

## *Dr Rafał Demkowicz - Dobrzański*

Opracowanie utworu pod tytułem:

*„Informatyka kwantowa: teoretyczne zagadnienia kwantowej kryptografii i teorii estymacji” w ramach kursu zaawansowanego organizowanego dniach 29.06 – 24.07.09 będącego kontynuacją szkoleń z zakresu zarządzania dużą infrastrukturą badawczą organizowanego przez Narodowe Laboratorium Technologii Kwantowych*



**INNOWACYJNA GOSPODARKA**

NARODOWA STRATEGIA SPÓJNOŚCI

**UNIA EUROPEJSKA**  
EUROPEJSKI FUNDUSZ  
ROZWOJU REGIONALNEGO



Projekt współfinansowany ze środków Europejskiego Funduszu Rozwoju Regionalnego w ramach Programu Operacyjnego Innowacyjna Gospodarka



**INNOWACYJNA  
GOSPODARKA**  
NARODOWA STRATEGIA SPÓJNOŚCI

**UNIA EUROPEJSKA**  
EUROPEJSKI FUNDUSZ  
ROZWOJU REGIONALNEGO



# Estimate, Clone and Eavesdrop! - what you can do with an unknown quantum state?

Rafał Demkowicz-Dobrzański

September 3, 2009



Narodowe  
Laboratorium  
Technologii  
Kwantowych

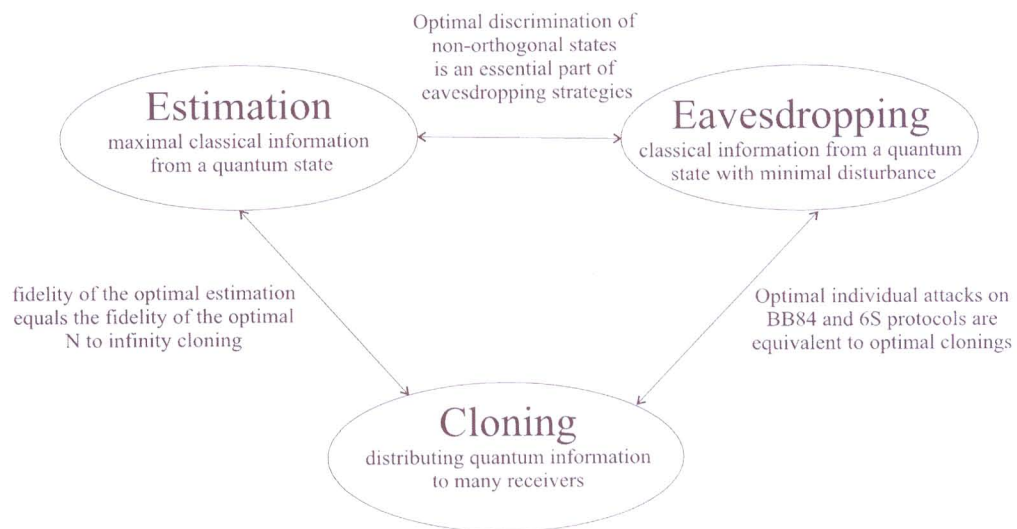


Figure 1: Interplay between cloning, estimation and eavesdropping



# Chapter 1

## Elements of classical information

### 1.1 Shannon entropy – Data compression

#### 1.1.1 Motivation

Imagine you want to transmit a message where each letter can be one of 4 symbols  $a, b, c, d$  via a binary channel – encoding the message as a sequence 01111010.... How to encode the symbols to encode the message using the smallest number of bits?

If symbols  $a, b, c, d$  appear with equal frequency  $p(a) = p(b) = p(c) = p(d) = 1/4$  you will probably assign two digit encodings to them:

symbols	$a$	$b$	$c$	$d$	(1.1)
probability	$1/4$	$1/4$	$1/4$	$1/4$	
encoding	00	01	10	11	

You need two bits per letter transmitted. Imagine now that in the message symbols appear with not equal frequencies e.g.  $p(a) = 1/2, p(b) = 1/4, p(c) = 1/8, p(d) = 1/8$ . What is the most economical way to encode the symbols? You may try this:

symbols	$a$	$b$	$c$	$d$	(1.2)
probability	$1/2$	$1/4$	$1/8$	$1/8$	
encoding	0	10	110	111	

On average you use  $1/2 \cdot 1 + 1/4 \cdot 2 + 1/8 \cdot 3 + 1/8 \cdot 3 = 1.75$ bits per letter transmitted!. The optimal transmission rate is quantified by Shannon





entropy. Notice that these are instantaneous codes, we know when a given codeword ends and can decode it without a reference to future codewords.

### 1.1.2 Definition

Let  $X$  be a random variable, with possible outcomes  $x \in \underbrace{\{0, 1, 2, \dots\}}_{\mathcal{X}}$ .

Let  $p(x)$  be a probability of outcome  $x$ . Shannon entropy is defined as:

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log_2 p(x). \quad (1.3)$$

$H(X)$  is given in bits. Notice that  $H(X) \geq 0$ .  $H(X)$  should intuitively be understood as a measure of randomness of random variable  $X$ . Alternatively one can regard it as an amount of information one gets once learning the exact outcome of the variable. Some simple examples if we consider only binary outcomes  $\mathcal{X} = \{0, 1\}$ :

- $p(0) = 1/2, p(1) = 1/2$  – complete randomness:  $H(X) = 1$
- $p(0) = 0, p(1) = 1$  – no randomness:  $H(X) = -0 \log_2 0 = 0$
- $p(0) = p, p(1) = 1 - p, H(X) = -p \log_2 p - (1 - p) \log_2 (1 - p)$

Let us calculate Shannon entropy for examples presented in previous subsection. In the case of equal frequencies we have  $H(X) = -4 \cdot 1/4 \log_2 1/4 = 2$ , in the second example we have

$$H(X) = -1/2 \log_2 1/2 - 1/4 \log 1/4 - 2 \cdot 1/8 \log 1/8 = 1.75. \quad (1.4)$$

Let us denote  $h(p) = -p \log_2 p$ . The function is plotted in Fig. ?? This function is concave which means that for any weights  $w_i \geq 0$ , which sum up to one we have:  $\sum_i w_i h(p_i) \leq h(\sum_i w_i p_i)$ . Since  $H(X) = \sum_{x_i \in \mathcal{X}} h(p(x_i))$ , and the sum of concave functions is concave we have:

$$\frac{1}{|\mathcal{X}|} H(X) = \frac{1}{|\mathcal{X}|} \sum_{x_i \in \mathcal{X}} h(p(x_i)) \leq \mathcal{H} \left( \frac{1}{|\mathcal{X}|} \sum_i p(x_i) \right) \leq h(1/|\mathcal{X}|), \quad (1.5)$$

where  $|\mathcal{X}|$  denotes the number of elements in the set  $\mathcal{X}$ . Hence we have  $H(X) \leq \log_2 |\mathcal{X}|$ .

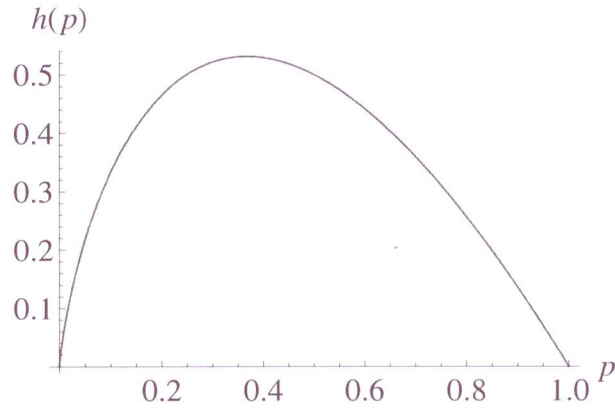


Figure 1.1: The function  $h(p) = -p \log_2 p$ . It is visible it is concave

Given many variables  $X_1, \dots, X_N$  which joint probability distribution is  $p(x_1, \dots, x_N)$  their entropy is defined:

$$H(X_1, \dots, X_N) = \sum_{x_1, \dots, x_N} -p(x_1, \dots, x_N) \log_2 p(x_1, \dots, x_N). \quad (1.6)$$

Notice the inequality:

$$H(X_i) \leq H(X_1, \dots, X_N) \leq \sum_i H(X_i), \quad (1.7)$$

which means that the entropy of the full system is larger than each of sub-systems, and that correlations decrease the entropy.

In particular given  $N$  independent realization of random variable  $X$ , we have:

$$H(X^N) = NH(X). \quad (1.8)$$

### 1.1.3 Relative entropy

For future use we introduce here the concept of relative entropy. The relative entropy  $D(p \parallel q)$  of probability distribution  $p$  with respect to  $q$  is defined:

$$D(p \parallel q) = \sum_i p_i \log_2 \frac{p_i}{q_i} \quad (1.9)$$



It is a measure of distinguishability between two probability distributions. Notice, however, that this is not a symmetric function and hence cannot be regarded as a proper measure of a distance between probability distributions. The relative entropy is always positive:

$$D(p \parallel q) = - \sum_i p_i \log_2 \frac{q_i}{p_i} \leq \log_2 \left( \sum_i p_i \frac{q_i}{p_i} \right) = 0, \quad (1.10)$$

where we have used concavity of  $-\log_2 t$  function.

