Journal of Cryptology manuscript No.
(will be inserted by the editor)

A Proof of the Security of Quantum Key Distribution*

Eli Biham[1], **Michel Boyer**[2], **P. Oscar Boykin**[3], **Tal Mor**[4], **Vwani Roychowdhury**[5]

[1] Computer Science Department, Technion,
Haifa 32000, Israel.
e-mail: biham@cs.Technion.AC.IL
[2] DIRO, Université de Montréal,
CP 6128, Succ. Centre-Ville,
Montréal, H3C 3J7, Canada.
e-mail: boyer@IRO.UMontreal.CA
[3] Dept. of Electrical Engineering, UCLA,
Los Angeles, CA 90095-1594, USA.
e-mail: boykin@ee.ucla.edu
[4] Computer Science Department, Technion,
Haifa 32000, Israel.
e-mail: talmo@cs.technion.ac.il
[5] Dept. of Electrical Engineering, UCLA,
Los Angeles, CA 90095-1594, USA.
e-mail: vwani@ee.ucla.edu

July 22, 2013

Abstract We prove the security of theoretical quantum key distribution against the most general attacks which can be performed on the channel, by an eavesdropper who has unlimited computation abilities, and the full power allowed by the rules of classical and quantum physics. A key created that way can then be used to transmit secure messages such that their security is also unaffected in the future.

Key words Quantum key distribution, Quantum information, Information vs. disturbance, Quantum Security, BB84.

* A shortened version of this paper is published in *STOC'2000*, and a preliminary full version appears on the Los-Alamos archive http://arxiv.org/abs/quant-ph/9912053 [6].

1 Introduction

Quantum key distribution [3,2] uses the power of quantum mechanics to suggest the distribution of a key that is secure against an adversary with unlimited computation power. Such a task is beyond the ability of classical information processing; thus, it is the main success of the original idea of Wiesner [34] who suggested using quantum mechanics to perform cryptographic tasks. The extra power gained by the use of quantum bits (quantum two-level systems, "qubits") is due to the fact that the state of such a system cannot be cloned. [Of course, one could use higher level quantum systems as well.] On the other hand, the security of conventional key distribution is based on the (unproven) existence of various one-way functions, and mainly on the difficulty of factoring large numbers, a problem which is assumed to be difficult for a classical computer, and is proven to be easy for a hypothetical quantum computer [32].

The quantum key distribution (QKD) scheme considered in our work is the protocol of Bennett and Brassard [3], known as the BB84 protocol. The legitimate users of this (actually, of any) QKD protocol are conventionally called Alice (the sender) and Bob (the receiver). Their aim is to create and share a secret key.

There are several classes of attacks (see for instance [8,7]) on quantum key distribution that can be performed by an eavesdropper having full control of the channel. The simplest ones are known as individual-particle attacks [17] in which the transmitted qubits are attacked separately, so that the eavesdropper can be left with some optimal classical information about each transmitted quantum bit. The eavesdropper can use this classical information in order to learn some information about the final secret key. In contrast, in the most general attack called the "joint attack", all transmitted quantum particles are attacked together, and the eavesdropper's goal is to learn as much information as possible about the final key, rather than about each transmitted qubit. A special class of the joint attack, the "collective attack" [8] was shown to provide more information to the eavesdropper than an individual-particle attack [5]. We further explain the differences between the individual-particle attacks, the collective attacks, and the most general attacks (the joint attacks) in Subsection 2.2, when we describe the two steps of Eve's attack. Various proofs of security were previously obtained against collective attacks [8,9, 7,29] (which is a most important subclass of the joint attack), and we continue this line of research here to prove the ultimate security of QKD, against any attack (under the conventional assumptions of theoretical QKD, as explained below). Note that the eavesdropper is assumed to have unlimited technology (e.g., unlimited computing power, a quantum memory, a quantum computer), while the legitimate users use practical tools (or more precisely, simplifications of practical tools). Such assumptions are required since the aim of the invention of quantum key distribution is to obtain a *practical* key distribution scheme, which is proven secure against *any* attack, even one which is far from being practical with current technology.

To prove security against such a super-strong eavesdropper, conventionally called Eve, we develop some important technical tools and we reach some novel results: We obtain a new *information versus disturbance* result, where the power of quantum information theory is manifested in an intuitive and clear way. We show

explicitly how the randomness of the choice of bases, and the randomness of the choice of test-bits provides the desired security of QKD. We adopt and generalize sophisticated tools invented in [7]: "Purifications" which simplify Eve's states, a bound on accessible information (using Trace-Norm-Difference of density matrices) which avoids any complicated optimization of Eve's possible measurements, and a connection between Eve's accessible information and the error-rate she induces. We add some more simplifications (which were not required in the analysis of collective attacks in [7]): a reduction to a scheme in which all qubits are used by Alice and Bob, and a symmetrization of Eve's attack.

This paper complements the result of Bennett, Mor, and Smolin [5]: That paper shows that individual particle attacks are strictly weaker (less informative to the eavesdropper) than joint attacks[1], and the current paper shows that security can still be obtained even when the eavesdropper applies the strongest joint attacks. The current paper also complements the work of Bennett, Brassard, Crépeau, and Maurer [4]: That paper shows that privacy amplification provides security when the eavesdropper is restricted to perform only individual particle attacks, and the current paper shows that privacy amplification provides security when the eavesdropper is not restricted, and can apply any joint attack on the particles.

Two other security proofs [26,27], and [24,23] were reported just prior to ours [6]. The security result of Lo and Chau [24] [note that some of the details were completed or improved in [23]] uses novel techniques and is very important, but it is somewhat limited: The QKD protocol which is analyzed in [24] requires that the legitimate users have quantum memories and fault tolerant quantum computers, technologies which are not yet available to the legitimate users, and are not expected within the next ten or twenty years, while the QKD protocol which is analyzed here, the BB84 protocol, is now demonstrated with some partial success in many labs (see many references in Gisin's reviews [36,20]). Some of the ideas used in [24] appeared earlier, [e.g., the quantum privacy amplification [16], and the quantum repeaters [28,29], and the use of fault tolerance quantum error correction for performing quantum privacy amplification [28,29] but Lo and Chau succeeded in using them to yield a novel proof of security from classical random sampling techniques. The security result of Mayers [26,27] is similar to ours in the sense that it proves the security of a much more realistic protocol against an unrestricted eavesdropper, and provides explicit bounds on the eavesdropper's information. It continues earlier works such as a solution to the error-free case [35].

Our proof is different from Mayers, was derived independently, and may shed more light on the subject. We analyze the density matrices which are available to the eavesdropper and we prove that it is extremely rare that these density matrices carry non-negligible information about the secret key, and at the same time, Alice and Bob agree to form a secret key. In other words, it is extremely rare that Alice and Bob agree to form a secret key about which these density matrices reveal non-negligible information.

[1] Many of the leading researchers in experimental quantum cryptography are unfamiliar with this work of Bennett Mor and Smolin, and still wrongly state that individual particle attacks could be as strong as collective/joint attacks.

Two additional proofs were announced more recently [33,1]. Shor and Preskill's proof [33] proposes a way to extend Lo and Chau's proof so that it becomes applicable to a more practical protocol, hence bypasses the main limitation of Lo and Chau's proof. A written draft of the proof of Ben-Or is expected in the near future [1].

We base our work on standard assumptions of QKD: 1) We assume the correctness of quantum theory and its relativistic generalizations, as these were verified with incredible accuracy in many experiments. 2) Alice and Bob share an unjammable classical channel. This assumption is usually replaced by the demand that the classical channel is "unforgeable"; an unforgeable channel can be modified by an eavesdropper but Alice and Bob will notice that, with probability exponentially close to 1. If Alice and Bob share a much shorter secret key to be used for authenticating a standard classical channel, they can indeed obtain an unforgeable channel (hence the protocol is then a quantum key expansion protocol, although everyone still call it QKD). 3) Eve cannot attack Alice's and Bob's laboratories. She can only attack the quantum channel and listen to all transmissions on the classical channel. 4) Alice sends quantum bits, i.e. two level systems. This assumption cannot be fully met in any experimental scenario, but can only be approximated.

We prove, under those assumptions, the security of the BB84 protocol [3], against any attack allowed by the rules of quantum physics. We prove security for instances in which the error rate in the transmission from Alice to Bob is up to 7.56%.

Although experimental QKD is very common (see for instance Gisin's reviews [36,20]), at the present time no experimental system whatsoever is proven unconditionally secure. Some security analyses which take into account corrections due to having more than two levels in the quantum systems have been provided ([12,11]), but research in this area is still in its early stages. In fact, many experimental systems are totally insecure due to the photon-number-splitting attack [11].

Quantum cryptography [34,3] is described in several publications, some of which also introduce the notations in a more expository way. Readers unfamiliar with the basics of quantum information processing are referred to any recently published textbook on the subject, e.g., [30,21]. Here we focus on QKD [3,2] and specifically on the BB84 protocol [3].

In BB84 we let

$$
\begin{aligned}
&|0\rangle_0 \equiv |0\rangle;\\
&|1\rangle_0 \equiv |1\rangle;\\
&|0\rangle_1 \equiv \tfrac{1}{\sqrt{2}}(|0\rangle + |1\rangle);\\
&|1\rangle_1 \equiv \tfrac{1}{\sqrt{2}}(|0\rangle - |1\rangle),
\end{aligned}
$$

define four states, such that the first two are orthogonal in one basis (known as the computation basis, or the "z" basis), and the other two are orthogonal in another basis (the "x" basis). [Using these "spin" notations the bases are $|\rangle_0 \equiv |\rangle_z$, and $|\rangle_1 \equiv |\rangle x$.] Note that the two bases are conjugate, namely, applying a measurement in one basis on a state belonging to the other basis gives a fully random

outcome. In the BB84 protocol Alice and Bob use these four possible quantum states. Therefore, we shall refer to these states as the BB84 states.

The quantum part of the communication in the BB84 protocol contains one step — Alice sends Bob a string of qubits, each in one of the four BB84 states (chosen randomly by Alice). To simplify the analysis, we assume all qubits are sent to Eve, and then Eve sends all qubits to Bob[2].

The rest of the protocol involves sending classical communication via the unjammable channel. First Alice sends Bob the basis used for each photon. By comparing bases after Alice sends such a state for each qubit and Bob receives the qubit, a common key can be created in instances when Alice and Bob used the same basis. Comparing the bases must be performed after Bob receives the qubits, so that the eavesdropper cannot benefit from having this knowledge while still holding the qubits. The common key obtained from the above steps is known as the "sifted key". A final key is then obtained from the sifted key, after performing several more steps: testing the error rate on some test bits, chosen at random; throwing away these test bits, while Alice and Bob can now have some good estimation of the error-rate on the remaining shared bits (called information bits); correcting errors on these information bits, and amplifying the privacy, by creating a shorter final key.

Alternatively, if Bob has a memory where he can keep his qubits unchanged after receiving them (we call such a memory "a quantum memory"), a simpler protocol for obtaining a sifted key is obtained: Bob waits with the received qubits till he learns the basis, and then measures in the right bases. The sifted key is twice as big in this case or the initial string of qubits can be shortened to half, if the final length of the sifted key is to remain the same.

We prove here the security of that simplified protocol in which only the bits relevant for the sifted key are discussed; we call it the "used-bits-BB84". We formally describe the used-bits protocol (in detail) in the next section. The proof of the security of the original BB84 protocol (in which Bob does not have a quantum memory) easily follows due to a simple reduction, as we show in Appendix A.

In the most general attack on the channel, Eve attacks the qubits in two steps. First, she lets all qubits pass through a device that weakly probes their state via a quantum unitary transformation. Then, after receiving all the classical data, she measures the probe. Eve's goal is to learn as much information as possible about the final key without causing Alice and Bob to abort the protocol due to a failure of the test. We consider here any attack chosen by Eve, described by these two steps, and we prove security against any such attack. We formally explain Eve's most general attack in the next section.

The issue of the security criteria is non-trivial since one obvious security criterion, namely that *"Eve's information given that the test passed, is negligible"*, does not work; this criterion cannot be proven, as a counter example exists[3] An-

[2] In case Eve can only hold each qubit for a short time and must release it before she gets the next, she is less powerful, so our proof of security covers that case as well.

[3] Namely, there is an attack such that Eve's information is large even when the test is passed (although in such cases the test is passed very rarely); Such attacks are studied in Section 2.3.

other natural security criterion saying that *"either Eve's average information is negligible or the probability that the test is passed is negligible"*, also does not work (for a similar reason). The criterion that we shall prove in this work says that *"the event where the test is passed AND Eve's information is not negligible, is extremely rare"*. This security criterion is formally presented in the next section.

We will moreover show that the final key is reliable: the keys distilled by Alice and Bob (after error correction and privacy amplification) are identical except for some exponentially small probability.

Section 2 provides a formal description of the used-bits-BB84 protocol, the most general attacks, and the security and reliability criteria. The rest of the paper contains three main steps leading to the desired proof of security: In Section 3 we reduce the problem of proving security to a simpler problem of optimizing over all attacks symmetric to the bit values 0 and 1. In Section 4 we analyze the information bits in the bases actually used by Alice and Bob, and we prove our main *information versus disturbance* theorem for symmetric attacks; the eavesdropper information about the final key is bounded by the probability of errors induced in the *other* bases (namely, errors induced if the other bases were used by Alice and Bob). We then obtain in Section 5 an exponentially small bound on Eve's information, proving that the security criterion (2.1) described in Section 2 is always satisfied in QKD, provided a good code for error correction and privacy amplification is used. Finally, we analyze a specific code, the random linear code, and we prove security for instances in which the error rate in the transmission from Alice to Bob is up to 7.56%. We also analyze the conditions under which this code can provide data relevant to experimentalists who choose some parameters (such as the number of photons used for the communication) and would like to obtain bounds on Eve's information, on the probability of errors in the final key, and on the resulting bit-rate of the protocol. Such explicit bounds are presented here for any error rate equal to or smaller than 5.50%. We summarize these results in Table 5.1.

We conclude the paper by summarizing the tools used here, and by suggesting that some of them could be relevant for other proofs as well. Various technical details and proofs of several lemmas are provided in the appendices.

2 Notations, the Protocol, Eve's Attack, the Security Criteria, and the Main Results

2.1 The used-bits BB84 protocol

Let us describe the used-bits protocol in detail, splitting it into *creating the sifted key* and *creating the final key from the sifted key*. This simplified protocol assumes that Bob has a quantum memory.

I. Creating the sifted key:

1. Alice and Bob choose a large integer $n \gg 1$. The protocol uses $2n$ bits.
2. Alice randomly selects two $2n$-bit strings, b and i and sends Bob, via a quantum communication channel, the string of $2n$ qubits
$$|i\rangle_b = |i_1\rangle_{b_1}|i_2\rangle_{b_2}\ldots|i_{2n}\rangle_{b_{2n}}$$
3. Bob tells Alice when he receives the qubits. [If he received less than $2n$ qubits he adds any missing qubit, but in an arbitrary state. If he received more than $2n$ qubits he ignores any extra qubit. E.g., if qubit number 17 did not arrive Bob will add it (by choosing its value and basis at random), and if two qubits arrived instead of one when Bob expects qubit number 17, then Bob will ignore one of them. Obviously, such cases will contribute to the error rate, $p_{\rm test}$.]
4. Alice publishes the bases she used, b; this step should be performed only after Bob received all the qubits.
Bob measures the qubits in Alice's bases to obtain a $2n$-bit string j.
We shall refer to the resulting $2n$-bit string as the sifted key, and it would be the same for Alice and Bob, i.e. $j = i$, if natural errors and eavesdropping did not exist.

II. Creating the final key from the sifted key:

1. Alice chooses at random a $2n$-bit string s which has exactly n zeroes and n ones. There are $\binom{2n}{n}$ such strings to choose from.
2. From the $2n$ bits, Alice selects a subset of n bits, determined by the zeros in s, to be the test bits. Alice publishes the string s, along with the values of the test bits (given by an n-bit string i_T). The values of Bob's bits on the test bits are given by j_T.
The other n bits are the information bits (given by an n-bit string i_I). They are used for deriving a final key via error correction codes (ECC) and privacy amplification (PA) techniques.
Later on, Alice will send the ECC and PA information to Bob, hence Bob needs to correct his errors using the ECC data, and to obtain a final secret key equal to Alice's using the PA data.
3. Bob verifies that the error rate $p_{\text{test}} = |i_T \oplus j_T|/n$ in the test bits is lower than some pre-agreed allowed error-rate p_{allowed}, and aborts the protocol if the error rate is larger. The maximal possible allowed error-rate is found in Section 5.4.
4. Bob also publishes the values of his test bits (j_T). This is not crucial for the protocol, but it is done to simplify the proof.

5. Alice selects an (n, k, d) linear error correcting code $\mathcal{C}$ with 2^k code words of n bits and a minimal Hamming distance d between any two words, along with the ECC parities on the information bits. The strategy is that Alice announces an $r \times n$ parity check matrix $P_{\mathcal{C}}$ of $\mathcal{C}$ by announcing its $r = n - k$ rows of n bits $v_1, \ldots, v_r$. This means that the code contains any i such that $i \cdot v_q = 0$ for any $q \in \{1 \ldots r\}$. Formally speaking, $\mathcal{C} = \{i \in \{0,1\}^n \mid iP_{\mathcal{C}}^\top = 0\}$, with $P_{\mathcal{C}}^\top$ the transpose of $P_{\mathcal{C}}$. Alice then also announces the r-bit string $\xi = i_I P_{\mathcal{C}}^\top$ whose bits are the parities of her (random) information string i_I with respect to the parity check matrix (so the q-th bit ξ_q of ξ is $\xi_q = i_I \cdot v_q$ for all $1 \leq q \leq r$). Bob doesn't announce anything.
   We now explain how the code $\mathcal{C}$ is chosen. The condition on $\mathcal{C}$ is that it corrects $t \geq (p_{\text{allowed}} + \epsilon_{\text{rel}})n$ errors, for some positive (pre-determined) reliability parameter ϵ_{rel}. If an ECC has Hamming weight $d \geq 2t+1$ it will always correct t errors, and thus the condition $d \geq 2(p_{\text{allowed}} + \epsilon_{\text{rel}})n + 1$ is sufficient. Meaning that, any code satisfying this criterion is good for Alice and Bob.
   For Random Linear Codes a better bound exists, and $d \geq (p_{\text{allowed}} + \epsilon_{\text{rel}})n + 1$ is also sufficient as noted in [27]; It is not promised that such a code always corrects t errors, but it is promised that it corrects t errors with probability as close to 1 as we want (provided we choose a sufficiently large n).
6. Bob performs the correction on his information bits j_I as follows: he finds the n-bit string j^{Bob} such that $j^{\text{Bob}} P_{\mathcal{C}}^\top = \xi$ and such that the Hamming distance between j^{Bob} and j_I is minimal. As long as there are at most t errors in j_I (i.e. $|j_I \oplus i_I| \leq t$) the obtained string is unique, and Bob finds the right string, namely $j^{\text{Bob}} = i_I$. Note that we are not concerned here with the efficiency of finding j^{Bob}, but a practical protocol ought to be efficient as well.
7. Alice selects a privacy amplification function ($\mathcal{PA}$) and publishes it. The PA strategy is to publish m strings, of length n each. These *privacy-amplification parity-check strings* $v_{r+1}, \ldots, v_{r+m}$ shall be used as the rows of an $m \times n$ parity matrix $P_{\mathcal{PA}}$ so that the final secret key is $a \equiv i_I P_{\mathcal{PA}}^\top$, with $a_t = i_I \cdot v_{r+t+1}$ (for $0 \leq t \leq m-1$)). This strategy is similar to error correction except that the m-bit string (namely, the final key) $i_I P_{\mathcal{PA}}^\top$ is kept secret.
   The PA strings must be chosen such that the minimal distance $\hat{v}$ between any PA parity string v and any string in the span of their union with the parity-check-strings of the ECC (the dual to the code) is at least $\hat{v} \geq 2(p_{\text{allowed}} + \epsilon_{\text{sec}})\, n$. [This is important for preventing Eve from learning much from the error-correcting procedure, and furthermore from learning something about the correlations between the bits of the final key.] Note that, by definition, the minimal distance of the space spanned by the ECC and PA strings $v_1, \ldots, v_{r+m}$, which we shall denote $d^\perp$, is less than the distance $\hat{v}$; hence if we demand $d^\perp \geq 2(p_{\text{allowed}} + \epsilon_{\text{sec}})\, n$, the above desired criterion, $\hat{v} \geq 2(p_{\text{allowed}} + \epsilon_{\text{sec}})\, n$, is automatically satisfied (due to $\hat{v} \geq d^\perp$).
8. Bob calculates $a = i_I P_{\mathcal{PA}}^\top$ to finally get the key.

2.2 Eavesdropping

In the most general attack on the channel, Eve attacks the qubits in two steps. First she lets all qubits pass through a device that weakly probes their state via a quantum unitary transformation. Then, after receiving all the classical data, she measures the probe. Note that Eve can gain nothing by measuring the probe earlier, or by measuring the qubits while passing through her. Any such measurement can also be performed by attaching a probe, applying a unitary transformation, and measuring the probe (or part of it) at a later stage. Since there is no gain in performing a measurement before learning all the classical information that is transmitted throughout the protocol, the optimal attack (WLoG) is to perform all measurements after receiving all classical information. Furthermore, Eve gains nothing by sending Bob a state that is not a $2n$ qubit state, so without loss of generality, we assume she sends exactly $2n$ photons: If Eve sends less than $2n$ qubits, Bob will add the missing qubits in an arbitrary state (see item I-3 in the protocol), so Eve could have done it herself. If Eve sends more than $2n$ qubits, Bob ignores the extra qubits, and again Eve could have done it herself. [An important remark though: the allowed error rate in these cases must still be limited as described in this work. However, in real applications the natural losses of qubits become very high due to transmission across long distances. If one does not wish to limit the distance too much, and wishes to have security even if losses are much higher than p_{allowed}, then this is still possible. See a brief explanation in Appendix A.]

It is important to enable an analysis of Eve's most general attack. Thus we formally split Eve's attack into her transformation U and her measurement $\mathcal{E}$.

Eve's transformation, U: Eve attacks the qubits while they are in the channel between Alice and Bob. Eve can perform any attack allowed by the laws of physics, the most general one being any unitary transformation U on Alice's qubits and Eve's probe (an ancilla initially in a state $|0\rangle_{\mathrm{E}}$).
We are generous to Eve, allowing her to attack all the qubits together (in practice, she usually needs to release the preceding qubit towards Bob before she has access to the next one).
Without loss of generality we assume that all the noise on the qubits is caused by Eve's transformation.
A remark: In individual-particle attacks and in collective attacks Eve's transformation is restricted so that each transmitted qubit is attacked using a separate, unentangled probe, so that the analysis of U is much simplified. In collective attacks the next step is as general as it is for the joint attacks (so that Eve can measure all probes together). In contrast, in individual-particle attacks Eve is only allowed to measure each probe separately from the others.
Eve's measurement, $\mathcal{E}$: Eve keeps the probe in a quantum memory, meaning that she keeps its state unchanged. After Eve receives *all* the classical information from Alice and Bob, including the bases of all bits b, the choice of test bits s, the test bits values, i_T and j_T, the ECC, the ECC parities ξ, and the PA, she tries to guess the final key using her best strategy of measurement. The measurement can be done by adding a second ancilla, and performing a standard projection measurement on Eve's probe and the ancilla. This measure-

ment is alternatively described (without the need for this second ancilla) by the so called "generalized measurement" or "POVM", $\mathcal{E}$, which is a set of positive operators $\mathcal{E}_e$ such that $\sum_e \mathcal{E}_e = 1$. When the measurement is applied onto a density matrix ρ the outcome e is obtained with probability $p(e) = \mathrm{Tr}(\rho\mathcal{E}_e)$. We fix [4] the set of possible outcomes e, so that it is the same for all the POVMs used by Eve after she learns i_T, j_T, b, s and ξ.
For more information about POVMs and their connection to standard projection measurements in an enlarged Hilbert space, see [31,30].

Eve's goal is to learn as much information as possible about the final key without causing Alice and Bob to abort the protocol due to a failure of the test. The task of finding Eve's optimal operation in these two steps is very difficult. Luckily, to prove security that task *need not* be solved, and it is enough to find bounds on Eve's optimal information (via any operation she could have done): In order to analyze her optimal transformation we find bounds for *any* transformation U she could perform, and in order to analyze her optimal measurement we find bounds for *any* measurement $\mathcal{E}$ she could perform.

2.3 What does security mean?

We consider here any attack chosen by Eve, described by U and $\mathcal{E}$. Let us explain what we mean by saying that security shall be proven.

As we already mentioned in the introduction, the issue of the security criteria is non-trivial. One obvious security criterion, namely that *"Eve's information given that the test passed, is negligible"*, can be proven wrong (for QKD), and furthermore, another natural security criterion saying that *"either Eve's average information is negligible or the probability that the test is passed is negligible"*, also does not work.

The criterion that we shall prove here says that *"the event where the test is passed AND Eve's information is not negligible, is extremely rare"*.

To be more precise we formally present now these security criteria. We first provide some relevant information-theoretic notations (for some more basic definitions see Appendix B.1). Let $\mathbf{A}$ be the random variable whose values are Alice's final key, $a = i_I P_{\mathcal{PA}}^{\top}$, and $\mathbf{E}$ be a random variable whose values e are the outputs of Eve's measurement $\mathcal{E}$. Note that e are outcomes of a measurement that itself is a function of all the classical data provided to Eve, the ECC and PA (that can be given to Eve in advance), and also i_T, j_T, b, s, and ξ. However, we usually consider *any* attack, therefore for any fixed parameters of the attack, $\{U, \mathcal{E}\}$, the resulting e are regular classical values of a regular classical random variable $\mathcal{E}$, so all standard rules of classical information theory (as described in Appendix B.1) apply to them. Note that our proof never needs to assume that the ECC data $P_{\mathcal{C}}^{\top}$ and the PA data $P_{\mathcal{PA}}^{\top}$ are random, or even that these are initially unknown to Eve. Therefore these can be chosen in advance and be considered as fixed parameters of the protocol.

