NORTHWESTERN UNIVERSITY

Quantum Detection and Coding with Applications to Quantum Cryptography

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Electrical Engineering and Computer Science

By

Ranjith Nair

EVANSTON, ILLINOIS

December 2006

ABSTRACT

Quantum Detection and Coding with Applications to Quantum Cryptography

Ranjith Nair

Analytical lower and upper bounds for the average error probability in M-ary quantum detection are derived. The upper bound is valid when the state ensemble, which can consist of pure or mixed states, satisfies a certain linear independence condition. The lower bound is generally valid and also has a classical interpretation. The quantum direct encryption protocol called $\alpha\eta$ is described, focusing on its equivalence under individual identical measurements by the eavesdropper to a classical *random* cipher. A new characterization of classical random ciphers is given that focuses on their potential security against known-plaintext attacks. The concept of a quantum random cipher is defined. The quantum random cipher characteristics of $\alpha\eta$ against phase and heterodyne measurements are elucidated. The derived lower and upper bounds on error probability are applied to the problem of security of $\alpha\eta$ under known-plaintext and ciphertext-only joint attacks. The system is shown to be insecure against known-plaintext attacks for sufficiently large known-plaintext regardless of the values of all system parameters. The eavesdropper's error probability is upper and lower bounded by expressions involving respectively the

minimum and the maximum distance of the code generated by the underlying linear feedback shift register. For ciphertext-only attack, it is seen that the upper bound is not applicable to $\alpha\eta$, leaving open the questions of security of $\alpha\eta$ and also the related key generation protocol $\alpha\eta$-KG.

Table of Contents

List of Figures

CHAPTER 1

Introduction and Background

There are two threads of ideas in the work of this thesis: The *first* thread relates to standard private-key cryptography - specifically, the cryptographic objective of data encryption. While standard stream ciphers can provide the maximum possible information-theoretic security against ciphertext-only attacks, their security against known-plaintext attacks is assumed in practice (but not proved) on the grounds of complexity alone. In fact, we can prove that any nonrandom, i.e., standard cipher satisfying a nondegeneracy condition is broken at a certain length of plaintext under known-plaintext attack, and therefore, cannot have information-theoretic security. On the other hand, this result does not hold for *random* ciphers, which are ciphers in which the sender adds an additional randomness into the ciphertext that does not prevent the receiver from decrypting it but can confuse the attacker. This potential advantage of random ciphers has been qualitatively understood for some time but has not been systematically studied. We make some progress in this direction by defining a characteristic of random ciphers that is related to their security under known-plaintext attack.

Our study of random ciphers was motivated by the data encryption protocol called $\alpha\eta$ developed by Yuen [**1**]. This protocol uses coherent states of laser light as signaling states that, under the KCQ principle described in [**1**], yield noisier observations to an eavesdropper making any individual identical measurement on each signal in a sequence of signals. It was realized that, once the eavesdropper's measurement is fixed, this excess noise essentially makes the system a classical random cipher to her. Thus, the possibility of information-theoretic security of some level is opened up, as in the case of classical random ciphers mentioned above. In this quantum situation, however, the eavesdropper's capability is enhanced by allowing her to make *joint* measurements on the entire sequence of signals rather than a separate measurement on each signal. With such capability, analysis of the eavesdropper's error probability involves the use of *quantum detection theory*, i.e. M-ary quantum hypothesis testing, which is the *second* thread in this work. As we will see, the exact calculation, even numerically, of the M-ary optimum quantum measurement, and hence, the optimum error probability, is a difficult task. An explicit solution is known only for the case of $M = 2$. Therefore, our goal was to obtain upper and lower bounds for the M-ary error probability directly, without going through the calculation of the optimum quantum measurement operators. Indeed, in trying to prove security or insecurity of a system, we are more concerned with the eavesdropper's optimal performance and not with the explicit measurement which achieves it. At the same time, however, it is important that the bounds derived be *rigorous* in order to apply to a real security *proof*. The reason is that, unlike in communication system design, no amount of testing or simulation of attacks on the system (unless one simulates the optimal attack, which is unknown in our case) can establish its security when a real eavesdropper is present, since she is assumed to be able to launch the optimal attack.

Our progress in addressing these problems is described in this thesis, whose organization is as follows: In this section, we devote a subsection to the necessary background in private-key cryptography, and one to the language of quantum states and measurements.