#### 1.1.4 Shannon source coding theorem

Given the random variable  $X$  with Shannon entropy  $H(X)$  the minimal average length of a codeword  $L$  we need to use is bounded by:

$$H(X) \leq L \leq H(X) + 1 \quad (1.11)$$

*Lemma. Kraft inequality* We first proof the Kraft inequality. Suppose we want to have a code with codewords of length  $l_i$ . If the code is to be instantaneous we have a constraint on the lengths of the codewords. Consider a tree, where at each node a branch splits into two. Let  $l_{\max}$  be the maximum length of a codeword, which corresponds to the number of levels in the tree. We have total  $2^{l_{\max}}$  leaves and hence this many different codewords. If a given codeword has length  $l_i$  which is shorter than  $l_{\max}$ , however, it automatically exclude  $2^{l_{\max}-l_i}$  codewords from being used (all leaves having stemming from this branch). Since the total number of codewords is  $2^{l_{\max}}$  we have an inequality:  $\sum_i 2^{l_{\max}-l_i} \leq 2^{l_{\max}}$ , which leads to:

$$\sum_i 2^{-l_i} \leq 1 \quad (1.12)$$

called Kraft inequality. A codeword of given lengths exist iff the Kraft inequality is satisfied.

*Proof.* Given probabilities  $p_i$ , we construct codewords of length:

$$\log_2 \frac{1}{p_i} \leq l_i \leq \log_2 \frac{1}{p_i} + 1. \quad (1.13)$$

Note that the first inequality is equivalent to  $p_i \geq 2^{-l_i}$ , and the Kraft inequality is satisfied. The average codeword length reads:

$$H(X) \leq \sum_i p_i l_i \leq H(X) + 1. \quad (1.14)$$



What remains to be proven is that one can do no better than  $H(X)$ . Let us define probability distribution  $q_i = 2^{-l_i} / \sum_i 2^{-l_i}$ . Positivity of relative entropy  $D(p \parallel q)$  leads to:

$$-\sum_i p_i (\log_2 p_i - \log_2 q_i) = -\sum_i p_i (\log_2 p_i - l_i - \log_2 \sum_j 2^{-l_j}) \geq 0 \quad (1.15)$$

$$\sum_i p_i l_i \geq H(X) + \sum_i p_i \log_2 \left( \sum_j 2^{-l_j} \right) \quad (1.16)$$

Thanks to the Kraft inequality the second term on the right hand side above is nonnegative. Finally we have

$$\sum_i p_i l_i \geq H(X). \quad (1.17)$$

Which ends the proof ■. Actually this inequality holds also for non-instantaneous codes the proof is a bit more complicated then.

Instead of single letter encoding consider block encoding when we use codewords encoding  $N$  letter words. If  $L_N$  denotes average codeword length for  $N$ -letter block encoding, using Eqs (1.8,1.11) we have:

$$NH(X) \leq L_N \leq NH(X) + 1 \quad (1.18)$$

hence

$$H(X) \leq \frac{L_N}{N} \leq H(X) + \frac{1}{N}. \quad (1.19)$$

As a result asymptotically  $L_N/N \lim_{N \rightarrow \infty} = H(X)$ , so the rate of transmission given by Shannon entropy can be saturated in the limit of large  $N$ .

### 1.1.5 Typical sequences

How intuitively understand that optimal data compression of  $N$  bits can be done using  $NH(X)$  bits. Let  $X$  be binary random variable, taking value 0 with probability  $q$  and 1 with probability  $1 - q$ . Let us take a long sequence of  $N$  realizations of  $X$ . If sequence is long we will usually have sequence with approximately  $qN$  bits 0 and  $(1 - q)N$  bits 1. Probability of a given sequence is

$$p(x_1, \dots, x_N) = q^{qN} (1 - q)^{(1-q)N} \quad (1.20)$$







Hence:

$$\log_2 p(x_1, \dots, x_N) = -NH(X) \quad (1.21)$$

$$p(x_1, \dots, x_N) = 2^{-NH(X)}. \quad (1.22)$$

Since these are approximately only sequences that happen, we have  $2^{NH(X)}$  typical sequences. In compression when encoding large blocks in order to transmit  $2^N$  sequences we need only use  $2^{NH(X)}$  codewords. This is an intuitive understanding of the result from previous section.

## 1.2 Shannon mutual information – Communication over noisy channel

Consider a channel which is noisy and can flip transmitted bits with some probability. Let  $X$  be input random variable and  $Y$  be a random variable describing the output of the channel. Conditional probability  $p(y_j|x_i)$  describes the action of the channel. One would like to know how to protect transmitted information against errors and what is maximal number of logical bits that can be transmitted per one physical bit sent (channel capacity). We start by quantify correlations between two random variables  $X$  and  $Y$ .

### 1.2.1 Conditional entropy

Let joint probability of  $X$  and  $Y$  be  $p(x_i, y_j)$ . One quantifies the amount of randomness of random variable  $Y$  provided one knows the value  $x_i$  of random variable  $X$  using conditional entropy:

$$H(Y|x_i) = - \sum_j p(y_j|x_i) \log_2 p(y_j|x_i) \quad (1.23)$$

On average the conditional entropy reads:

$$H(Y|X) = - \sum_{ij} p(x_i)p(y_j|x_i) \log_2 p(y_j|x_i) = - \sum_{ij} p(x_i, y_j) \log_2 p(y_j|x_i). \quad (1.24)$$

This characterizes randomness of random variable  $Y$  provided variable  $X$  is known.



## 1.2.2 Mutual Information

Let us introduce the measure of correlation between random variables  $X$  and  $Y$  which will represent: How much do we learn about variable  $Y$  once we learn the value of variable  $X$ :

$$I(X : Y) = H(Y) - H(Y|X) \quad (1.25)$$

and is called *mutual information*. It is symmetric in  $X, Y$  since:

$$I(X : Y) = H(X) + H(Y) - H(X, Y). \quad (1.26)$$

It is zero iff  $X$  and  $Y$  are uncorrelated. Notice that if  $p(x, y)$  is the joint probability distribution, and by  $p_x(x), p_y(y)$  we denote its marginal distributions we have:

$$I(X : Y) = D(p(x, y) \parallel p_x(x)p_y(y)), \quad (1.27)$$

which reflect the fact that mutual information measures in some sense the distance between  $p(x, y)$  and uncorrelated probability distribution with the same marginals.

## 1.2.3 Channel capacity

Let us consider a channel, which action is described by  $p(y_j|x_i)$ . Let the input random variable be  $X$ . Let us transmit  $N$  bits via the channel using typical sequences. Transmitting a sequence of  $N$  bits, there are on average  $2^{H(Y|X)}$  typical error sequences (similar argument as in Sec. 1.1.5). . At the output we have  $2^{NH(Y)}$  typical sequence. For reliable transmission we can use at most  $2^{NH(Y)}/2^{NH(Y|X)} = 2^{NI(X:Y)}$  different inputs – we cannot use all typical sequences. Hence we can transmit at most  $I(X : Y)$  logical bits per one physical bit transmitted. Moreover this rate can be achieved. .

Channel capacity is defined (Shannon noisy channel theorem):

$$C = \max_{p(x)} I(X : Y). \quad (1.28)$$

Where we maximize over input probability distribution. For binary symmetric channel optimal choice for input probability distribution is  $p(0) = p(1) = 1/2$ .





## Chapter 2

# Quantum state estimation - Quantum->Classical channel

### 2.1 State discrimination - two states

Imagine you are given one of two states  $|\psi_1\rangle, |\psi_2\rangle$ . Your goal is to perform a measurement in order to determine which state we received.

The task is simple provided states are orthogonal  $\langle\psi_1|\psi_2\rangle = 0$ . One simply performs projective measurement in  $\{|\psi_1\rangle, |\psi_2\rangle\}$  basis. Distinguishability is perfect. Let us now take nonorthogonal, nonidentical states  $0 < |\langle\psi_1|\psi_2\rangle| < 1$ . Assume for simplicity that both states are equiprobable.

#### 2.1.1 Minimizing probability of error

We want to minimize probability of error in discrimination. Let  $p(j|i)$  be the probability that we guess the state  $j$  when the actual state is  $i$ . We need to find the optimal measurement that will minimize the error. Since this is a two outcome measurement we model our measurement with two POVM:  $M_1, M_2$ , ( $\sum_i M_i = \mathbb{1}$ ). Probability distribution is given by  $p(j|i) = \text{Tr}(M_j|\psi_i\rangle\langle\psi_i|) = \langle\psi_i|M_j|\psi_i\rangle$ . Error we want to minimize is given by

$$E = \frac{1}{2}\langle\psi_1|M_2|\psi_1\rangle + \frac{1}{2}\langle\psi_2|M_1|\psi_2\rangle. \quad (2.1)$$

We want to minimize  $E$  over  $\{M_i\}$ . Substituting  $M_2 = \mathbb{1} - M_1$  we have:

$$E = \frac{1}{2} + \frac{1}{2}\text{Tr}[M_1(|\psi_2\rangle\langle\psi_2| - |\psi_1\rangle\langle\psi_1|)]. \quad (2.2)$$



Without losing generality we may take:

$$|\psi_1\rangle = \sin \frac{\theta}{2} |0\rangle + \cos \frac{\theta}{2} |1\rangle = \begin{pmatrix} \sin \frac{\theta}{2} \\ \cos \frac{\theta}{2} \end{pmatrix} \quad (2.3)$$

$$|\psi_2\rangle = -\sin \frac{\theta}{2} |0\rangle + \cos \frac{\theta}{2} |1\rangle = \begin{pmatrix} -\sin \frac{\theta}{2} \\ \cos \frac{\theta}{2} \end{pmatrix}. \quad (2.4)$$