[4] This fixing is allowed due to Davies' theorem [15].

Let $\mathbf{T}$ be the random variable presenting whether the test passed or failed ($\mathbf{T}$ is "pass" if $|i_T \oplus j_T| \leq np_a$ and is "fail" otherwise, with p_a denoting the allowed error rate $p_a \equiv p_{\text{allowed}}$). Let $c_T = i_T \oplus j_T$ and $c_I = i_I \oplus j_I$ be the error syndromes on the test and the information bits. Let $I(\mathbf{A};\mathbf{E})$ be the mutual information between Alice's final key and the results of Eve's measurement. Since some classical data is given to Eve, let $\mathbf{I}_{Eve} \equiv I(\mathbf{A};\mathbf{E} \mid i_T, j_T, b, s, \xi)$ be the information Eve has about the key given a particular PA, ECC (that remain fixed parameters), i_T, j_T, b, s and ξ (the parity string on the information bits, $\xi = i_I P_{\mathcal{C}}^{\top}$). This information might be large for some specific values (for instance, if b is fixed, and Eve has accidently guessed all the bases correctly), but on average it ought to be negligible in order for the key to be secret. The average information obtained by Eve if a key was always created by Alice is $\langle \mathbf{I}_{Eve} \rangle \equiv I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi})$, where $\mathbf{I}_T$, $\mathbf{J}_T$, $\mathbf{B}$, $\mathbf{S}$ and $\mathbf{\Xi}$ are the random variables associated to the random outputs i_T, j_T, b, s and $\xi = i_I P_{\mathcal{C}}^{\top}$. This information cannot be proven to be small, because the fact that the test must be passed is not taken into consideration.

We can now formally present our security criteria. In order to get a better intuition of what security really means, we also formally present in Appendix B.2 the two security criteria mentioned above, criteria that are not met by the QKD protocol. We even prove via counter examples, the *SWAP attack* and the *half-SWAP attack*, that these security criteria indeed don't work[5]. The SWAP and the half-SWAP examples motivate a more precise definition of security (first used in [27]) that does work properly, and shall be used in the current work.

2.3.1 The security criterion: We show in this paper that the event where the test is passed *and* Eve obtains meaningful information about the key is extremely unlikely. This is proven here for any attack $\{U, \mathcal{E}\}$. Formally, our security criterion is:

$$P\left[(\mathbf{T} = \text{pass}) \wedge (\mathbf{I}_{Eve} \geq A_{\text{info}}\, e^{-\beta_{\text{info}} n})\right] \leq A_{\text{luck}}\, e^{-\beta_{\text{luck}} n}\ , \qquad (2.1)$$

with A_{info}, β_{info}, A_{luck} and β_{luck} positive constants. Note that this is a criterion for exponential security, and a less strict criterion can be defined if one is willing to accept polynomial security (say, with a huge polynomial such as n^{1000}). However, exponential criteria are preferable when possible, and we succeed to prove here an exponential security criterion.

2.3.2 An alternative security criterion: Let us define $\mathbf{I}'_{Eve}$ to be equal to $\mathbf{I}_{Eve}$ when $\mathbf{T} = \text{pass}$ and to be equal to 0 otherwise. Then, the event $\left[(\mathbf{T} = \text{pass}) \wedge (\mathbf{I}_{Eve} \geq A_{\text{info}}\, e^{-\beta_{\text{info}} n})\right]$ is identical to the event $[\mathbf{I}'_{Eve} \geq A_{\text{info}} e^{-\beta_{\text{info}} n}]$. The security criterion can now be written more concisely as

$$P[\mathbf{I}'_{Eve} \geq A_{\text{info}} e^{-\beta_{\text{info}} n}] \leq A_{\text{luck}} e^{-\beta_{\text{luck}} n}\ .$$

[5] If Eve is applying the SWAP attack, her information given that the test is passed will not be small, and the first criterion is not satisfied; If Eve is applying the Half-SWAP attack, she gets a lot of information (half the bits on average), and yet passes the test with high probability, so the second criterion is not satisfied. In contrast, the criteria we use in this paper are satisfied by any attack whatsoever.

The expectancy of $\mathbf{I}'_{Eve}$ which is

$$\langle \mathbf{I}'_{Eve} \rangle = \sum_{i_T, j_T, b, s, \xi} \mathbf{I}'_{Eve}(i_T, j_T, b, s, \xi) p(i_T, j_T, b, s, \xi) \ ,$$

can now be used to define an important security condition:

$$\langle \mathbf{I}'_{Eve} \rangle \leq A e^{-\beta n} \ , \tag{2.2}$$

with A and β positive constants. As the following lemma shows, the security criterion, Eq. (2.1) is implied by this security condition.

Lemma 2.1 *If* $\langle \mathbf{I}'_{Eve} \rangle \leq A e^{-\beta n}$ *for* $A > 0$ *then*

$$P[\mathbf{I}'_{Eve} \geq A_{\text{info}} e^{-\beta_{\text{info}} n}] \leq A_{\text{luck}} e^{-\beta_{\text{luck}} n}$$

for all A_{info}, A_{luck}, β_{info}, β_{luck} *such that* $A_{\text{info}} A_{\text{luck}} = A$, $\beta_{\text{info}} + \beta_{\text{luck}} = \beta$ *and* $A_{\text{luck}} > 0$.

[Note that the security criterion 2.1 is therefore implied since the event $\left[(\mathbf{T} = \text{pass}) \wedge (\mathbf{I}_{Eve} \geq A_{\text{info}}\, e^{-\beta_{\text{info}} n})\right]$ is identical to the event $[\mathbf{I}'_{Eve} \geq A_{\text{info}} e^{-\beta_{\text{info}} n}]$.]

Proof $\mathbf{I}'_{Eve}$ is never negative. Therefore, by Markov's inequality [10] (that is $P[X \geq \alpha] \leq \langle X \rangle / \alpha$ for any non-negative random variable X),

$$P[\mathbf{I}'_{Eve} \geq A_{\text{info}} e^{-\beta_{\text{info}} n}] \leq \frac{\langle \mathbf{I}'_{Eve} \rangle}{A_{\text{info}} e^{-\beta_{\text{info}} n}} \leq \frac{A e^{-\beta n}}{A_{\text{info}} e^{-\beta_{\text{info}} n}} = A_{\text{luck}} e^{-\beta_{\text{luck}} n} \qquad \square$$

We gain two things by using this alternative security criteria. The first is some additional intuition about the security parameter, and the second is a final form of the criterion which is the one we actually prove here in the paper.

By definition, $\langle \mathbf{I}'_{Eve} \rangle = \sum_{i_T, j_T, b, s, \xi} \mathbf{I}'_{Eve}(i_T, j_T, b, s, \xi) p(i_T, j_T, b, s, \xi)$ is equal to $\sum_{i_T, j_T : |i_T \oplus j_T| \leq n\, p_a} \sum_{b,s,\xi} I(\mathbf{A}; \mathbf{E} \mid i_T, j_T, b, s, \xi) p(i_T, j_T, b, s, \xi)$, thus, it is easy to calculate that

$$\langle \mathbf{I}'_{Eve} \rangle = I(\mathbf{A}; \mathbf{E} \mid \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}, \mathbf{T} = \text{pass}) P[\mathbf{T} = \text{pass}] \ , \tag{2.3}$$

(see Appendix B.3.1 for the details of this calculation). This expression provides some intuition regarding the security criterion, Eq.(2.2): It says that if either the probability to pass the test is negligible or Eve's information given that the test is passed is negligible, then security is promised.

Using c_T (the error syndrome on the test bits) and using the random variable $\mathbf{C}_T \equiv \mathbf{I}_T \oplus \mathbf{J}_T$ (the random variable corresponding to the error syndrome), we can also write

$$\langle \mathbf{I}'_{Eve} \rangle = \sum_{c_T | \mathbf{T} = \text{pass}} P\left[\mathbf{C}_T = c_T\right] \ I(\mathbf{A}; \mathbf{E} | \mathbf{I}_T, \mathbf{C}_T = c_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}) \tag{2.4}$$

(see Appendix B.3.1 for the details of this calculation as well). This is true since the random variable $\mathbf{C}$ is equivalent to the random variable $\mathbf{J}$ when the random

variable $\mathbf{I}$ is given, and since summing over all the events $\{c_T|\mathbf{T}=\text{pass}\}$ provides exactly the event $\{\mathbf{T}=\text{pass}\}$.

This last expression, Eq.(2.4), tells us that the security criterion (2.2) is satisfied if:

$$\sum_{c_T|\mathbf{T}=\text{pass}} P\left[\mathbf{C}_T=c_T\right]\ I(\mathbf{A};\mathbf{E}|\mathbf{I}_T,\mathbf{C}_T=c_T,\mathbf{B},\mathbf{S},\mathbf{\Xi}) \leq Ae^{-\beta n}\ . \qquad (2.5)$$

Thus, this last equation is yet another form of the security criteria. Indeed, in Lemmas 5.3 and 5.4 in Section 5 we obtain an exponentially small bound on $\langle \mathbf{I}'_{Eve}\rangle$. This inequality then implies that the security criterion (2.1) is satisfied, for all attacks without any restriction whatsoever, therefore proving the security of the used-bits-BB84 and the original BB84 protocols.

To improve the intuition about the different security criteria (those that work for QKD and also those that do not work) we prove in Appendix B.3.2 that the Half-SWAP attack can easily be dealt with, once we use our security criteria; meaning that the security criteria are still satisfied.

2.4 The main result: a security proof

In this paper we provide a proof of the security of the used-bits BB84 protocol against any attack on the channel.

Formally we prove the following:

If the allowed error-rate p_a, some positive number ϵ_{sec}, and the ECC+PA codes are chosen such that $p_a+\epsilon_{\text{sec}} \leq \hat{v}/2n$ with $\hat{v}=\min_{r'=r+1}^{r+m} d_H(v_{r'},V_{r'}^{\text{exc}})$ where d_H is the Hamming distance, $v_{r'}$ a parity-check string, and $V_{r'}^{\text{exc}}$ the 2^{r+m-1} space which is the span of the ECC and PA excluding $v_{r'}$ (namely, the span of $v_1,\ldots,v_{r'-1},v_{r'+1},\ldots,v_{r+m}$), then for any $A_{\text{info}}>0$, $A_{\text{luck}}>0$ such that $A_{\text{info}}A_{\text{luck}}=2m$ and any β_{info} and β_{luck} such that $\beta_{\text{info}}+\beta_{\text{luck}}=\epsilon_{\text{sec}}^2/4$,

$$P\left[(\mathbf{T}=\text{pass})\wedge(\mathbf{I}_{Eve}\geq A_{\text{info}}\ e^{-\beta_{\text{info}}n})\right] \leq A_{\text{luck}}\ e^{-\beta_{\text{luck}}n} \qquad (2.6)$$

where $\mathbf{T}=\text{pass}$ iff $|c_T|\leq np_a$ and $\mathbf{I}_{Eve}=I(\mathbf{A};\mathbf{E}\mid i_T,j_T,b,s,\xi)$.

2.5 Reliability

It will moreover be shown here that if the ECC corrects $P_a+\epsilon_{\text{rel}}$ errors then the final m-bit key is reliable: The keys distilled by Alice and Bob are identical except for some exponentially small probability $A_{\text{rel}}\ e^{-\beta_{\text{rel}}n}$, with $A_{\text{rel}}=1$ and $\beta_{\text{rel}}=\epsilon_{\text{rel}}^2/2$.

We shall eventually present here an example of a family of ECC+PA codes such that the final key is secure and reliable, as long as the error rate p_a is less than 7.56%, and such that the bit-rate approaches one when the error-rate approaches zero. Furthermore, we present a different range of these codes such that for large

enough[6] but reasonable n the final key is secure and reliable, as long as the allowed error rate p_a is less than 5.50%; in Table 5.1 we provide some specific numbers that might be interesting to experimentalists who design a QKD protocol.

[6] Namely, not asymptotically large. For instance, n of the order of 10^4 or 10^5.

3 Eve's Attack

In the used-bits BB84 protocol Alice encodes a string i in the bases of her choice b in the state $|i\rangle_b$ which she sends to Bob via a quantum channel; Bob measures a string j using the same set of bases. In order to perform her attack, Eve prepares a probe, E, in a known (ancillary) state, which W.L.G. can be written as a vector $|0\rangle_{\rm E}$ and performs a unitary transformation U on the state

$$|0\rangle_{\rm E}|i\rangle_b$$

where $|i\rangle_b$ is assumed to have been intercepted by Eve. The resulting state $U|0\rangle_{\rm E}|i\rangle_b$ can be expressed in a unique way as a sum

$$U|0\rangle_{\rm E}|i\rangle_b = \sum_j |E'_{i,j}\rangle_b|j\rangle_b \tag{3.1}$$

where the vectors $|E'_{i,j}\rangle_b$ are non normalized vectors in Eve's probe space.

$$|E'_{i,j}\rangle_b = {}_b\langle j|U|0\rangle_{\rm E}|i\rangle_b \tag{3.2}$$

Eve then sends the disturbed qubits to Bob, keeping her probe in her hands. We call the state above

$$|\psi'_i\rangle \equiv \sum_j |E'_{i,j}\rangle_b|j\rangle_b \tag{3.3}$$

"Eve-Bob's state", because it is the state in the hands of Eve and Bob together.

Of course, Eve does not know the basis b when she performs her attack U with initial probe $|0\rangle_{\rm E}$. Actually, Eve-Bob's state is not known to any of the players: Alice knows i and b, Eve knows U (namely, the set of states $|E'_{i,j}\rangle_b$) but she knows neither i, nor j nor b, while Bob knows nothing prior to obtaining b from Alice (except his knowledge of the protocol). In the next steps Alice sends b to Bob (and Eve), and Bob measures and obtains his sifted key j. Then Alice sends s to Bob (and Eve) and both Alice and Bob disclose the test bits i_T and j_T. The information bits are still kept secret.

This section deals with two issues. 1.— symmetrizing Eve's attack; 2.— the attack on all bits versus the attack induced on the information bits.

Subsection 3.1 presents the symmetrized attack. Subsection 3.2 presents important properties of the symmetric attack. Subsection 3.3 proves that symmetric attacks are at least as good for Eve as any other attack can be. Subsection 3.4 distills the attack on the information bits, and finally, Subsection 3.5 analyzes the symmetrized attacks, when test bits and information bits are treated separately.

3.1 Symmetrizing Eve's attack

For any attack $\{U, \mathcal{E}\}$, we shall now define a different attack $\{U^{\rm sym}, \mathcal{E}^{\rm sym}\}$, which can be at least as good (for Eve) as the attack $\{U, \mathcal{E}\}$, it is symmetric to bit flips, and it is simpler to analyze. The symmetric attack $\{U^{\rm sym}, \mathcal{E}^{\rm sym}\}$ is obtained by enlarging Eve's probe, adding a second probe, M, containing $2n$ qubits in a state

$(1/2^n)\sum_m |m\rangle_{\rm M}$, and transforming it and measuring it as described below. The attack is "symmetric" in a sense that it is unaffected by the choice of i by Alice, and this is true for any basis b.

The symmetrization is done here in a physical way, namely, as a process that Eve can actually do if she wants to[7]. The symmetrization process can be done in a way that is always beneficial for Eve, and therefore, any attack, no matter how good it is, is no better than its optimal symmetrization. Thus, W.L.G., it is sufficient to prove security against all symmetric attack. In order to intuitively understand the design of these symmetric attacks (starting from any attack), we note that for the original attack, applying the attack (U) to a state $i \oplus m$ gives $U|0\rangle_{\rm E}|i\oplus m\rangle_b = \sum_{j'} |E'_{i\oplus m,j'}\rangle_b |j'\rangle_b = \sum_j |E'_{i\oplus m,j\oplus m}\rangle_b |j\oplus m\rangle_b$ with $j = j' \oplus m$. The symmetrization is achieved by Eve in practice in several steps.

We first present the symmetrization as if Eve knows the bases b: When the additional ancilla state is $|m\rangle_{\rm M}$ she applies her original attack after "shifting" i by m (namely XORing i with m, via bitwise Controlled-NOT gates): $U|0\rangle_{\rm E}|i\oplus m\rangle_b|m\rangle_{\rm M} = \sum_j |E'_{i\oplus m,j\oplus m}\rangle_b |j\oplus m\rangle_b|m\rangle_{\rm M}$. Now we can see that averaging the original attack over i is equivalent to averaging the shifted attack over all values m. The averaging over m is easily obtained due to starting with a quantum state which is an equal superposition of all values of m, $|0_x\rangle_{\rm M} \equiv (1/2^n)\sum_m |m\rangle_{\rm M}$. Then Eve could always measure m and continue with the same POVMs (where each POVM is a function of the values of $i_T, b, \ldots$) as in the original attack obtaining her original asymmetric attack up to a shift of all values by XORing them with m. Let us refer to this attack as the "trivial symmetric attack" $\{U^{\rm sym}, \mathcal{E}^{\rm trivial}\}$. We can also define a slightly stronger and more general attack in which Eve measures m on her additional probe, but continues with any POVM she finds appropriate. We call this attack the "simple symmetrized attack". Obviously, for a given U (and its modified attack, $U^{\rm sym}$), the optimal *simple* symmetrical attack is better than the *trivial* symmetric attack, because potentially more informative POVMs are chosen. The most general symmetric attack $\{U^{\rm sym}, \mathcal{E}^{\rm sym}\}$ generalizes this *simple* symmetric attack, as Eve can choose any measurement (rather than measuring m first). Clearly, the optimal symmetric attack (for a given U) is therefore at least as informative as the *trivial* and the *simple* symmetric attacks.

Note that in the *trivial* symmetric attack, when Eve's second probe is measured yielding an outcome m, we get back the original attack, up to a shift by m. If the error rate in the original attack U is averaged over all i and the error rate in the new attack is averaged over all m, the resulting average error rate is the same. Thus, the *trivial* symmetric attack induces the same error-rate, and gives Eve the same information as the original attack. However, as we just explained, in the symmetrized attack $\{U^{\rm sym}, \mathcal{E}^{\rm sym}\}$ Eve can also use the state $|m\rangle$ in other ways than just measuring m. This modification cannot change the error-rate due to causality (Eve's measurement can be done after Alice and Bob completed their protocol). On the other hand, the optimal symmetrization (optimal POVM, $\mathcal{E}^{\rm sym}$,

[7] One can also view the symmetrization as a virtual process. This makes some differences, but we do not consider this case here.

for each value of $i_T, b, \ldots$) will be at least as good as the trivial one, meaning that for any value of $i_T, b, \ldots$, it would not decrease Eve's information, while it could increase it. As a result of these two intuitive observations dealing with symmetrized attacks is sufficient, and any other attack cannot be better for Eve. We render these observations formally sound later on in Subsection 3.3, but we first must deal with the general case in which the basis b is not known to Eve by the time she performs the symmetrization.

The fact that Alice's state is also defined by a basis b which is unknown to Eve makes the required symmetrization slightly more complex, because we would like to obtain $i \oplus m$ no matter what the basis is. This is done as follows: We define the new attack in terms of a previously fixed basis; we will choose the computational basis, i.e. the basis $\{|i\rangle_0\}$ (for $b = 0$, the zero string). For each qubit sent by Alice, Eve attaches a new ancillary bit; her new ancilla (Eve's second probe, M) is thus a $2n$ qubit register, whose basis states are called $|m\rangle_{\rm M}$. She then applies independently to each pair of qubits (Alice's qubit plus the attached qubit from the probe M) the unitary transform satisfying the equalities $S|0\rangle_0|0\rangle = |0\rangle_0|0\rangle$, $S|1\rangle_0|0\rangle = |1\rangle_0|0\rangle$, $S|0\rangle_0|1\rangle = |1\rangle_0|1\rangle$ and $S|1\rangle_0|1\rangle = -|0\rangle_0|1\rangle$ (if the computational basis is $|0_z\rangle, |1_z\rangle$ then this corresponds to performing a controlled $\sigma_x\sigma_z$ transformation on each of Alice's qubits using the corresponding ancillary bit as control bit). If we evaluate S on basis vectors of the alternate basis $|0\rangle_1 \equiv \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|1\rangle_1 \equiv \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, we get immediately $S|0\rangle_1|0\rangle = |0\rangle_1|0\rangle$, $S|1\rangle_1|0\rangle = |1\rangle_1|0\rangle$, $S|0\rangle_1|1\rangle = -|1\rangle_1|1\rangle$ and $S|1\rangle_1|1\rangle = |0\rangle_1|1\rangle$; as a consequence, for each such pair of qubits, we can summarize the effect of S on basis states by the equality (where i, m and b are 0 or 1)

$$S|i\rangle_b|m\rangle = (-1)^{(i\oplus b)m}|i \oplus m\rangle_b|m\rangle \ .$$

On $2n$ such pairs of qubits, the exponents simply add up and, for any string i, m and b of $2n$ bits we get

$$S_{{\rm A}M}|i\rangle_b|m\rangle_{\rm M} = (-1)^{(i\oplus b)\cdot m}|i \oplus m\rangle_b|m\rangle_{\rm M} \tag{3.4}$$

$$S^\dagger_{{\rm A}M}|i\rangle_b|m\rangle_{\rm M} = (-1)^{(i\oplus b\oplus m)\cdot m}|i \oplus m\rangle_b|m\rangle_{\rm M} \tag{3.5}$$

where the subscript for S means it acts on Alice's qubits (A) and the second probe (M), where the second equation is deduced from the first by using the fact that $S^\dagger S = \mathbf{1}$, and with S being a $2^{4n} \times 2^{4n}$ matrix.

The symmetrized attack is therefore defined by the initial state of the additional probe $|0_x\rangle_{\rm M} \equiv (1/2^n)\sum_m |m\rangle_{\rm M}$, and by the unitary transform

$$U^{\rm sym} \equiv (\mathbf{1}_{\rm E} \otimes S^\dagger_{{\rm A}M})(U_{{\rm E}A} \otimes \mathbf{1}_{\rm M})(\mathbf{1}_{\rm E} \otimes S_{{\rm A}M}) \tag{3.6}$$

where $U_{{\rm E}A}$ is Eve's original attack on Alice's qubits (A) and Eve's first probe (E), S is applied onto Alice's qubits and Eve's second probe, and $\mathbf{1}_{\rm E}$ and $\mathbf{1}_{\rm M}$ are the identity on Eve's first and second probe space respectively. This completes the definition of the symmetrized attack.

3.2 Some basic properties of symmetric attacks

3.2.1 The "Basic Lemma of Symmetrization": For any attack U, and for any basis b, we write $U^{\rm sym}$ slightly differently now by defining $|E^{\rm sym\,\prime}_{i,j}\rangle_b$ via

$$U^{\rm sym}|0\rangle_{\rm Eve}|i\rangle_b = U^{\rm sym}|0\rangle_{\rm E}|0_x\rangle_{\rm M}|i\rangle_b \equiv \sum_j |E^{\rm sym\,\prime}_{i,j}\rangle_b|j\rangle_b$$

where both probes $|0\rangle_{\rm E}$ and $|0_x\rangle_{\rm M}$ have been put together (adjacent to each other) on the left side, to clarify the definition of these $|E^{\rm sym\,\prime}_{i,j}\rangle_b$.

Given any attack U, with its $|E'_{i,j}\rangle_b$ the symmetrization leads to these $E^{\rm sym\,\prime}_{i,j}$s that can now be described via the original $E'_{i,j}$s as follows:

Lemma 3.1 *For any basis string b*

$$|E^{\rm sym\,\prime}_{i,j}\rangle_b = 2^{-n}\sum_m (-1)^{(i\oplus j)\cdot m}|E'_{i\oplus m,j\oplus m}\rangle_b|m\rangle \tag{3.7}$$

We refer to this Lemma as *the Basic Lemma of Symmetrization*.

Proof In order to calculate smoothly, we write (again) $|0\rangle_{\rm E}|i\rangle_b|0_x\rangle_{\rm M}$ (instead of $|0\rangle_{\rm E}|0_x\rangle_{\rm M}|i\rangle_b$) in the order the Hilbert spaces appear in equation (3.6) defining $U^{\rm sym}$:

$$\begin{aligned}
&U^{\rm sym}|0\rangle_{\rm E}|i\rangle_b|0\rangle_{\rm M} = \\
&\quad 2^{-n}(\mathbf{1}_{\rm E}\otimes S)^\dagger(U\otimes\mathbf{1}_{\rm M})(\mathbf{1}_{\rm E}\otimes S)\left[\sum_m |0\rangle_{\rm E}|i\rangle_b|m\rangle\right] \\
&= 2^{-n}(\mathbf{1}_{\rm E}\otimes S)^\dagger(U\otimes\mathbf{1}_{\rm M})\left[\sum_m (-1)^{(i\oplus b)\cdot m}|0\rangle_{\rm E}|i\oplus m\rangle_b|m\rangle\right] \\
&= 2^{-n}(\mathbf{1}_{\rm E}\otimes S)^\dagger\left[\sum_{m,j} (-1)^{(i\oplus b)\cdot m}|E'_{i\oplus m,j\oplus m}\rangle_b|j\oplus m\rangle_b|m\rangle\right] \\
&= 2^{-n}\sum_{m,j}(-1)^{(i\oplus b)\cdot m}(-1)^{(j\oplus m\oplus b\oplus m)\cdot m}|E'_{i\oplus m,j\oplus m}\rangle_b|j\rangle_b|m\rangle \\
&= 2^{-n}\sum_j\sum_m (-1)^{(i\oplus j)\cdot m}|E'_{i\oplus m,j\oplus m}\rangle_b|j\rangle_b|m\rangle
\end{aligned}$$

which proves the lemma. □

The Lemma tells us (intuitively) that Eve gets a similar replacement of $E'_{i,j}$ by $E'_{i\oplus m,j\oplus m}$ whether she symmetrizes with respect to the computational basis or with respect to any other basis. This means that symmetrization with respect to the output bits 0 or 1 results also in some form of symmetry with respect to the bases.

