p_0(x)$ and $p_1(x)$. All the probability distributions in the expression of the mutual information can be calculated from $p_0(x)$, $p_1(x)$, so we can define another function SD, Shannon Distinguishability, $SD(p_0(x), p_1(x)) \equiv I(B; X)$ (restricted to binary input with equal probabilities).

Suppose we are given a state (density matrix) ρ. The most general measurement giving a result x in some set X of possible outputs is given by a POVM [6] indexed by X, i.e. a family $\mathcal{E} = (E_x)_{x\in X}$ of Hermitian operators E_x with non negative eigenvalues such that $\sum_{x\in X} E_x = \mathbb{1}$. The probability of occurrence of x given the state ρ is then equal to $p^{\mathcal{E}}(x) = Tr(\rho E_x)$. Given two equally likely states ρ_0 and ρ_1, and a measurement procedure $\mathcal{E}$, let $p_b^{\mathcal{E}}(x) = Tr(\rho_b E_x)$; for any given $\mathcal{E}$ let $SD^{\mathcal{E}}(\rho_0,\rho_1) \equiv SD(p_0^{\mathcal{E}}(x), p_1^{\mathcal{E}}(x))$. The maximum information we can get regarding the state we are facing is given by the optimal Shannon Distinguishability $SD(\rho_0,\rho_1) \equiv \sup_{\mathcal{E}}[SD^{\mathcal{E}}(\rho_0,\rho_1)]$ where the supremum is taken over all POVM's on all possible sets X.

Unfortunately, there is no known analytic formula giving optimal mutual information. In what follows, we will present two bounds which are simple to state and to derive, and which will be found very useful.

Theorem 1. — If $\widetilde{\rho_0}$ and $\widetilde{\rho_1}$ are two density matrices defined on some space $\mathcal{H}_1 \otimes \mathcal{H}_2$ and $\rho_i = Tr_2(\widetilde{\rho_i})$ are the density matrices on $\mathcal{H}_1$ obtained by tracing-out $\mathcal{H}_2$, then,

$$SD(\rho_0,\rho_1) = SD(Tr_2(\widetilde{\rho_0}), Tr_2(\widetilde{\rho_1})) \leq SD(\widetilde{\rho_0},\widetilde{\rho_1}) \ . \quad (1)$$

Proof. — If $\mathcal{E} = (E_x)_{x\in X}$ is a POVM on $\mathcal{H}_1$ then $\mathcal{E}\otimes\mathbb{1}_{\mathcal{H}_2} \equiv (E_x\otimes\mathbb{1}_{\mathcal{H}_2})_{x\in X}$ is a POVM on $\mathcal{H}_1\otimes\mathcal{H}_2$ and $Tr_1(Tr_2(\widetilde{\rho_i})E_x) = Tr(\widetilde{\rho_i}(E_x\otimes\mathbb{1}_{\mathcal{H}_2}))$. Consequently $SD^{\mathcal{E}}(Tr_2(\widetilde{\rho_0}), Tr_2(\widetilde{\rho_1})) = SD^{\mathcal{E}\otimes\mathbb{1}_{\mathcal{H}_2}}(\widetilde{\rho_0},\widetilde{\rho_1})$. By definition of the optimization process $\sup_{\mathcal{E}}[SD^{\mathcal{E}\otimes\mathbb{1}_{\mathcal{H}_2}}(\widetilde{\rho_0},\widetilde{\rho_1})] \leq \sup_{\widetilde{\mathcal{E}}}[SD^{\widetilde{\mathcal{E}}}(\widetilde{\rho_0},\widetilde{\rho_1})]$, and thus, $SD(Tr_2(\widetilde{\rho_0}), Tr_2(\widetilde{\rho_1})) \leq SD(\widetilde{\rho_0},\widetilde{\rho_1})$. The $\widetilde{\rho_b}$ will be called a *lift-up* of ρ_p, and it is known as *purification* if it is a pure state.

This theorem (proven independently in [8]) actually states that tracing out cannot increase information. It provides a useful upper bound on the mutual information that can be obtained about mixed states ρ_i, if we can find appropriate states $\widetilde{\rho_i}$. A similar idea which says that mixing cannot improve information was used in [2] to obtain a more limited security result.

For any two density matrices ρ_0 and ρ_1 we can define $Tr|\rho_0-\rho_1|$ the trace-norm of the Hermitian operator $\rho_0-\rho_1$. In our context where we only consider Hermitian matrices, $Tr|A|$ is nothing but the sum of the absolute values of the eigenvalues of A. It is relatively easy to calculate the trace-norm of $\rho_0-\rho_1$. Therefore, the following upper bound is very important.