Introducing

$$W = |\psi_2\rangle\langle\psi_2| - |\psi_1\rangle\langle\psi_1| = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = |+\rangle\langle+| - |-\rangle\langle-|, \quad (2.5)$$

where  $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ ,  $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$ . We see that the problem amounts to finding  $M_1$  such that  $\text{Tr}WM_1$  is maximal. Keeping constraints  $M_1 \geq 0$ , and  $M_2 \geq 0$  (which means  $M_1 \leq 1$  - has to have all eigenvalues no bigger than 1). Notice that  $\text{Tr}WM_1 \in [-1, 1]$  since for every  $|\psi\rangle$  we have  $\text{Tr}M_i|\psi\rangle\langle\psi| \in [0, 1]$ . Hence the optimal choice is to take:  $M_1 = |+\rangle\langle+|$  since it gives  $\text{Tr}WM_1 = 1$ . The optimal POVMs and minimal discrimination error thus read:

$$E = \frac{1}{2}(1 + |\langle\psi_2|+\rangle|^2 - |\langle\psi_1|+\rangle|^2) = \frac{1}{2}(1 - \sin \theta) \quad (2.6)$$

Notice that the optimal measurement is an von Neumann projection measurement (this is also true for discrimination of  $N$  linearly independent states). Only for  $\theta = \pi/2$  we have  $E = 0$ , hence perfect discrimination is possible only when states are orthogonal. This has profound consequences and leads to Quantum Cryptography. Replacing  $\sin \theta$  with a function of scalar product between two states we get:

$$E = \frac{1}{2} \left( 1 - \sqrt{1 - |\langle\psi_1|\psi_2\rangle|^2} \right). \quad (2.7)$$

This formula can be used easily for arbitrary states.

**Many copies** Imagine that as above you are given with probability 1/2 either  $|\psi_1\rangle$ , or  $|\psi_2\rangle$ , but this time not a single copy but  $N$  copies. So in fact you are given either  $|\psi_1\rangle^{\otimes N}$ , or  $|\psi_2\rangle^{\otimes N}$ . Using Eq. 2.7 we get:

$$E_N = \frac{1}{2} \left( 1 - \sqrt{1 - |\langle\psi_1|\psi_2\rangle|^{2N}} \right). \quad (2.8)$$



Notice that if  $\langle \psi_1 | \psi_2 \rangle < 1$ ,  $E_{N+1} < E_N$  so more copies we have the better is distinguishability. In particular

$$\lim_{N \rightarrow \infty} E_N = 0, \quad (2.9)$$

which means one can distinguish quantum states perfectly once one has an arbitrary large number of copies.

### 2.1.2 Unambiguous discrimination

We again face the problem of discriminating between  $|\psi_1\rangle$ ,  $|\psi_2\rangle$ , but this time we only tell which state we received when we are sure. Otherwise we say we do not know. The goal is to find a measurement strategy that will minimize the probability of cases when we do not know.

The measurement will be described by three POVMs (it is clear that von Neuman measurement cannot describe this):  $M_1$ ,  $M_2$ ,  $M_?$ , corresponding to the result that leads us to guess correctly that the state was  $|\psi_1\rangle$ , result that leads us to guess correctly that the state was  $|\psi_2\rangle$ , and the result when we say we do not know.

Unambiguity conditions read:

$$\langle \psi_2 | M_1 | \psi_2 \rangle = 0 \quad \langle \psi_1 | M_2 | \psi_1 \rangle = 0, \quad (2.10)$$

since  $M_1 \geq 0$ ,  $M_2 \geq 0$ , this leads to  $M_1 = \xi_1 |\psi_2^\perp\rangle\langle \psi_2^\perp|$ ,  $M_2 = \xi_2 |\psi_1^\perp\rangle\langle \psi_1^\perp|$ . Symmetry between  $|\psi_1\rangle$ ,  $|\psi_2\rangle$  allows us to take  $\xi_1 = \xi_2 =: \xi$ . Positivity of  $M_i$  requires  $\xi \geq 0$ . However, we have additional constraint, namely  $M_? = \mathbb{1} - M_1 - M_2 \geq 0$ . Using parametrization of states given in Eq. 2.3 this condition can be written as:

$$M_? = \mathbb{1} - \xi(|\psi_2^\perp\rangle\langle \psi_2^\perp| + |\psi_1^\perp\rangle\langle \psi_1^\perp|) = \begin{pmatrix} 1 - \xi(1 + \cos \theta) & 0 \\ 0 & 1 - \xi(1 - \cos \theta) \end{pmatrix} \quad (2.11)$$

This means that  $\xi$  is limited by  $\xi \leq 1/(1 + \cos \theta)$ . Taking  $\xi = 1/(1 + \cos \theta)$  we find that the optimal probability of successful discrimination reads:

$$1 - \frac{1}{2} \text{Tr} [M_?( |\psi_1\rangle\langle \psi_1| + |\psi_2\rangle\langle \psi_2| )] = 1 - |\langle \psi_1 | \psi_2 \rangle|. \quad (2.12)$$



### 2.1.3 Mutual information in ambiguous and unambiguous discrimination

What is better strategy ambiguous or unambiguous strategy if one aims at optimizing mutual information between classical values  $X = 0, 1$  encoded in states  $|\psi_1\rangle, |\psi_2\rangle$  and values  $Y$  obtained in the measurement.

For optimal ambiguous discrimination we have random variables  $X = \{0, 1\}$ ,  $Y = \{0, 1\}$ :

$$p(1|1) = p(0|0) = p_a \quad (2.13)$$

$$p(1|0) = p(0|1) = 1 - p_a, \quad \text{where} \quad (2.14)$$

$$p_a = \frac{1}{2} \left( 1 + \sqrt{1 - |\langle \psi_1 | \psi_2 \rangle|^2} \right) = \frac{1}{2} (1 + \sin \theta) \quad (2.15)$$

Thus mutual information  $I_a(X : Y)$  reads:

$$I_a(X : Y) = 1 + p_a \log_2 p_a + (1 - p_a) \log_2 (1 - p_a) \quad (2.16)$$

In unambiguous discrimination we have  $X = \{0, 1\}$ ,  $Y = \{0, 1, ?\}$

$$p(1|1) = p(0|0) = p_u \quad (2.17)$$

$$p(?|0) = p(?|1) = 1 - p_u \quad \text{where} \quad (2.18)$$

$$p_u = 1 - |\langle \psi_1 | \psi_2 \rangle| = 1 - \sqrt{|\cos \theta|} \quad (2.19)$$

And the mutual information reads:

$$I_u(X : Y) = p_u \quad (2.20)$$

Fig. 2.1 presents mutual information in both ambiguous and unambiguous discrimination.

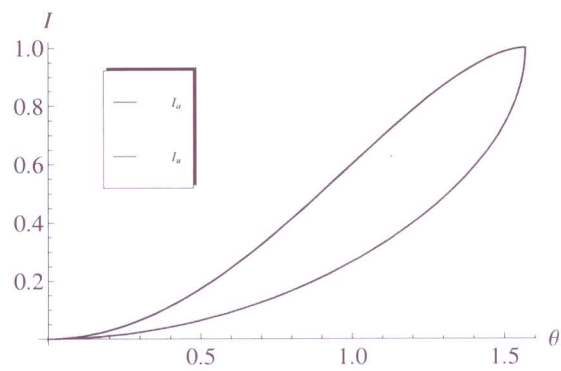


Figure 2.1: Comparison of mutual information obtained in the optimal ambiguous  $I_a$  and unambiguous  $I_u$  discrimination of two states





## Chapter 3

# Quantum cloning - Quantum- $\rightarrow$ Quantum channel

### 3.1 No-cloning theorems

Classical information in its digital form can be copied perfectly. Can you copy a unknown quantum state that is given to you. First of all you may grow suspicious since you have learned before that one can not discriminate nonorthogonal quantum states perfectly and hence one can not just measure and then reprepare more copies of the state, as this will induce unavoidable errors. But can you just copy an unknown state without measuring it? The answer is again no!

The general framework for cloning is the following. Consider the Hilbert space  $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_X$ , where  $\mathcal{H}_1$  is the space supporting the state of a system to be copied,  $\mathcal{H}_2$  supports the states of the system which is our “blank page”, and  $\mathcal{H}_X$  supports all other degrees of freedom including the copying machine and the rest of the universe. We say that the operation  $U$  (according to quantum theory should be unitary) performs cloning of a state  $|\psi\rangle$  iff:

$$U|\psi\rangle \otimes |0\rangle \otimes |X_0\rangle = |\psi\rangle \otimes |\psi\rangle \otimes |X_\psi\rangle. \quad (3.1)$$

In other words should produce a state  $|\psi\rangle$  in both systems 1 and 2 while the remaining degrees of freedom can change depending on the cloned state. Notice that the output state is a product state – there is no entanglement between subsystems. It has to be so, otherwise clones when inspected independently would be in mixed states.





### 3.1.1 Linearity $\rightarrow$ No-cloning of linearly dependent states

*Theorem.* If there is a quantum machine that can copy two quantum states  $|\psi_1\rangle, |\psi_2\rangle$  perfectly, then it cannot copy the state which is their linear superposition  $|\phi\rangle = a|\psi_1\rangle + b|\psi_2\rangle$ .