3.2.2 Symmetrization and the error-rate: For any attack (symmetric or not) the probability that Bob measures the string j in basis b if Alice sent i is given by $p(j \mid i,b) = \langle E'_{i,j}|E'_{i,j}\rangle_b$. In particular, for symmetric attacks $p^{\rm sym}(j \mid i,b) = \langle E^{\rm sym\,\prime}_{i,j}|E^{\rm sym\,\prime}_{i,j}\rangle_b$. As a consequence of the Basic Lemma of Symmetrization (Lemma 3.1) we can now establish a link between $p^{\rm sym}(j \mid i,b)$, the probability that (under the symmetrized attack) Bob measures j in basis b if Alice sent i, with $p(j \mid i,b)$, the corresponding probability for the original attack. For a given b and i, the probability of some specific $j = i \oplus c$ becomes the probability of c. Thus we can also conclude a link between $p^{\rm sym}(c \mid i,b)$ and $p(c \mid i,b)$. The two main conclusions of the forthcomming lemma are that (a) — the probability (in the symmetrized attack) $p^{\rm sym}(c \mid i,b)$ for a given i, is actually independent of i, as it is equal to $p(c \mid b)$, and (b) — the probability (in the symmetrized attack) $p^{\rm sym}(c \mid b)$ is equal to the probability in the original attack, as it is equal to $p(c \mid b)$.

Lemma 3.2 *For any i chosen by Alice and for any $j = i \oplus c$*

$$p^{\rm sym}(j \mid i,b) \equiv p^{\rm sym}(i \oplus c \mid i,b) = 2^{-2n} \sum_{i'} p(i' \oplus c \mid i',b)\ , \tag{3.8}$$

$$p^{\rm sym}(c \mid i,b) = p^{\rm sym}(i \oplus c \mid i,b) = p^{\rm sym}(j \mid i,b) = p(c \mid b)\ , \tag{3.9}$$

$$p^{\rm sym}(c \mid b) = p(c \mid b) \tag{3.10}$$

Proof Using the fact that the states $|m\rangle$ are orthonormal, we get

$$\begin{aligned} p^{\rm sym}(j \mid i,b) &= \langle E^{\rm sym\,\prime}_{i,j}|E^{\rm sym\,\prime}_{i,j}\rangle_b \\ &= 2^{-2n} \sum_m \langle E'_{i\oplus m,j\oplus m}|E'_{i\oplus m,j\oplus m}\rangle_b \qquad \text{by Eq. (3.7)} \\ &= 2^{-2n} \sum_m p(j \oplus m \mid i \oplus m,b)\ , \end{aligned}$$

By assigning $i' = i \oplus m$ this gives $p^{\rm sym}(j \mid i,b) = 2^{-2n} \sum_{i'} p(j \oplus i' \oplus i \mid i',b)$. With $c = i \oplus j$ we finally get $p^{\rm sym}(i \oplus c \mid i,b) = 2^{-2n} \sum_{i'} p(i' \oplus c \mid i',b)$. This completes the first part of the Lemma.

By definition, the averaging over all i' means that $2^{-2n} \sum_{i'} p(i' \oplus c \mid i',b) \equiv p(c \mid b)$, so we get $p^{\rm sym}(i \oplus c \mid i,b) = p(c \mid b)$. We conclude that $p^{\rm sym}(i \oplus c \mid i,b)$ is actually independent of i, namely, $p^{\rm sym}(j \mid i,b) = p(c \mid b)$. For a given b and i, $p^{\rm sym}(j \mid i,b) = p^{\rm sym}(i \oplus c \mid i,b) = p^{\rm sym}(c \mid i,b)$. This completes the proof of the second part of the lemma.

We now start with $p^{\rm sym}(i \oplus c \mid i,b) = p(c \mid b)$. Then, averaging $p^{\rm sym}(i \oplus c \mid i,b)$ over all i means that $2^{-2n} \sum_i p^{\rm sym}(i \oplus c \mid i,b) \equiv p^{\rm sym}(c \mid b)$. However, the summation is over equal terms $[p(c \mid b)]$, so we finally get $p^{\rm sym}(c \mid b) = p^{\rm sym}(i \oplus c \mid i,b)$, proving the last part of the Lemma. □

3.3 Symmetric attacks are optimal for the eavesdropper

We now show that for any attack $\{U, \mathcal{E}\}$, the attack $\{U^{\rm sym}, \mathcal{E}^{\rm trivial}\}$ leaves the same average error rate and also provides the same information to Eve as the original

attack. The optimal symmetric attack (for a given U), in which the optimization is over all the possible measurements $\mathcal{E}^{\text{sym}}$ leaves the same average error rate and provides information to Eve that is equal or larger than that of the original attack U. These results imply (see Lemma 5.2) that if the security criterion is satisfied for all symmetric attacks, then it is satisfied for *all attacks*. Let us recall that due to causality Bob's outcome will be the same whatever measurement Eve performs. Since symmetrization in one basis yields symmetrization at any basis, we may assume (W.L.G.) that Eve performed her symmetrization with respect to the basis used by Alice and Bob. In that context, if Eve uses the *trivial* symmetrized attack, and measures $|m\rangle$ in the standard basis, this is simply a replacement of i by $i \oplus m$ and j by $j \oplus m$ with respect to the original attack. Continuing by a POVM as in the original attack, now yields the same information as the original attack, while clearly Eve could do better, as earlier explained.

In the following subsections we make the above intuition mathematically solid. [Recall that the string s (where a position equal to 1 corresponds to an information bit in i whilst a 0 indicates a test bit) determines two substrings of i, namely i_I (information bits) and i_T (test bits); after s is published by Alice we may identify $|i\rangle_b$ with $|i_T\rangle_b|i_I\rangle_b = |i_T\rangle_b \otimes |i_I\rangle_b$ (this isomorphism depends on s, and is just a permutation of bits); note that the same modification applies to $|j\rangle_b$.]

3.3.1 Symmetrization does not affect the average error-rate: As a corollary of Lemma 3.2, when s is known, we get

Corollary 3.1

$$P^{\text{sym}}[c_I, c_T \mid b, s] = P[c_I, c_T \mid b, s] \tag{3.11}$$

$$P^{\text{sym}}[c_T \mid b, s] = P[c_T \mid b, s]\ . \tag{3.12}$$

The first equation is a slight modification of the third part of Lemma 3.2 (due to s being published), and the second equation is obtained from the first by summing over all c_I.

These results prove that the average error-rate is not changed when an attack U is replaced by any symmetric attack U^{sym}.

3.3.2 Eve's information is not decreased by symmetrization: Let $\mathbf{E}^{\text{sym}}$ be the random variable whose values e are the output of Eve's measurement $\mathcal{E}^{\text{sym}}$, and note that the measurement is fixed at the end of the protocol, hence depends on the value of $\{i_T, c_T, b, s, \xi\}$. For any particular attack U and particular value $\{i_T, c_T, b, s, \xi\}$, the *maximal value* of $I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid i_T, \mathbf{C}_T = c_T, b, s, \xi)$ corresponding to Eve's symmetrized attack and *optimal measurement* is larger than or equal to that obtained if she restricts herself to performing the *trivial* symmetric attack (namely, to measuring the $|m\rangle$ probe in the standard basis, and repeat the POVM of the original attack).

Let us denote $(\mathbf{E}', \mathbf{M})$ the (multivariate) random variable where for each particular value of m, $\mathbf{E}'$ are the random outputs of the *trivial* symmetric attack. Then, we have by the very definition of the optimal measurement that

$$\max_{\{\mathcal{E}^{\text{sym}}\}} I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid i_T, \mathbf{C}_T = c_T, b, s, \xi) \geq I(\mathbf{A}; \mathbf{E}', \mathbf{M} \mid i_T, \mathbf{C}_T = c_T, b, s, \xi)\ ,$$

where $\{\mathcal{E}^{\rm sym}\}$ does not stand for one POVM but for a set of POVMs, one for each value of i_T, c_T, b, s, ξ. We would like to bound

$$I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) = \sum_{i_T,\xi} P[i_T, \xi \mid c_T, b, s] I(\mathbf{A};\mathbf{E} \mid i_T, \mathbf{C}_T = c_T, b, s, \xi) \, .$$

We must note the important fact that the POVM is only fixed at the end of the protocol, hence a different POVM $\mathcal{E}$ is chosen for each fixed value of i_T, ξ (as the other parameters are fixed here). The same is true for the *trivial* symmetrized attack

$$\begin{aligned} &I(\mathbf{A};\mathbf{E}', \mathbf{M} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) \\ &\quad = \sum_{i_T,\xi} P[i_T, \xi \mid c_T, b, s] I(\mathbf{A};\mathbf{E}', \mathbf{M} \mid i_T, \mathbf{C}_T = c_T, b, s, \xi) \, , \end{aligned}$$

and the same is true for the optimal symmetrized attack (for a given U)

$$\begin{aligned} &\max I(\mathbf{A};\mathbf{E}^{\rm sym} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) \\ &\quad \equiv \sum_{i_T,\xi} P[i_T, \xi \mid c_T, b, s] \max_{\{\mathcal{E}^{\rm sym}\}} I(\mathbf{A};\mathbf{E}^{\rm sym} \mid i_T, \mathbf{C}_T = c_T, b, s, \xi) \, . \end{aligned} \quad (3.13)$$

With that definition we are promised that symmetrization is optimal for each particular value of $\{i_T, c_T, b, s, \xi\}$ and the resulting information is optimal also after summing over i_T, ξ:

$$\max I(\mathbf{A};\mathbf{E}^{\rm sym} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) \geq I(\mathbf{A};\mathbf{E}', \mathbf{M} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi})$$

Now we are ready to present the main result of this subsection. An optimal symmetrization of U will not decrease the information accessible to Eve in the following sense:

Lemma 3.3 *For any fixed U, c_T, b, s,*

$$\max I(\mathbf{A};\mathbf{E}^{\rm sym} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) \geq I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) \quad (3.14)$$

Proof For any given U, the optimal symmetric attack is at least as good as the *trivial* symmetric attack for each value of c_T, b, s, i_T, ξ, and therefore also after summing over i_T, ξ.

Proving formally that the *trivial* symmetric attack is as good as the original attack is less trivial[8]. Actually, for simplicity, we only prove the relevant direction, namely, that the *trivial* symmetric attack is at least as good as the original attack:

$$I(\mathbf{A};\mathbf{E}', \mathbf{M} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) \geq I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) \quad (3.15)$$

For the details of that proof, see Appendix C.1. □

The above result means that we can use a bound on Eve's average information in the case of a symmetrized attack to apply to the unsymmetrized case.

[8] Still, it is somewhat similar to the argument given when we analyzed the case in which Eve knows the bases.

3.4 Eve-Bob's state after the basis and the test bits are known

When the strings b and s are given to Bob (and to Eve) then Eve-Bob's state (Eq.3.3) ought to be modified. The sifted keys $|i\rangle_b$ and $|j\rangle_b$, the resulting error syndrome $c = i \oplus j$, Eve's attack U, and Eve's unnormalized states $E'_{i,j}$ are now expressed differently, so that the test bits and information bits are written separately. Equation (3.1) can thus be rewritten as

$$U|0\rangle_{\rm E}|i_T\rangle_b|i_I\rangle_b = \sum_j |E'_{i_T,i_I,j_T,j_I}\rangle_b|j_T\rangle_b|j_I\rangle_b \tag{3.16}$$

where the right-hand side corresponds to Eve-Bob's state ($|\psi'_i\rangle$) for a given $i = i_T i_I$, and where

$$|E'_{i_T,i_I,j_T,j_I}\rangle_b = {}_b\langle j_T|_b\langle j_I|U|0\rangle_{\rm E}|i_T\rangle_b|i_I\rangle_b \ . \tag{3.17}$$

The probability that Bob measures $|j_T\rangle_b|j_I\rangle_b$ is

$$p(j_T, j_I|i_T, i_I, b, s) = \langle E'_{i_T,i_I,j_T,j_I}|E'_{i_T,i_I,j_T,j_I}\rangle_b \ . \tag{3.18}$$

Once i_T is also given to Eve and Bob, it is considered as a fixed parameter instead of a variable in the equation above. When j_T is measured, the right-hand states $\sum_j |E'_{i_T,i_I,j_T,j_I}\rangle_b|j_T\rangle_b|j_I\rangle_b$ are projected onto the particular j_T obtained by the measurement on the test bits, and 2^n basis states are left in the summation, corresponding to the 2^n possible values of the n information qubits in Bob's hands. Formally, the projection is described via $\langle j_T|\psi'_i\rangle = \sum_{j_I} |E'_{i_T,i_I,j_T,j_I}\rangle_b|j_I\rangle_b$. The projection should now be followed by a normalization of the state, thus modifying Eve-Bob's state to become

$$|\psi_{i_I}\rangle = \sum_{j_I} \frac{1}{\sqrt{p(j_T|i_T, i_I, b, s)}} |E'_{i_T,i_I,j_T,j_I}\rangle_b|j_I\rangle_b \ . \tag{3.19}$$

With $p(j_T|i_T, i_I, b, s) = \sum_{j_I} p(j_T, j_I|i_T, i_I, b, s)$ and using Eq.(3.18) we get that the normalization factor (due to the projection on j_T) is the square root of

$$p(j_T|i_T, i_I, b, s) = \sum_{j_I} \langle E'_{i_T,i_I,j_T,j_I}|E'_{i_T,i_I,j_T,j_I}\rangle_b \ . \tag{3.20}$$

Let us now define[9]

$$|E_{i_I,j_I}\rangle_{b,s} \equiv \frac{1}{\sqrt{p(j_T|i_T, i_I, b, s)}} |E'_{i_T,i_I,j_T,j_I}\rangle_b \ , \tag{3.21}$$

so that the resulting Eve-Bob's state can be written more economically in the form

$$|\psi_{i_I}\rangle = \sum_{j_I} |E_{i_I,j_I}\rangle_{b,s}|j_I\rangle_b \ . \tag{3.22}$$

[9] The expression $|E_{i_I,j_I}\rangle_{b,s}$ is also a function of the parameters i_T and j_T (which are known to Eve by now), but, writing the expression as $|E_{i_I,j_I}\rangle_{b,s,i_T,j_T}$ looks cumbersome; therefore, for convenience, we did not write them in the expression, while we keep b, s to remind us that the bases and the test are known.

From Eqs. (3.18, 3.21) and the conditional probability formula [$p(ab)/p(a) = p(b|a)$] we get

$$\langle E_{i_I,j_I}|E_{i_I,j_I}\rangle_{b,s} = p(j_I \mid i_I, i_T, j_T, b, s)\ ; \tag{3.23}$$

with $c = i \oplus j$ this gives $\langle E_{i_I,i_I\oplus c_I}|E_{i_I,i_I\oplus c_I}\rangle_{b,s} = p(i_I \oplus c_I \mid i_I, i_T, j_T, b, s) = p(c_I \mid i_I, i_T, j_T, b, s)$.

3.5 Symmetrization — its impact on the test and information bits:

We first prove that for symmetrized attacks various expressions become independent of i_I:

Lemma 3.4

$$p^{\text{sym}}(j_T \mid i_T, i_I, b, s) = p^{\text{sym}}(j_T \mid i_T, b, s)\ . \tag{3.24}$$

Proof As an immediate corollary of Lemma 3.2 (that says that $p^{\text{sym}}(c \mid i, b) = p(c \mid b)$) when s is known, we get

$$p^{\text{sym}}(c_T, c_I \mid i_T, i_I, b, s) = p[c_T, c_I \mid b, s]\ .$$

Recalling that $c_T = i_T \oplus j_T$ and $c_I = i_I \oplus j_I$, this implies that for any m'_I

$$p^{\text{sym}}(j_T, j_I \oplus m'_I \mid i_T, i_I \oplus m'_I, b, s) = p^{\text{sym}}(j_T, j_I \mid i_T, i_I, b, s)\ .$$

If we sum both sides of this equality over j_I we get $p^{\text{sym}}(j_T \mid i_T, i_I \oplus m_I, b, s) = p^{\text{sym}}(j_T \mid i_T, i_I, b, s)$ which means that the probability is independent of i_I,

$$p^{\text{sym}}(j_T \mid i_T, i_I, b, s) = p^{\text{sym}}(j_T \mid i_T, b, s)\ .$$

□

As a corollary of the above Lemma, notice that for symmetric attacks,

Corollary 3.2

$$p^{\text{sym}}(i_I|i_T, j_T, b, s) = 1/2^n. \tag{3.25}$$

Indeed, using the Bayes rule (on $\{j_T; i_I\}$)

$$p^{\text{sym}}(i_I \mid i_T, j_T, b, s) = \frac{p^{\text{sym}}(j_T \mid i_I, i_T, b, s)}{p^{\text{sym}}(j_T \mid i_T, b, s)} p^{\text{sym}}(i_I \mid i_T, b, s) = \frac{1}{2^n}$$

where the last equality results from Eq. (3.24) and the fact that all bits of i, b and s are chosen independently [so $p^{\text{sym}}(i_I \mid i_T, b, s) = \frac{1}{2^n}$].

Another important consequence of Lemma 3.4 is:

Lemma 3.5 *For the information bits:*

1. $\langle E^{\text{sym}}_{i_I,i_I\oplus c_I}|E^{\text{sym}}_{i_I\oplus k_I,i_I\oplus c_I\oplus k_I}\rangle_{b,s}$ *is independent of* i_I.
2. $\sum_j \langle E^{\text{sym}}_{i_I,j_I}|E^{\text{sym}}_{i_I\oplus k_I,j_I\oplus k_I}\rangle_{b,s}$ *is independent of* i_I.

The proof is given in Appendix C.2.

The next step is to show that for symmetrized attacks various expressions are independent also of b_I. We proved in Lemma 3.4 that the normalizing factor for fixed i_T, j_T, b and s is the same for all the indices i_I. In addition, that normalizing factor does not depend on b_I either:

Lemma 3.6

$$p^{\text{sym}}(j_T \mid i_T, b, s) \equiv p^{\text{sym}}(j_T \mid i_T, b_I, b_T, s) = p^{\text{sym}}(j_T \mid i_T, b_T, s) \; . \qquad (3.26)$$

Proof In fact, Eq.(3.26) is true for any attack (symmetrized or not):

$$p(j_T \mid i_T, b, s) \equiv p(j_T \mid i_T, b_I, b_T, s) = p(j_T \mid i_T, b_T, s) \; . \qquad (3.27)$$

Intuitively, the fact that i_I is *not* a given parameter actually means that we average over it (as $p(a) = \sum_b p(a,b) = \sum_b p(b)p(a|b)$). Once we average over it, the relevant quantum bits are traced out, causing independence of b_I as well. Thus, in general, j of one subset (such as j_T) is independent of b of another subset (such as b_I). This is formally proven in Appendix C.3. Thus follows $p^{\text{sym}}(j_T \mid i_T, b, s) = p^{\text{sym}}(j_T \mid i_T, b_T, s)$. □

As a trivial Corollary of Lemmas 3.4 and 3.6 we get the following:

Corollary 3.3 *For symmetrized attacks, the probability of j_T satisfies*

$$p^{\text{sym}}(j_T \mid i_T, i_I, b, s) = p^{\text{sym}}(j_T \mid i_T, b_T, s) \; , \qquad (3.28)$$

and therefore, Eq.(3.21) is simplified to

$$|E^{\text{sym}}_{i_I, j_I}\rangle_{b,s} = \frac{1}{\sqrt{p^{\text{sym}}(j_T|i_T, b_T, s)}} |E^{\text{sym}\,\prime}_{i_T, i_I, j_T, j_I}\rangle_b \; . \qquad (3.29)$$

4 Information vs. Disturbance

In this section we analyze the information bits alone (for a given symmetric attack U^{sym}, a given input i_T and outcome j_T on the test bits, and given bases b and choice of test bits s). When no ambiguity arises, the indices b and s will be dropped; $|i\rangle$ will denote $|i\rangle_b$, $|i_I\rangle$ will denote $|i_I\rangle_{b_I}$ and $|E^{\text{sym}}_{i_I,j_I}\rangle_{b,s}$ will be denoted $|E_{i_I,j_I}\rangle$. Our result here applies for any U^{sym}, hence in particular *for the optimal one*. The optimization over Eve's measurement is avoided by using the fact that trace norm of the difference of two density matrices provides an upper bound on the accessible information one could obtain *via any measurement* when having the two density matrices as the possible inputs.

4.1 Eve's state

When Alice sends a state $|i_I\rangle \equiv |i_I\rangle_{b_I}$ for the information bits (where b_I is the string actually used by her and Bob to fix the bases on information bits), the state of Eve and Bob together, $|\psi_{i_I}\rangle = \sum_{j_I} |E_{i_I,j_I}\rangle|j_I\rangle$ is fully determined by Eve's attack and by the data regarding the test bits. Eve's state in that case is fully determined by tracing-out Bob's subsystem $|j_I\rangle$ from Eve-Bob's state, and it is

$$\rho^{i_I} = \sum_{j_I} |E_{i_I,j_I}\rangle\langle E_{i_I,j_I}| \,, \tag{4.1}$$

calculated given i_T and j_T. This state in Eve's hands is a mixed state.

4.2 Purification and a related basis

We can "purify" the state while giving more information to Eve by assuming she keeps the state

$$|\phi_{i_I}\rangle = \sum_{j_I} |E_{i_I,j_I}\rangle|i_I \oplus j_I\rangle \tag{4.2}$$

where we introduce another subsystem for the "purification". Notice that the indices of ϕ and of E are always information bits (n-bit strings). As a consequence, we could as well have written without ambiguity $|\phi_i\rangle = \sum_j |E_{i,j}\rangle|i \oplus j\rangle$ where the sum is taken over all n-bit strings j that can serve as index in $|E_{i,j}\rangle$. We will do this when expressions do not involve test bits. The term purification means different things in different papers, thus we explain it a bit more: A mixed state can also be obtained from a pure state in an enlarged system (the original system plus an ancilla), once the ancilla is traced out; the pure state of the enlarged system (or its density matrix) is called a purification of the mixed state. In a more general case, the state in the enlarged system is not necessarily pure, and then we refer to it as a "lift-up" [7] of the state of the original system.

The resulting purified state (i.e., any purification or any lift-up of Eve's states, for instance, the purification $\rho^i = |\phi_i\rangle\langle\phi_i|$), is at least as informative to Eve as ρ^{i_I} (of Eq. 4.1) is. This is because the density matrix ρ^{i_I} is exactly the same as Eve's

state would be if Eve ignored the $i_I \oplus j_I$ register of ϕ. Thus, any information Eve can obtain from her mixed state is bounded by the information she could get if the purified state was available to her.

Note that the overlap between these purified states satisfies

$$\begin{aligned}\langle \phi_l | \phi_{l\oplus k} \rangle &= \sum_j \sum_{j'} \langle E_{l,j} | E_{l\oplus k, j'} \rangle \langle l \oplus j | l \oplus k \oplus j' \rangle \\ &= \sum_j \langle E_{l,j} | E_{l\oplus k, j\oplus k} \rangle \ , \end{aligned} \tag{4.3}$$

where all the indices are n-bit strings.

As a consequence of Lemma 3.5 we immediately get for the information bits that $\langle \phi_l | \phi_{l\oplus k} \rangle$ is independent of l [meaning, independent of i_I, see Eq. (4.2)]. Thus, it is only a function of k (namely, k_I), and we can write this as

Corollary 4.1

$$\Phi_k \equiv \langle \phi_l | \phi_{l\oplus k} \rangle \ .$$

For the 2^n Hilbert-space spanned by the purified states $|\phi_l\rangle$ (corresponding to information bits), we define a Fourier basis $\{|\eta\rangle\}$, and show that it is possible to compute a bound on Eve's information about the information bits, once the purified states are expressed in this basis.

Definition 4.1

$$|\eta_i\rangle = \frac{1}{2^n} \sum_l (-1)^{i\cdot l} |\phi_l\rangle \ ; \ d_i^2 = \langle \eta_i | \eta_i \rangle \ ; \ \hat{\eta}_i = \eta_i / d_i$$

Using the above definitions and $(1/2^n)\sum_l (-1)^{(i\oplus j)\cdot l} = \delta_{ij}$, Eve's purified state can be rewritten as:

$$|\phi_i\rangle = \sum_l (-1)^{i\cdot l} |\eta_l\rangle = \sum_l (-1)^{i\cdot l} d_l |\hat{\eta}_l\rangle \ . \tag{4.4}$$

Note that $\langle \eta_i | \eta_i \rangle = \frac{1}{2^{2n}} \sum_l \sum_k (-1)^{i\cdot k} \langle \phi_l | \phi_{l\oplus k} \rangle = \frac{1}{2^n} \sum_k (-1)^{i\cdot k} \Phi_k$. In terms of Eve's states we can write

$$d_i^2 = \langle \eta_i | \eta_i \rangle = \frac{1}{2^{2n}} \sum_l \sum_k (-1)^{i\cdot k} \sum_j \langle E_{l,j} | E_{l\oplus k, j\oplus k} \rangle_{b,s} \ . \tag{4.5}$$

Proposition 4.1 *For symmetrized attacks,* $\langle \eta_j | \eta_i \rangle = 0$ *if* $i \neq j$.

Proof Note that $\langle \eta_j | \eta_i \rangle = \frac{1}{2^{2n}} \sum_l (-1)^{(i\oplus j)\cdot l} \sum_k (-1)^{i\cdot k} \langle \phi_l | \phi_{l\oplus k} \rangle$.

Since $\langle \phi_l | \phi_{l\oplus k} \rangle \equiv \Phi_k$ is independent of l, we see that:

$$\begin{aligned}\langle \eta_j | \eta_i \rangle &= \frac{1}{2^{2n}} \sum_l (-1)^{(i\oplus j)\cdot l} \sum_k (-1)^{i\cdot k} \Phi_k \\ &= \frac{1}{2^n} \delta_{i,j} \sum_k (-1)^{i\cdot k} \Phi_k \\ &= \delta_{i,j} \langle \eta_i | \eta_i \rangle \quad \square \end{aligned}$$

The above proposition is used to prove Lemma 4.2.