Theorem 2. — For any two density matrices ρ_0 and ρ_1,

$$SD(\rho_0,\rho_1) \leq \frac{1}{2}Tr|\rho_0-\rho_1| \ . \quad (2)$$

Proof. — In order to prove this equation (see also [7]) let us first fix some measurement procedure $\mathcal{E} = (E_x)_{x\in X}$. Then $SD^{\mathcal{E}}(\rho_0,\rho_1) = I(B;X) = \sum_{x\in X} p(x) I_2(p_x(0))$, where (from the Bayes formula) $p_x(0) = p(B = 0)p(X = x|B = 0)/p(x) = (1/2)p_0^{\mathcal{E}}(x)/p(x)$. Knowing that $I_2(r) \leq |2r-1|$ for $0 \leq r \leq 1$ we conclude that $SD^{\mathcal{E}}(\rho_0,\rho_1) \leq \sum_{x\in X} p(x)|2p_x(0)-1|$. Assigning $p_x(0)$ into the last expression [and using $p(x) = (p_0^{\mathcal{E}}(x)+p_1^{\mathcal{E}}(x))/2$ in the following equality], we obtain $SD^{\mathcal{E}}(\rho_0,\rho_1) \leq \sum_{x\in X} p(x)|2[p_0^{\mathcal{E}}(x)/2p(x)]-1| = \frac{1}{2}\sum_{x\in X}|p_0^{\mathcal{E}}(x)-p_1^{\mathcal{E}}(x)|$. Now, since $\rho_0-\rho_1$ is Hermitian, it can be diagonalized and consequently written in the form $\rho_0-\rho_1 = \sum\lambda_j|j\rangle\langle j|$ where $|j\rangle$ is an orthonormal basis and $Tr|\rho_0-\rho_1| = \sum|\lambda_j|$. Clearly $Tr(|j\rangle\langle j|E_x) = \langle j|E_x|j\rangle$ and so $p_0^{\mathcal{E}}(x)-p_1^{\mathcal{E}}(x) = Tr((\rho_0-\rho_1)E_x) = \sum_j\lambda_j\langle j|E_x|j\rangle$. Using the last expression for SD and using $\langle j|E_x|j\rangle \geq 0$ (since E_x is positive definite), we can now deduce $SD^{\mathcal{E}}(\rho_0,\rho_1) \leq \frac{1}{2}\sum_j|\lambda_j|\sum_{x\in X}\langle j|E_x|j\rangle = \frac{1}{2}Tr|\rho_0-\rho_1|$. Since $\mathcal{E}$ is arbitrary, we choose the one which optimizes SD and this concludes the proof.

c. Error versus information Let us assume that Eve is powerful enough to control the natural noise. Without loss of generality, we assume that Eve's probes are in some arbitrary but fixed initial (tensor product) pure state, and that each probe is in a state $|E\rangle$. In the *collective attack*, the state $|E\rangle\otimes|b\rangle$ is subjected to Eve's unitary transformation $\mathcal{U}$ that changes the state $|b\rangle$ sent by Alice to the final global state

$$|0\rangle_z \mapsto |E_{0,0}^z\rangle|0\rangle_z + |E_{0,1}^z\rangle|1\rangle_z \equiv |\phi_0^z\rangle \quad (3a)$$
$$|1\rangle_z \mapsto |E_{1,0}^z\rangle|0\rangle_z + |E_{1,1}^z\rangle|1\rangle_z \equiv |\phi_1^z\rangle \quad (3b)$$

where the $|E_{i,j}^z\rangle$ are Eve's non normalized states. Implicitly, this description corresponds to restricting natural noise to follow the spirit of the collective attacks. It is reasonable to suspect that more general noise models would not be to Eve's advantage.