*Proof.* If an operation  $U$  clones two states  $|\psi_1\rangle, |\psi_2\rangle$  we have:

$$U|\psi_1\rangle \otimes |0\rangle \otimes |X_0\rangle = |\psi_1\rangle \otimes |\psi_1\rangle \otimes |X_{\psi_1}\rangle \quad (3.2)$$

$$U|\psi_2\rangle \otimes |0\rangle \otimes |X_0\rangle = |\psi_2\rangle \otimes |\psi_2\rangle \otimes |X_{\psi_2}\rangle. \quad (3.3)$$

Thanks to linearity of Quantum Mechanics:

$$U|\phi\rangle \otimes |0\rangle \otimes |X_0\rangle = aU|\psi_1\rangle \otimes |0\rangle \otimes |X_0\rangle + bU|\psi_2\rangle \otimes |0\rangle \otimes |X_0\rangle = a|\psi_1\rangle \otimes |\psi_1\rangle \otimes |X_{\psi_1}\rangle + |\psi_2\rangle \otimes |\psi_2\rangle \otimes |X_{\psi_2}\rangle \quad (3.4)$$

whereas for cloning of  $|\phi\rangle$  we would like to have at the output:

$$|\phi\rangle \otimes |\phi\rangle \otimes |X_\phi\rangle = (a|\psi_1\rangle + b|\psi_2\rangle) \otimes (a|\psi_1\rangle + b|\psi_2\rangle) \otimes |X_\phi\rangle \quad (3.5)$$

Clearly cloning of  $|\phi\rangle$  is impossible.

The above proof only made use of linearity of transformation and not unitarity. In particular the above proof does not forbid cloning of two nonorthogonal states. But this will come... Nevertheless the proof basing on linearity is useful since it also forbids probabilistic cloning of linearly dependent states – probabilistic transformation need not be unitary but are always linear.

### 3.1.2 Unitarity $\rightarrow$ No-cloning of non-orthogonal states

*Theorem* There is no deterministic cloning transformation (unitary) performing cloning for two nonorthogonal state

*Proof.* Let  $|\psi_1\rangle, |\psi_2\rangle$  be two different nonorthogonal states:  $0 < |\langle\psi_1|\psi_2\rangle| < 1$ . Assume the cloning is possible:

$$U|\psi_1\rangle \otimes |0\rangle \otimes |X_0\rangle = |\psi_1\rangle \otimes |\psi_1\rangle \otimes |X_{\psi_1}\rangle \quad (3.6)$$

$$U|\psi_2\rangle \otimes |0\rangle \otimes |X_0\rangle = |\psi_2\rangle \otimes |\psi_2\rangle \otimes |X_{\psi_2}\rangle. \quad (3.7)$$

Thanks to unitarity scalar product of input states should be equal to scalar product of output states:

$$\langle\psi_1|\psi_2\rangle\langle 0|0\rangle\langle X_0|X_0\rangle = \langle\psi_1|\psi_2\rangle\langle\psi_1|\psi_2\rangle\langle X_{\psi_1}|X_{\psi_2}\rangle \quad (3.8)$$



this leads to:

$$\langle \psi_1 | \psi_2 \rangle (1 - \langle \psi_1 | \psi_2 \rangle \langle X_{\psi_1} | X_{\psi_2} \rangle) = 0 \quad (3.9)$$

which is only possible for  $\langle \psi_1 | \psi_2 \rangle = 0$  or  $\langle X_{\psi_1} | X_{\psi_2} \rangle = 0$ , hence we arrive at contradiction and conclude that cloning of nonorthogonal states is impossible.

## 3.2 Optimal cloning

Since the perfect cloning is impossible except for very limited cases, we would like to investigate what is the best quality of copies that can be obtained. Let us consider a general problem of producing  $M$  imperfect copies out of  $N$  perfect originals  $|\psi\rangle^{\otimes N}$ . Let us consider unitary transformation  $U$  acting on the Hilbert space  $\mathcal{H}^{\otimes M} \otimes \mathcal{H}_X$ , where  $\mathcal{H}_X$  represents the space of the cloning machine:

$$U : |\psi\rangle^{\otimes N} \otimes |0\rangle^{\otimes (M-N)} \otimes |X\rangle \mapsto |\Psi\rangle_{M,X}. \quad (3.10)$$

The output  $|\Psi\rangle_{M,X}$  is in general an entangled states of all  $M$  copies and the cloning machine. Tracing out the space  $X$  we obtain the state of  $M$  copies:

$$\rho_M = \text{Tr}_X(|\Psi\rangle\langle\Psi|) \quad (3.11)$$

It should be remembered that this state typically will contain correlations between copies. Comparing the obtained state with the perfect  $M$  copies state  $|\psi\rangle^{\otimes M}$  can be done using *global fidelity* figure of merit:

$$F_G = \langle \psi |^{\otimes M} \rho_M | \psi \rangle^{\otimes M}. \quad (3.12)$$

If instead we are only interested in *single copy fidelity*, we can calculate single copy fidelity for the  $i$ -th copy:

$$F_i = \langle \psi | \rho_i | \psi \rangle, \quad (3.13)$$

where

$$\rho_i = \text{Tr}_{1,\dots,i-1,i+1,\dots,M}(\rho_M) \quad (3.14)$$

is obtained after tracing out all the copies instead of the  $i$ -th one. Provided all  $F_i$  are equal we call the cloning *symmetric*, otherwise we call it *asymmetric*.



### 3.2.1 Optimal 1 → 2 qubit asymmetric cloning cloning

We present below an intuitive construction of optimal 1 → 2 universal asymmetric cloning, without the proof for its optimality. Consider the following unitary operation acting on three qubits, denoted  $A$ ,  $B$ , and  $X$ :

$$V : |i\rangle_A \otimes |j\rangle_B \otimes |k\rangle_X \mapsto |i \oplus j \oplus k\rangle_A \otimes |i \oplus j\rangle_B \otimes |i \oplus k\rangle_X \quad (3.15)$$

Notice the following properties of operation  $V$ . If the system  $A$  is prepared in an unknown state  $|\psi\rangle_A$ , while subsystems  $BX$  are prepared in  $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$  we get:

$$V : |\psi\rangle_A \otimes |\Phi^+\rangle_{BX} \mapsto |\psi\rangle_A \otimes |\Phi^+\rangle_{BX}, \quad (3.16)$$

and hence the state  $|\psi\rangle_A$  remains where it was. On the other hand if at the input we take subsystems  $BX$  in the state  $|0\rangle_B \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)_X$  the transformation read:

$$V : |\psi\rangle_A \otimes |\Phi^+\rangle_{BX} \mapsto |\psi\rangle_B \otimes |\Phi^+\rangle_{AX}, \quad (3.17)$$

and as a result the state is “teleported” to subsystem  $B$ .

Since we want to have imperfect copies in both  $A$  and  $B$  subsystems it is natural to consider the transformation where initially we prepare systems  $BX$  in a superposition of  $|\Phi^+\rangle_{BX}$  and  $|0\rangle_B \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)_X$ . The transformation then reads:

$$V : |\psi\rangle_A \otimes \left( a|\Phi^+\rangle_{BX} + b|0\rangle_B \otimes \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)_X \right) \mapsto a|\psi\rangle_A \otimes |\Phi^+\rangle_{BX} + b|\psi\rangle_B \otimes |\Phi^+\rangle_{AX}. \quad (3.18)$$

Calculating the output single copy reduced density matrices we get:

$$\rho_A = \left[ \frac{(a+b)^2}{2} + \frac{a^2}{2} \right] |\psi\rangle\langle\psi| + \frac{b^2}{2} |\psi_\perp\rangle\langle\psi_\perp| \quad (3.19)$$

$$\rho_B = \left[ \frac{(a+b)^2}{2} + \frac{b^2}{2} \right] |\psi\rangle\langle\psi| + \frac{a^2}{2} |\psi_\perp\rangle\langle\psi_\perp| \quad (3.20)$$

where  $|\psi_\perp\rangle$  is the orthogonal state to  $|\psi\rangle$ . The corresponding fidelities read:

$$F_A = 1 - \frac{b^2}{2} \quad (3.21)$$

$$F_B = 1 - \frac{a^2}{2} \quad (3.22)$$



The *symmetric* cloning corresponds to the choice  $a = b$ , which together with normalization constraint implies  $a = b = 1/\sqrt{3}$  and leads to the optimal cloning fidelity  $F = 5/6$ .





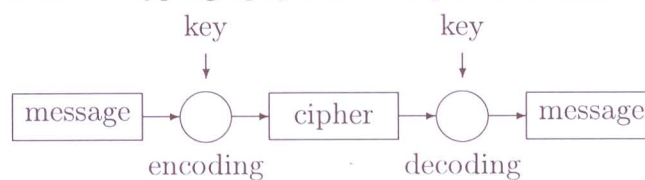
## Chapter 4

# Quantum eavesdropping - Quantum- $\rightarrow$ Quantum- $\rightarrow$ Classical channel

### 4.1 Cryptography

The science of cryptography is about transmitting a messages in the way that no illegitimate party can learn its meaning. One of the earliest cryptographic method was *Cesar cipher* in which a letter in a message was replaced by a letter  $k$  places further in the alphabet. If we took  $k = 3$  then CEASAR would be encoded as FADVDU. Such a code can be broken easily once one knows that the message was encoded using Cesar cipher. One simply has to check all possible values of  $k$ , which is the number of letters in the alphabet – 26, which is not a great amount of work.