4.3 Eve's state and probability of errors induced on information bits

In this subsection we show that the probability of any error string Eve would have induced if the conjugate basis was used for the information bits, is a simple function of the d_is (of Definition 4.1), hence a function of the overlap of Eve's purified states. For any attack (i_T and j_T being fixed once and for all), any b and s, we have

$$P[\mathbf{C}_I = c_I \mid i_I, i_T, j_T, b, s] = \langle E_{i_I, i_I \oplus c_I} | E_{i_I, i_I \oplus c_I} \rangle_{b,s} \ . \tag{4.6}$$

See Eq. (3.23).

For any symmetrized attack and any b and s the error distribution in the information bits is

$$\begin{aligned} & P^{\text{sym}}[\mathbf{C}_I = c_I \mid i_T, j_T, b, s] \\ & = \sum_{i_I} P^{\text{sym}}[\mathbf{C}_I = c_I \mid i_I, i_T, j_T, b, s] p^{\text{sym}}(i_I \mid i_T, j_T, b, s) \\ & = \frac{1}{2^n} \sum_{i_I} \langle E^{\text{sym}}_{i_I, i_I \oplus c_I} | E^{\text{sym}}_{i_I, i_I \oplus c_I} \rangle_{b,s} \ , \end{aligned} \tag{4.7}$$

namely, the average probability of an error syndrome c_I on the information bits (when the test bits, basis and sequence are given). The first equality is derived using standard probability theory ($p(a) = \sum_b p(a|b)p(b)$) and the second is due to Eq. (3.25) and Eq. (4.6).

Identity (4.7) applies for all strings b and s and, in particular, for $b^0 = b \oplus s$ we get

$$P^{\text{sym}}[\mathbf{C}_I = c_I \mid i_T, j_T, b^0, s] = \frac{1}{2^n} \sum_{i_I} \langle E^{\text{sym}}_{i_I, i_I \oplus c_I} | E^{\text{sym}}_{i_I, i_I \oplus c_I} \rangle_{b^0, s} \ . \tag{4.8}$$

The basis b^0 is a basis where the basis for the test bits is the same as b, but the basis for each information bit is opposite. With a little algebra, as shown in Appendix C.4, we can express $|E^{\text{sym}}_{i_I, i_I \oplus c_I}\rangle_{b^0,s}$ in terms of the $|E^{\text{sym}}_{i_I, i_I \oplus c_I}\rangle_{b,s}$. Then, doing this for the right-hand side of Eq. (4.8) we get the right-hand side of Eq. (4.5) with $i = c_I$; this means that we get the following

Lemma 4.1

$$P^{\text{sym}}\left[\mathbf{C}_I = c_I \mid i_T, j_T, b^0, s\right] = d^2_{c_I} \ . \tag{4.9}$$

The proof is presented in Appendix C.4. Note that the d_i used here are those of the symmetrized attack.

Put differently, the term $d^2_{c_I}$ defined in terms of the actual bases used by Alice and Bob is equal to the probability of the error syndrome c_I on information bits had Alice and Bob used the conjugate bases on information bits. As we shall soon see, these d_is actually provide a measure of the information Eve could get from her purified states, therefore leading to a novel *information versus disturbance* result.

4.4 Bounds on Eve's information – the one-bit key case

In this subsection we much improve upon a result obtained in [7] (the result was derived for the collective attack). Eve's information about a particular bit of the final key (even if all other bits of the final key are given to her) is bounded. We take into consideration the error-correction data that is given to Eve, and we do it more efficiently than in [7], hence we obtain a much better threshold for the allowed error-rate.

Let us first discuss a one-bit final key a, defined to be the parity of a substring of the input i_I. The substring is defined using a mask v, meaning that the secret key is $a = v \cdot i_I$. (In the general case, the key is defined as the string $a = i_I P_{\mathcal{PA}}^T$ where $P_{\mathcal{PA}}$ is an $m \times n$ matrix; c.f. subsection 2.1, item II. 7). Bob first corrects his errors using the error correction code data, hence he learns Alice's string i_I. Eve does not know i_I, but she learns the error correcting code $\mathcal{C}$ used by Alice and Bob as well as v and the parity bits ξ sent by Alice to help Bob correct the sequence he received. All the possible inputs i_I that have the correct parities ξ for the code $\mathcal{C}$ form a set denoted $\mathcal{C}_\xi = \{i_I \mid i_I P_{\mathcal{C}}^\top = \xi\}$.

When the purification of Eve's state is given by $|\phi_i\rangle$ the density matrix is $\rho^i = |\phi_i\rangle\langle\phi_i|$. In order to guess the key $a = v \cdot i$, Eve must now distinguish between two *ensembles* of states: The ensemble of equally likely states ρ^i (these states are equally likely due to Corollary 3.2), with $i_I \in \mathcal{C}_\xi$ (i.e. $i_I P_{\mathcal{C}}^\top = \xi$) and key $a = i_I \cdot v = 0$, and the ensemble of (equally likely) states ρ^i with $i_I \in \mathcal{C}_\xi$ and key $a = i_I \cdot v = 1$. For $a \in \{0, 1\}$ these ensembles are represented by the density matrices $\rho_0 = \rho_0(v, \xi)$ and $\rho_1 = \rho_1(v, \xi)$ defined by:

$$\rho_a(v, \xi) = \frac{1}{2^{n-(r+1)}} \sum_{\substack{i_I P_{\mathcal{C}}^\top = \xi \\ i_I \cdot v = a}} \rho^i \tag{4.10}$$

and Eve's goal is to distinguish between those two. Note that the two density matrices $\rho_a(v, \xi)$ are the lift-ups of the density matrices really known to Eve, namely, matrices in which the sum is over the states of Eq. (4.1) rather than a sum over their purifications.

A good measure for the distinguishability of $\rho_0(v, \xi)$ and $\rho_1(v, \xi)$ is the optimal mutual information (known as the accessible information) that one could get if one needs to guess the bit a by performing an optimal measurement to distinguish between the two density matrices, when the two are given with equal probability (of half). This information will be called the *Shannon Distinguishability* ($SD = SD(\rho_0, \rho_1)$) to emphasize that it is a distinguishability measure. If v is the string used to define the one-bit key $\mathbf{A}$ sent by Alice, then, due to the optimality of SD, we get (for any symmetric attack)

$$I(\mathbf{A}; \mathbf{E}^{\text{sym}} | i_T, j_T, b, s, \xi) \leq SD(\rho_0(v, \xi), \rho_1(v, \xi)) \tag{4.11}$$

where $\mathbf{E}^{\text{sym}}$ is the random variable corresponding to Eve's actual measurement in the symmetrized attack.

Let $v_1, \ldots, v_r$ be the rows of the $r \times n$ parity check matrix $P_{\mathcal{C}}$ of the (n, k, d) code $\mathcal{C}$ where $r = n - k$. The matrix $P_{\mathcal{C}}$ is assumed of rank r and so, the r "parity-check strings" $v_1, v_2, \ldots v_r$ (that are known to Eve) are linearly independent. Let V_r be the r-dimensional linear space generated by $\{v_1, \ldots, v_r\}$. Then, $V_r = \{v_{\mathbf{s}} \mid \mathbf{s} \in \{0,1\}^r\}$ where, by definition[10] $v_{\mathbf{s}} = \sum_{l=1}^{r} s_l v_l$. For any $v_{\mathbf{s}} \in V_r$, Eve knows $i_I \cdot v_{\mathbf{s}}$ because she knows all the ξ_l and $i_I \cdot v_{\mathbf{s}} = \xi_{\mathbf{s}}$ where $\xi_{\mathbf{s}} = \sum_{l=1}^{r} s_l \xi_l$. As a consequence, Eve has total knowledge of the key if $a = i_I \cdot v_{\mathbf{s}}$ for $v_{\mathbf{s}} \in V_r$. Notice that V_r is nothing but the dual code $\mathcal{C}^{\perp}$ of $\mathcal{C}$ which can be viewed as the set of all the parity strings for $\mathcal{C}$.

For any $v \in \{0,1\}^n$, let $\hat{v}$ be the minimum Hamming distance $d_H(v, \mathcal{C}^{\perp})$ between v and all the strings in $\mathcal{C}^{\perp}$. This means that

$$\hat{v} = \min_{v' \in \mathcal{C}^{\perp}} d_H(v, v') = \min_{v' \in \mathcal{C}^{\perp}} |v \oplus v'| \, .$$

The value $\hat{v}$ will prove to be a security parameter. We use here, as in [7], Eve's purified states $|\phi_i\rangle = \sum_l (-1)^{i \cdot l} d_l |\hat{\eta}_l\rangle$, and the resulting density matrices of Eq. (4.10).

We now show that

Lemma 4.2 *For any $\xi \in \{0,1\}^r$, any (n, k, d) code $\mathcal{C}$ with $r \times n$ parity check matrix $P_{\mathcal{C}}$ of rank $r = n - k$ and any $v \notin \mathcal{C}^{\perp}$ the Shannon distinguishability between the parity 0 and the parity 1 of the information bits over the PA string, v, is bounded above by the following inequality:*

$$SD(\rho_0(v, \xi), \rho_1(v, \xi)) \leq 2 \sqrt{\sum_{|l| \geq \frac{\hat{v}}{2}} d_l^2} \, , \tag{4.12}$$

where $\hat{v} = d_H(v, \mathcal{C}^{\perp})$ is the minimum Hamming distance between v and $\mathcal{C}^{\perp}$ and $\rho_b(v, \xi)$ is defined by Eq. (4.10).

See proof in Appendix D.2. As that proof was developed from methods used in [7] we present in Appendix D.1 the preliminary analysis we did for the joint attack, an analysis that was based on using the tools of [7]. Appendix D.2 then presents improved tools leading to the result described in Lemma 4.2. Appendix D.2 is self contained yet reading Appendix D.1 may help the reader to better understand the motivation and the development of the tools used for this proof.

The result of Lemma 4.2 gives an upper bound for Eve's information about the bit defined by this privacy amplification string v. In order to get a useful result, namely, an *information versus disturbance* result, we now prove a proposition in which the bound on Eve's information is expressed in terms of the probability of error on the information bits *in the conjugate basis*.

Proposition 4.2 *For any $\xi \in \{0,1\}^r$, any (n, k, d) code $\mathcal{C}$ with $r \times n$ parity check matrix $P_{\mathcal{C}}$ of rank $r = n - k$ and any $v \notin \mathcal{C}^{\perp}$*

$$I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid i_T, j_T, b, s, \xi) \leq 2 \sqrt{P^{\text{sym}} \left[|\mathbf{C}_I| \geq \frac{\hat{v}}{2} \mid i_T, j_T, b^0, s \right]} \tag{4.13}$$

[10] Note that the vector $\mathbf{s}$ is used now to define the possible vectors $v_{\mathbf{s}}$ in the span of the parity-check strings [this is in addition to s being used as the $2n$-bit string defining the test bits and the information bits]; The bit s_l is the l'th bit of $\mathbf{s}$.

where $\hat{v} = d_H(v, \mathcal{C}^\perp)$ *is the minimum Hamming distance between* v *and* $\mathcal{C}^\perp$, $c_I = i_I \oplus j_I$, $\xi = i_I P_{\mathcal{C}}^\top$, *the key is* $a = i_I \cdot v$ *and* $b^0 = b \oplus s$.

Proof

$$\begin{aligned} I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid i_T, j_T, b, s, \xi) &\leq SD(\rho_0(v,\xi), \rho_1(v,\xi)) && \text{by Eq. (4.11)} \\ &\leq 2\sqrt{\sum_{|l| \geq \frac{\hat{v}}{2}} d_l^2} && \text{by Lemma (4.2)} \\ &= 2\sqrt{\sum_{|l| \geq \frac{\hat{v}}{2}} P^{\text{sym}}[\mathbf{C}_I = l \mid i_T, j_T, b^0, s]} && \text{by Lemma (4.1)} \\ &= 2\sqrt{P^{\text{sym}}\left[|\mathbf{C}_I| \geq \frac{\hat{v}}{2} \mid i_T, j_T, b^0, s\right]}\,. \end{aligned}$$ □

Notice that the bound obtained in the previous proposition holds for all ξ, that is, it is the same whatever is the syndrome sent by Alice to Bob to help him correct his information bits.

Equation (4.13) bounds the information of Eve (about a one-bit key) using the probability of the error strings in the other basis, and it completes the basic *information versus disturbance* result of our proof. Previous security proofs (for simpler attacks), such as [17,9,7] are also based on various *information versus disturbance* arguments, since the non-classicality of QKD is manifested via such arguments.

The result is expressed using classical terms: Eve's information is bounded using the probability of error strings with large Hamming weight. If only error strings with low weight have non-zero probability, Eve's information becomes zero. Such a result is a "low weight" property and it resembles a similar result with this name which was derived by Yao [35] for the security analysis of the error-free quantum oblivious transfer (and QKD).

4.5 Bounds on Eve's information – the m-bit key case

The case of an m-bit key a is closely related to the one-bit case. The only differences are that the upper bound is multiplied by m, and that $\hat{v}$ is defined differently in order to take into account the privacy amplification code (in addition to the error-correction code).

In terms of the bound [the R.H.S. of Eq. (4.13)], the case of an m bit key a follows from that of a one-bit key if we use the following lemma:

Lemma 4.3 *Let* $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_m)$ *be defined by* m *random variables. Let* $\mathbf{E}$ *be any random variable. If* $I(\mathbf{A}_1; \mathbf{E}) \leq F$ *and for all* j, $1 \leq j \leq m-1$ *and all* $a_1, \ldots, a_j$, $I(\mathbf{A}_{j+1}; \mathbf{E} \mid a_1 \ldots a_j) \leq F$ *then* $I(\mathbf{A}; \mathbf{E}) \leq mF$.

Proof Note that

$$I(\mathbf{A}_{j+1}; \mathbf{E} \mid \mathbf{A}_1 \ldots \mathbf{A}_j) = \sum_{a_1 \ldots a_j} P(a_1, \ldots, a_j) I(\mathbf{A}_{j+1}; \mathbf{E} \mid a_1 \ldots a_j)$$

$$\leq \sum_{a_1 \ldots a_j} P(a_1, \ldots, a_j) F \;\; \leq \;\; F \; .$$

The lemma follows from the above and the chain rule for information (see Appendix B.1),

$$I(\mathbf{A}; \mathbf{E}) = I(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_m; \mathbf{E}) = \sum_{j=1}^{m} I(\mathbf{A}_j; \mathbf{E} \mid \mathbf{A}_1, \ldots, \mathbf{A}_{j-1}) \; .$$

□

Next, in the particular case at hand, we want to bound Eve's information about the m-bit key given the values i_T, j_T, b, s and ξ she learned. This means we want to bound $I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid i_T, j_T, b, s, \xi)$ where $\mathbf{A}$ is the m-bit key. This is nothing but a mutual information between $\mathbf{A}$ and $\mathbf{E}^{\text{sym}}$ for some fixed (known) values of random outputs, and the above lemma thus applies. More precisely, it tells us that if some number F is an upper bound for $I(\mathbf{A}_{j+1}; \mathbf{E}^{\text{sym}} \mid i_T, j_T, b, s, \xi, a_1 \ldots a_j)$ then mF will be an upper bound for $I(\mathbf{A}; \mathbf{E}^{\text{sym}}; i_T, j_T, b, s, \xi)$. Announcing ξ and $a_1 \ldots a_j$ is announcing publicly the bits $v_1 \cdot i_I$, ..., $v_{r+j} \cdot i_I$, which is just the same as using the $r + j$ strings $v_1, \ldots v_{r+j}$ as parity strings of a code for which Proposition 4.2 applies. More formally,

Proposition 4.3 *Let $v_1, \ldots, v_{r+m}$ be $r+m$ linearly independent n-strings and $V_{r'}$ be the subspace of $\{0,1\}^n$ spanned by $\{v_1, \ldots, v_{r'}\}$ ($1 \leq r' \leq r+m$). Let $P_{\mathcal{C}}$ be the matrix whose rows are $v_1, \ldots, v_r$ and $P_{\mathcal{PA}}$ the one with rows $v_{r+1}, \ldots, v_{r+m}$. Then for any $\xi \in \{0,1\}^r$*

$$I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid i_T, j_T, b, s, \xi) \leq 2m \sqrt{P^{\text{sym}} \left[|\mathbf{C}_I| \geq \frac{\hat{v}}{2} \mid i_T, j_T, b^0, s \right]} \qquad (4.14)$$

where $\hat{v} = \min_{r \leq r' < r+m} d_H(v_{r'+1}, V_{r'})$, $c_I = i_I \oplus j_I$, $\xi = i_I P_{\mathcal{C}}^{\top}$, $a = i_I P_{\mathcal{PA}}^{\top}$ and $b^0 = b \oplus s$.

Proof See Appendix C.5.

If we modify $\hat{v}$ to any value that is less than or equal to the minimum over all the Hamming distances $d_H(v_{r'+1}, V_{r'})$ then equation (4.14) is satisfied with the modified $\hat{v}$ as well, as only the RHS increases. In particular this is true if we follow the definition given in Subsection 2.1 in item II. 7; thus we define $\hat{v}$ to be (from now on) the minimal distance between any string v in the set of PA parity-check strings, and any string v' in the span of their union with the parity-check-strings of the ECC (the dual to the code). This formally means:

Corollary 4.2 *Let $v_1, \ldots, v_{r+m}$ be $r + m$ linearly independent n-strings. Let $P_{\mathcal{C}}$ be the matrix whose rows are $v_1, \ldots, v_r$ and $P_{\mathcal{PA}}$ the one with rows $v_{r+1}, \ldots, v_{r+m}$. Let $V_{r'}^{\text{exc}}$ be the 2^{r+m-1}-dimensional subspace of $\{0,1\}^n$*

spanned by a subset of the $r+m-1$ *parity strings which excludes the PA string* $v_{r'}$ *(namely, the span of* $v_1,\ldots,v_{r'-1},v_{r'+1},\ldots,v_{r+m}$*). Then for any* $\xi\in\{0,1\}^r$

$$I(\mathbf{A};\mathbf{E}^{\text{sym}}\mid i_T,j_T,b,s,\xi)\leq 2m\sqrt{P^{\text{sym}}\left[|\mathbf{C}_I|\geq\frac{\hat{v}}{2}\mid i_T,j_T,b^0,s\right]} \tag{4.15}$$

where $\hat{v}=\min_{r+1\leq r'\leq r+m} d_H(v_{r'},V_{r'}^{\text{exc}})$, $c_I=i_I\oplus j_I$, $\xi=i_IP_{\mathcal{C}}^{\top}$, $a=i_IP_{\mathcal{PA}}^{\top}$ *and* $b^0=b\oplus s$.

[First remark: in fact, for binary linear codes, the two $\hat{v}$ defined above, the one used in Proposition 4.3 and the one used in Proposition4.2 are equal, but this fact is irrelevant for our paper.
Second remark: we could even follow a stricter definition and replace $\hat{v}$ by $d^{\perp}$, the minimum (non-zero) distance of the code V_{r+m} of dimension $r+m$ (the space spanned by the ECC and PA strings $v_1,\ldots,v_{r+m}$, see Subsection 2.1, item II. 7). Notice that the rows of the generator matrix of this code are those of $P_{\mathcal{C}}$ and $P_{\mathcal{PA}}$.]

5 Completing the Security Proof

In this section we analyze the attack on the test and information qubits together (cf Eq. 3.16). For these states, we bound the weighted average of Eve's information $\langle \mathbf{I}'_{Eve} \rangle$, used in the alternative security criteria [see Eq. (2.4)]:

$$\sum_{c_T | \mathbf{T}=\text{pass}} P\left[\mathbf{C}_T = c_T\right] \; I(\mathbf{A}; \mathbf{E} | \mathbf{I}_T, \mathbf{C}_T = c_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}) \; .$$

We show that the above bound is exponentially small and therefore Lemma 2.1 promises us that security is achieved. We generalize here previous (and more limited) proofs [5,8,7] that information about parity bits is exponentially small, to be applicable for the most general attack on the channel — the joint attack. [A remark: We freely switch below between c_T and j_T whenever i_T is given.]

5.1 Applying the bounds to all attacks

The maximum error rate that still passes the test is denoted p_a (or p_{allowed}). This means that $\mathbf{T} =$ pass if and only if $|c_T| \leq n p_a$. For $\hat{v}$ as defined in Corollary 4.2, and making use of that corollary we get, for fixed b and s:

Lemma 5.1

$$\sum_{|c_T| \leq n p_a} P^{\text{sym}}\left[\mathbf{C}_T = c_T | b, s\right] I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi})$$
$$\leq 2m \sqrt{P^{\text{sym}}\left[(|\mathbf{C}_I| > \frac{\hat{v}}{2}) \wedge (\frac{|\mathbf{C}_T|}{n} \leq p_a) \mid b^0, s \right]} \; .$$

The proof is given in Appendix C.6.

Let U (and $\mathcal{E}$) be some arbitrary attack and $\{U^{\text{sym}}, \mathcal{E}^{\text{sym}}\}$ an arbitrary symmetrized attack resulting from U. As the Lemma above is true for any symmetric attack, it is also true for any $\{U^{\text{sym}}, \mathcal{E}^{\text{sym}}\}$ and in particular for the optimal one (in which the optimal POVM is performed for each value of $i_T, b, \ldots$) Thus, we immediately get from Lemma 5.1

Corollary 5.1

$$\sum_{|c_T| \leq n p_a} P^{\text{sym}}\left[\mathbf{C}_T = c_T | b, s\right] \max I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi})$$
$$\leq 2m \sqrt{P^{\text{sym}}\left[(|\mathbf{C}_I| > \frac{\hat{v}}{2}) \wedge (\frac{|\mathbf{C}_T|}{n} \leq p_a) \mid b^0, s \right]}$$

with the maximum [$\max I(\;)$] defined in Eq. (3.13).

We now prove that the above bound, with the same definition of $\hat{v}$, also applies to the original unsymmetrized attack (b and s still fixed).

Lemma 5.2

$$\sum_{|c_T|\leq np_a} P\left[\mathbf{C}_T = c_T|b,s\right] I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) \leq 2m\sqrt{P\left[(|\mathbf{C}_I| > \frac{\hat{v}}{2}) \wedge (\frac{|\mathbf{C}_T|}{n} \leq p_a) \mid b^0, s\right]}$$

Proof This follows from Lemma 3.3, Corollary 5.1 and equations (3.11, 3.12) from Corollary 3.1:

$$\begin{aligned}
&\sum_{|c_T|\leq np_a} P\left[\mathbf{C}_T = c_T|b,s\right] I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) \\
&\quad = \sum_{|c_T|\leq np_a} P^{\text{sym}}\left[\mathbf{C}_T = c_T|b,s\right] I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) && \text{by Eq. (3.12)} \\
&\quad \leq \sum_{|c_T|\leq np_a} P^{\text{sym}}\left[\mathbf{C}_T = c_T|b,s\right] \max I(\mathbf{A};\mathbf{E}^{\text{sym}} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) && \text{by Lemma 3.3} \\
&\quad \leq 2m\sqrt{P^{\text{sym}}\left[(|\mathbf{C}_I| > \frac{\hat{v}}{2}) \wedge (\frac{|\mathbf{C}_T|}{n} \leq p_a) \mid b^0, s\right]} && \text{by Corollary 5.1}
\end{aligned}$$

By Eq. (3.11), $P^{\text{sym}}\left[\mathbf{C}_I = c_I, \mathbf{C}_T = c_T \mid b, s\right] = P\left[\mathbf{C}_I = c_I, \mathbf{C}_T = c_T \mid b, s\right]$ for any basis string, in particular b^0; this concludes the proof. □

From now on, there will be no restriction of symmetry on the attacks. The results will hold for any attack whatsoever.

5.2 Exponentially-small bound on Eve's information

For any ϵ_{sec} and p_a, such that $\hat{v} \geq 2n(p_a + \epsilon_{\text{sec}})$ Lemma 5.2 leaves the following bound:

Corollary 5.2

$$\sum_{|c_T|\leq np_a} P\left[\mathbf{C}_T = c_T|b,s\right] I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) \leq 2m\sqrt{P\left[(\frac{|\mathbf{C}_I|}{n} > p_a + \epsilon_{\text{sec}}) \wedge (\frac{|\mathbf{C}_T|}{n} \leq p_a) \mid b^0, s\right]}$$

.

Thus far, there is nothing that causes the bound on the right hand side to be a small number. The result above is true even if Eve is told in advance the bases of Alice and Bob (the string b), or if she is told in advance which are the test bits and which are the used bits (the string s), two cases in which Eve easily obtains full information about the secret key a.

Only Eve's lack of knowledge regarding the random strings b and s provides an exponentially small number at the right hand side. Since Eve must fix her attack *before* she knows the basis or the test-bits choice, we compute the average information for a fixed attack over all bases b and test-bits choice s. Averaging over b means that we sum over all b's and multiply each term by the constant $p(b) = 1/2^{2n}$. The averaging over b removes the dependence on the particular basis [due to $\sum_b p(z|b)p(b) = \sum_b p(z,b) = p(z)$].

Averaging over s means that we sum over all s's and multiply each term by the constant $p(s) = 1/\binom{2n}{n}$. The averaging over s removes the dependence on the particular choice of which bits are the test bits [due to $\sum_s p(z|s)p(s) = \sum_s P(z,s) = p(z)$].