Bob's error probability in the z basis [measuring $|0\rangle_z$ if $|1\rangle_z$ was sent etc.] is $p_e^z = (1/2)[\langle E_{0,1}^z|E_{0,1}^z\rangle + \langle E_{1,0}^z|E_{1,0}^z\rangle]$. Alice can also use the alternate basis x, and then the transformation $\mathcal{U}$ can also be expressed in the x basis (replacing everywhere z by x) to yield $p_e^x = (1/2)[\langle E_{0,1}^x|E_{0,1}^x\rangle + \langle E_{1,0}^x|E_{1,0}^x\rangle]$. Since Alice uses both bases with the same probability, Bob's overall probability of error is $p_e = \frac{1}{2}(p_e^x+p_e^z)$ and so $p_e^x \leq 2p_e$ and $p_e^z \leq 2p_e$. Due to linearity of the transformation $\mathcal{U}$ we obtain $|E_{0,1}^x\rangle = \frac{1}{2}[(|E_{0,0}^z\rangle - |E_{1,1}^z\rangle) + (|E_{1,0}^z\rangle - |E_{0,1}^z\rangle)]$ and $|E_{1,0}^x\rangle = \frac{1}{2}[(|E_{0,0}^z\rangle - |E_{1,1}^z\rangle) - (|E_{1,0}^z\rangle - |E_{0,1}^z\rangle)]$. If we expand p_e^x in terms of the vectors in the z basis we get $p_e^x = (1/4)[\langle E_{0,0}^z - E_{1,1}^z|E_{0,0}^z - E_{1,1}^z\rangle + \langle E_{1,0}^z - E_{0,1}^z|E_{1,0}^z - E_{0,1}^z\rangle]$. Since $\mathcal{U}$ preserves inner products, the states $|\phi_0^z\rangle$ and $|\phi_1^z\rangle$ have norm 1. Therefore, $\langle E_{0,0}^z|E_{0,0}^z\rangle + \langle E_{0,1}^z|E_{0,1}^z\rangle = 1$ and $\langle E_{1,0}^z|E_{1,0}^z\rangle + \langle E_{1,1}^z|E_{1,1}^z\rangle = 1$, which we use to get

$$p_e^x = \frac{1}{2}[1 - Re\{\langle E_{0,0}^z|E_{1,1}^z\rangle + \langle E_{1,0}^z|E_{0,1}^z\rangle\}] \ . \quad (4)$$

Eve's view is obtained by tracing-out Bob from the states ϕ_b^z (if the z basis was used): $\rho_0^z(E) = |E_{0,0}^z\rangle\langle E_{0,0}^z| + |E_{0,1}^z\rangle\langle E_{0,1}^z|$ and $\rho_1^z(E) = |E_{1,0}^z\rangle\langle E_{1,0}^z| + |E_{1,1}^z\rangle\langle E_{1,1}^z|$.

Many other pure states (purifications) also yield the same reduced density matrices for Eve. In particular

$$|\psi_0^z\rangle = |E_{0,0}^z\rangle|0\rangle_z + |E_{0,1}^z\rangle|1\rangle_z \quad (5a)$$
$$|\psi_1^z\rangle = |E_{1,1}^z\rangle|0\rangle_z + |E_{1,0}^z\rangle|1\rangle_z \quad (5b)$$

will prove useful since the angle between them is zero if there is no disturbance. While these states have only virtual existence, they will be used in Theorem 1 to yield the desired bound, since Eve's states are the trace-out of these pure states. They live in some Hilbert space $\mathcal{H}^E \otimes \mathcal{H}^2$, with $\mathcal{H}^2$ two-dimensional Hilbert space. They are normalized, and consequently $|\langle\psi_0^z|\psi_1^z\rangle| = cos(2\alpha_z)$ for some angle $0 \le \alpha_z \le \pi/4$. Moreover, there is some phase angle θ such that $e^{i\theta}\langle\psi_0^z|\psi_1^z\rangle = |\langle\psi_0^z|\psi_1^z\rangle|$. Let $|\Psi_0^z\rangle = |\psi_0^z\rangle$ and $|\Psi_1^z\rangle = e^{i\theta}|\psi_1^z\rangle$. One can now find two (normalized) orthogonal states $|0_{\mathcal{H}}^z\rangle$ and $|1_{\mathcal{H}}^z\rangle$ (spanning a two dimensional subspace $\mathcal{H}$ of $\mathcal{H}^E\otimes\mathcal{H}^2$) such that $|\Psi_0^z\rangle = \cos(\alpha_z)|0_{\mathcal{H}}^z\rangle+\sin(\alpha_z)|1_{\mathcal{H}}^z\rangle$ and $|\Psi_1^z\rangle = \cos(\alpha_z)|0_{\mathcal{H}}^z\rangle-\sin(\alpha_z)|1_{\mathcal{H}}^z\rangle$. From $1-2\sin^2(\alpha_z) = \langle\Psi_0^z|\Psi_1^z\rangle = |\langle\psi_0^z|\psi_1^z\rangle| = |\langle E_{0,0}^z|E_{1,1}^z\rangle + \langle E_{1,0}^z|E_{0,1}^z\rangle|$ we deduce that $Re\{\langle E_{0,0}^z|E_{1,1}^z\rangle + \langle E_{1,0}^z|E_{0,1}^z\rangle\} \le 1-2\sin^2(\alpha_z)$. Using (4) we get that Eve's state is a partial trace of one of the pure states $|\Psi_b^z\rangle$ with angle satisfying $\sin(\alpha_z) \le (p_e^x)^{1/2}$.