In the latter subsection, we give special attention to the M-ary quantum detection problem that is central throughout the thesis. In Chapter 2, we review the existing results on M-ary quantum detection and develop the novel upper and lower bounds for this problem. In Chapter 3, we first give our novel characterization of classical random ciphers and extend it to the quantum case. We then describe the $\alpha\eta$ direct encryption protocol and present its random cipher characteristics. In Chapter 4, we apply the bounds of Chapter 2 to $\alpha\eta$ security under both ciphertext-only and known-plaintext attacks and suggest future directions of work.

1.1. Symmetric-Key Cryptography

1.1.1. Basics

We review the basics of symmetric-key data encryption. Further details can be found in, e.g., [**2, 3**]. Throughout the paper, random variables will be denoted by *upper-case* letters such as K, X_1 etc. It is sometimes necessary to consider explicitly sequences of random variables $(X_1, X_2, \ldots, X_n)$. We will denote such *vector* random variables by a *boldface* upper-case letter $\mathbf{X}_n$ and, whenever necessary, indicate the length of the vector (n in this case) as a subscript. Confusion with the n-th component X_n of $\mathbf{X}_n$ should not arise as the latter is a boldface vector. Particular values taken by these random variables will be denoted by similar *lower-case* alphabets. Thus, particular values taken by the key random variable K are denoted by k, k' etc. Similarly, a particular value of $\mathbf{X}_n$ can be denoted $\mathbf{x}_n$. The plaintext alphabet will be denoted $\mathcal{X}$, the set of possible key values $\mathcal{K}$ and the ciphertext alphabet $\mathcal{Y}$. Thus, for example, the sequences $\mathbf{x}_n \in \mathcal{X}^n$. In most nonrandom ciphers, $\mathcal{X}$ is simply the set $\{0, 1\}$ and $\mathcal{Y} = \mathcal{X}$.

With the above notations, the n-symbol long *plaintext* (i.e., the message sequence that needs to be encrypted) is denoted by the random vector $\mathbf{X}_n$, the *ciphertext* (i.e., the output of the encryption mechanism) is denoted by $\mathbf{Y}_n$ and the secret key used for encryption is denoted by K. In this paper, we will often call the legitimate sender of the message 'Alice', the legitimate receiver 'Bob', and the attacker (or eavesdropper) 'Eve'. Note that although the secret key is typically a sequence of bits, we do not use vector notation for it since the bits constituting the key will not need to be singled out separately in our considerations in this paper. In standard cryptography, one usually deals with *nonrandom ciphers.* These are ciphers for which the ciphertext is a function of only the plaintext and key. In other words, there is an encyption function $E_k(\cdot)$ such that:

$$\mathbf{y}_n = E_k(\mathbf{x}_n). \tag{1.1}$$

There is a corresponding decryption function $D_k(\cdot)$ such that:

$$\mathbf{x}_n = D_k(\mathbf{y}_n). \tag{1.2}$$

In such a case, the X_i and $Y_i, i = 1, \ldots, n$ are usually taken to be from the same alphabet.

In contrast, a *random cipher* makes use of an additional random variable R called the *private randomizer* [**2**], generated by Alice while encrypting the plaintext and known only to her, if at all. Thus the ciphertext is determined as follows:

$$\mathbf{y}_n = E_k(\mathbf{x}_n, r). \tag{1.3}$$

Because of the additional randomness in the ciphertext, it typically happens that the ciphertext alphabet $\mathcal{Y}$ needs to be larger than the plaintext alphabet $\mathcal{X}$ (or else, $\mathbf{Y}$ is a longer sequence than $\mathbf{X}$, as in homophonic substitution). It may even be a continuous infinite alphabet, e.g. an analog voltage value. However, we still require, as in [**2**], that Bob be able to decrypt with just the ciphertext and key (i.e., without knowing R), so that there exists a function $D_k(\cdot)$ such that Eq.(1.2) holds. We note that random ciphers are called ‘privately randomized ciphers’ in Ref. [**2**] – we will however use the shorter term ‘random cipher’ (Note that ‘random cipher’ is used in a completely different sense by Shannon [**4**]).

We note that the presence or absence of the private randomizer R may be indicated using the conditional Shannon entropy (We assume a basic familiarity with Shannon entropy and conditional entropy. See any information theory textbook, e.g., [**5**].). For nonrandom ciphers, we have from Eq.(1.1) that

$$H(\mathbf{Y}_n|K\mathbf{X}_n) = 0. \tag{1.4}$$

On the other hand, a *random cipher* satisfies

$$H(\mathbf{Y}_n|K\mathbf{X}_n) \neq 0, \tag{1.5}$$

due to the randomness supplied by the private randomizer R. The decryption condition Eqs.(1.2) for both random and nonrandom ciphers has the entropic characterization:

$$H(\mathbf{X}_n|K\mathbf{Y}_n) = 0. \tag{1.6}$$

Note that this characterization of a random cipher is problematic when the ciphertext alphabet is continuous, as could be the case with $\alpha\eta$, because then the Shannon entropy is not defined. It may be argued that the finite precision of measurement forces the ciphertext alphabet to be discrete. Indeed, in Sec. 3.1, we define a parameter Λ that characterizes the "degree of randomness" of a random cipher. In any case, the definition makes sense, similar to Eq. (1.5), only when the ciphertext alphabet is finite, or at most discrete.