Lemma 5.3 Let $\mathbf{T}$ = pass iff $|c_T| \leq np_a$, and let $\mathbf{I}'_{Eve}$ be the random variable equal to $\mathbf{I}_{Eve} = I(\mathbf{A};\mathbf{E} \mid i_T, j_T, b, s, \xi)$ when $\mathbf{T}$ = pass and $\mathbf{I}'_{Eve} = 0$ otherwise. Then for any $\epsilon_{\rm sec}$ and p_a such that $p_a + \epsilon_{\rm sec} \leq \hat{v}/2n$ we get

$$\langle \mathbf{I}'_{Eve} \rangle \leq 2m \sqrt{P\left[(\frac{|\mathbf{C}_I|}{n} > p_a + \epsilon_{\rm sec}) \wedge (\frac{|\mathbf{C}_T|}{n} \leq p_a) \right]}\,.$$

Proof We already proved (Eq. 2.4) that

$$\langle \mathbf{I}'_{Eve} \rangle = \sum_{c_T | \mathbf{T}=\text{pass}} P\left[\mathbf{C}_T = c_T\right]\; I(\mathbf{A};\mathbf{E}|\mathbf{I}_T, \mathbf{C}_T = c_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi})$$

where $\mathbf{T}$ = pass iff $|c_T| \leq np_a$. Expanding the right-hand side, we get

$$\langle \mathbf{I}'_{Eve} \rangle = \sum_{b,s} p(b,s) \sum_{|c_T| \leq np_a} P\left[\mathbf{C}_T = c_T | b, s\right] I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}).$$

Using Corollary 5.2 we obtain the first bound below; then using the fact that $\sum_i p_i \sqrt{x_i} \leq \sqrt{\sum_i p_i x_i}$, and that $p(b,s) = p(b^0,s) = 2^{-2n} p(s)$ (b and s being chosen independently) we get the second bound; finally noting that summing over b is the same as summing over b^0, we get the third bound:

$$\begin{aligned}
\langle \mathbf{I}'_{Eve} \rangle &\leq \sum_{b,s} p(b,s) 2m \sqrt{P\left[(\frac{|\mathbf{C}_I|}{n} > p_a + \epsilon_{\rm sec}) \wedge (\frac{|\mathbf{C}_T|}{n} \leq p_a) \mid b^0, s \right]} \\
&\leq 2m \sqrt{\sum_{b,s} 2^{-2n} p(s) P\left[(\frac{|\mathbf{C}_I|}{n} > p_a + \epsilon_{\rm sec}) \wedge (\frac{|\mathbf{C}_T|}{n} \leq p_a) \mid b^0, s \right]} \\
&= 2m \sqrt{P\left[(\frac{|\mathbf{C}_I|}{n} > p_a + \epsilon_{\rm sec}) \wedge (\frac{|\mathbf{C}_T|}{n} \leq p_a) \right]}\,.
\end{aligned}$$

□

For a long string, the test bits and the information bits should have a similar number of errors if the test is picked at random. The probability that they have different numbers of errors should go to zero exponentially fast as shown in the following lemma.

Lemma 5.4 For any $\epsilon > 0$, $P\left[(\frac{|\mathbf{C}_I|}{n} > p_a + \epsilon) \wedge (\frac{|\mathbf{C}_T|}{n} \leq p_a)\right] \leq e^{-\frac{1}{2}n\epsilon^2}$.

Proof This follows directly from Hoeffding's law of large numbers [22]. The details are given in Appendix C.7.

5.3 The main results

We are now in a position to state and prove our main results.

Proposition 5.1 *If p_a and ϵ_{sec} and the ECC+PA codes are such that $p_a + \epsilon_{\text{sec}} \leq \hat{v}/2n$ with $\hat{v} = \min_{r'=r+1}^{r+m} d_H(v_{r'}, V_{r'}^{\text{exc}})$ where d_H is the Hamming distance, $v_{r'}$ is a parity-check string, and $V_{r'}^{\text{exc}}$ is the 2^{r+m-1} space which is the span of $v_1, \ldots, v_{r'-1}, v_{r'+1}, \ldots, v_{r+m}$, then*

$$\langle \mathbf{I}'_{Eve} \rangle \leq 2m\sqrt{e^{-\frac{1}{2}n\epsilon_{\text{sec}}^2}}$$

where $\mathbf{I}'_{Eve} = \mathbf{I}_{Eve}$ if $|c_T| = |i_T \oplus j_T| \leq np_a$ (test passed) and $\mathbf{I}'_{Eve} = 0$ otherwise.

Proof This follows immediately from Lemma 5.3 and Lemma 5.4.

Theorem 5.1 *If p_a and ϵ_{sec} and the ECC+PA codes are such that $p_a + \epsilon_{\text{sec}} \leq \hat{v}/2n$ with $\hat{v} = \min_{r'=r+1}^{r+m} d_H(v_{r'}, V_{r'}^{\text{exc}})$ where d_H is the Hamming distance, $v_{r'}$ is a parity-check string, and $V_{r'}^{\text{exc}}$ is the 2^{r+m-1} space which is the span of $v_1, \ldots, v_{r'-1}, v_{r'+1}, \ldots, v_{r+m}$, then for any $A_{\text{info}} > 0$, $A_{\text{luck}} > 0$ such that $A_{\text{info}} A_{\text{luck}} = 2m$ and any β_{info} and β_{luck} such that $\beta_{\text{info}} + \beta_{\text{luck}} = \epsilon_{\text{sec}}^2/4$,*

$$P\left[(\mathbf{T} = \text{pass}) \wedge (\mathbf{I}_{Eve} \geq A_{\text{info}}\, e^{-\beta_{\text{info}} n})\right] \leq A_{\text{luck}}\, e^{-\beta_{\text{luck}} n} \tag{5.1}$$

where $\mathbf{T} = \text{pass}$ iff $|c_T| \leq np_a$ and $\mathbf{I}_{Eve} = I(\mathbf{A}; \mathbf{E} \mid i_T, j_T, b, s, \xi)$.

Proof This follows from Proposition 5.1 if we let $A = 2m$ and $\beta = \epsilon_{\text{sec}}^2/4$ in Lemma 2.1.

Let us recall that, in addition to the security, one must also guarantee the reliability of the final key. Namely we need to make sure that Alice's final key and Bob's final key are (almost always) identical. Note that Lemma 5.4 can be rewritten:

$$P\left[(\mathbf{T} = \text{pass}) \wedge (|\mathbf{C}_I| > (p_a + \epsilon_{\text{rel}})n)\right] \leq e^{-\frac{1}{2}\epsilon_{\text{rel}}^2}$$

This also means that

Corollary 5.3 *The probability that the test is passed and that there are more than $(p_a + \epsilon_{rel})n$ errors in the information string is exponentially small; it is bounded by*

$$h = e^{-\frac{1}{2}n\epsilon_{rel}^2}\ .$$

Once the ECC is chosen such that $(p_a + \epsilon_{\text{rel}})n$ errors in the information string are corrected, Alice's and Bob's final keys identical except for an exponentially small probability bounded by h. This result means that $A_{\text{rel}} = 1$ and $\beta_{\text{rel}} = \epsilon_{\text{rel}}^2/2$, in the reliability criterion of Subsection 2.5.

5.4 The existence of codes that provide security and reliability

The above bound on Eve's information is exponentially small, provided there is a family of good linear ECCs satisfying also the requirement that $\hat{v} \geq 2n(p_a + \epsilon_{\rm sec})$ when PA strings are added. What we formally need is a family of (linear) ECC+PA codes satisfying the following two conditions:

(1) The ECC can correct up to $t = p_{\rm allowed} + \epsilon_{\rm rel}$ errors. For this to happen, we demand that the minimum distance d between the code words of the ECC satisfy $d \geq 2t+1$. Hence, a $d \geq 2t+1 = 2n(p_{\rm allowed}+\epsilon_{\rm rel})+1$ is sufficient. This code can correct all the errors in the information string, except for an exponentially small probability bounded by h (of Corollary 5.3) of having more errors in the information string than expected.
(2) The minimum distance $d^{\perp}$, of the code words in the span of the dual code and the PA strings (hence, the augmented dual code is of dimension $r+m$) should have a minimum distance $d^{\perp} \geq 2n(p_{\rm allowed} + \epsilon_{\rm sec})$.

We discuss below the class of linear codes called random linear codes. Such codes cannot be easily decoded hence their practical usefullness is limited. It may well be that such codes can be replaced by the much more practical codes — the Reed-Solomon codes, without losing the security and reliability proven below. However, analyzing Reed-Solomon codes is beyond the scope of this work.

For random linear codes (RLC's) the two requirements mentioned above can easily be satisfied. We can generate an m-bit secret key if we pick an $(n, n-r)$ RLC, where r and m satisfy

$$H_2(2p_a + 2\epsilon_{\rm rel} + 1/n) < r/n$$
$$H_2(2p_a + 2\epsilon_{\rm sec}) + H_2(2p_a + 2\epsilon_{\rm rel} + 1/n) < 1 - R_{\rm secret}\ ,$$

with H_2 the entropy, and $R_{\rm secret} \equiv m/n$ the bit-rate (namely, the efficiency of the QKD scheme). If these conditions are not met then the random linear code provides neither reliability nor security; see Appendix E. At the limit of large n and ϵ's close to zero we get as a bound $2H_2(2p_a) < 1$. Then, $p_{\rm allowed} < 5.50\%$ satisfies the bound and hence this is our first threshold [see Appendix E for the detailed calculation]. It is the threshold in the case in which we want to have an exact bound on Eve's information and on the reliability of the final key, as a function of parameters chosen by the designer of the QKD protocol. This is important for a designer who needs to choose a sufficiently large n (that is not assumed to go to infinity); then Eve's information is bounded as in Proposition 5.1 and the reliability is bounded as in Corollary 5.3.

Note that if we let $p_{\rm allowed}$ be sufficiently close to zero then (for sufficiently large n and small ϵ's) a bit-rate $R_{\rm secret}$ close to one can be obtained. Specific values of Eve's information, the probability of error in the final key, and the resulting bit-rate are provided in Table 5.1; this is done by choosing $\epsilon_{\rm sec} = \epsilon_{\rm rel} = \epsilon$ (for the sake of simplicity). As the parameters n, ϵ, and $p_{\rm allowed}$ can be chosen by the designer of the protocol, we present here 3 values of the reliability/security parameter, and we

		$\epsilon = 0.5\%$	$\epsilon = 1\%$	$\epsilon = 2\%$
Reliability	$n = 12500$		0.54	1/12
Bound (h)	$n = 50000$	0.54	1/12	1/22026
	$n = 200000$	1/12	1/22026	$4 \cdot 10^{-18}$
	$n = 800000$	1/22026	$4 \cdot 10^{-18}$	$\approx 10^{-70}$
	$n = 3200000$	$4 \cdot 10^{-18}$	$\approx 10^{-70}$	

		$\epsilon = 0.5\%$	$\epsilon = 1\%$	$\epsilon = 2\%$
Rate ($R_{\text{secret}} = m/n$)	$P_{\text{allowed}} = 2.0\%$	41.7%	33.5%	18.5%
	$P_{\text{allowed}} = 3.5\%$	18.5%	11.7%	0.007%*
	$P_{\text{allowed}} = 5.0\%$	0.007%*	†	†

† Out of the allowed range (negative rate)

* For the case of $2P_{\text{allowed}} + 2\epsilon = 11.0\%$ we calculate R_{secret} by solving $H_2(2p_a + 2\epsilon) + H_2(2p_a + 2\epsilon + 1/n) = 0.9999 - R_{\text{secret}}$. Here, security and reliability can be obtained only with $n > 10^6$ or so

Table 5.1 Summary of the characteristics of a QKD protocol that uses RLC: The "Reliability Bound", h, is calculated according to Corollary 5.3, and the maximal bit rate R_{secret} is calculated by solving $H_2(2p_a + 2\epsilon) + H_2(2p_a + 2\epsilon + 1/n) = 0.99 - R_{\text{secret}}$ (with two exceptions, denoted with * in the table). The parameters in this table are closely related to the parameters used in experiments: n is related to the number of photons obtained by Bob; $2n$ photons are used according to the used-bits-BB84 protocol and slightly more than $4n$ in the conventional BB84. The error rate considered here is achieved in many experimental setups, but might limit the distance of transmission. A photon rate of 1000 photons per second (if we count the photons obtained by Bob) was also reported in various experiments, so the resulting secret-key bit-rate R_{secret} can be sufficient for many practical usages.

then calculate[11] the reliability as a function of n, and we calculate[12] the maximal bit-rate as a function of P_{allowed}.

The "Reliability Bound" h is calculated according to Corollary 5.3, and (due to the equal ϵ's) we can then get the bound on Eve's information (according to Proposition 5.1), which is exactly $2m\sqrt{h}$. We consider the numbers we got for the "Reliability Bound" in the table to be "Good" when the probability of error is $1/22026$ or below. However, with $h = 1/22026$, Eve's information is $2m$ times $1/148$ which means that the users cannot really enjoy the allowed bit-rate, and must use a much smaller value for m, as Eve could then learn too much. When the "Reliability Bound" is $4 \cdot 10^{-18}$ or $\approx 10^{-70}$ there is clearly no problem at all with Eve's information, and m can be as large as the allowed bit-rate enables.

For RLC one can actually obtain a better threshold for the allowed error rate (as first noticed by Mayers [27]), by modifying requirement (1) so that:

[11] The term $1/n$ that appears in the parameter $[2p_{\text{allowed}} + 2\epsilon + 1/n]$ is negligible except in the two cases where the entire term approaches 11.0%.

[12] We choose a maximal bit rate by solving $H_2(2p_a + 2\epsilon) + H_2(2p_a + 2\epsilon + 1/n) = 0.99 - R_{\text{secret}}$.

(1') The ECC can correct up to $p_{\rm allowed} + \epsilon_{\rm rel}$ errors, with probability as close to 1 as we wish.

Namely, for any $\hat{\delta}$, the ECC can correct up to $p_{\rm allowed} + \epsilon_{\rm rel}$ errors, with probability smaller than $\hat{\delta}$. For RLC this is true (due to Shannon's bound, see for instance [25]) for any code having a minimum distance $d \geq t+1 = n(p_{\rm allowed} + \epsilon_{\rm rel}) + 1$ (rather than $d \geq 2t+1$, that promises the success of correcting all errors), provided that $r/n > H_2(p_{\rm allowed} + \epsilon_{\rm rel})$, and that a sufficiently large n is chosen.

We show in Appendix E that requirements (1') and (2) can be satisfied and one can generate an m-bit secret key, if one picks an $(n, n-r)$ RLC, where r and m satisfy the following:

$$H_2(p_a + \epsilon_{\rm rel} + 1/n) < r/n$$
$$H_2(2p_a + 2\epsilon_{\rm sec}) + H_2(p_a + \epsilon_{\rm rel} + 1/n) < 1 - R_{\rm secret}\ ,$$

where $R_{\rm secret} \equiv m/n$. In the limit of large n and ϵ's close to zero we get as a bound $H_2(2p_a) + H_2(p_a) < 1$. Then, $p_{\rm allowed} < 7.56\%$ satisfies the bound and hence this is our improved threshold (which is identical to the threshold calculated by Mayers [27]). Note that Eve's information is still bounded to be exponentially small due to Theorem 5.1, but the reliability is now bounded only asymptotically as we did not find an explicit formula for the probability $\hat{\delta}$ of having an error (as a function of n) when the distance is $d > t+1$.

Asymptotically, with a rate $R_{\rm secret} < 1 - H_2(p_a) - H_2(2p_a)$ the final key is secure and reliable for the given ECC+PA. Note, as p_a goes to zero, $R_{\rm secret}$ goes to 1, which means that (asymptotically) almost all the information bits are secret.

This threshold is based on the properties of the code, and other codes might give worse thresholds, but might have other desired properties. Random linear codes are not so useful as their decoding cannot be done efficiently. It is possible to make use of methods for approximate decoding (in which we are not always promised that the closest code word is chosen after the error correction), but the bound on reliability then need some adjustments. It might be better to replace the RLC by a code that can be decoded efficiently (e.g., Reed-Solomon concatenated code, with a random seed), and add random PA strings. The Hamming distance between the PA check-strings and the ECC check-strings is still bounded below in the same way as for the RLC (see [27]).

Finally, it is interesting to note that the bound $H_2(p_a) + H_2(p_a) < 1$ (which was neither reported by us nor by Mayers) leads to the threshold of 11%, and such threshold was reported and proven by Shor and Preskill [33]. This probably means that the alternative proof presented there can, in some sense, modify requirement (2) in a way similar to the modification done here to change from (1) to (1') above. However, we could not see how the same modification could apply to our proof.

A well-known way to improve the threshold further is to allow two-way communication as part of the ECC+PA process. This technique is known as key distillation, see the basic idea described in [13]. The analysis of Eve's density matrices

becomes much more complicated in such a case, and we do not yet know if our proof can easily be adjusted to allow that[13].

6 Summary

We proved the security of the Bennett-Brassard (BB84) protocol for quantum key distribution. Our proof is based on analyzing Eve's reduced density matrices, on a novel *information versus disturbance* result, on the optimality of symmetric attacks, on laws of large numbers, and on various techniques that simplify the analysis of the problem.

Many of the ideas and the tools developed here can be found relevant when proving the security of other QKD schemes: the analysis of Eve's reduced density matrices, the purifications of her states, the usage of that purification for finding a relevant information versus disturbance bound, the use of Hoeffding's law of large numbers, the trace-norm-difference bound, etc. Other tools, such as the reduction to the used-bits-BB84 protocol, and the extensive usage of symmetry could also provide some important insight, but are somewhat more specific to the BB84 scheme.

7 Acknowledgement

The work of T.M. is supported in part by the Israel MOD Research and Technology Unit. The work of M.B. is supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada. The work of E.B. is supported in part by the European Commission through the IST Programme under contract IST-1999-11234. The work of P.O.B., T.M., and V.R., is supported in part by the Defense Advanced Research Projects Agency (DARPA) project MDA972–99–1–0017, by the U.S. Army Research Office/DARPA DAAD19–00–1–0172, by Grant No. 530-1415-01 from the DARPA Ultra program, and by Grant No. 961360 from the Jet Propulsion Lab.

[13] After the submission of our paper, Gottesman and Lo proved that the Shor-Preskill proof of security can be adjusted to deal with such a key distillation, yielding an improved threshold for p_{allowed}; see quant-ph/0105121).

References

1. M. Ben-Or, Talk given in NEC workshop on quantum cryptography, 1999.
2. C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, Experimental quantum cryptography, *Journal of Cryptology*, vol. 5 (1992), pp. 3–28.
3. C. H. Bennett and G. Brassard, Quantum cryptography: Public key distribution and coin tossing, in *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India*, IEEE, 1984, pp. 175–179.
4. C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, Generalized privacy amplification, *IEEE Transactions on Information Theory*, vol. IT-41 (1995), pp. 1915–1923.
5. C. H. Bennett, T. Mor, and J. A. Smolin, Parity bit in quantum cryptography, *Physical Review. A.*, vol. 54, no 4 (1996), pp. 2675–2684.
6. E. Biham, M. Boyer, P. O. Boykin, T. Mor and V. Roychowdhury, A proof of the security of quantum key distribution, Proceedings of the 32nd Ann. ACM Symposium on the Theory of Computing (STOC'00), ACM press, New-York, 2000, pp.715–724. See also Quant-ph/9912053.
7. E. Biham, M. Boyer, G. Brassard, J. van de Graaf, and T. Mor, Security of quantum key distribution against all collective attacks, *Algorithmica*, vol. 34 (2002), pp. 372–388. See also Quant-ph/9801022.
8. E. Biham and T. Mor, Security of quantum cryptography against collective attacks, *Physical Review Letters*, vol. 78, no 11 (1997), pp. 2256–2259.
9. E. Biham and T. Mor, Bounds on information and the security of quantum cryptography, *Physical Review Letters*, vol. 79, no 20 (1997), pp. 4034–4037.
10. P. Billingsley, *Probability and measure*, John Wiley & Sons Inc., New York, second edition, 1986.
11. G. Brassard, N. Lütkenhaus, T. Mor, and B. C. Sanders, Limitations of practical quantum cryptography, *Physical Review Letters*, vol. 85, (2000), pp. 1330–1333. See also: Security aspects of practical quantum cryptography, Advances in Cryptology-EuroCrypt'2000, LNCS vol. 1807, Springer-Verlag (2000), pp. 289–298.
12. G. Brassard, T. Mor, and B. C. Sanders, Quantum cryptography via parametric down-conversion, *Proceedings of the Quantum Communication, Computing, and Measurement 2 (QCM'98)* conference, Evanston, Ill., USA, Aug., 1998; Kluwer Academic/Plenum Publishers, New-York, 2000, pp. 381–386. Quant-ph/9906074.
13. G. Brassard and L. Salvail, Secret-key reconciliation by public discussion, Advances in Cryptology-EuroCrypt'1993, LNCS vol. 765, Springer-Verlag (1994), pp. 410–423.
14. T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley and Sons, New York, 1991.
15. E. B. Davies Information and quantum measurement, *IEEE Transactions on Informormation Theory*, vol. IT-24 (1978), pp. 596-599.
16. D. Deutsch, A. Ekert, R. Jozsa, C. Macchiavello, S. Popescu, and A. Sanpera, Quantum privacy amplification and the security of quantum cryptography over noisy channels, *Physical Review Letters*, vol. 77, no 13 (1996), pp. 2818–2821.
17. C. A. Fuchs, N. Gisin, R. B. Griffiths, C.-S. Niu, and A. Peres, Optimal eavesdropping in quantum cryptography. I. Information bound and optimal strategy, *Physical Review. A.*, vol. 56, no 2 (1997), pp. 1163–1172.
18. C. A. Fuchs and J. van de Graaf, Cryptographic distinguishability measures for quantum-mechanical states, *IEEE Transactions on Information Theory*, vol. IT-45, no 4 (1999), pp. 1216–1227.
19. R. C. Gallagher, *Low-density parity-check codes*, The M.I.T. Press, Cambridge, MA, 1963, Chapter 2.

20. N. Gisin, G. G. Ribordy, W. Tittel, and H. Zbinden, Quantum cryptography, *Review of Modern Physics* vol. 74 (2002), pp. 145-195.
21. J. Gruska, *Quantum Computing*, McGraw-Hill Publishers, Berkshire, England, 1999.
22. W. Hoeffding, Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association*, vol. 58 (1963), pp. 13–20.
23. H.-K. Lo, A simple proof of the unconditional security of quantum key distribution, *Yournal of Physics A*, vol 34 (2001), pp. 6957–6968. Quant-ph/9904091, 1999.
24. H.-K. Lo and H. F. Chau, Unconditional security of quantum key distribution over arbitrarily long distances, *Science*, vol. 283 (1999), pp. 2050–2056.
25. F. J. MacWilliams and N. J. A. Sloane, The theory of error correcting codes, North Holland Mathematical Library, Elsevier Science Publishers (1977), Amsterdam, The Netherlands.
26. D. Mayers, Quantum key distribution and string oblivious transfer in noisy channel, Advances in cryptology - CRYPTO'96, LNCS vol. 1109, Springer-Verlag, Berlin (1996), pp. 343–357.
27. D. Mayers, Unconditional security in quantum cryptography, *J. of the ACM* vol. 48 no. 3 (2001), pp. 351–406. See also Quant-ph/9802025.
28. T. Mor, Reducing quantum errors and improving large scale quantum cryptography, Quant-ph/9608025, 1996.
29. T. Mor, Quantum Memory in Quantum Cryptography, Ph.D. Thesis (Technion, Israel, 1997); Quant-ph/9906073.
30. M. Nielsen and I. Chuang, *Quantum computation and quantum information*, Cambridge University Press, Cambridge, 2000.
31. A. Peres, *Quantum theory: concepts and methods*, Kluver Academic Publishers, 1993.
32. P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM Journal on Computing*, vol. 26, no 5 (1997), pp. 1484–1509.
33. P. W. Shor and J. Preskill, Simple Proof of Security of the BB84 Quantum Key Distribution Protocol, *Physical Review Letters*, vol. 85, no 2 (2000), pp. 441–444, See also Quant-ph/0003004.
34. S. Wiesner, Conjugate coding, *Sigact News*, vol. 15 (1983), pp. 77–88.
35. A. C.-C. Yao, Security of quantum protocols against coherent measurements, in *Proceedings of the 26th ACM Symposium on the Theory of Computing*, ACM, 1995, pp. 67–75.
36. H. Zbinden, N. Gisin, B. Huttner, A. Muller, and W. Tittel Practical aspects of quantum key distribution, *J. of Cryptology* vol. 13, no. 2 (2000), pp. 207-220.

A Security of BB84

In the paper we prove that used-bits-BB84 is secure. Let us now present the original BB84 protocol and prove, by reduction, that its security follows immediately from the security of the used-bits-BB84 protocol.

The differences between the protocols are only in the first part. The first part of the BB84 protocol is as follows:

I. Creating the sifted key:
1. Alice and Bob choose a large integer $n \gg 1$, and a number δ_{num}, such that $1 \gg \delta_{\text{num}} \gg 1/\sqrt{(2n)}$. The protocol uses $n'' = (4 + \delta_{\text{num}})n$ bits.
2. Alice randomly selects two n''-bit strings, b'' and i'', which are then used to create qubits: The string b'' determines the basis $0 \equiv z$, and $1 \equiv x$ of the qubits. The string i'' determines the value (0 or 1) of each of the n'' qubits (in the appropriate bases).
3. Bob randomly selects an n''-bit string, b''^{Bob}, which determines Bob's later choice of bases for measuring each of the n'' qubits.
4. Alice generates n'' qubits according to her selection of b'' and i'', and sends them to Bob via a quantum communication channel.
5. After receiving the qubits, Bob measures in the basis b''^{Bob}.
6. Alice and Bob publish the bases they used; this step should be performed only after Bob received all the qubits.
7. All qubits with different bases are discarded by Alice and Bob. Thus, Alice and Bob finally have $n' \approx n''/2$ bits for which they used the same bases b'. The n'-bit string would be identical for Alice and Bob if Eve and natural noise do not interfere.
8. Alice selects the first $2n$ bits from the n'-bit string, and the rest of the n' bits are discarded. If $n' < 2n$ the protocol is aborted (a fake random key can be chosen in this case via the unjammable classical channel, so that the key is not secret; however the probability for this to happen is exponentially small).
   We shall refer to the resulting $2n$-bit string as the sifted key.

The second part of the protocol is identical to the second part of the used-bits-BB84 protocol. To prove that BB84 is secure let us modify BB84 by a few steps in a way that each step can only be helpful to Eve, and the final protocol is the used-bits-BB84. Each item below describes a different protocol, obtained by modifying the previous protocol.

Recall that Alice and Bob choose their strings of basis b'' and b''^{Bob} in advance. Recall that the two strings are random. Thus, the first modification below has no influence at all on the security or the analysis of the BB84 protocol. Note that after the first modification Alice knows the un-used bits in advance. The second and the third modifications are done in a way that Eve can only gain, hence security of the resulting protocol provides the security of BB84. The last modification is only "cosmetic", in order to derive precisely the used-bits-BB84 protocol. This modification changes nothing in terms of Eve's ability.