Everything that has been said about $|\psi_b^z\rangle$ and p_e^x holds by symmetry for replacing the bases, yielding $\sin(\alpha_x) \le (p_e^z)^{1/2}$. Using $p_e^x \le 2p_e$ and $p_e^z \le 2p_e$, we obtain $\sin(\alpha_z) \le (2p_e)^{1/2}$ and $\sin(\alpha_x) \le (2p_e)^{1/2}$. In the sequel we will simply drop the indices x and z, taking as a convention that we are dealing with the actual basis that Alice and Bob agreed upon (and which become known to Eve only after she retransmitted the particle towards Bob).

d. The state in Eve's hands We now look at the n' remaining qubits after Alice and Bob discard those bits where the bases did not agree. Some bits are used to verify that $p_{test} \le p_{allowed}$, to be left with n-bit string $\mathbf{x}$. From the previous paragraph, we know that after retransmitting the i-th bit (namely, x_i) to Bob, the purification of Eve's state is $|\Psi_{x_i}\rangle = \cos(\alpha_i)\ |0\rangle_i + (-1)^{x_i}\sin(\alpha_i)\ |1\rangle_i$, where x_i is either 0 or 1 ($1 \le i \le n$) according to the bit which Alice sent to Bob, and $|b\rangle_i$ would be $|b_{\mathcal{H}_i}\rangle$ in the notations of the previous paragraph. Moreover $\sin(\alpha_i) \le (2p_i)^{1/2}$, where p_i is Bob's probability of error on the i-th bit (averaged over the four possible input states), which is completely determined by Eve's transformation. The global state of Eve's probes is, thanks to the properties of the trace, a partial trace of $|\Psi_{\mathbf{x}}\rangle$, the tensor product of the $|\Psi_{x_i}\rangle$.

To expand $|\Psi_{\mathbf{x}}\rangle$ we first need some notations. Boldface letters like $\mathbf{j}$, $\mathbf{x}$ are used to denote strings in $\{0,1\}^n$ that are interpreted as n-vectors on the binary field. Boldface letters are also used in kets, with the following understanding: if $\mathbf{j} = j_1 \dots j_n$ is concatenation of n bits then $|\mathbf{j}\rangle = |j_1\rangle_1 \dots |j_n\rangle_n$ where $|b\rangle_i$ are the basis vectors of the purifications of Eve's i'th qubit. The state $|\Psi_{\mathbf{x}}\rangle = \bigotimes_{i=1}^n[(\cos(\alpha_i)\ |0\rangle_i + (-1)^{x_i}\sin(\alpha_i)\ |1\rangle_i)]$ can be written as $|\Psi_{\mathbf{x}}\rangle = \sum_{\mathbf{j}\in\{0,1\}^n} d_{\mathbf{j}}(-1)^{\mathbf{x}\cdot\mathbf{j}}\ |\mathbf{j}\rangle$, where $d_{\mathbf{j}} = d_{j_1}\dots d_{j_n}$ with $d_{j_i} = \cos\alpha_i$ if $j_i = 0$ and $d_{j_i} = \sin\alpha_i$ if $j_i = 1$, and where $\mathbf{x}\cdot\mathbf{j}$ is by definition $\mathbf{x}\cdot\mathbf{j} = \sum_{i=0}^n x_i j_i\ mod\ 2$. For instance, $|\Psi_{01}\rangle = \cos\alpha_1\cos\alpha_2|00\rangle - \cos\alpha_1\sin\alpha_2|01\rangle + \sin\alpha_1\cos\alpha_2|10\rangle - \sin\alpha_1\sin\alpha_2|11\rangle$. We let $\mathbf{j}\oplus\mathbf{k}$ be the string obtained by adding $\mathbf{j}$ and $\mathbf{k}$ bit by bit with the understanding that $1\oplus1 = 0$. Then [using $(-1)^{\mathbf{x}\cdot\mathbf{j}}(-1)^{\mathbf{x}\cdot\mathbf{k}} = (-1)^{\mathbf{x}\cdot(\mathbf{j}\oplus\mathbf{k})}$], the lift-up of Eve's density matrix is

$$\widetilde{\rho_{\mathbf{x}}} = |\Psi_{\mathbf{x}}\rangle\langle\Psi_{\mathbf{x}}| = \sum_{\mathbf{j},\mathbf{k}\in\{0,1\}^n} d_{\mathbf{j}}d_{\mathbf{k}}(-1)^{\mathbf{x}\cdot(\mathbf{j}\oplus\mathbf{k})}\ |\mathbf{j}\rangle\langle\mathbf{k}|\ , \quad (6)$$

for any string $\mathbf{x}$ sent by Alice.