In the cryptography literature, the characterization of a general random cipher is limited to that given by Eqs. (1.3) and (1.5). See, e.g., [**2**]. In the next section, we will see that the purposes of cryptographic security suggest a sharper quantitative definition of a random cipher involving a pertinent security parameter Γ. This new definition, unlike (1.5), will be meaningful irrespective of whether the ciphertext alphabet is discrete or continuous. Before we discuss the above new definition of random ciphers, we conclude this section with some important cryptographic terminology.

By *standard cryptography*, we shall mean that Eve and Bob both observe the same ciphertext random variable, i.e., $\mathbf{Y}_n^{\mathrm{E}} = \mathbf{Y}_n^{\mathrm{B}} = \mathbf{Y}_n$. Thus, standard cryptography includes usual mathematical private-key (and also public-key) cryptography but excludes quantum cryptography and classical-noise cryptography [**6**]. For a standard cipher, random or nonrandom, one can readily prove from the above definitions the following result known as the *Shannon limit* [**2, 4**]:

$$H(\mathbf{X}_n|\mathbf{Y}_n) \leq H(K). \tag{1.7}$$

This result may be thought of as saying that no matter how long the plaintext sequence is, the attacker's uncertainty on it *given the ciphertext* cannot be greater than that of the key.

By *information-theoretic security* (or *IT security*) on the data, we mean that Eve cannot, even with unlimited computational power, pin down uniquely the plaintext from the ciphertext, i.e.,

$$H(\mathbf{X}_n|\mathbf{Y}_n) \neq 0. \tag{1.8}$$

The level of such security may be quantified by $H(\mathbf{X}_n|\mathbf{Y}_n)$. Shannon has defined *perfect security* [**4**] to mean that the plaintext is statistically independent of the ciphertext, i.e.,

$$H(\mathbf{X}_n|\mathbf{Y}_n) = H(\mathbf{X}_n). \tag{1.9}$$

With the advent of quantum cryptography, the term 'unconditional security' has come to be used, unfortunately in many possible senses. By *unconditional security*, we shall mean near-perfect information-theoretic security against all attacks consistent with the known laws of quantum physics.

Incidentally, note that the Shannon limit Eq. (1.7) immediately shows that perfect security can be attained only if $H(\mathbf{X}_n) \leq H(K)$, so that, in general, the key needs to be as long as the plaintext.

1.1.2. Classes of Attacks on Ciphers

In this section, we summarize some relevant terminology and general results on the key security of both random and nonrandom ciphers. We first present an overview of the various possible cryptographic attacks possible on a cipher and some early results on the subject. We also present our result on the security of a nonrandom cipher under known-plaintext attacks. In the process, we define the important term 'unicity distance' coined by Shannon and broaden it to include the notion of 'unicity distance under known-plaintext attack' for both random and nonrandom ciphers. We also define the important concept of 'nondegeneracy' for both random and nonrandom ciphers that is needed to make the concept of unicity distance meaningful. Finally, we discuss how random ciphers may enhance security against known-plaintext attacks.

The following terminology in regard to cryptographic attacks has been used in this paper, as in our paper [**7**]. This terminology is not standard, however. In the cryptography literature, what we call statistical attacks are sometimes referred to as ciphertext-only attacks (See, e.g., [**3**], Ch. 2) but are also often lumped together with known-plaintext attacks.

By a *ciphertext-only attack (CTA)*, we refer to the case where the probability distribution $p(\mathbf{X}_n)$ is completely uniform, i.e., $p(\mathbf{X}_n) = 2^{-n}$ to Eve, so that her attack cannot exploit input frequencies or correlations and must be based only on the ciphertext in her possession. By a *statistical attack (STA)*, we refer to the case where the probability distribution $p(\mathbf{X}_n)$ is nonuniform, so that Eve may in principle exploit input frequencies or correlations to launch a better attack. Such an attack is typical when the plaintext is in a language such as English. It is also the attack that obtains when the $\{X_i\}$ are independent and identically distributed (i.i.d.) but each $p(X_i)$ is nonuniform. By a *known-plaintext attack (KPA)* we mean the case where Eve knows *exactly* some length m of plaintext

$\mathbf{x}_m$. Finally, by a *chosen-plaintext attack (CPA)*, we mean a KPA where the data $\mathbf{x}_m$ is chosen by Eve.