– Let Bob have a quantum memory. Let Alice choose b''^{Bob} instead of Bob at step 3. When Bob receives the qubits at step 5, let him keep the qubits in a

memory, and tell Alice he received them. In step 6, let Alice announce b''^{Bob} to Bob, and Bob measures in bases b''^{Bob}.

From the announcements of b'' and b''^{Bob} Bob knows which are the used and the un-used bits, as determined in steps 7 and 8. Now, at the end of step 8, Alice and Bob know all the un-used bits, so they ignore them, to be left with $2n$ bits.

Note that in this modified protocol, Alice can calculate which are the un-used bits already at step 3 (if she wishes to know this).

– Let Alice calculate the un-used bits and announce them already at the end of step 3. Let her also announce their bases ($b^{\mathrm{Alice}}_{\text{un-used}}$ and $b^{\mathrm{Bob}}_{\text{un-used}}$) and bits-values $i_{\text{un-used}}$. Obviously, such announcements can only help Eve to gain more information (and maybe even to chose a better attack). Thus this step only reduces the security, so if the protocol defined here is secure, so is the original BB84 protocol.
– Let Alice generate and send to Bob only the used bits in step 4, and let her ask Eve to send the un-used bits (by telling her which these are, and also the preparation data for the relevant subsets, that is—$b^{\mathrm{Alice}}_{\text{un-used}}$ and $i_{\text{un-used}}$). Knowing which are the used bits, and knowing their bases and values can only help Eve in designing her attack, thus security can only be reduced by this step.

  Since Bob never uses the values of the unused bits in the protocol (he only ignores them), he doesn't care if Eve doesn't provide him these bits or provide them to him without following Alice's preparation request.

  After Bob receives the used and unused bits, let him give Eve the unused qubits (without measuring them), and ask her to measure them in bases $b^{\mathrm{Bob}}_{\text{un-used}}$. Having these qubits can only help Eve in designing her optimal final measurement, thus security can only be reduced by this step.

  Since Bob never use the values of the unused bits in the rest of the protocol, he doesn't care if Eve doesn't provide him these values correctly or at all.
– Since Alice and Bob never made any use of the unused bits, Eve could have them as part of her ancilla to start with, and Alice could just create $2n$ bits, send them to Bob, and then tell him the bases.

  The protocol obtained after this reduction, is a protocol in which Eve has full control on her qubits and on the unused qubits. Alice and Bob have control on the preparation and measurement of the used bits only. This is the used-bits BB84, for which we prove security in the text.

One important remark is that the exponentially small probability that $n' < 2n$ in Step 8 (so that the protocol is aborted due to insufficient number of bits in the sifted key) now becomes a probability that Eve learns the key.

Another important remark is that the issue of high loss rate of qubits (e.g., due to losses in transmission or detection) can also be handled via the same reduction. Thus, our proof could apply also to a more practical BB84 protocol where high losses are allowed. The required modification to the protocol then is that Bob now will not add missing qubits, in step I.3 of the used-bits BB84 protocol, and in an additional step (prior to step I.4.) he will inform Alice of the bits he did not obtain.

By the way, another practical aspect is imperfect sources (in which the created states are not described by a two-level system). This subject is the issue of recent

subtlety regarding the security of practical schemes [12, 11], and it is not discussed in this current work.

B Information Theoretic Basics and Results

B.1 Basics of information theory [14]

Let $\mathbf{X}$ and $\mathbf{Y}$ be random variables whose values are indexed by x and y respectively, appearing with probabilities $p(x)$ and $p(y)$. The entropy of a random variable is $H(\mathbf{X}) = -\sum_x p(x)\log_2 p(x)$. For two variables $H(\mathbf{X}|y) = -\sum_x p(x|y)\log_2 p(x|y)$ and $H(\mathbf{X}|\mathbf{Y}) \equiv \sum_y p(y)H(\mathbf{X}|y)$. For any two random variables $\mathbf{X}$ and $\mathbf{Y}$, the mutual information $I(\mathbf{X};\mathbf{Y}) = H(\mathbf{X}) - H(\mathbf{X} \mid \mathbf{Y})$ describes the decrease in the entropy of $\mathbf{X}$ due to learning $\mathbf{Y}$; This function I is symmetric to swapping $\mathbf{X}$ and $\mathbf{Y}$.

For three random variables $\mathbf{A}$, $\mathbf{E}$, and $\mathbf{X}$ given to be x, the conditional mutual information is $I(\mathbf{A};\mathbf{E} \mid x) = H(\mathbf{A} \mid x) - H(\mathbf{A} \mid \mathbf{E}, x)$ Then, the conditional mutual information for the three random variables is $I(\mathbf{A};\mathbf{E} \mid \mathbf{X}) \equiv \sum_x p(x) I(\mathbf{A};\mathbf{E} \mid x)$. Another case which is relevant is with four random variables $\mathbf{A}$, $\mathbf{E}$, $\mathbf{X}$ and $\mathbf{Y}$ given to be equal to y, $I(\mathbf{A};\mathbf{E} \mid \mathbf{X}, y) = \sum_x p(x|y) I(\mathbf{A};\mathbf{E} \mid x, y)$.

An important tool is the chain rule $I(\mathbf{A},\mathbf{B};\mathbf{C}) = I(\mathbf{A};\mathbf{C}) + I(\mathbf{B};\mathbf{C} \mid \mathbf{A})$. As a corollary from the chain rule and the positivity of mutual information, one gets $I(\mathbf{A},\mathbf{B};\mathbf{C}) \geq I(\mathbf{B};\mathbf{C} \mid \mathbf{A})$.

B.2 Bad Security Criteria

B.2.1 A first bad security criterion and the SWAP attack: What one might like to obtain as a security criterion is that Eve's information given that the test is passed, is negligible. Formally, this puts a restriction on the values of j_T: for any i_T, only j_T such that $|j_T \oplus i_T| \leq np_a$ are allowed. Then, the criterion is

$$I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}, \mathbf{T} = \text{pass}) \leq A\, e^{-\beta n} \tag{B.1}$$

with A and β positive constants, and $I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}, \mathbf{T} = \text{pass}) = \sum_{i_T, j_T, b, s, \xi} p(i_T, j_T, b, s, \xi \mid \mathbf{T} = \text{pass}) I(\mathbf{A};\mathbf{E} \mid i_T, j_T, b, s, \xi, \mathbf{T} = \text{pass})$, with $c_T = i_T \oplus j_T$, and $\mathbf{T} = \text{pass}$ meaning that $c_T \leq np_a$.

Unfortunately, the above bound is too demanding and is *not* satisfied in quantum cryptography. Given that the test is passed, Eve can still have full information. Consider the *SWAP attack*: Eve takes Alice's qubits and puts them into a quantum memory. She sends random BB84 states to Bob. Eve measures the qubits she kept after learning their bases, hence gets full information about Alice's final key. In this case, Bob will almost always abort the protocol because it is very unlikely that his bits will pass the test. However, in the rare event when the test is passed, Eve has full information about Alice's key. So, given the test is passed (a rare event), information is still m bits, and the above criterion cannot be satisfied.

B.2.2 A second bad security criteria and the half-SWAP attack: Another potential security criterion says the following: "For any attack, either Eve's average information is negligible or the probability that the test is passed is negligible". Namely, if Eve tries an attack that would give her non-negligible information about a final key, she has to be extremely lucky in order to pass the test. This security criterion can be formally written as $\langle \mathbf{I}_{Eve} \rangle P(\mathbf{T} = \text{pass}) \leq A\, e^{-\beta n}$ with A and β positive constants. This suggested security criterion is different from the previously suggested one, and it is satisfied by the SWAP attack mentioned above.

Unfortunately, as observed in an earlier (archive) version of [27], this criterion is also inappropriate. Consider the *half-SWAP attack* in which Eve does nothing with probability half, and performs the SWAP attack with probability half. This half-SWAP attack gives an average information of exactly m/2, and it passes the test with probability larger than half. Obviously these two cases, getting a non-negligible information, and passing the test with high probability, will not happen in the same event, hence security can still be achieved, but it must be defined via less demanding criteria, such as those two used in the paper.

B.3 Alternative Security Criteria

B.3.1 Finding different expressions for $\langle \mathbf{I}'_{Eve} \rangle$: First, we prove Eq.(2.3) namely, that $\langle \mathbf{I}'_{Eve} \rangle = I(\mathbf{A}; \mathbf{E} \mid \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}, \mathbf{T} = \text{pass}) P[\mathbf{T} = \text{pass}]$.

By expanding of $\langle \mathbf{I}'_{Eve} \rangle$ we get:

$$\begin{aligned}
\langle \mathbf{I}'_{Eve} \rangle &= \sum_{i_T, j_T : |i_T \oplus j_T| \leq n\, p_a} \sum_{b,s,\xi} I(\mathbf{A}; \mathbf{E} \mid i_T, j_T, b, s, \xi) p(i_T, j_T, b, s, \xi) \\
&= \sum_{i_T, j_T, b, s, \xi} I(\mathbf{A}; \mathbf{E} \mid i_T, j_T, b, s, \xi) p(i_T, j_T, b, s, \xi \mid \mathbf{T} = \text{pass}) P(\mathbf{T} = \text{pass}) \\
&= \sum_{i_T, j_T, b, s, \xi} I(\mathbf{A}; \mathbf{E} \mid i_T, j_T, b, s, \xi, \mathbf{T} = \text{pass}) p(i_T, j_T, b, s, \xi \mid \mathbf{T} = \text{pass}) P(\mathbf{T} = \text{pass}) \\
&= [\sum_{i_T, j_T, b, s, \xi} I(\mathbf{A}; \mathbf{E} \mid i_T, j_T, b, s, \xi, \mathbf{T} = \text{pass}) p(i_T, j_T, b, s, \xi \mid \mathbf{T} = \text{pass})] P(\mathbf{T} = \text{pass}) \\
&= I(\mathbf{A}; \mathbf{E} \mid \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}, \mathbf{T} = \text{pass}) P[\mathbf{T} = \text{pass}]
\end{aligned}$$

Indeed, $p(i_T, j_T, b, s, \xi \mid \text{pass}) p(\text{pass}) = p(i_T, j_T, b, s, \xi, \text{pass})$ and this value is equal to $p(i_T, j_T, b, s, \xi)$ if $|i_T \oplus j_T| \leq n p_a$ and is 0 otherwise. When the value is not 0, then the condition pass is automatically satisfied and can be put in the right-hand side of the mutual information.

Second, we prove in full details Eq.(2.4) namely, that $\langle \mathbf{I}'_{Eve} \rangle = \sum_{|c_T| \leq n p_a} P[\mathbf{C}_T = c_T] I(\mathbf{A}; \mathbf{E} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi})$.

Note that $\mathbf{I}'_{Eve}$ is the random variable equal to $I(\mathbf{A}; \mathbf{E} \mid i_T, j_T, b, s, \xi)$ when $|i_T \oplus j_T| \leq n p_a$ (i.e. when $\mathbf{T} = \text{pass}$) and to 0 otherwise. As a consequence,

$$\langle \mathbf{I}'_{Eve} \rangle = \sum_{i_T, j_T, b, s, \xi} \mathbf{I}'_{Eve}(i_T, j_T, b, s, \xi) p(i_T, j_T, b, s, \xi)$$

$$= \sum_{|c_T|\leq np_a} \sum_{i_T,b,s,\xi} I(\mathbf{A};\mathbf{E} \mid i_T, \mathbf{C}_T = c_T, b, s, \xi) P\left[i_T, \mathbf{C}_T = c_T, b, s, \xi\right]$$
$$= \sum_{|c_T|\leq np_a} \sum_{i_T,b,s,\xi} I(\mathbf{A};\mathbf{E} \mid i_T, \mathbf{C}_T = c_T, b, s, \xi) P\left[i_T, b, s, \xi \mid c_T\right] P\left[\mathbf{C}_T = c_T\right]$$
$$= \sum_{|c_T|\leq np_a} P[\mathbf{C}_T = c_T] I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi})$$

B.3.2 Security against the Half-SWAP Attack: In the half-SWAP attack Eve has a probe $|p\rangle$ where p is a $2n$-bit string. With probability half she applies the unitary transform $U_0|p\rangle|i\rangle_b = |p\rangle|i\rangle_b$ (she does nothing and then sends $|i\rangle_b$ to Bob) and with probability half she applies the unitary transform $U_1|p\rangle|i\rangle_b = |i\rangle_b|p\rangle$ (swap) and sends $|p\rangle$ to Bob, keeping the probe in the state $|i\rangle_b$. We can present a fully-quantum attack, and let Eve use an additional single-qubit probe $|e_0\rangle$ initially in the state $H|0\rangle$, so that her full probe contains $2n+1$ qubits. Her attack is defined by the unitary transform

$$U|0\rangle|p\rangle|i\rangle_b = |0\rangle|p\rangle|i\rangle_b$$
$$U|1\rangle|p\rangle|i\rangle_b = |1\rangle|i\rangle_b|p\rangle$$

which means that she uses her additional qubit $|e_0\rangle$ to decide whether she swaps or not (using $2n$ Controlled-SWAP gates). Let us describe Eve's measurement: she measures her new bit e_0 in the standard basis and then, if she gets $e_0 = 1$, she measures the "probe" $|e_1\rangle = |i\rangle_b$ in the basis b and gets i, else she measures her original probe $|p\rangle$ in the standard basis and gets p. Her two outputs (e_0, e_1), equal to either $(0, p)$ or $(1, i)$, define the random variable $\mathbf{E} = (\mathbf{E}_0, \mathbf{E}_1)$ (respectively). Formulated that way, the half-SWAP attack fits better our framework. Notice that p and a (Alice's final key) are completely uncorrelated and that i determines completely a after the ECC and PA steps are completed.

Now let us look at our security criteria, and observe $I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}, \text{pass}) P[\mathbf{T} = \text{pass}]$. Of course $p(\text{pass}) = 1/2$. It is however a big mistake to believe that $I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}, \text{pass})$ is equal to m or $m/2$. Eve's information is equal to m if the following **two conditions** are satisfied:

– the test is passed
– she applied the SWAP attack

otherwise, she gets 0 information. So Eve's information is m times the probability that both the test is passed and she applied the SWAP attack, which is equal to 1/2 times the probability of passing the test when she swaps. This is exponentially small.

In order to make this intuitive reasoning formal, let us use (a particular case of) the chain rule for mutual information (see Appendix B.1):

$$I(\mathbf{E};\mathbf{A}) \equiv I(\mathbf{E}_0, \mathbf{E}_1; \mathbf{A}) = I(\mathbf{E}_0;\mathbf{A}) + I(\mathbf{E}_1;\mathbf{A} \mid \mathbf{E}_0)$$

Now, $\mathbf{E}_0$ corresponds to a random bit generated by Eve, independently of i and thus independently of a. As a consequence $I(\mathbf{E}_0;\mathbf{A} \mid i_T, j_T, b, s, \xi) = 0$ and thus

$I(\mathbf{E}_0; \mathbf{A} \mid \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}, \text{pass}) = 0$. This implies that

$$I(\mathbf{E}; \mathbf{A} \mid \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}, \text{pass}) p(\text{pass}) = I(\mathbf{E}_1; \mathbf{A} \mid \mathbf{E}_0, \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}, \text{pass}) p(\text{pass})$$

Now

$$I(\mathbf{E}_1; \mathbf{A} \mid \mathbf{E}_0, i_T, j_T, b, s, \xi) = \sum_{e_0} I(\mathbf{E}_1; \mathbf{A} \mid e_0, i_T, j_T, b, s, \xi) p(e_0 \mid i_T, j_T, b, s, \xi)$$

If $e_0 = 0$ then $\mathbf{E}_1$ is just the dummy output that is independent of a and and as a consequence $I(\mathbf{E}_1; \mathbf{A} \mid e_0, i_T, j_T, b, s, \xi) = 0$. On the other hand, if $e_0 = 1$ (written "swap" hereunder) then, Eve gets full information, i.e. the m bits of the key. We are thus left with the equality

$$I(\mathbf{E}; \mathbf{A} \mid i_T, j_T, b, s, \xi) = m\, p(\text{swap} \mid i_T, j_T, b, s, \xi)$$

where, of course, Bob's outputs j_T will depend heavily on the swap! We can now expand

$$\begin{aligned} I(\mathbf{E}; \mathbf{A} \mid \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}, \text{pass}) p(\text{pass}) &= m \sum_{|i_T \oplus j_T| \le n p_a} \sum_{b,s,\xi} p(\text{swap} \mid i_T, j_T, b, s, \xi) p(i_T, j_T, b, s, \xi) \\ &= m\, p(\text{swap} \wedge \text{pass}) \\ &= m\, p(\text{pass} \mid \text{swap}) p(\text{swap}) \\ &= \frac{m}{2} p(\text{pass} \mid \text{swap}) \end{aligned}$$

which is exponentially small.

In fact, the half-SWAP attack does not even make $I(\mathbf{E}; \mathbf{A} \mid \mathbf{I}_T, \mathbf{J}_T, \mathbf{B}, \mathbf{S}, \mathbf{\Xi}, \text{pass})$ large since this is equal to

$$\frac{m}{2} p(\text{pass} \mid \text{swap}) \frac{1}{p(\text{pass})} = m\, p(\text{pass} \mid \text{swap})$$

meaning that the first inappropriate security criteria is actually satisfied correctly if the Half-SWAP attack is used.

C A Few Technical Lemmas

C.1 A Proof of Lemma 3.3

We prove here Eq.3.15. It is actually possible to prove equality[14], but for our purpose inequality is as good, so we do not bother with proving equality.

[14] This is done by proving that $I(\mathbf{A}; \mathbf{M} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \mathbf{\Xi}) = 0$. See the chain rule used in the first inequality below.

Proof Using the chain rule described in Appendix B.1, we get

$$\begin{aligned}
&I(\mathbf{A};\mathbf{E}',\mathbf{M} \mid \mathbf{I}_T,\mathbf{C}_T = c_T,b,s,\mathbf{\Xi})\\
&\geq I(\mathbf{A};\mathbf{E}' \mid \mathbf{M},\mathbf{I}_T,\mathbf{C}_T = c_T,b,s,\mathbf{\Xi})\\
&= \sum_{i_T,\xi,m} P'\left[i_T,\xi,m \mid c_T,b,s\right] I(\mathbf{A};\mathbf{E}' \mid m,i_T,c_T,b,s,\xi)\\
&= \sum_{i_T,\xi,m} P'\left[i_T,\xi \mid c_T,b,s,m\right] I(\mathbf{A};\mathbf{E}' \mid m,i_T,c_T,b,s,\xi)p(m)
\end{aligned}$$

For any fixed m, the effect of the symmetrizing transformation S is to replace i by $i \oplus m$, (c_T remaining fixed). In particular i_T becomes $i_T \oplus m_T$ and i_I becomes $i_I \oplus m_I$ and so ξ becomes $(i_I \oplus m_I)P_{\mathcal{C}}^{\top} = \xi \oplus m_I P_{\mathcal{C}}^{\top}$ and so

$$\begin{aligned}
P'(i_T,\xi \mid c_T,b,s,m) &= P(i_T \oplus m_T, \xi \oplus m_I P_{\mathcal{C}}^{\top} \mid c_T,b,s)\\
I(\mathbf{A};\mathbf{E}' \mid m,i_T,\mathbf{C}_T = c_T,b,s,\xi) &= I(\mathbf{A};\mathbf{E} \mid i_T \oplus m_T,\mathbf{C}_T = c_T,b,s,\xi \oplus m_I P_{\mathcal{C}}^{\top})
\end{aligned}$$

If we let $i'_T = i_T \oplus m_T$, $\xi' = \xi \oplus m_I P_{\mathcal{C}}^{\top}$ and use the fact that the same value of ξ' is obtained 2^{n-r} times, we get

$$\begin{aligned}
&I(\mathbf{A};\mathbf{E}',\mathbf{M} \mid \mathbf{I}_T,\mathbf{C}_T = c_T,b,s,\mathbf{\Xi})\\
&\geq \sum_{i_T,\xi,m} P'[i_T,\xi \mid c_T,b,s,m] I(\mathbf{A};\mathbf{E}' \mid m,i_T,\mathbf{C}_T = c_T,b,s,\xi)p(m)\\
&= 2^{n-r} \sum_{i_T,\xi,i'_T,\xi'} P[i'_T,\xi' \mid c_T,b,s] I(\mathbf{A};\mathbf{E} \mid i'_T,\mathbf{C}_T = c_T,b,s,\xi')p(m)\\
&= 2^{n-r}2^{n+r} \sum_{i'_T,\xi'} P[i'_T,\xi' \mid c_T,b,s] I(\mathbf{A};\mathbf{E} \mid i'_T,\mathbf{C}_T = c_T,b,s,\xi')2^{-2n}\\
&= \sum_{i'_T,\xi'} P[i'_T,\xi' \mid c_T,b,s] I(\mathbf{A};\mathbf{E} \mid i'_T,\mathbf{C}_T = c_T,b,s,\xi')\\
&= I(\mathbf{A};\mathbf{E} \mid \mathbf{I}_T,\mathbf{C}_T = c_T,b,s,\mathbf{\Xi}) \quad \square
\end{aligned}$$

C.2 A Proof of Lemma 3.5

Using the Basic Lemma of Symmetrization (Eq. 3.7) and the fact that the $|m\rangle$ form an orthonormal basis,

$$\langle E^{\mathrm{sym}\,\prime}_{i,j} | E^{\mathrm{sym}\,\prime}_{i',j'} \rangle_b = 2^{-2n} \sum_m (-1)^{(i\oplus j\oplus i'\oplus j')\cdot m} \langle E'_{i\oplus m,j\oplus m} | E'_{i'\oplus m,j'\oplus m} \rangle_b \,. \tag{C.1}$$

By replacing i,j,i' and j' by $i \oplus u$, $j \oplus u$, $i' \oplus u$ and $j' \oplus u$ in this formula, we get $\langle E^{\mathrm{sym}\,\prime}_{i\oplus u,j\oplus u} | E^{\mathrm{sym}\,\prime}_{i'\oplus u,j'\oplus u} \rangle_b = 2^{-2n} \sum_m (-1)^{(i\oplus j\oplus i'\oplus j')\cdot m} \langle E'_{i\oplus u\oplus m,j\oplus u\oplus m} | E'_{i'\oplus u\oplus m,j'\oplus u\oplus m} \rangle_b$.
Defining $w = u \oplus m$ we get $\langle E^{\mathrm{sym}\,\prime}_{i\oplus u,j\oplus u} | E^{\mathrm{sym}\,\prime}_{i'\oplus u,j'\oplus u} \rangle_b =$

$2^{-2n}\sum_w (-1)^{(i\oplus j\oplus i'\oplus j')\cdot w\oplus u}\langle E'_{i\oplus w,j\oplus w}|E'_{i'\oplus w,j'\oplus w}\rangle_b$, and using (C.1) we finally get

$$\langle E^{\text{sym}\,\prime}_{i\oplus u,j\oplus u}|E^{\text{sym}\,\prime}_{i'\oplus u,j'\oplus u}\rangle_b = (-1)^{(i\oplus j\oplus i'\oplus j')\cdot u}\langle E^{\text{sym}\,\prime}_{i,j}|E^{\text{sym}\,\prime}_{i',j'}\rangle_b\ . \tag{C.2}$$

Considering information and test bits, if we let $u = u_I u_T$ with $u_T = 0$ and use the fact (Lemma 3.4) that the normalizing factor for a symmetrized attack depends only on i_T, j_T, b and s (so we can divide both sides by the same normalization factor), we deduce from (C.2) the identity

$$\langle E^{\text{sym}}_{i_I\oplus u_I,j_I\oplus u_I}|E^{\text{sym}}_{i'_I\oplus u_I,j'_I\oplus u_I}\rangle_{b,s} = (-1)^{(i_I\oplus j_I\oplus i'_I\oplus j'_I)\cdot u_I}\langle E^{\text{sym}}_{i_I,j_I}|E^{\text{sym}}_{i'_I,j'_I}\rangle_{b,s}\ . \tag{C.3}$$

For any n-bit string u_I, we get by Eq. (C.3), by letting $i'_I = i_I \oplus k_I$, $j'_I = j_I \oplus k_I$ that $(i_I \oplus j_I \oplus i'_I \oplus j'_I)\cdot u_I = 0$ and so

$$\langle E^{\text{sym}}_{i_I\oplus u_I,j_I\oplus u_I}|E^{\text{sym}}_{i_I\oplus k_I\oplus u_I,j_I\oplus k_I\oplus u_I}\rangle_{b,s} = \langle E^{\text{sym}}_{i_I,j_I}|E^{\text{sym}}_{i_I\oplus k_I,j_I\oplus k_I}\rangle_{b,s}\ .$$

By writing $j = i \oplus c$ we get

$$\langle E^{\text{sym}}_{i_I\oplus u_I,j_I\oplus c_I\oplus u_I}|E^{\text{sym}}_{i_I\oplus k_I\oplus u_I,i_I\oplus c_I\oplus k_I\oplus u_I}\rangle_{b,s} = \langle E^{\text{sym}}_{i_I,i_I\oplus c_I}|E^{\text{sym}}_{i_I\oplus k_I,i_I\oplus c_I\oplus k_I}\rangle_{b,s}\ ,$$

so that the first part of the Lemma is proven [$\langle E^{\text{sym}}_{i_I,i_I\oplus c_I}|E^{\text{sym}}_{i_I\oplus k_I,i_I\oplus c_I\oplus k_I}\rangle$ is independent of i_I.]

Summing over c_I and changing back to j_I we get that $\sum_j \langle E^{\text{sym}}_{i_I,j_I}|E^{\text{sym}}_{i_I\oplus k_I,j_I\oplus k_I}\rangle$ is also independent of i_I.