e. The parity bit In order to encode one key-bit b (0 or 1) using a substring of the n bits she sent, Alice proceeds as follows: she chooses some string $\mathbf{v}\in\{0,1\}^n$ to define the relevant (privacy amplification) substring, and announces it to Bob; Bob understands that the key-bit sent is $b = \mathbf{x}\cdot\mathbf{v}$, and can calculate the final bit b. Eve now knows $\mathbf{v}$ (but not $\mathbf{x}$) and has to guess $b = \mathbf{x}\cdot\mathbf{v}$. Only strings $\mathbf{x}$ such that $\mathbf{x}\cdot\mathbf{v} = b$ shall contribute to $\widetilde{\rho_b^{\mathbf{v}}} \equiv 2^{-n+1}\sum_{\{\mathbf{x}|\mathbf{x}\cdot\mathbf{v}=b\}}|\Psi_{\mathbf{x}}\rangle\langle\Psi_{\mathbf{x}}|$. To learn b Eve needs to distinguish between the two density matrices (in her hands) $\rho_b^{\mathbf{v}}$ for which $\widetilde{\rho_b^{\mathbf{v}}}$ are lift-ups. For convenience let us define $\Delta^{\mathbf{v}} \equiv \widetilde{\rho_0^{\mathbf{v}}} - \widetilde{\rho_1^{\mathbf{v}}}$, and in the following we evaluate the trace-norm of $\Delta^{\mathbf{v}}$.

Using $(-1)^b = (-1)^{\mathbf{x}\cdot\mathbf{v}}$ and (6) we get $\Delta^{\mathbf{v}} = (-1)^0\widetilde{\rho_0^{\mathbf{v}}} + (-1)^1\widetilde{\rho_1^{\mathbf{v}}} = 2^{-n+1}\sum_{\mathbf{j},\mathbf{k}} d_{\mathbf{j}}d_{\mathbf{k}}\sum_{\mathbf{x}}(-1)^{\mathbf{x}\cdot(\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v})}\ |\mathbf{j}\rangle\langle\mathbf{k}|$. We now simplify the preceding sum using a technique similar to the one of [5]. If $\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v} \ne \mathbf{0}$, there is some string $\mathbf{y}$ such that $(\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v})\cdot\mathbf{y} = 1$. If we let $\mathbf{x}' = \mathbf{x}\oplus\mathbf{y}$ then $(-1)^{\mathbf{x}'\cdot(\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v})} + (-1)^{\mathbf{x}\cdot(\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v})} = 0$ and since $\mathbf{x}\ne\mathbf{x}'$ (because $\mathbf{y}\ne\mathbf{0}$) *all* the coefficients of $|\mathbf{j}\rangle\langle\mathbf{k}|$ cancel in pairs. If $\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v} = \mathbf{0}$, then $(-1)^{\mathbf{x}\cdot(\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v})} = 1$ for all $\mathbf{x}$ and since there are 2^n such strings, we get

$$\Delta^{\mathbf{v}} = 2\sum_{\mathbf{i}\oplus\mathbf{j}=\mathbf{v}} d_{\mathbf{i}}d_{\mathbf{j}}\ |\mathbf{i}\rangle\langle\mathbf{j}| = 2\sum_{\mathbf{j}\in\{0,1\}^n} d_{\mathbf{j}}d_{\mathbf{j}\oplus\mathbf{v}}\ |\mathbf{j}\rangle\langle\mathbf{j}\oplus\mathbf{v}|. \quad (7)$$

If $\mathbf{i}\oplus\mathbf{j} = \mathbf{v}$ then clearly $\mathbf{j}\oplus\mathbf{i} = \mathbf{v}$. Therefore, $\Delta^{\mathbf{v}}$ is a sum of 2^{n-1} Hermitian matrices $d_{\mathbf{i}}d_{\mathbf{j}}|\mathbf{i}\rangle\langle\mathbf{j}| + d_{\mathbf{j}}d_{\mathbf{i}}|\mathbf{j}\rangle\langle\mathbf{i}| = d_{\mathbf{i}}d_{\mathbf{j}}[|\mathbf{i}\rangle\langle\mathbf{j}| + |\mathbf{j}\rangle\langle\mathbf{i}|]$. For each of them the Trace-norm is $2d_{\mathbf{i}}d_{\mathbf{j}} = d_{\mathbf{i}}d_{\mathbf{j}} + d_{\mathbf{j}}d_{\mathbf{i}}$. Using this result and the triangle inequality (which is satisfied by any norm) we obtain

$$Tr|\Delta^{\mathbf{v}}| \le 2\sum_{\mathbf{i}\oplus\mathbf{j}=\mathbf{v}} d_{\mathbf{i}}d_{\mathbf{j}} = 2\sum_{\mathbf{j}} d_{\mathbf{j}}d_{\mathbf{j}\oplus\mathbf{v}}\ . \quad (8)$$