In standard cryptography, one typically does not worry about ciphertext-only attack on nonrandom ciphers. The reason is that, under CTA, Eq. (1.7) is satisfied with equality for large n for the designed key length $|K| = H(K)$ under a certain 'nondegeneracy' condition [**8**] that is readily satisfied. Thus, in practice, the data security is assumed to be sufficient if $H(K)$ is chosen large enough by adjusting the key length. However, it follows from (1.7) that no meaningful lower bound on $H(\mathbf{X}_n|\mathbf{Y}_n)$ exists for $n \gg |K|$. A new fundamental treatment of data security in symmetric-key ciphers has to be developed separately. Under CTA, it is also the case for nonrandom nondegenerate ciphers that [**8**]

$$H(K|\mathbf{Y}_n) = H(K), \tag{1.10}$$

i.e., the key is *statistically independent* of the ciphertext. Thus, no attack better than pure guessing can be launched on the key.

The above two results do not hold for statistical and known-plaintext attacks. Eve can indeed launch an attack on the key and use her resulting information on the key to get at future and past data. In fact, it is such attacks that are the focus of concern for standard ciphers such as the Advanced Encryption Standard (AES). For STAs, Shannon [**4**] characterized the security by the so-called unicity distance. The *unicity distance* n_0 of a cipher is the smallest input data length for which $H(K|\mathbf{Y}_{n_0}) = 0$. In other words, if a plaintext sequence of length n_0 is encrypted by the cipher, the ciphertext contains enough information to fix the key (and hence, the plaintext) uniquely – the cipher has no information-theoretic security. For nonrandom ciphers defined by Eq. (1.4), Shannon, in [**4**], derived in terms of the data entropy an estimate on n_0 that is independent of the cipher. This estimate is actually *not* a rigorous bound. Indeed, it can be shown that one of the inequalities used in the derivation goes in the wrong direction. Even so, the estimate works well empirically for English language plaintexts, for which $n_0 \sim 25$ characters are found to be sufficient to break many ciphers.

We now consider, in some detail, security against known-plaintext attacks. Here, a natural quantity to consider is $H(K|\mathbf{X}_n\mathbf{Y}_n)$, since it provides a measure of key uncertainty when both plaintext and ciphertext are known to the attacker. Before we state the main result, we define the notion of nondegeneracy distance. The reader can readily convince himself that a finite unicity distance exists only if, for some n, there is no *redundant key use* in the cryptosystem, i.e., no plaintext sequence $\mathbf{x}_n$ is mapped to the same ciphertext $\mathbf{y}_n$ by more than one key value. With redundant key use, one cannot pin down the key but it seems that this may not enhance the system security either, and so is merely wasteful. In any case, we call a cipher *nondegenerate* if it has no redundant key use for some finite n or for $n \to \infty$. Under the condition

$$\lim_{n\to\infty} H(\mathbf{Y}_n|\mathbf{X}_n) = H(K), \tag{1.11}$$

which is similar but not identical to the definition of a 'nondegenerate' cipher given in [**8**], one may show that, when Eq. (1.4) also holds, one has

$$\lim_{n\to\infty} H(K|\mathbf{X}_n, \mathbf{Y}_n) = 0, \tag{1.12}$$

so that the system is asymptotically broken under a known-plaintext attack. More generally, for a nonrandom cipher, we define a *nondegeneracy distance* n_d to be the smallest

n such that

$$H(\mathbf{Y}_n|\mathbf{X}_n) = H(K) \tag{1.13}$$

holds, with $n_d = \infty$ if (1.11) holds and there is no finite n satisfying (1.13). Thus, a nonrandom cipher is nondegenerate in our sense if it has a nondegeneracy distance, finite or infinite. In general, of course, the cipher may be *degenerate*, i.e., it has no nondegeneracy distance. We can readily show (see Appendix A of [**9**]) that, under known-plaintext attack, a nonrandom nondegenerate cipher is broken at data length $n = n_d$, in the sense that

$$H(K|\mathbf{X}_{n_d}\mathbf{Y}_{n_d}) = 0. \tag{1.14}$$

More generally, for both random and nonrandom ciphers, we define the *unicity distance under known-plaintext attacks*, denoted by n_1, to be the smallest integer such that

$$H(K|\mathbf{X}_{n_1}\mathbf{Y}_{n_1}) = 0. \tag{1.15}$$

If no such integer exists, the unicity distance under KPA is taken to be infinite if

$$\lim_{n\to\infty} H(K|\mathbf{X}_n\mathbf{Y}_n) = 0. \tag{1.16}$$

Thus, n_1 is the minimum length of data needed to break the cipher for *any* possible known-plaintext $\mathbf{X}_n$. For a nonrandom cipher, it is equal to the nondegeneracy distance.