C.3 A Proof of Eq. (3.27)

We show that $p(j_T \mid i_T, b_I, b_T, s) = p(j_T \mid i_T, b'_I, b_T, s)$ for any choice of basis b'_I on information bits. For any basis b'_I, the change of basis between b'_I and b_I is expressed by a unitary matrix $U = (u_{i'_I,i_I})$ such that $|i'_I\rangle_{b'_I} = \sum_{i_I} u_{i'_I,i_I}|i_I\rangle_{b_I}$, $|i_I\rangle_{b_I} = \sum_{i'_I} u^\dagger_{i_I,i'_I}|i'_I\rangle_{b'_I}$ and, of course, $UU^\dagger = U^\dagger U = 1$. From the defining equation $|E'_{i_T,i_I,j_T,j_I}\rangle_b = {}_b\langle j_T|_b\langle j_I|U|0\rangle_{\text{E}}|i_T\rangle_b|i_I\rangle_b$ (Eq. 3.17) and the above, we get

$$|E'_{i_T,i'_I,j_T,j'_I}\rangle_{b_T,b'_I} = \sum_{i_I,j_I} u_{i'_I,i_I}u^\dagger_{j_I,j'_I}|E'_{i_T,i_I,j_T,j_I}\rangle_{b_T,b_I} \tag{C.4}$$

For any b, we have $p(j_T \mid i_T, b, s) = \sum_{i_I} p(j_T \mid i_T, i_I, b, s)p(i_I \mid i_T, b, s)$. As $p(i_I \mid i_T, b, s) = 1/2^n$ (since these values are chosen at random by Alice) we can deduce, using Eq. (3.20) $p(j_T|i_T, i_I, b, s) = \sum_{j_I}\langle E'_{i_T,i_I,j_T,j_I}|E'_{i_T,i_I,j_T,j_I}\rangle_b$, that

$$p(j_T \mid i_T, b, s) = \frac{1}{2^n}\sum_{j_I,i_I}\langle E'_{i_T,i_I,j_T,j_I}|E'_{i_T,i_I,j_T,j_I}\rangle_b \tag{C.5}$$

If we apply Eq. (C.5) in the particular case where the basis is b'_I, b_T, and we expand its right-hand side using Eq. (C.4), then, because of the unitarity of U, the

six sums reduce to two, yielding a term that is exactly equal to the right-hand side of Eq. (C.5) with basis $b = b_I, b_T$. That is:

$$p(j_T \mid i_T, b'_I, b_T, s) = \frac{1}{2^n} \sum_{j_I, i_I} \langle E'_{i_T, i_I, j_T, j_I} | E'_{i_T, i_I, j_T, j_I} \rangle_b \tag{C.6}$$

□

C.4 A Proof of Lemma 4.1

We start from Eq. (4.8), namely

$$P^{\text{sym}}\left[\mathbf{C}_I = c_I \mid i_T, j_T, b^0, s\right] = \frac{1}{2^n} \sum_{i'_I} \langle E^{\text{sym}}_{i'_I, i'_I \oplus c_I} | E_{i'_I, i'_I \oplus c_I} \rangle_{b^0, s}. \tag{C.7}$$

with $b^0 = b \oplus s$. From Hadamard, we know that the unitary matrix $U = (u_{i'_I, i_I})$ used to express $|i'_I\rangle_{\bar{b}_I}$ in terms of the $|i_I\rangle_{b_I}$ is defined by $u_{i'_I, i_I} = 2^{-n/2}(-1)^{i'_I \cdot i_I}$ and, for that particular choice of b'_I, Eq. (C.4) reduces to

$$|E^{\text{sym}\,\prime}_{i_T, i'_I, j_T, j'_I}\rangle_{b_T, \bar{b}_I} = \frac{1}{2^n} \sum_{i_I, j_I} (-1)^{i'_I \cdot i_I} (-1)^{j_I \cdot j'_I} |E^{\text{sym}\,\prime}_{i_T, i_I, j_T, j_I}\rangle_{b_T, b_I}$$

Due to Corollary 3.3 $p^{\text{sym}}(j_T | i_T, b_T, s)$ is independent of b_I, so both sides can be divided by the same normalization factor, and this implies that

$$|E^{\text{sym}}_{i'_I, j'_I}\rangle_{b^0, s} = \frac{1}{2^n} \sum_{i_I, j_I} (-1)^{(i'_I \cdot i_I + j_I \cdot j'_I)} |E^{\text{sym}}_{i_I, j_I}\rangle_{b, s} \ .$$

Then, going back to Eq. (C.7) and replacing $|E^{\text{sym}}_{i'_I, i'_I \oplus c_I}\rangle_{b^0, s}$ by those values, leaves

$$\begin{aligned}
&P^{\text{sym}}[\mathbf{C}_I = c_I \mid i_T, j_T, b^0, s] \\
&= \frac{1}{2^n} \sum_{k_I} \sum_{i_I, j_I} \sum_{i'_I, j'_I} \frac{1}{2^{2n}} (-1)^{(i_I \oplus i'_I) \cdot k_I \oplus (j_I \oplus j'_I) \cdot (k_I \oplus c_I)} \langle E^{\text{sym}}_{i_I, j_I} | E^{\text{sym}}_{i'_I, j'_I} \rangle_{b, s} \\
&= \frac{1}{2^{3n}} \sum_{i_I, i'_I, j_I, j'_I} \left(\sum_{k_I} (-1)^{k_I \cdot (i_I \oplus i'_I \oplus j_I \oplus j'_I)} \right) (-1)^{c_I \cdot (j_I \oplus j'_I)} \langle E^{\text{sym}}_{i_I, j_I} | E^{\text{sym}}_{i'_I, j'_I} \rangle_{b, s}
\end{aligned}$$

The sum over k_I is non zero only when $i_I \oplus i'_I = j_I \oplus j'_I \stackrel{\Delta}{=} h_I$,
and then it is 2^n, so

$$\begin{aligned}
&= \frac{1}{2^{2n}} \sum_{i_I, j_I, h_I} (-1)^{c_I \cdot h_I} \langle E^{\text{sym}}_{i_I, j_I} | E^{\text{sym}}_{i_I \oplus h_I, j_I \oplus h_I} \rangle_{b, s} \\
&= \langle \eta_{c_I} | \eta_{c_I} \rangle = d^2_{c_I}
\end{aligned}$$

where the last equalities are due to the calculation of the norm of η in Eq. (4.5).

C.5 A Proof of Proposition 4.3

We prove here Proposition 4.3 that claims a bound on the m-bit key given a bound on 1-bit key.

Proof Let $F(x) = 2\sqrt{P^{\text{sym}}[|\mathbf{C}_I| \geq \frac{x}{2} \mid i_T, j_T, b^0, s]}$. For any r' such that $r \leq r' < r+m$, let $\mathcal{C}'$ be the code whose parity check matrix $P_{\mathcal{C}'}$ has the rows $v_1, \ldots, v_{r'}$. Then $P_{\mathcal{C}'}$ has rank r' and $\mathcal{C}'$ is an (n, k', d') code with $k' = n - r'$. Moreover $v_{r'+1} \notin \mathcal{C}'^{\perp} = V_{r'}$. As a consequence, Proposition 4.2 applies and gives that

$$I(\mathbf{A}'; \mathbf{E}^{\text{sym}} \mid i_T, j_T, b, s, \xi') \leq F(\hat{v}_{r'+1})$$

for $a' = v_{r'+1} \cdot i_I = a_{j+1}$ with $j = r' - r$, $\xi' = i_I P_{\mathcal{C}'}^{\top} = \xi_1 \ldots \xi_r a_1 \ldots a_j$, $\hat{v}_{r'+1} = d_H(v_{r'+1}, V_{r'})$ and $b^0 = b \oplus s$. This can be rewritten

$$I(\mathbf{A}_{j+1}; \mathbf{E}^{\text{sym}} \mid i_T, j_T, b, s, \xi, a_1 \ldots a_j) \leq F(\hat{v}_{r'+1})$$

and the result follows from Lemma 4.3 by taking $F = \max_{r \leq r' < r+m} F(\hat{v}_{r'+1}) = F(\hat{v})$ for $\hat{v} = \min_{r \leq r' < r+m} \hat{v}_{r'+1}$. □

C.6 A Proof of Lemma 5.1

The Lemma says:

$$\sum_{|c_T| \leq np_a} P^{\text{sym}}[\mathbf{C}_T = c_T | b, s]\, I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \boldsymbol{\Xi})$$
$$\leq 2m \sqrt{P^{\text{sym}}\left[(|\mathbf{C}_I| > \frac{\hat{v}}{2}) \wedge (\frac{|\mathbf{C}_T|}{n} \leq p_a) \mid b^0, s\right]}$$

Proof If we expand $\mathbf{I}_T$ and $\boldsymbol{\Xi}$ in the expression $I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \boldsymbol{\Xi})$ then we get

$$\begin{aligned}
&\sum_{|c_T| \leq np_a} P^{\text{sym}}[\mathbf{C}_T = c_T | b, s]\, I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid \mathbf{I}_T, \mathbf{C}_T = c_T, b, s, \boldsymbol{\Xi}) \\
&\quad = \sum_{|c_T| \leq np_a, i_T, \xi} p^{\text{sym}}(i_T, C_T = c_T, \xi \mid b, s) I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid i_T, c_T, b, s, \xi) \\
&\quad = \sum_{|i_T \oplus j_T| \leq np_a, i_T, \xi} p^{\text{sym}}(i_T, j_T, \xi \mid b, s) I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid i_T, j_T, b, s, \xi) \\
&\quad = \sum_{|i_T \oplus j_T| \leq np_a, i_T, \xi} p^{\text{sym}}(i_T, j_T \mid b^0, s) 2^{-r} I(\mathbf{A}; \mathbf{E}^{\text{sym}} \mid i_T, j_T, b, s, \xi)\,.
\end{aligned}$$

The last equality requires a detailed explanation: First, notice that $p^{\text{sym}}(j_T \mid i_T, b, s, \xi) = p^{\text{sym}}(j_T \mid i_T, b, s)$ because the probability $p^{\text{sym}}(j_T \mid i_T, b, s, i_I)$ is independent of i_I by Lemma (3.4) and the condition $\boldsymbol{\Xi} = \xi$ means $i_I P_{\mathcal{C}}^{\top} = \xi$, which is a condition on i_I. As a consequence, using the fact that (for any attack)

$p(\xi \mid b,s) = 2^{-r}$, $p(i_T \mid b,s) = 2^{-n}$ and $p(i_T,\xi \mid b,s) = p(i_T \mid b,s)p(\xi \mid b,s)$ [so that $p(i_T,\xi \mid b,s) = 2^{-(n+r)}$], we get

$$\begin{aligned}
p^{\text{sym}}(i_T,j_T,\xi \mid b,s) &= p^{\text{sym}}(j_T \mid i_T,b,s,\xi)p^{\text{sym}}(i_T,\xi \mid b,s) \\
&= p^{\text{sym}}(j_T \mid i_T,b,s,\xi)2^{-(n+r)} \\
&= p^{\text{sym}}(j_T \mid i_T,b,s)2^{-(n+r)} && \text{by the above} \\
&= p^{\text{sym}}(j_T \mid i_T,b^0,s)2^{-(n+r)} && \text{by Lemma (3.6)} \\
&= 2^{-(n+r)}[p^{\text{sym}}(i_T,j_T \mid b^0,s)/p^{\text{sym}}(i_T \mid b^0,s)] && \text{by definition of } p(A \mid B) \\
&= p^{\text{sym}}(i_T,j_T \mid b^0,s)2^{-r} && \text{due to } p(i_T \mid b^0,s) = 2^{-n}
\end{aligned}$$

The result

$$\begin{aligned}
&\sum_{|c_T| \le np_a} P^{\text{sym}}\left[\mathbf{C}_T = c_T | b,s\right] I(\mathbf{A};\mathbf{E}^{\text{sym}} \mid \mathbf{I}_T,\mathbf{C}_T = c_T,b,s,\mathbf{\Xi}) \\
&\quad = \sum_{|i_T \oplus j_T| \le np_a, i_T,\xi} p^{\text{sym}}(i_T,j_T \mid b^0,s)2^{-r} I(\mathbf{A};\mathbf{E}^{\text{sym}} \mid i_T,j_T,b,s,\xi) \\
&\quad \le \sum_{|i_T \oplus j_T| \le np_a, i_T,\xi} 2^{-r} p^{\text{sym}}(i_T,j_T \mid b^0,s) 2m \sqrt{P^{\text{sym}}\left[|\mathbf{C}_I| \ge \frac{\hat{v}}{2} \mid i_T,j_T,b^0,s\right]}
\end{aligned}$$

now follows immediately from corollary (4.2). Using the fact that square-root is a convex function $\sum p_i\sqrt{x_i} \le \sqrt{\sum p_i x_i}$ so we get

$$\begin{aligned}
&\sum_{|c_T| \le np_a} P^{\text{sym}}\left[\mathbf{C}_T = c_T | b,s\right] I(\mathbf{A};\mathbf{E}^{\text{sym}} \mid \mathbf{I}_T,\mathbf{C}_T = c_T,b,s,\mathbf{\Xi}) \\
&\quad \le 2m \sqrt{\sum_{|i_T \oplus j_T| \le np_a, i_T,\xi} 2^{-r} p^{\text{sym}}(i_T,j_T \mid b^0,s) P^{\text{sym}}\left[|\mathbf{C}_I| \ge \frac{\hat{v}}{2} \mid i_T,j_T,b^0,s\right]}
\end{aligned}$$

Finally, we get rid of the 2^{-r} factor by summing over ξ (each equally likely), and we complete the proof using

$$\begin{aligned}
&\sum_{|i_T \oplus j_T| \le np_a, i_T,\xi} P^{\text{sym}}\left[|\mathbf{C}_I| \ge \frac{\hat{v}}{2} \mid i_T,j_T,b^0,s\right] p^{\text{sym}}(i_T,j_T \mid b^0,s)2^{-r} \\
&\qquad = P^{\text{sym}}\left[|\mathbf{C}_I| \ge \frac{\hat{v}}{2}, |\mathbf{C}_T| \le np_a \mid b^0,s\right] \qquad \square
\end{aligned}$$

C.7 A Proof of Lemma 5.4

Let

$$P\left[\left(\frac{|\mathbf{C}_I|}{n} > p_a + \epsilon\right) \wedge \left(\frac{|\mathbf{C}_T|}{n} \le p_a\right)\right] = \sum_b p(b) h_b(p_a,\epsilon)$$

with

$$h_b(p_a,\epsilon) = P\left[\left(\frac{|\mathbf{C}_I|}{n} > p_a+\epsilon\right) \wedge \left(\frac{|\mathbf{C}_T|}{n} \leq p_a\right) \mid b\right]$$

This $h_b(p_a,\epsilon)$ is the probability that the information bits have ϵ more than the allowed error rate, when the test bits have less than the allowed error rate averaged over all choices of test and information bits, for a particular basis b, and is given by

$$\sum_c P\left[\left(\frac{|\mathbf{C}_I|}{n} > p_a+\epsilon\right) \wedge \left(\frac{|\mathbf{C}_T|}{n} \leq p_a\right) \mid \mathbf{C}=c,b\right] P\left[\mathbf{C}=c \mid b\right]$$

where c is over all possible error strings on all bits, test and information. Note that in principle $P\left[\mathbf{C}=c \mid b\right]$, can be calculated but we shall soon see that there is no need for it.

Now we must note that

$$P\left[\left(\frac{|\mathbf{C}_I|}{n} > p_a+\epsilon\right) \wedge \left(\frac{|\mathbf{C}_T|}{n} \leq p_a\right) \mid \mathbf{C}=c,b\right]$$

does not depend on the attack. And in fact, in the aforementioned expression, the basis b is superfluous. Once the error string c is fixed, the values $\frac{|c_I|}{n}$ and $\frac{|c_T|}{n}$ depend uniquely on the random string s. In fact $\frac{|c_I|}{n}$ is the average of a random sampling without replacement of n bits taken from the $2n$ bits c whose mean μ is $\frac{|c|}{2n}$. From Hoeffding [22] we know that

$$P\left[\frac{|\mathbf{C}_I|}{n} - \mu \geq \frac{\epsilon}{2} \mid c,b\right] \leq e^{-\frac{1}{2}n\epsilon^2} \tag{C.8}$$

By definition $|c| = |c_I| + |c_T|$ and so

$$\mu = \frac{|c|}{2n} = \frac{|c_I|}{2n} + \frac{|c_T|}{2n}$$

Replacing μ by its value in (C.8) and simplifying, equation (C.8) becomes

$$P\left[\frac{|\mathbf{C}_I|}{n} \geq \frac{|\mathbf{C}_T|}{n} + \epsilon \mid \mathbf{C}=c,b\right] \leq e^{-\frac{1}{2}n\epsilon^2} \tag{C.9}$$

Now, since

$$(\frac{|c_I|}{n} > p_a+\epsilon) \wedge (\frac{|c_T|}{n} \leq p_a) \Longrightarrow \frac{|c_I|}{n} \geq \frac{|c_T|}{n} + \epsilon$$

we deduce from (C.9) that

$$P\left[\left(\frac{|\mathbf{C}_I|}{n} > p_a+\epsilon\right) \wedge \left(\frac{|\mathbf{C}_T|}{n} \leq p_a\right) \mid \mathbf{C}=c,b\right] \leq e^{-\frac{1}{2}n\epsilon^2}$$

and consequently,

$$h_b(p_a,\epsilon) = P\left[\left(\frac{|\mathbf{C}_I|}{n} > p_a+\epsilon\right) \wedge \left(\frac{|\mathbf{C}_T|}{n} \leq p_a\right) \mid b\right] \leq e^{-\frac{1}{2}n\epsilon^2}$$

and

$$P\left[\left(\frac{|\mathbf{C}_I|}{n} > p_a+\epsilon\right) \wedge \left(\frac{|\mathbf{C}_T|}{n} \leq p_a\right)\right] \leq e^{-\frac{1}{2}n\epsilon^2}$$

D Eve's Information Versus the disturbance

In this appendix we do not prove Lemma 4.2 immediately. We prove it later on, in the second subsection (the tight bound). For simplicity of the presentation, we first prove another Lemma which leads to a loose bound (with an additional factor of 2^r), for which the derivation is simpler. The bulk of the loose bound was derived in [7], and is adapted here to the analysis of the joint attack. The tight bound is an improvement over that derivation yielding a much better threshold for p_{allowed}: The loose bound leads to a threshold of less than 1%, while the threshold for the tight bound is 7.56%. One can skip directly to the second subsection if desired.

Both the loose and the tight bound are derived using the fact that the Shannon distinguishability between the parity 0 density matrix, ρ_0, and the parity 1 density matrix, ρ_1, is bounded ([7,18]) by the trace norm of $\rho_0 - \rho_1$ and using the fact that we can easily calculate this trace norm when the purified states are given by Eq. (4.4).

D.1 The Loose Bound

Exploiting the techniques developed in [7] (to prove security against any collective attack) we now present a bound which is applicable to the joint attack.

We have already defined a purification of Eve's state: $|\phi_{i_I}\rangle = \sum_l (-1)^{i_I \cdot l} |\eta_l\rangle$ The density matrix for such a $|\phi_{i_I}\rangle$ is

$$\rho^{i_I} = |\phi_{i_I}\rangle\langle\phi_{i_I}| = \sum_{l,l'} (-1)^{i_I \cdot (l \oplus l')} d_l d_{l'} |\hat{\eta}_l\rangle\langle\hat{\eta}_{l'}| \tag{D.1}$$

Recall that the final key is computed as $b = v \cdot i_I$. Eve does not know i_I, but she knows v, and she knows (from the announced ECC parity string ξ) that i_I is in the coset $\mathcal{C}_\xi$. Hence, in order to know the key, Eve must distinguish between the states $i_I = i_\xi \oplus c$ in $\mathcal{C}_\xi$ that give parity $b = 0$ and the states $i_I = i_\xi \oplus c$ in $\mathcal{C}_\xi$ that give parity $b = 1$. For $b \in \{0,1\}$ the reduced density matrix is

$$\begin{aligned}
\rho_b(v,\xi) &= \frac{1}{2^{n-(r+1)}} \sum_{\substack{c \in \mathcal{C} \\ v \cdot (i_\xi \oplus c) = b}} \rho^{i_\xi \oplus c} \\
&= \frac{1}{2^{n-(r+1)}} \sum_{\substack{c \in \mathcal{C} \\ v \cdot (i_\xi \oplus c) = b}} \sum_{l,l'} (-1)^{(i_\xi \oplus c) \cdot (l \oplus l')} d_l d_{l'} |\hat{\eta}_l\rangle\langle\hat{\eta}_{l'}|
\end{aligned}$$

where the sum is over values c that satisfy both the condition of being a code word, and the condition of leading to the particular parity b for the PA.

Lemma D.1 *Let $\mathcal{C}$ be any linear code in $\{0,1\}^n$ and $a \in \{0,1\}^n$ be such that $a \notin \mathcal{C}^\perp$ then*

$$\sum_{c \in \mathcal{C}} (-1)^{c \cdot a} = 0 \tag{D.2}$$

Proof Let $\{w_1, \ldots, w_k\}$ be a basis of $\mathcal{C}$. Define $t \in \{0,1\}^k$ by $t_\alpha = w_\alpha \cdot a,\ 1 \leq \alpha \leq k$; $a \notin \mathcal{C}^\perp$ means that t is not the zero string. Let now $h : \{0,1\}^k \to \mathcal{C}$ be defined by $h(s) = \sum_{1 \leq \alpha \leq k} s_\alpha w_\alpha$; then $h(s) \cdot a = \sum s_\alpha w_\alpha \cdot a = \sum s_\alpha t_\alpha = s \cdot t$ and so

$$\sum_{c \in \mathcal{C}} (-1)^{c \cdot a} = \sum_s (-1)^{h(s) \cdot a} = \sum_s (-1)^{s \cdot t} = 0$$

□

Lemma D.2 *For any (n,k,d) code $\mathcal{C}$ with $r \times n$ parity check matrix $P_\mathcal{C}$ of rank $r = n-k$, any $\xi \in \{0,1\}^r$ and any $v \in \{0,1\}^n$ the Shannon distinguishability $SD(\rho_0(v,\xi), \rho_1(v,\xi))$ where*

$$\rho_b(v,\xi) = \frac{1}{2^{n-(r+1)}} \sum_{\substack{i_I P_\mathcal{C}^\top = \xi \\ i_I \cdot v = b}} \rho^i$$

between the parity 0 and the parity 1 of the information bits over any PA string, v, is bounded above by the following inequality:

$$SD(\rho_0(v,\xi), \rho_1(v,\xi)) \leq 2^{r+1} \sqrt{\sum_{|l| \geq \frac{\hat{v}}{2}} d_l^2}\,, \tag{D.3}$$

where $\hat{v}$ is the minimum distance between v and the code $\mathcal{C}^\perp$, i.e. the minimum weight of $v \oplus v'$ for any $v' \in \mathcal{C}^\perp$.

Proof The Shannon distinguishability between the parity 0 and the parity 1 is bounded by the trace norm of $\rho_0(v,\xi) - \rho_1(v,\xi)$, see [7,18]. Let us calculate the required bound:

$$\begin{aligned}
&\rho_0(v,\xi) - \rho_1(v,\xi) \\
&= \frac{1}{2^{n-(r+1)}} \sum_{c \in \mathcal{C}} (-1)^{(i_\xi \oplus c) \cdot v} \sum_{l,l'} (-1)^{(i_\xi \oplus c) \cdot (l \oplus l')} d_l d_{l'} |\hat{\eta}_l\rangle\langle\hat{\eta}_{l'}| \\
&= \frac{1}{2^{n-(r+1)}} \sum_{l,l'} \left(\sum_{c \in \mathcal{C}} (-1)^{(i_\xi \oplus c) \cdot (l \oplus l' \oplus v)} \right) d_l d_{l'} |\hat{\eta}_l\rangle\langle\hat{\eta}_{l'}| \\
&= \frac{1}{2^{n-(r+1)}} \sum_{l,l'} (-1)^{i_\xi \cdot (l \oplus l' \oplus v)} \left(\sum_{c \in \mathcal{C}} (-1)^{c \cdot (l \oplus l' \oplus v)} \right) d_l d_{l'} |\hat{\eta}_l\rangle\langle\hat{\eta}_{l'}|
\end{aligned}$$

From equation (D.2) we know the sum over $\mathcal{C}$ is zero except when $l \oplus l' \oplus v \in \mathcal{C}^\perp = V_r$, i.e. when $l' = l \oplus v \oplus v_\mathbf{s}$ for some $v_\mathbf{s} \in V_r$. As a consequence:

$$\rho_0(v,\xi) - \rho_1(v,\xi) = 2 \sum_{v_s \in V_r} (-1)^{i_\xi \cdot v_s} \sum_l d_l d_{l \oplus v \oplus v_s} |\hat{\eta}_l\rangle\langle\hat{\eta}_{l \oplus v \oplus v_s}|$$

As already said, the trace norm of this matrix serves as a bound on the information Eve receives [7,18].