The general scheme in cryptography can be depicted as follows:



In case of the Cesar cipher the message is CEASAR, the cipher is FADVDU and the key is  $k = 3$ .

A more general cipher is the substitution cipher, where each letter is mapped onto another letter. The Cesar cipher is an example of substitution



cipher. In a general substitution cipher we have  $26!$  possibilities. Hence, such a cipher can not be broken by checking all possible letter substitutions, and thus is more secure than the Caesar cipher. Nevertheless, it can be broken by letter frequency analysis, since each language has its particular letter frequency pattern, and one can quickly find out which letter was substituted to which one by investigating frequencies in which they appear.

Actually almost all used ciphers can theoretically be broken, and their strength stems from the practical difficulties of doing so. Nevertheless, if one really wants to use a cipher which is *proven* to be secure then there is such a cipher: *the one time pad*. Write your message in the binary form, take the key which is a completely random sequence of 0 and 1 of the same length as the message and perform bitwise XOR operations to obtain the cipher

message	0	1	1	0	1	1	0
key	0	0	0	1	0	1	1
cipher = message $\oplus$ key	0	1	1	1	1	0	1

Notice that since the key is completely random so is the message. More formally let  $K$  be the random variable associated with the key. Let the key have length  $n$ . Complete randomness means that all binary sequences are equally probable:  $p(K) = 1/2^n$ . The cipher is obtained as  $C = M \oplus K$ , where  $M$  is the message. What means perfect security? It means that the cipher carries no information about the message for someone who does not know the key. This corresponds to the statement that mutual information between  $M$  and  $C$  is zero:  $I(M : C) = 0$ .

*Proof security of one time pad.* Let  $p(M)$  be probability distribution of messages transmitted. The conditional probability  $p(C|M)$  that a cipher  $C$  is obtained from message  $M$  reads:

$$p(C|M) = \sum_K p(K) \delta_{C, M \oplus K} = 1/2^n. \quad (4.1)$$

Hence obviously we have  $p(C) = 1/2^n$ . This means the cipher is completely random. The mutual information  $I(M : C)$  reads:

$$I(M : C) = H(C) - H(C|M) = n - \sum_{C,M} p(C|M) p(M) \log_2 p(C|M) = 0. \quad (4.2)$$

Hence the one time pad is secure.





The one time pad has one drawback which makes it impractical for real life communication: it must be the same length as the transmitted message, otherwise the mutual information  $I(M : C)$  is not zero and the cipher could in principle be broken. The main obstacle is thus the effective distribution of the random key to legitimate parties. One of the most promising techniques for doing this is *quantum key distribution* (QKD).

## 4.2 Quantum key distribution

We have observed in previous chapters that nonorthogonal quantum states cannot be distinguished nor cloned perfectly. This *inaccessibility* of quantum states which seems only a nuisance at a first glance proves to be the key to secure information transmission.

### 4.2.1 BB84 protocol

Let us describe here the most famous protocol proposed by Bennet and Brassard in 1984 (BB84). Consider two parties  $A$ ,  $B$ , which are connected by a quantum channel allowing for transmission of qubits (e.g. an optical fiber in through which single photons are sent), and a classical channel (e.g. telephone). We assume that both channels are insecure and can be subjected to eavesdropping. We only assume that classical channel is authenticated i.e.  $A$  and  $B$  know that they talk to each other and their classical messages although potentially tapped cannot be altered.

$A$  and  $B$  will use photon polarization for qubits transmitted via the quantum channel.  $A$  will send to  $B$  one of four states:  $|\leftrightarrow\rangle, |\updownarrow\rangle, |\nearrow\rangle, |\searrow\rangle$  randomly with equal probabilities. We will say that the first two states form basis 1, and the last two basis 2.  $A$  and  $B$  assign logical values to these states as follows:

basis 1		basis 2	
$ \leftrightarrow\rangle$	$ \updownarrow\rangle$	$ \nearrow\rangle$	$ \searrow\rangle$
0	1	0	1

(4.3)

$B$  measures the polarization state of an incoming photon randomly in one of two bases. If he measures in the correct basis his results should be perfectly correlated with bits sent, whereas when he measures in the incorrect basis his results will be completely uncorrelated with that of  $A$ . After the transmission took place  $B$  communicates to  $A$  via the classical channel in which basis he



performed the measurement in a given run. He does not reveal, however, the actual results obtained. After this communication  $A$  i  $B$  keep only bits measured in compatible basis (approximately half). We call this a sifting stage. In ideal situation  $A$  and  $B$  should have perfectly correlated bits.

$A$	$\leftrightarrow$	$\nearrow$	$\nearrow$	$\searrow$	$\updownarrow$	$\updownarrow$	$\leftrightarrow$	$\nearrow$
$B$	+	×	+	+	+	×	×	×
compatible?	✓	✓			✓			✓
key	0	0			1			0

Now the quantum features enter the game. How  $A$  and  $B$  can be sure that they share bits that nobody else know about – i.e. that they have a one time pad. Put in one sentence this can be stated as follows:

You can not distinguish perfectly between 4 states used in BB84, and moreover you cannot learn anything about their identity without introducing disturbance.

Hence,  $A$  and  $B$  can make themselves sure that nobody have eavesdropped on their communication, by revealing part of their bits on classical channel (e.g. 100 bits), and checking whether they all agree. If there are no errors they can be sure with high degree of confidence (the higher the more bits they have revealed) that nobody had eavesdropped. If all bits agree, the revealed bits are of course discarded, while the remaining ones are kept and constitute the one time pad. If there are errors in bits revealed, however,  $A$  and  $B$  it suggest a presence of an eavesdropper and hence they abort communication and try again.

The above scenario is oversimplified, and also impractical. In reality there will always be errors in communication even if there is no eavesdropped, but which result from noise in the channel, imperfect detectors etc. Thus we need a more sophisticated approach: What is the tolerable error rate below which we can in some way distill a one time pad that will have no errors and will be secure i.e. no third party will have any information on it. This can be done using classical methods of error correction and privacy amplification.



## 4.3 Classical tools

### 4.3.1 Error correction

**Asymptotic considerations.** Imagine  $A$  and  $B$  share correlated bit strings of length  $n$ , where  $p(A, B)$  is the probability distribution of the strings. We assume that  $p(A, B) = p(a_1, b_1) \cdots p(a_n, b_n)$ , which means each pair of bits is independently identically distributed. From chapter 1 we know that the mutual information  $I(A : B) = nI(a : b)$  where by  $I(a : b)$  we mean the mutual information corresponding to single bit random variables. Error correction is the process in which we allow  $A$  and  $B$  to exchange additional  $m$  bits of information in order to correct all errors and have perfectly correlated strings i.e.  $I'(A : B) = n$ . If strings are long we can adapt Shannon typical sequences technique for proving capacity of noisy channels. If  $I(A : B) = nI(a : b)$  it means there is approximately  $2^{n(1-I(a:b))}$  typical sequences of  $B$  that could in principle have been created from a given sequence of  $A$  and vice versa. Hence in order to identify uniquely a sequence in  $B$  with a sequence in  $A$  one needs to send

$$m = n(1 - I(a : b)). \quad (4.4)$$

In other words one needs  $n(1 - I(a : b))$  additional bits in order to correct all errors. This is of course a theoretical bound. Real schemes will perform usually worse, yet the longer is the sequence the closer they can achieve the bound. More explicitly, if  $q$  is the probability of single bit error, then  $I(a : b) = 1 - (-q \log_2 q)$  and hence the number of bits needed to be communicated is:

$$m = -nq \log_2 q. \quad (4.5)$$

**Error correction in practice.** One grasp the intuition of error correction by considering the simplest example when  $n = 2$ . Let  $p$  be the probability that a given bit of  $A$  is the same as the bit of  $B$ . Let  $A$  take her two bits, calculate XOR function on them and communicate the result to  $B$ .  $B$  checks whether the XOR function of  $A$  bits agrees with XOR function of his bits. If this is true they keep their bits unchanged and if not  $A$  send the value of the first bit to  $B$  – hence she effectively has sent all two bits. Notice that the only possibility that bits of  $A$  and  $B$  disagree is that they were error on both bits. Hence after this error correction probability that  $A$  and  $B$  bits agree is  $p' = 1 - (1 - p)^2 = p(2 - p) \geq p$  (for  $p \in [0, 1]$ ). In this operations





$A$  has to send to  $B$  on average:  $m = 1 + 2p(1 - p)$  bits. This is usually than the theoretical bound.