If v_i, the i^{th} bit of $\mathbf{v}$ equals 1, then the product of the i^{th} factor of $d_{\mathbf{j}}$ by the i^{th} bit of $d_{\mathbf{v}\oplus\mathbf{j}}$ is $\cos\alpha_i\sin\alpha_i$, since either [$d_{j_i} = \cos\alpha_i$ and $d_{j_i\oplus v_i} = \sin\alpha_i$] or alternatively [$d_{j_i} = \sin\alpha_i$ and $d_{j_i\oplus v_i} = \cos\alpha_i$], since $j_i\oplus1 = \text{not}(j_i)$. The contribution of such terms is $(\sin 2\alpha_i)$ since the sum is over all $\mathbf{j}$ so the term $d_{\mathbf{j}}d_{\mathbf{j}\oplus\mathbf{v}}$ contributes twice. If v_i, the $i^{\rm th}$ bit of $\mathbf{v}$ equals 0, then the product of the $i^{\rm th}$ factor of $d_{\mathbf{j}}$ by the $i^{\rm th}$ bit of $d_{\mathbf{v}\oplus\mathbf{j}}$ is either $\cos^2\alpha_i$ or $\sin^2\alpha_i$. When summing over all $\mathbf{j}$, such terms sum up to yield 1.

As result the sum reduces to $\sum_{\mathbf{j}} d_{\mathbf{j}} d_{\mathbf{j}\oplus\mathbf{v}} = \prod_{v_i=1}\sin(2\alpha_i)\prod_{v_i=0} 1 = \prod_{v_i=1}\sin(2\alpha_i)$. If we look at $\mathbf{v}$ as the characteristic function of a set also denoted $\mathbf{v}$, one can write $i\in\mathbf{v}$ instead of $v_i=1$ and

$$Tr|\Delta^{\mathbf{v}}| \le 2\prod_{i\in\mathbf{v}}\sin(2\alpha_i)\ . \qquad (9)$$

Due to (2) and (1) we get $SD(\rho_0^{\mathbf{v}},\rho_1^{\mathbf{v}}) \le SD(\widetilde{\rho_0^{\mathbf{v}}},\widetilde{\rho_1^{\mathbf{v}}}) \le Tr|\Delta^{\mathbf{v}}| \le 2\prod_{i\in\mathbf{v}}\sin(2\alpha_i)$, if the error correction data is unknown to Eve.

f. Error correction For error correction, a number of linear constraints are imposed on the bits of $\mathbf{x}$. More precisely, Alice chooses a system $\mathbf{E}=\{\mathbf{v}_1\cdot\mathbf{x}=b_1, \mathbf{v}_2\cdot\mathbf{x}=b_2,\ldots,\mathbf{v}_r\cdot\mathbf{x}=b_r\}$ of r linear equations such that the $r+1$ strings $\mathbf{v}$, $\mathbf{v}_1$, …, $\mathbf{v}_r$ are linearly independent and each b_i is either 0 or 1. We write $\mathbf{E}(\mathbf{x})$ to mean that $\mathbf{x}$ satisfies the system (that is, $\mathbf{x}$ is a code word).