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

Using the above and making use of the triangle inequality for the trace norm, the following is obtained (where $SD(\rho_0(v,\xi),\rho_1(v,\xi))$ is denoted SD_v for short):

$$\begin{aligned}
SD_v &\leq Tr\left|\sum_{v_s\in V_r}(-1)^{i_\xi\cdot v_s}\sum_l d_l d_{l\oplus v\oplus v_s}|\hat{\eta}_m\rangle\langle\hat{\eta}_{m\oplus v\oplus v_s}|\right| \\
&= \frac{1}{2}Tr\left|\sum_{v_s\in V_r}(-1)^{i_\xi\cdot v_s}\sum_l d_l d_{l\oplus v\oplus v_s}\left(|\hat{\eta}_l\rangle\langle\hat{\eta}_{l\oplus v\oplus v_s}|+|\hat{\eta}_{l\oplus v\oplus v_s}\rangle\langle\hat{\eta}_l|\right)\right| \\
&\leq \sum_{v_s\in V_r}\sum_l d_l d_{l\oplus v\oplus v_s}\left(\frac{1}{2}Tr\left|\,|\hat{\eta}_l\rangle\langle\hat{\eta}_{l\oplus v\oplus v_s}|+|\hat{\eta}_{l\oplus v\oplus v_s}\rangle\langle\hat{\eta}_l|\,\right|\right) \\
&= \sum_{v_s\in V_r}\sum_l d_l d_{l\oplus v\oplus v_s}\sqrt{1-[\Im(\langle\hat{\eta}_l|\hat{\eta}_{l\oplus v\oplus v_s}\rangle)]^2} \\
&\leq \sum_{v_s\in V_r}\sum_l d_l d_{l\oplus v\oplus v_s}
\end{aligned}$$

where the sign $\Im$ means the imaginary part. In the above, we made use of the fact that the trace norm is exactly computable for needed matrix. Now we will concern ourselves with bounding each of the terms $\sum_l d_l d_{l\oplus w_s}$, where $w_s = v\oplus v_s$.

$$\begin{aligned}
\sum_l d_l d_{l\oplus w_s} &= \sum_{|l|>\frac{|w_s|}{2}} d_l d_{l\oplus w_s}+\sum_{|l|\leq\frac{|w_s|}{2}} d_l d_{l\oplus w_s} \\
&= \sum_{|l|>\frac{|w_s|}{2}} d_l d_{l\oplus w_s}+\sum_{|l'\oplus w_s|\leq\frac{|w_s|}{2}} d_{l'\oplus w_s} d_{l'}
\end{aligned}$$

If $|l'\oplus w_s|\leq\frac{|w_s|}{2}$ then $|w_s| = |l'\oplus w_s\oplus l'|\leq|l'\oplus w_s|+|l'|\leq\frac{|w_s|}{2}+|l'|$ and so $|l'|\geq\frac{|w_s|}{2}$. Therefore,

$$\begin{aligned}
\sum_{|l|>\frac{|w_s|}{2}} d_l d_{l\oplus w_s}+\sum_{|l'\oplus w_s|\leq\frac{|w_s|}{2}} d_{l'\oplus w_s} d_{l'} &\leq \sum_{|l|\geq\frac{|w_s|}{2}} d_l d_{l\oplus w_s}+\sum_{|l'|\geq\frac{|w_s|}{2}} d_{l'\oplus w_s} d_{l'} \\
&= 2\sum_{|l|\geq\frac{|w_s|}{2}} d_l d_{l\oplus w_s} \\
&= \frac{1}{\alpha}\sum_{|l|\geq\frac{|w_s|}{2}} 2d_l(\alpha d_{l\oplus w_s}) \\
&\leq \frac{1}{\alpha}\sum_{|l|\geq\frac{|w_s|}{2}}[d_l^2+\alpha^2 d_{l\oplus w_s}^2] \\
&= \alpha\sum_{|l|\geq\frac{|w_s|}{2}} d_{l\oplus w_s}^2+\frac{1}{\alpha}\sum_{|l|\geq\frac{|w_s|}{2}} d_l^2
\end{aligned}$$

where the last three steps are true for any real α, and real $d_l, d_{l\oplus w_s}$.

Due to the fact that the d_l^2 form a probability distribution, any sum of them is less than or equal to unity.

$$\begin{aligned}\sum_l d_l d_{l\oplus w_s} &\leq \alpha + \frac{1}{\alpha} \sum_{|l|\geq \frac{|w_s|}{2}} d_l^2 \\ &\leq \alpha + \frac{1}{\alpha} \sum_{|l|\geq \frac{\hat{v}}{2}} d_l^2\end{aligned}$$

where $\hat{v} = \min_{v_s} |v \oplus v_s|$ (remember that $w_s = v \oplus v_s$). Summing over all $v_s \in V_r$ and setting $\alpha = \sqrt{\sum_{|l|\geq \frac{\hat{v}}{2}} d_l^2}$ now leaves:

$$SD_v \leq 2^{r+1} \sqrt{\sum_{|l|\geq \frac{\hat{v}}{2}} d_l^2} \tag{D.4}$$

□

Following the proof of the above Lemma, one can guess that it is not a tight bound since we sum over 2^r terms while most of them do not contribute to the sum (or contribute negligible values). This understanding led us to reach a tighter bound.

D.2 Eve's Information about one bit – Tight Bound

We will now make a finer analysis of Eve's state after she learns the parity matrix and parity string ξ. We start again from the equality:

$$|\phi_{i_I}\rangle = \sum_l (-1)^{i_I \cdot l} |\eta_l\rangle \tag{D.5}$$

Let $v_1, \ldots, v_r$ be the rows of $P_{\mathcal{C}}$, and $v_{r+1} = v$. It is assumed that the sequence $v_1, \ldots, v_{r+1}$ is linearly independent; it can thus be extended to a basis $v_1, \ldots, v_n$ of $\{0,1\}^n$. For any r' let $V_{r'}$ be the span of $\{v_1, \ldots, v_{r'}\}$ and $V_{r'}^c$ be the span of $\{v_{r'+1}, \ldots, v_n\}$. For all r', the spaces $V_{r'}$ and $V_{r'}^c$ are complementary; this means that any element $l \in \{0,1\}^n$ has a *unique representation* $l = m \oplus n$ with $m \in V_{r'}^c$ and $n \in V_{r'}$.

For $\xi \in \{0,1\}^r$, let i_ξ denote some fixed n-bit string such that $i_\xi P_{\mathcal{C}}^\top = \xi$ (existence is guaranteed by the fact that $P_{\mathcal{C}}$ has maximal rank). For any $i_I \in \mathcal{C}_\xi$ we have $(i_I - i_\xi) P_{\mathcal{C}}^\top = \xi - \xi = 0$ and so $i_I - i_\xi \in \mathcal{C}$ and thus, for any $n \in V_r = \mathcal{C}^\perp$,$(i_I - i_\xi) \cdot n = 0$ i.e. $i_I \cdot n = i_\xi \cdot n$.

Putting those remarks together we get:

$$\begin{aligned}|\phi_{i_I}\rangle &= \sum_{m \in V_r^c} \sum_{n \in V_r} (-1)^{i_I \cdot (m \oplus n)} |\eta_{m\oplus n}\rangle \\ &= \sum_{m \in V_r^c} (-1)^{i_I \cdot m} \sum_{n \in V_r} (-1)^{i_I \cdot n} |\eta_{m\oplus n}\rangle\end{aligned}$$

$$= \sum_{m \in V_r^c} (-1)^{i_I \cdot m} \sum_{n \in V_r} (-1)^{i_\xi \cdot n} |\eta_{m \oplus n}\rangle$$
$$= \sum_{m \in V_r^c} (-1)^{i_I \cdot m} |\eta'_m\rangle$$

where η'_m is defined, for each $m \in V_r$, by

$$|\eta'_m\rangle = \sum_{n \in V_r} (-1)^{i_\xi \cdot n} |\eta_{m \oplus n}\rangle \tag{D.6}$$

Let us write

$$\eta'_m = d'_m \hat{\eta}'_m$$

with the $\hat{\eta'}_m$s normalized so that $d'^2_m = \langle \eta'_m | \eta'_m \rangle$, and the density matrix for $|\phi_{i_I}\rangle$ reduces to:

$$\rho^{i_I} = |\phi_{i_I}\rangle\langle\phi_{i_I}|$$
$$= \sum_{m,m' \in V_r^c} (-1)^{i_I \cdot (m \oplus m')} d'_m d'_{m'} |\hat{\eta}'_m\rangle\langle\hat{\eta}'_{m'}|$$

Due to Proposition 4.1 (the orthogonality of the η_ms), we get that $\langle \eta_{m \oplus n_1} | \eta_{m \oplus n_2} \rangle = 0$ except when $n_1 \oplus n_2 = 0$. Together with Eq. (D.6) this implies

$$d'^2_m = \sum_{n \in V_r} d^2_{m \oplus n} \,. \tag{D.7}$$

Recall that the final key is computed as $b = v \cdot i_I$. Of course, Eve does not know i_I, but she knows v and she knows (from the announced ECC parity string ξ) that $i_I \in \mathcal{C}_\xi = \{i_\xi \oplus c \mid c \in \mathcal{C}\}$. Eve wants to determine b. For $b \in \{0, 1\}$ the reduced density matrix is

$$\rho_b(v, \xi) = \frac{1}{2^{n-(r+1)}} \sum_{\substack{c \in \mathcal{C} \\ (i_\xi \oplus c) \cdot v = b}} \rho^{i_\xi \oplus c}$$
$$= \frac{1}{2^{n-(r+1)}} \sum_{\substack{c \in \mathcal{C} \\ (i_\xi \oplus c) \cdot v = b}} \sum_{m,m' \in V_r^c} (-1)^{(i_\xi \oplus c) \cdot (m \oplus m')} d'_m d'_{m'} |\hat{\eta}'_m\rangle\langle\hat{\eta}'_{m'}|$$

We can now prove

Lemma 4.2 *The Shannon distinguishability between the parity 0 and the parity 1 of the information bits over any PA string, v, is bounded above by the following inequality:*

$$SD(\rho_0(v, \xi), \rho_1(v, \xi)) \leq 2 \sqrt{\sum_{|l| \geq \frac{\hat{v}}{2}} d_l^2} \,, \tag{D.8}$$

where $\hat{v} = d_H(v, V_r)$ is the minimum weight of $v \oplus v_s$ for any $v_s \in V_r$.

Proof The Shannon distinguishability between the parity 0 and the parity 1 is bounded by the trace norm of $\rho_0(v,\xi) - \rho_1(v,\xi)$:

$$\rho_0(v,\xi) - \rho_1(v,\xi) =$$
$$\frac{1}{2^{n-(r+1)}} \sum_{c\in\mathcal{C}} (-1)^{(i_\xi \oplus c)\cdot v} \sum_{m,m'\in V_r^c} (-1)^{(i_\xi\oplus c)\cdot(m\oplus m')} d'_m d'_{m'} |\hat{\eta}'_m\rangle\langle\hat{\eta}'_{m'}| =$$
$$\frac{1}{2^{n-(r+1)}} \sum_{m,m'\in V_r^c} \left(\sum_{c\in\mathcal{C}} (-1)^{(i_\xi\oplus c)\cdot(m\oplus m'\oplus v)} \right) d'_m d'_{m'} |\hat{\eta}'_m\rangle\langle\hat{\eta}'_{m'}| =$$
$$\frac{1}{2^{n-(r+1)}} \sum_{m,m'\in V_r^c} (-1)^{i_\xi\cdot(m\oplus m'\oplus v)} \left(\sum_{c\in\mathcal{C}} (-1)^{c\cdot(m\oplus m'\oplus v)} \right) d'_m d'_{m'} |\hat{\eta}'_m\rangle\langle\hat{\eta}'_{m'}|$$

Applying equality (D.2) the sum indexed by c is zero except when $m\oplus m'\oplus v \in \mathcal{C}^\perp = V_r$. But $m\oplus m'\oplus v \in V_r^c$ because m, m' and $v \in V_r^c$. This implies $m\oplus m'\oplus v \in V_r \cap V_r^c = \{0\}$ and thus $m' = m\oplus v$. Of course, with $m\oplus m'\oplus v = 0$, the sum indexed by c is $2^k = 2^{n-r}$ and the coefficient $(-1)^{i_\xi\cdot(m\oplus m'\oplus v)}$ is 1. Therefore $\rho_0(v,\xi) - \rho_1(v,\xi)$ takes the very simple form:

$$\rho_0(v,\xi) - \rho_1(v,\xi) = 2 \sum_{m\in V_r^c} d'_m d'_{m\oplus v} |\hat{\eta}'_m\rangle\langle\hat{\eta}'_{m\oplus v}| \tag{D.9}$$

We now claim that

$$V_r^c = V_{r+1}^c \cup \{m\oplus v \mid m \in V_{r+1}^c\} \qquad \text{(disjoint union)} \tag{D.10}$$

$$\text{if } d_H(m,V_r) < \frac{\hat{v}}{2} \text{ then } d_H(m\oplus v, V_r) \geq \frac{\hat{v}}{2} \quad \text{for any } m \in \{0,1\}^n \tag{D.11}$$

Claim (D.10) follows from the fact that $v_{r+1} = v$, V_r^c is the span of $\{v_{r+1},\ldots,v_n\}$ and V_{r+1}^c is the span of $\{v_{r+2},\ldots,v_n\}$, and that those elements are all linearly independent. As for claim (D.11) if $d_H(m,V_r) < \hat{v}/2$ and $d_H(m\oplus v, V_r) < \hat{v}/2$, then there is n and n' in V_r such that $|m\oplus n| < \hat{v}/2$ and $|m\oplus v\oplus n'| < \hat{v}/2$. This implies that $|m\oplus n\oplus m\oplus v\oplus n'| < \hat{v}$. However $m\oplus n\oplus m\oplus v\oplus n' = n\oplus n'\oplus v$ and $n\oplus n' \in V_r$ and this contradicts the fact that $\hat{v} = d_H(v,V_r)$

Now, using claim (D.10), we can rewrite Eq (D.9):

$$\rho_0(v,\xi) - \rho_1(v,\xi) = 2 \sum_{m\in V_{r+1}^c} d'_m d'_{m\oplus v} \left\{ |\hat{\eta}'_m\rangle\langle\hat{\eta}'_{m\oplus v}| + |\hat{\eta}'_{m\oplus v}\rangle\langle\hat{\eta}'_m| \right\}$$

As usual, the trace norm of this matrix serves as a bound on the information Eve receives. It is

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

Writing SD_v instead of $SD(\rho_0(v,\xi),\rho_1(v,\xi))$ for short:

$$SD_v \leq Tr \left| \sum_{m\in V_{r+1}^c} d'_m d'_{m\oplus v} \left\{ |\hat{\eta}'_m\rangle\langle\hat{\eta}'_{m\oplus v}| + |\hat{\eta}'_{m\oplus v}\rangle\langle\hat{\eta}'_m| \right\} \right|$$

$$\begin{aligned}
&\leq \sum_{m\in V_{r+1}^c} d'_m d'_{m\oplus v} Tr\left|\,|\hat{\eta}'_m\rangle\langle\hat{\eta}'_{m\oplus v}| + |\hat{\eta}'_{m\oplus v}\rangle\langle\hat{\eta}'_m|\,\right| \\
&= \sum_{m\in V_{r+1}^c} 2d'_m d'_{m\oplus v}\sqrt{1-[\Im(\langle\hat{\eta}'_m|\hat{\eta}'_{m\oplus v}\rangle)]^2} \\
&\leq \sum_{m\in V_{r+1}^c} 2d'_m d'_{m\oplus v}
\end{aligned}$$

where the sign $\Im$ means the imaginary part. Now we wish to give a bound in terms of the original values d_l. Using the fact that for any $\alpha > 0$ and any x, y (which are real numbers), $0 \leq (\alpha^{\frac{1}{2}}x - \alpha^{-\frac{1}{2}}y)^2 = \alpha x^2 + y^2/\alpha - 2xy$, we get the general inequality $2xy \leq \alpha x^2 + \frac{1}{\alpha}y^2$ and so

$$\begin{aligned}
SD_v &\leq \sum_{m\in V_{r+1}^c} 2d'_m d'_{m\oplus v} \\
&\leq \sum_{\substack{m\in V_{r+1}^c \\ d_H(m,V_r)\geq\hat{v}/2}} 2d'_m d'_{m\oplus v} + \sum_{\substack{m\in V_{r+1}^c \\ d_H(m,V_r)<\hat{v}/2}} 2d'_m d'_{m\oplus v} \\
&\leq \sum_{\substack{m\in V_{r+1}^c \\ d_H(m,V_r)\geq\hat{v}/2}} \left[\alpha d'^2_{m\oplus v} + \frac{1}{\alpha}d'^2_m\right] + \sum_{\substack{m\in V_{r+1}^c \\ d_H(m,V_r)<\hat{v}/2}} \left[\alpha d'^2_m + \frac{1}{\alpha}d'^2_{m\oplus v}\right] \\
&\leq \alpha \sum_{m\in V_r^c} d'^2_m + \frac{1}{\alpha} \sum_{\substack{m\in V_r^c \\ d_H(m,V_r)\geq\hat{v}/2}} d'^2_m \qquad \text{by Eqs. (D.10,D.11)} \\
&\leq \alpha \sum_{l\in V_r^c\oplus V_r} d_l^2 + \frac{1}{\alpha} \sum_{\substack{m\in V_r^c, n\in V_r \\ d_H(m,V_r)\geq\hat{v}/2}} d^2_{m\oplus n} \qquad \text{by Eq. (D.7)} \\
&\leq \alpha + \frac{1}{\alpha} \sum_{|l|\geq\frac{\hat{v}}{2}} d_l^2
\end{aligned}$$

Now we fix $\alpha = \sqrt{\sum_{|l|\geq\frac{\hat{v}}{2}} d_l^2}$ and obtain:

$$SD_v \leq 2\sqrt{\sum_{|l|\geq\frac{\hat{v}}{2}} d_l^2} \qquad \text{(D.12)}$$

□

Note that $\hat{v} = d_H(v, V_r)$ where r is the number of parity check strings.

E Existence of Codes for Both Reliability and Security

Choosing a code which is good when n is large (for constant error rate) is not a trivial problem in ECC. A Random Linear Code (RLC) is one such code, however,

it does not promise us that the distances are as required, but only gives the desired distances with probability as close to one as we want. With RLC, we find that the threshold below which a secure key can be obtained is $p_{\text{allowed}} \leq 7.56\%$.

In order to correct t errors with certainty, a code must have a minimal Hamming distance between the code words $d \geq 2t+1$ so that all original code words, even when distorted by t errors, can still be identified correctly. For any c_T which passes the test, we are promised (due to Lemma 5.4) that the probability of having $t = |c_I| > n(p_{\text{allowed}} + \epsilon_{\text{rel}})$ errors is smaller than $h = e^{-\frac{1}{2}n\epsilon_{\text{rel}}^2}$.

Thus, we need to choose a RLC that promises a Hamming distance at least d such that $p_{\text{allowed}} + \epsilon_{\text{rel}} < t/n = \frac{d-1}{2n}$, and then the t errors are corrected except for a probability smaller than $h = e^{-\frac{1}{2}n\epsilon_{\text{rel}}^2}$. However, RLC can never promise a specific minimal distance with certainty, but can only promise it with probability exponentially close to one: For any $n, r = n-k$, and for δ such that $H_2(\delta) < r/n$, an arbitrary *random linear code* (n,k,d) satisfies $d/n \geq \delta$, except for a probability (see [19], Theorem 2.2)

$$P[d/n < \delta] \leq \frac{c(\delta)}{\sqrt{n}} 2^{n(H_2(\delta)-r/n)} \stackrel{\Delta}{=} g_1 \tag{E.1}$$

where $c(\delta) = \frac{1}{1-2\delta}\sqrt{\frac{1-\delta}{2\pi\delta}}$.

If we choose $\delta = 2(p_{\text{allowed}}+\epsilon_{\text{rel}})+1/n$ then we are promised that the errors are corrected, except for some probability (bounded by h) that the error rate is larger than expected, and some probability (bounded by g_1) that a bad random code was chosen.

Using such a code, ϵ_{rel} is now a function of δ so that $\epsilon_{\text{rel}} = \delta/2 - 1/(2n) - p_{\text{allowed}}$ and therefore,

$$h = e^{-(n/8)(\delta-\frac{1}{n}-2p_{\text{allowed}})^2} \tag{E.2}$$

and almost all such codes correct all the errors. One could conclude that the code is reliable except for a probability $g_1 + h$, but this is not the case here; although the code is randomly produced, it can still be checked in advance, and used only if it satisfies the condition on d. Thus the term g_1 does not need to be added[15] to the reliability bound, and the bound is then given by h alone.

Recall that we choose ϵ_{sec} such that $|v| \geq 2n(p_{\text{allowed}} + \epsilon_{\text{sec}})$. Let $|v|$ be the minimal distance between one PA string and any other parity check string (or linear combination) taken from ECC and PA. Clearly, the Hamming weight of the dual code of the ECC, once the PA is also added, provides a lower bound on $|v|$. Thus, it is sufficient to demand $d^{\perp} \geq 2n(p_{\text{allowed}} + \epsilon_{\text{sec}})$ in order to prove security. Choosing a RLC for the ECC and PA, one cannot be completely sure that the distance indeed satisfies the constraint, but this shall be true [19] with probability exponentially close to one (and can be checked in advance). We use the dual code

[15] We can still add the term g_1 and this saves us the need to find the minimal distance of the code.

$(n, r^\perp, d^\perp)$, where $r^\perp = n - r - m$. Such codes satisfy $d^\perp/n \geq \delta^\perp$, except for a fraction of

$$P[d^\perp/n < \delta^\perp] \leq \frac{c(\delta^\perp)}{\sqrt{n}} 2^{n(H_2(\delta^\perp)-(n-r-m)/n)} = g_2 \tag{E.3}$$

with $\delta^\perp = 2(p_{\text{allowed}} + \epsilon_{\text{sec}})$.

Assuming that Eve gets full information (namely, m bits) when the code fails we get due to the above and Proposition 5.1

$$\langle \mathbf{I}'_{Eve} \rangle \leq m \left(2e^{-\frac{1}{4} n \epsilon_{\text{sec}}^2} + g_2 \right) \tag{E.4}$$

but we can get rid of g_2 by checking the code in advance[16]. If we demand that

$$\begin{aligned} H_2(\delta) - r/n &< 0 \\ H_2(\delta^\perp) + r/n + m/n - 1 &< 0 \,, \end{aligned}$$

then both g_1 and g_2 are exponentially small. Written another way:

$$\begin{aligned} H_2(2p_{\text{allowed}} + 2\epsilon_{\text{rel}} + 1/n) &< r/n \\ H_2(2p_{\text{allowed}} + 2\epsilon_{\text{sec}}) + r/n &< 1 - R_{\text{secret}} \end{aligned}$$

where $R_{\text{secret}} \equiv m/n$.

In order to find the threshold on p_{allowed} we combine these two equations together

$$H_2(2p_{\text{allowed}} + 2\epsilon_{\text{sec}}) + H_2(2p_{\text{allowed}} + 2\epsilon_{\text{rel}} + 1/n) < 1 - R_{\text{secret}} \,. \tag{E.5}$$

In the limit of large n and the two ϵ's close to zero, we get that $p_{\text{allowed}} < 5.50\%$ satisfies the bound and hence this is our threshold. [We can then chose the appropriate r/n so that both g_1 and g_2 functions are exponentially small.]

Asymptotically, a final key with a bit-rate $R_{\text{secret}} < 1 - H_2(2p_a) - H_2(2p_a)$ is secure and reliable for the given ECC+PA chosen at random. Note, as p_a goes to zero, R_{secret} goes to 1, which means all the information bits are secret (asymptotically).

The above result can be improved (as noticed first by Mayers [27]) by taking RLC with distance $d = t+1$ instead of $d = 2t+1$. Namely, $d-1 \geq n(p_{\text{allowed}}+\epsilon_{\text{rel}})$ (without the factor of 2). Due to Shannon's bound [25] such a code can also correct $t = n(p_{\text{allowed}} + \epsilon_{\text{rel}})$ errors with probability of failure smaller than $\hat{\delta}$ (for any $\hat{\delta}$). This is true provided that $r/n > H_2(p_{\text{allowed}} + \epsilon_{\text{rel}})$, and that a sufficiently large n is chosen, but we did not find an explicit connection between n and $\hat{\delta}$, as we did with the other probabilities g_1, g_2, and h.

The above is true except for an exponentially small probability g'_1 that the code got the wrong distance [19], and an exponentially small probability h' that the code is fine yet there are more errors in the information bits than expected.

[16] Or we can add that term to Eve's information and this saves us the need to find the minimal distance of the dual code.

Choosing now $\delta = p_{\text{allowed}} + \epsilon_{\text{rel}} + 1/n$, the term g_1' is still the same as before, but with a different δ then before. The condition for g_1' to be exponentially small becomes now

$$H_2(p_{\text{allowed}} + \epsilon_{\text{rel}} + 1/n) < r/n \; .$$

The term h' (telling us the probability of having more errors on the information bits than expected from the test results) is

$$h' = e^{-(n/2)(\delta - \frac{1}{n} - p_{\text{allowed}})^2} \; .$$

One could conclude that the code is reliable except for a probability $g_1' + h' + \hat{\delta}$, but (again) the term g_1' can be removed if we check the code in advance to make sure it has the right distance. The bound is thus given by $h' + \hat{\delta}$. However, we do not have an exponentially small expression for $\hat{\delta}$ (as a function of n) and it is only known that we can render the error as small as we want by taking a sufficiently large n.

For the security proof we choose ϵ_{sec} such that $|v| \geq 2n(p_{\text{allowed}} + \epsilon_{\text{sec}})$, and we demand $d^{\perp} \geq 2n(p_{\text{allowed}} + \epsilon_{\text{sec}})$. Choosing a RLC for the ECC and PA, one cannot be completely sure that the distance indeed satisfies the constraint, but this shall be true with probability exponentially close to one (and can be checked in advance). As before, we use the dual code $(n, r^{\perp}, d^{\perp})$, where $r^{\perp} = n - r - m$. Such codes satisfy $d^{\perp}/n \geq \delta^{\perp}$, except for a fraction of

$$P[d^{\perp}/n < \delta^{\perp}] \leq \frac{c(\delta^{\perp})}{\sqrt{n}} 2^{n(H_2(\delta^{\perp}) - (n-r-m)/n)} = g_2' \qquad \text{(E.6)}$$

with $\delta^{\perp} = 2(p_{\text{allowed}} + \epsilon_{\text{sec}})$. As before, we can get rid of g_2' by checking the code in advance.

In order for g_2' to be exponentially small we demand

$$H_2(\delta^{\perp}) + r/n + m/n - 1 < 0 \; ,$$

so finally:

$$H_2(p_{\text{allowed}} + \epsilon_{\text{rel}} + 1/n) < r/n$$
$$H_2(2p_{\text{allowed}} + 2\epsilon_{\text{sec}}) + H_2(p_{\text{allowed}} + \epsilon_{\text{rel}} + 1/n) < 1 - R_{\text{secret}}$$

where $R_{\text{secret}} \equiv m/n$.

In the limit of large n and ϵ's close to zero, $p_{\text{allowed}} < 7.56\%$ satisfies the bound and hence this is our imporved threshold. With this threshold we have an explicit bound on Eve's information, but only an asymptotic bound for the probability of failing in terms of reliability.