For larger  $n$  one could also use a strategy using pairs of bits but this strategy is not very efficient – usually to many bits have to be communicated (at least  $n/2$ ). The following strategy is more efficient. Let us denote by  $q$  the approximate fraction of errors between strings of  $A$  and  $B$ . In QKD  $e$  is estimated from the revealed part of shared bits. Usually before error correction  $A$  and  $B$  apply a common random permutation of their bits in order not to distinguish any of them. After that they choose a block length  $k$  such that  $kq < 1$  i.e. such that it will be a very rare case that there is more than one error in the block, and divide their sequences in blocks of length  $k$ .  $A$  transmits the XOR function of bits in each block and communicate this to  $B$ . If XOR values of a given block agrees in  $A$  and  $B$  sequences they keep them intact. If they differ, they divide the block by half and  $A$  transmits XOR value calculated on subblocks. The subblock in which XOR values agree they leave it intact and divide by half the subblock in which XOR values differ, and continue this procedure until errors are localized. Doing so, sending approximately  $(n/k) \log_2 k$  bits they can correct all errors provided there were no more than one error in a block of length  $k$ . Notice that when  $kq = 1$  the number of bits communicated in this phase equals the theoretical bound in Eq. 4.5. However, there still may be errors remaining, due to the fact that there could have been more than one error in each block. Hence one has to repeat the procedure but this time with larger blocks  $k'$ , since now the probability of an error is smaller  $q' < q$ . A few repetitions and checking XOR functions should eliminate most errors. When only a few errors are left dividing into blocks is not particularly effective we simply, take a random subset and calculate its XOR, if for e.g. 20 random subsets XOR values agree we have  $2^{-(20)}$  probability that an error is still there.

### 4.3.2 Privacy amplification

In cryptography what we really need to consider is the three party probability distribution  $p(A, B, E)$ , where  $E$  represents the data acquired by an eavesdropper. After the error correction procedure  $A$  and  $B$  have the same sequences i.e.  $I(A : B) = n$ , but most probably  $E$  also share some knowledge on them. If, however,  $E$  knowledge is not perfect i.e.  $I(A : E) < n$ , and  $I(B : E) < n$  then  $A$  and  $B$  can perform so called privacy amplification procedure reducing their number of bits from  $n$  to  $n'$  but making them



completely unknown for  $E$ .

Let us start with the simplest example when  $n = 2$ .  $A$  and  $B$  have two identical bit sequences of length 2. Let  $q$  be the probability of making an error for  $E$  when trying to deduce the value of a bit of  $A$  or  $B$ .  $A$  and  $B$  can perform XOR operation on their bits and keep its value, but unlike in error correction procedure they do not announce it. In effect they have shortened their sequence to one bit. What is the error probability  $q'$ ,  $E$  will make when guessing this value. She will not make an error only if she knew correct values of both bits or made errors in predicting value of both of them, hence:  $q' = 1 - q^2 - (1 - q)^2 = 2q(1 - q) \geq q$  (for  $q \in [0, 1/2]$ ). Hence her knowledge about bits of  $A$  and  $B$  will decrease.

In practice when  $n$  is large the above strategy will be applied to large blocks. Namely after assessing  $q$ , we take  $k$  such that  $k(1 - e) \approx 1$ , and apply a hashing function  $f : \{0, 1\}^n \mapsto \{0, 1\}^k$  (the hashing function takes a binary sequence of length  $n$  and returns a binary sequence of length  $k$ , where a bit in the output sequence is a XOR function of a random subset of input sequence) to bits of  $A$  and  $B$ . In this way their sequences will be shortened from  $n$  to  $k$  bits but the sequences will become completely unknown to  $E$ . The important question is how large  $k$  one can take and still be sure that  $E$  has no information on sequences of  $A$  and  $B$ . Intuitively of course the larger is the  $E$  information on  $A$  and  $B$  the smaller has to be  $k$ . This intuition is formalized in the Csiszár-Körner theorem, which combines both privacy amplification and error correction considerations.

### 4.3.3 Csiszár-Körner theorem

Using one way error-correction and privacy amplification, the number of secret bits  $k$ ,  $A$  and  $B$  can distill is bounded:

$$k \leq \max\{I(A : B) - I(A : E), I(A : B) - I(B : E)\} \quad (4.6)$$

hence provided that  $E$  is less correlated with either  $A$  or  $B$  than they are with each other distillation of secret key is possible.

## 4.4 Attacks on the QKD

In order to apply error correction and privacy amplification we need to know how much  $E$  could have possibly learned about bits of  $A$  and  $B$ . Judg-





ing by the qubit error rate estimated from the revealed part of the bits  $A$  and  $B$  should find out what is the optimal attack  $E$  could have performed which allowed her to gain largest possible amount of information.

At the moment we will restrict ourselves to a simple class of attacks called intercept and resend attacks, which are not optimal, and hence considering only them does not guarantee full security, but are often considered since they are the only realistic attacks under present technology.

#### 4.4.1 Intercept and resend attacks on BB84

In general, in intercept and resend attacks (IRA),  $E$  first measures incoming qubit in some basis and after learning result of the measurement and prepares a corresponding state which she sends to  $B$ . Ideally (for  $E$  of course) she would like to learn what state was sent and resend exactly the same state to  $B$  in order not to be detected.

In BB84, two basis are used for communication, basis 1:  $|\leftrightarrow\rangle, |\updown\rangle$ , and basis 2:  $|\nearrow\rangle, |\searrow\rangle$ . During transmission  $E$  does not know which basis she should measure in since this is revealed only after all qubits has been sent. Consider two strategies she may choose:

1. Measurement in a randomly chosen basis – with probability  $1/2$ ,  $E$  measures either in  $|\leftrightarrow\rangle, |\updown\rangle$  or in a  $|\nearrow\rangle, |\searrow\rangle$  basis
2. Measurement in an intermediate basis – every time  $E$  measures in  $|\updown\rangle, |\nearrow\rangle$ , which is an basis “in between” two basis used in BB84

Let us calculate what is the information gained by  $E$  in each of this attacks and what disturbance this attacks cause in the data of  $A$  and  $B$ .

**Random basis** In half of the cases  $E$  will measure in correct basis, hence will learn the state and transmit the state without any disturbance. In the second half, she will measure in the wrong basis. Since  $|\langle\leftrightarrow|\nearrow\rangle|^2 = |\langle\leftrightarrow|\searrow\rangle|^2 = 1/2$  and  $|\langle\updown|\nearrow\rangle|^2 = |\langle\updown|\searrow\rangle|^2 = 1/2$ , she will obtain a correct measurement result with probability  $1/2$ . She will resend, a state in the wrong basis, however, and consequently  $B$  has  $1/2$  probability of registering an error in communication even though his basis is set according with that of  $A$ . Summarizing  $B$  on average will observe qubit error rate (QBER)  $QBER = 1/4$ . Probability that  $E$  will measure an incorrect bit sent by  $A$  is





$1/2 \cdot 1/2 = 1/4$ , hence errors will be the same as between  $A$  and  $B$  (Notice also that  $E$  also with probability  $1/4$  has an error on bit of  $B$ ). Summarizing:

$$I(A : B) = 1 - h[1/4] \approx 0.189 \quad (4.7)$$

where  $h[x] = -x \log_2 x - (1 - x) \log_2(1 - x)$  is binary Shannon entropy, and the identical equation should hold for  $I(A : E)$  and  $I(B : E)$ . However, in the above considerations we have neglected an important fact, namely after  $A$  and  $B$  announce basis they have used,  $E$  knows the cases when she measured in the correct basis. When she measured in the correct basis she has full knowledge on the bit, while when she measured on the wrong basis she learns nothing, consequently the true mutual information reads:

$$I(A : E) = I(B : E) = 1/2. \quad (4.8)$$

Obviously if  $A$  and  $B$  measure  $QBER = 1/4$  they should abort their communication since  $E$  in principle could have gained more information than they. Let us now consider a more general situation in which  $E$  intercept only  $r$  fraction of incoming qubits. In this case,  $QBER = r/4$ , and consequently:

$$I(A : B) = 1 - h[r/4] \quad (4.9)$$

$$I(A : E) = I(B : E) = r/2 \quad (4.10)$$

For  $r \simeq 0.6821$  which  $I(A : B) = I(A : E)$ , this corresponds to  $QBER = 0.171$ . This tells us that if  $QBER \geq 0.171$ , we cannot distill any secret key since an eavesdropper could have obtained the same amount of information as we have using IRA in random basis. Taking a positive approach, if we assume an eavesdropper was restricted to perform IRA in random basis and we detect  $QBER < 0.171$  we can distill some secret key which maximal length is given by Ciszár-Körner criterion and reads:

$$k \leq 1 - h[r/4] - r/2 = 1 - h[QBER] - 2QBER. \quad (4.11)$$

**Intermediate basis attack** Using intermediate basis, probability that  $E$  measures a wrong bit value is

$$q = |\langle 22.5^\circ | \uparrow \rangle|^2 = 1/4(2 - \sqrt{2}) \simeq 0.146. \quad (4.12)$$

Notice that this error is smaller than average error in random basis attack. Such an attack induces  $QBER = 2q(1 - q) = 1/4$ :

$$I(A : B) = 1 - h[1/4] \approx 0.189 \quad (4.13)$$



Unlike in random basis attack, learning what basis was used in a given run does not provide  $E$  with any additional information, hence

$$I(A : E) = I(B : E) = 1 - h[q] \simeq 0.399, \quad (4.14)$$

which is smaller than in random basis attack. Even though average probability of error is smaller for  $E$  in intermediate basis attack the mutual information is smaller due to lack of certainty in which cases bits were correct and in which they were useless. Assume again that  $E$  intercept only a fraction  $r$  of incoming qubits, we get:

$$I(A : B) = 1 - h[r/4] \quad (4.15)$$

$$I(A : E) = I(B : E) = 1 - h[qr + (1 - r)/2], \quad (4.16)$$

When  $r = 1/3(4 - \sqrt{2}) \simeq 0.8619$ ,  $I(A : B) = I(A : E) = I(B : E)$ . This corresponds to  $QBER = 0.215$ , so considering intermediate basis attacks we get a bit higher QBER rate thresholds above which we cannot distill secret key. This is due to the fact that intermediate attacks are less efficient from the point of view of an eavesdropper than random basis attacks.