We can now define $\Delta^{\mathbf{E},\mathbf{v}}$ as in the previous section with $\widetilde{\rho_b^{\mathbf{E},\mathbf{v}}}$ an equal mixture of the states $|\Psi_{\mathbf{x}}\rangle\langle\Psi_{\mathbf{x}}|$ such that $\mathbf{x}$ satisfies the system $\{\mathbf{x}\cdot\mathbf{v} = b\}\cup\mathbf{E}$. This system has 2^{n-r-1} solutions and so $\Delta^{\mathbf{E},\mathbf{v}} = 2^{-n+r+1}\sum_{\mathbf{j},\mathbf{k}} d_{\mathbf{j}}d_{\mathbf{k}}\sum_{\mathbf{E}(\mathbf{x})}(-1)^{\mathbf{x}\cdot(\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v})}\ |\mathbf{j}\rangle\langle\mathbf{k}|$. As in the previous section, the expression for the Trace-norm can be simplified. For any $\mathbf{s}\in\{0,1\}^r$ let $\mathbf{v_s}$ denote the element $\sum_{i=1}^r s_i\mathbf{v}_i$. If $\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v} = \mathbf{v_s}$ and $\mathbf{x}$ is a solution of $\mathbf{E}$, then the exponent $\mathbf{x}\cdot(\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v})$ in the above expression for $\Delta^{\mathbf{E},\mathbf{v}}$ reduces to $(\sum_{i=1}^r s_i\mathbf{v}_i)\cdot\mathbf{x} = \sum_{i=1}^r s_i\ \mathbf{v}_i\cdot\mathbf{x} = \sum_{i=1}^r s_i b_i = \mathbf{s}\cdot\mathbf{b}$. This value is independent of $\mathbf{x}$ and so the coefficient of $|\mathbf{j}\rangle\langle\mathbf{k}| = |\mathbf{j}\rangle\langle\mathbf{k}\oplus\mathbf{v}\oplus\mathbf{v_s}|$ is $2d_{\mathbf{j}}d_{\mathbf{k}\oplus\mathbf{v}\oplus\mathbf{v_s}}(-1)^{\mathbf{s}\cdot\mathbf{b}}$ where $\mathbf{b}$ is the string $(b_i)_{1\le i\le r}$ of the parity bits in the equations of $\mathbf{E}$. If $\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v}$ is not in the span of $\{\mathbf{v}_1,\ldots,\mathbf{v}_r\}$ then there is a solution $\mathbf{y}$ to the system $\{(\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v})\cdot\mathbf{y}=1, \mathbf{v}_1\cdot\mathbf{y}=0,\ldots,\mathbf{v}_r\cdot\mathbf{y}=0\}$. For any $\mathbf{x}$ solution of $\mathbf{E}$, let $\mathbf{x}'$ denote $\mathbf{x}\oplus\mathbf{y}$. Clearly $\mathbf{x}'$ is also a solution of $\mathbf{E}$ and $(-1)^{\mathbf{x}\cdot(\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v})}+(-1)^{\mathbf{x}'\cdot(\mathbf{j}\oplus\mathbf{k}\oplus\mathbf{v})}=0$, and consequently the coefficient of $|\mathbf{j}\rangle\langle\mathbf{k}|$ is 0. Therefore

$$\Delta^{\mathbf{E},\mathbf{v}} = 2\sum_{\mathbf{j},\mathbf{s}} d_{\mathbf{j}}d_{\mathbf{j}\oplus\mathbf{v}\oplus\mathbf{v}_s}(-1)^{\mathbf{s}\cdot\mathbf{b}}\ |\mathbf{j}\rangle\langle\mathbf{j}\oplus\mathbf{v}\oplus\mathbf{v_s}|\ , \qquad (10)$$

generalizing Eq. (7) to contain the error correction data. Consequently $\Delta^{\mathbf{E},\mathbf{v}} = \sum_{\mathbf{s}\in\{0,1\}^r}(-1)^{\mathbf{s}\cdot\mathbf{b}}(\widetilde{\rho_0^{\mathbf{v}\oplus\mathbf{v_s}}} - \widetilde{\rho_1^{\mathbf{v}\oplus\mathbf{v_s}}})$. As before we define $\Delta^{\mathbf{v}\oplus\mathbf{v_s}}$ for the terms in the parenthesis, and these terms are given by Eq. (7) [and their Trace-norm is given by Eqs. (8) and (9)] once $\mathbf{v}$ there is replaced by $\mathbf{v}\oplus\mathbf{v_s}$. This gives $\Delta^{\mathbf{E},\mathbf{v}} = \sum_{\mathbf{s}\in\{0,1\}^r}(-1)^{\mathbf{s}\cdot\mathbf{b}}(\Delta^{\mathbf{v}\oplus\mathbf{v_s}})$, and due to the triangle inequality $Tr|\Delta^{\mathbf{E},\mathbf{v}}| \le \sum_{\mathbf{s}\in\{0,1\}^r} Tr|\Delta^{\mathbf{v}\oplus\mathbf{v_s}}|$. Using the set notation and Eq. (9) we finally get

$$Tr|\Delta^{\mathbf{E},\mathbf{v}}| \le 2\sum_{\mathbf{s}\in\{0,1\}^r}\ \prod_{i\in(\mathbf{v}\oplus\mathbf{v_s})}\sin(2\alpha_i)\ . \qquad (11)$$

Due to (2) and (1), we get $SD(\rho_0^{\mathbf{E},\mathbf{v}},\rho_1^{\mathbf{E},\mathbf{v}}) \le SD(\widetilde{\rho_0^{\mathbf{E},\mathbf{v}}},\widetilde{\rho_1^{\mathbf{E},\mathbf{v}}}) \le Tr|\Delta^{\mathbf{E},\mathbf{v}}| \le 2\sum_{\mathbf{s}\in\{0,1\}^r}\prod_{i\in(\mathbf{v}\oplus\mathbf{v_s})}\sin(2\alpha_i)$ when the error correction data is known to Eve. Using $\sin(2\alpha_i) \le 2\sin\alpha_i \le (8p_i)^{1/2}$ we finally get $SD(\rho_0^{\mathbf{E},\mathbf{v}},\rho_1^{\mathbf{E},\mathbf{v}}) \le 2\sum_{\mathbf{s}\in\{0,1\}^r}[\prod_{i\in(\mathbf{v}\oplus\mathbf{v_s})}(8p_i)]^{1/2}$.