#### 4.4.2 Optimal individual attack on BB84

Let us denote the four states used in BB84 protocols  $|+z\rangle, |-z\rangle, |+x\rangle, |-x\rangle$ , where  $|\pm z\rangle, |\pm x\rangle$  are the eigenstates of  $\sigma_x$  and  $\sigma_z$ . A general individual attack on BB84 is a unitary operation  $U$  acting on the space which describes states of the qubit send by  $A$  to  $B$  and the eavesdropper space:

$$U : |\pm k\rangle_B \otimes |0\rangle_E \mapsto \sqrt{F}|\pm k\rangle_B \otimes |\Psi_0^{\pm k}\rangle_E + \sqrt{D}|\mp k\rangle_B \otimes |\Psi_1^{\pm k}\rangle_E \quad (4.17)$$

where  $k = x, z$  represents the basis chosen,  $D$  is equal to the QBER, and  $F = 1 - D$ . Since we just have four different input states to consider, we can without losing generality limit the  $E$  space to 4 dimensions. Let us write the transformation when acting on the states from the  $z$  basis:

$$U : |+z\rangle_B \otimes |0\rangle_E \mapsto \sqrt{F}|+z\rangle_B \otimes |\Psi_0^{+z}\rangle_E + \sqrt{D}|-z\rangle_B \otimes |\Psi_1^{+z}\rangle_E \quad (4.18)$$

$$U : |-z\rangle_B \otimes |0\rangle_E \mapsto \sqrt{F}|-z\rangle_B \otimes |\Psi_0^{-z}\rangle_E + \sqrt{D}|+z\rangle_B \otimes |\Psi_1^{-z}\rangle_E \quad (4.19)$$

Unitarity implies the following constraint:

$$\langle \Psi_0^{+z} | \Psi_1^{-z} \rangle + \langle \Psi_1^{+z} | \Psi_0^{-z} \rangle = 0 \quad (4.20)$$

We assume the attack is symmetric which means that the output reduced density matrix has a Bloch vector which is shrunk but not rotated compared with the input one. Notice that this is not a limitation, since such rotation is useless as it increases the QBER while not providing any additional information for E, and can always be canceled by the eavesdropper. This implies that

$$\langle \Psi_0^{\pm z} | \Psi_1^{\pm z} \rangle = 0. \quad (4.21)$$

Now we look how the transformation acts on the state from the  $x$  basis. Recall that  $|\pm x\rangle = (|+z\rangle \pm |-z\rangle)/\sqrt{2}$ , hence the transformation reads:

$$\begin{aligned} U : | +x, 0 \rangle &\mapsto \sqrt{\frac{F}{2}}(|+z, \Psi_0^{+z}\rangle + |-z, \Psi_0^{-z}\rangle) + \sqrt{\frac{D}{2}}(|-z, \Psi_1^{+z}\rangle + |+z, \Psi_1^{-z}\rangle) = \\ &= \sqrt{F}|+x\rangle \otimes \frac{1}{2}(|\Psi_0^{+z}\rangle + \sqrt{\frac{D}{F}}|\Psi_1^{-z}\rangle + \sqrt{\frac{D}{F}}|\Psi_1^{+z}\rangle + |\Psi_0^{-z}\rangle) + \\ &+ \sqrt{D}|-x\rangle \otimes \frac{1}{2}(\sqrt{\frac{F}{D}}|\Psi_0^{+z}\rangle + |\Psi_1^{-z}\rangle - |\Psi_1^{+z}\rangle - \sqrt{\frac{F}{D}}|\Psi_0^{-z}\rangle) = \\ &= \sqrt{F}|+x\rangle \otimes |\Psi_0^{+x}\rangle + \sqrt{D}|-x\rangle \otimes |\Psi_1^{+x}\rangle \end{aligned} \quad (4.22)$$

and analogously for  $|-x\rangle$ . If the attack must treat the  $x$  basis in the same way as the  $z$  basis, we must have

$$\langle \Psi_0^{+x} | \Psi_1^{+x} \rangle = 0 \quad (4.23)$$

This implies:

$$\text{Re}\langle \Psi_0^{+z} | \Psi_1^{-z} \rangle - \text{Re}\langle \Psi_1^{+z} | \Psi_0^{-z} \rangle = 0 \quad (4.24)$$

$$\sqrt{\frac{F}{D}}\text{Im}\langle \Psi_0^{-z} | \Psi_0^{+z} \rangle + \sqrt{\frac{D}{F}}\text{Im}\langle \Psi_1^{+z} | \Psi_1^{-z} \rangle = 0 \quad (4.25)$$

Without losing generality, we can always redefine:  $|\Psi_1^{+z}\rangle \rightarrow e^{i\varphi}|\Psi_1^{+z}\rangle$ , such that  $\text{Im}\langle \Psi_1^{+z} | \Psi_1^{-z} \rangle = 0$ . From Eq. (4.25) it follows that  $\text{Im}\langle \Psi_0^{+z} | \Psi_0^{-z} \rangle = 0$ . Finally taking into account Eq. (4.24) together with Eq. (4.20) we arrive at:

$$\langle \Psi_0^{+z} | \Psi_1^{-z} \rangle = \langle \Psi_1^{+z} | \Psi_0^{-z} \rangle = 0. \quad (4.26)$$

As a result without losing generality we can parameterize the states as:

$$|\Psi_0^{+z}\rangle = [1, 0, 0, 0] \quad (4.27)$$

$$|\Psi_1^{+z}\rangle = [0, 1, 0, 0] \quad (4.28)$$

$$|\Psi_0^{-z}\rangle = [\cos\alpha, 0, \sin\alpha, 0] \quad (4.29)$$

$$|\Psi_1^{-z}\rangle = [0, \cos\beta, 0, \sin\beta] \quad (4.30)$$





Moreover, if the quality of the attack is to be the same in the  $x$  basis as in the  $z$  basis states  $|\Psi_0^{\pm x}\rangle, |\Psi_1^{\pm x}\rangle$  need to be normalized:

$$\langle \Psi_0^{+x} | \Psi_0^{+x} \rangle = \langle \Psi_1^{+x} | \Psi_1^{+x} \rangle = 1 \quad (4.31)$$

$$\langle \Psi_0^{-x} | \Psi_0^{-x} \rangle = \langle \Psi_1^{-x} | \Psi_1^{-x} \rangle = 1 \quad (4.32)$$

$$(4.33)$$

This implies

$$1 + \cos \alpha + \frac{D}{F}(\cos \beta + 1) = 2 \quad (4.34)$$

And finally recalling that  $D = 1 - F$ , the fidelity  $F$  reads:

$$F = \frac{1 + \cos \beta}{2 + \cos \beta - \cos \alpha} \quad (4.35)$$

The attack is thus parameterized with two real parameters  $\alpha, \beta$ .

Let us assume that the  $z$  basis was used in a given run of the protocol. The goal of the eavesdropper is to infer the value of the bit. Notice that the space spanned by  $|\Psi_1^{+z}\rangle, |\Psi_1^{-z}\rangle$  is orthogonal to the one spanned by  $|\Psi_0^{+z}\rangle, |\Psi_0^{-z}\rangle$ . This means that by projecting on one of these subspaces  $E$  knows for sure whether he inflicted an error in the transmission or not. After projecting on the subspaces  $E$  has to distinguish between  $|\Psi_1^{+z}\rangle$  and  $|\Psi_1^{-z}\rangle$  or between  $|\Psi_0^{+z}\rangle$  and  $|\Psi_0^{-z}\rangle$ . For this purpose  $E$  uses the optimal discrimination of two non-orthogonal states strategy from Sec.2.1.1. The information she gains can therefore be written in the form

$$I(A : E) = F \left( 1 - h \left[ \frac{1 + \sin \alpha}{2} \right] \right) + D \left( 1 - h \left[ \frac{1 + \sin \beta}{2} \right] \right). \quad (4.36)$$

Fixing  $F$  (i.e. fixing the QBER) the above information is maximal for  $\alpha = \beta$ . Finally we get that depending on  $\alpha$  the QBER reads:

$$QBER = \frac{1 - \cos \alpha}{2}, \quad (4.37)$$

while the mutual informations:

$$I(A : B) = 1 - h[QBER] \quad (4.38)$$

$$I(A : E) = I(B : E) = 1 - h \left[ \frac{1 + \sin \alpha}{2} \right] \quad (4.39)$$





Notice that  $I(A : E) = I(B : E)$  this follows simply from the fact that  $E$  always knows whether he inflicted an error in communication from  $A$  to  $B$ . Looking for the QBER for which  $I(A : B) = I(A : E)$  we obtain the QBER threshold:

$$QBER_{th} = \frac{1 - 1/\sqrt{2}}{2} \approx 14.6\%. \quad (4.40)$$

If the QBER is above this threshold the BB84 protocol is not safe. If it is below, the protocol is safe against individual attacks.