Let the "Hamming weight" $\hat{n}_{\mathbf{s}}$ (for each $\mathbf{s}$) be the number of one's in $\mathbf{v}\oplus\mathbf{v_s}$ [the number of factors in the product $\prod_{i\in(\mathbf{v}\oplus\mathbf{v_s})}$]. Also let $p_{\mathbf{s}} = [\sum_{i\in(\mathbf{v}\oplus\mathbf{v_s})}p_i]/\hat{n}_{\mathbf{s}}$ be the average error in any relevant subset $\mathbf{s}$. The geometrical mean of the p_i contributing to $p_{\mathbf{s}}$ is always less than their arithmetical mean so $[\prod_{i\in(\mathbf{v}\oplus\mathbf{v_s})}(8p_i)]^{1/2} \le [8p_{\mathbf{s}}]^{\hat{n}_{\mathbf{s}}/2}$, and thus $SD(\rho_0^{\mathbf{E},\mathbf{v}},\rho_1^{\mathbf{E},\mathbf{v}}) \le 2\sum_{\mathbf{s}\in\{0,1\}^r}[8p_{\mathbf{s}}]^{\hat{n}_{\mathbf{s}}/2}$.

Given that the test is passed $p_{test}\le p_{allowed}$ statistical analysis promise us that each of the $p_{\mathbf{s}}$ is bounded. Combining two laws of large numbers of Hoeffding [9], Theorem 2 (sums of independent random variables) and its extension in section 6 (sampling from a finite population), we are promised that $p_{n'}$, the average p_i of all n' relevant bits satisfies Prob$[p_{n'} > p_{test}+2\delta] \le 2e^{-2n_{test}\delta^2}$ (since the tested bits are picked at random). Once $p_{n'}$ is bounded we can bound $p_{\mathbf{s}}$ as follows: let n' be even [throw one bit if needed (before choosing the bits for the test)], and let Alice and Bob use $n_{test} = n'/2$ bits for the test. We then have $p_{\mathbf{s}} \le (n'/\hat{n}_{\mathbf{s}})p_{n'}$. Thus, Eve's information is generously bounded by $2\sum_{\mathbf{s}\in\{0,1\}^r}[(8n'/\hat{n}_{\mathbf{s}})(p_{test}+2\delta)]^{(\hat{n}_{\mathbf{s}}/2)}$, except with a probability of $p_{luck} = 2e^{-2n_{test}\delta^2}$. Recall that $n_{test} = n = n'/2$. Assuming (generously again) that in such a case of having *luck* Eve's information is maximal (one bit) her total information is bounded by $2\sum_{\mathbf{s}\in\{0,1\}^r}[(16n/\hat{n}_{\mathbf{s}})(p_{test}+2\delta)]^{(\hat{n}_{\mathbf{s}}/2)} + 2e^{-2n\delta^2}$, for any δ. Let $\alpha n = \hat{n} = \min_{\mathbf{s}}\hat{n}_{\mathbf{s}}$. Then, $SD(\rho_0^{\mathbf{E},\mathbf{v}},\rho_1^{\mathbf{E},\mathbf{v}}) \le 2^{r+1}[(16/\alpha)(p_{test}+2\delta)]^{\alpha n/2} + 2e^{-2n\delta^2}$. Entering into coding theory is beyond our aim in this letter and is left for the full paper: for error rates below 2%, many codes allow us to choose the parameters n, r, α and δ such that Eve's information is negligible [e.g., 2^{-100} of a bit].

[1] C. H. Bennett and G. Brassard, in *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing* (IEEE, Bangalore, India, 1984), pp. 175–179.

[2] E. Biham and T. Mor, *Phys. Rev. Lett.* **79**, 4034 (1997).

[3] D. Mayers, *Proc. of Crypto 96, LNCS* **1109**, 343 (1996).

[4] C. H. Bennett, T. Mor, and J. A. Smolin, *Phys. Rev. A***54**, 2675 (1996).

[5] E. Biham and T. Mor, *Phys. Rev. Lett.* **78**, 2256 (1997).

[6] A. Peres, "Quantum Theory: Concepts and Methods", Kluwer Academic Publishers, Dordrecht (1993).

[7] C. A. Fuchs and J. van der Graaf, quant-ph 9712042.

[8] C. A. Fuchs, Ph.D. Thesis; quant-ph 9601020.

[9] W. Hoeffding, *Amer. Stat.* **58**, 13 (1963).