## 4.5 Other QKD protocols

### 4.5.1 Six state protocol

A natural generalization of BB84 is to use also circularly polarized basis  $|\odot\rangle, |\ominus\rangle$ .  $A$  send with probability  $1/6$  one of six states  $|\leftrightarrow\rangle, |\updownarrow\rangle, |\nearrow\rangle, |\swarrow\rangle, |\odot\rangle, |\ominus\rangle$ , while  $B$  measures randomly in one of three basis. On average  $2/3$  of the bits will be discarded in the sifting phase. This protocol seems more secure than BB84, since it uses 3 different basis instead of 2, and hence make it harder for an eavesdropper it gain information under a given QBER. Nevertheless the fact that only  $1/3$  of bits is kept make this protocol less useful when one what to achieve higher transmission rates.

#### Intercept and resend attack on on 6S protocol

If the eavesdropper attack  $r$  fraction of the qubits, measures them randomly in one of three basis and resends the measured state the QBER he inflicts reads:  $QBER = r/3$ . The mutual informations read:

$$I(A : B) = 1 - h[QBER] \quad (4.41)$$

$$I(A : E) = r/3 \quad (4.42)$$

The QBER threshold corresponding to the situation when  $I(A : B) = I(A : E)$  reads:  $QBER = 22.7\%$ .

#### Individual attack on 6S using the optimal universal asymmetric cloning

The 6S protocol is easier to investigate in terms of security thanks it its higher symmetry, than that of BB84. Apart from “measurement” attacks



in which the eavesdropper measure qubits on-the-fly, a more sophisticated attack appear to be more powerful. Instead of measuring the qubit the eavesdropper can perform optimal  $1 \rightarrow 2$  cloning operation send one copy to  $B$  and keep one for himself. After basis reconciliation process has taken place the eavesdropper can measure his clone in order to gain information on transmitted bits.

Let the eavesdropper perform the optimal universal asymmetric cloning described in Sec. 3.2.1, parameterized by two real parameters  $a, b$  subject to normalization constraint  $a^2 + b^2 + ab = 1$  ( $b = (-a + \sqrt{4 - 3a^2})/2$ ). He keeps the first clone for himself and send the second one to  $B$ . Given that  $A$  had sent the state  $|\psi\rangle$ ,  $B$  and  $E$  obtain the following reduced states:

$$\rho_E = \left[ \frac{(a+b)^2}{2} + \frac{a^2}{2} \right] |\psi\rangle\langle\psi| + \frac{b^2}{2} |\psi_\perp\rangle\langle\psi_\perp| \quad (4.43)$$

$$\rho_B = \left[ \frac{(a+b)^2}{2} + \frac{b^2}{2} \right] |\psi\rangle\langle\psi| + \frac{a^2}{2} |\psi_\perp\rangle\langle\psi_\perp| \quad (4.44)$$

This implies that the QBER equals  $a^2/2$ . The eavesdropper wait until basis are announced and measures his copy in the correct basis. His probability of error equals  $b^2/2$ . The mutual informations  $I(A : B)$ ,  $I(A : E)$  read:

$$I(A : E) = 1 - h[a^2/2] \quad (4.45)$$

$$I(A : B) = 1 - h[b^2/2] \quad (4.46)$$

where  $h[x] = -x \log_2 x$ . The QBER threshold corresponds to the symmetric case  $a = b = 1/\sqrt{3}$ , which yields  $QBER = 1/6 = 16.7\%$ , and  $I(A : B) = I(A : E) = 0.35$ . One can see that this attack is much more powerful than prepare and resend strategy.

There is a subtle issue to be mentioned. In what was said above we have concentrated only on  $I(A : B)$ ,  $I(A : E)$  quantities, ignoring completely  $I(B : E)$ , which according to the Csiszar-Korner theorem plays equivalent role to  $I(A : E)$ . As a result of the optimal asymmetric cloning attack the two clones are in the state:

$$\rho_{EB} = \frac{1}{2} [(a+b)^2 |\psi, \psi\rangle\langle\psi, \psi| + (a|\psi, \psi_\perp\rangle + b|\psi_\perp, \psi\rangle)(a\langle\psi, \psi_\perp| + b\langle\psi_\perp, \psi|)] \quad (4.47)$$

After measuring in the correct basis (i.e.,  $|\psi\rangle, |\psi_\perp\rangle$ ) the mutual information  $I(B : E)$  reads:

$$I(B : E) = 1 - h[(a+b)^2/2] \quad (4.48)$$



One can check that this information is always less than  $I(A : E)$ , hence strictly speaking the attack is not that powerful as it might have seemed. Nevertheless it is possible to modify the cloning procedure in such a way to make it “symmetric” (do not confuse with symmetric cloning) with respect to  $I(A : E)$  and  $I(B : E)$ , without compromising clones quality.

#### 4.5.2 Optimal individual attack on 6S protocol

We can repeat the derivation of the optimal individual attack on BB84 protocol from Sec. 4.4.2. The only difference is the higher symmetry of the problem which requires that we need to consider also the attack on  $|\pm y\rangle$  states. This additional constraint leads to setting the parameter  $\beta = \pi/2$ . Hence the QBER as a function of  $\alpha$  reads:

$$QBER = \frac{1 - \cos \alpha}{2 - \cos \alpha} \quad (4.49)$$

And the mutual informations:

$$I(A : B) = 1 - h[QBER] \quad (4.50)$$

$$I(A : E) = I(B : E) = 1 - \frac{1}{2 - \cos \alpha} h \left[ \frac{1 + \sin \alpha}{2} \right] \quad (4.51)$$

Notice that because  $\beta = \pi/2$  then if  $E$  learned that he inflicted a mistake in the transmission he knows for sure what is the bit value ( $|\Psi_1^{+z}\rangle$  is orthogonal to  $|\Psi_1^{-z}\rangle$ ). Otherwise he performs the optimal two state discrimination. The QBER threshold corresponding to  $\alpha = 0.618686$ , equals

$$QBER_{th} = 15.6\% \quad (4.52)$$

Obviously it is larger than the one for BB84, since the 6 state protocol is more difficult to eavesdrop due to more states present.

#### Relation to the optimal asymmetric cloning

Notice that the result seems to indicate that the previously considered optimal universal symmetric cloning attack is not optimal, as it yielded a higher QBER threshold. The reason for this is that we chose the symmetric cloning as the one where we expect the mutual information  $I(A : E)$  and  $I(A : B)$  to be equal. However, we only equalized probabilities of errors i.e. clones





qualities. From the present analysis we see that comparing error probabilities and informations is not equivalent, since there are two cases, either error was inflicted in  $A$  to  $B$  transmission and  $E$  knows the bit perfectly, or  $E$  has to perform discrimination.

Let us take the optimal asymmetric universal cloning machine parameterized with  $a, b$  ( $b = (-a + \sqrt{4 - 3a^2})/2$ ). Using such a cloning we can obtain the following mutual informations

$$I(A : E) = \frac{b^2}{2} + \left(1 - \frac{b^2}{2}\right) \left(1 - h\left[\frac{1 - a^2/2 - b^2/2}{1 - b^2/2}\right]\right) \quad (4.53)$$

$$I(A : B) = 1 - h[b^2/2] \quad (4.54)$$

where  $\frac{b^2}{2}$  is the probability of  $E$  inflicting an error in  $A$  to  $B$  transmission (QBER), in which case  $E$  learns the bit perfectly otherwise with probability  $(1 - b^2/2)$  he learns the correct bit only with probability  $\frac{1 - a^2/2 - b^2/2}{1 - b^2/2}$  — on average the success probability of guessing the value of the bit is  $1 - a^2/2$  which is in agreement with the cloning fidelity. Looking for the QBER threshold when  $I(A : B) = I(A : E)$  we get:  $a = 0.595275$ ,  $b = 0.559238$ . Notice that this is asymmetric cloning — we need to give  $B$  a copy of a bit higher quality, in order that out informations are equal. The corresponding QBER

$$QBER_{th} = 1 - \frac{b^2}{2} = 15.6\%, \quad (4.55)$$

Which proves that using the optimal universal cloning in a proper way is equivalent to the optimal eavesdropping.

### 4.5.3 B92

A natural question arises, if two nonorthogonal states cannot be perfectly distinguished, then maybe one can construct a QKD protocol using only two states instead of four as used in BB84. Amazingly this is indeed possible.  $A$  sends either  $|\leftrightarrow\rangle$ , or  $|\swarrow\rangle$ .  $B$  measures either in  $|\leftrightarrow\rangle, |\updown\rangle$  or in  $|\swarrow\rangle, |\searrow\rangle$  basis. Unlike in BB84 he does not communicate the basis he used, but rather informs  $A$  about the cases in which he measured  $|\updown\rangle$  or  $|\searrow\rangle$  (without specifying which of them). This is an information that tells  $A$  that in this run  $B$  had a basis incompatible with the one she used. Hence if she denotes by 0 and 1 the cases when she sends  $|\leftrightarrow\rangle$  and  $|\swarrow\rangle$  respectively, and  $B$  denotes by 0 and 1 the cases when he used basis  $|\swarrow\rangle, |\searrow\rangle$  and  $|\leftrightarrow\rangle, |\updown\rangle$ , they will





have perfectly correlated bits at this positions. Notice also that there was no information revealed to an eavesdropper when  $B$  informed  $A$  about positions at which he measured  $|\uparrow\rangle$  or  $|\searrow\rangle$ . Moreover non perfect distinguishability of nonorthogonal states forces  $E$  to induce errors whenever she wants to learn something and thus makes the protocol secure.