Many ciphers including the one-time pad and LFSRs (linear feedback shift registers [**3**]) have finite n_d. Similar to the case of n_d for nonrandom ciphers, n_1 for a random cipher may not always exist. For our definition of n_1 to make sense for random ciphers, we will impose a 'nondegeneracy' restriction on random ciphers: A *random cipher* is said to be *nondegenerate* if and only if *each* nonrandom cipher resulting from an assignment $\mathbf{R} = \mathbf{r}$ of the randomizer is nondegenerate. Then we say it has *information-theoretic security against known-plaintext attacks* if

$$\inf_n H(K|\mathbf{X}_n, \mathbf{Y}_n) \neq 0, \tag{1.17}$$

i.e., if $H(K|\mathbf{X}_n, \mathbf{Y}_n)$ cannot be made arbitrarily small whatever n is. In other words, n_1 does not exist. The actual level of the information-theoretic security is quantified by the left side of (1.17). One major motivation to study random ciphers is the *possibility* that they possess such information-theoretic security. Some discussion on this point is also available in Appendix A of [**9**].

Even in the absence of information-theoretic security, nondegenerate random ciphers can be expected (we will discuss this further in Sec. 3.1) to have larger unicity distance n_1 under KPA compared to the case where the randomization is turned off. This would, as assumed in cryptography practice, increase the complexity of attacking the key significantly. If Eq. (1.14) holds when $\mathbf{X}_n$ is replaced by a specific $\mathbf{x}_n$, n defines the unicity distance corresponding to $\mathbf{x}_n$. The overall unicity distance under KPA may be defined by

$$\bar{n}_1 = \min_{H(K|\mathbf{X_n}=\mathbf{x_n},\mathbf{Y_n})=\mathbf{0}} n \text{ for some } \mathbf{x_n}. \tag{1.18}$$

The above result has not been given in the literature, perhaps because $H(K|\mathbf{X}_n\mathbf{Y}_n)$ has not been used previously to characterize known-plaintext attacks. Nevertheless, it is assumed to be true in cryptography practice that K would be pinned down for sufficiently long n in a nonrandom 'nondegenerate' cipher.

We now discuss the advantages that a random cipher provides as compared to nonrandom ciphers. For the case of STA on the key when the plaintext $\mathbf{X}_n$ has nonuniform but i.i.d. statistics, the so-called *homophonic substitution* method provides complete information-theoretic security, i.e. $H(K|\mathbf{Y}_n) = H(K)$ [**8**]. The original form of homophonic substitution involves assigning to each plaintext symbol a number of possible *sequences* of length l proportional to its a priori probability in such a way that all possible l-sequences are covered. Then, for every input symbol, if one of its assigned l-sequences is generated at random, the net effect is to generate l-sequences of plaintext with i.i.d. uniform statistics. These sequences may be passed through a non-degenerate cipher without revealing information on the key as per Eq. (1.10). To put it another way, a statistical attack has been converted to a ciphertext-only attack. A generalized homophonic substitution that allows each symbol to be coded into sequences of variable length is discussed in [**8**], for which it is shown that sometimes data compression instead of data expansion results.

Unfortunately, this reduction of a STA to a CTA does not work for known-plaintext attacks. However, we emphasize that there is *no result* on random ciphers analogous to Eq. (1.14) with n_d replaced by any definite n depending on the cipher, since under randomization, Eq. (1.4), and usually (1.13) also, does not hold for any n. We will say more on this in Sec. 3.1. In fact, the general problem of attacking a random cipher has received limited attention because they are *not used in practice* due to the associated reduction in effective bandwidth or data rate as is evident in homophonic substitution, due to the need for high speed random number generation, and also due to the uncertainty on the actual input statistics needed for, e.g., homophonic substitution randomization. Thus, the rigorous quantitative security of symmetric-key random ciphers against known-plaintext attacks is not known theoretically or empirically, although in principle random ciphers have actual and potential advantages just discussed.

1.2. Quantum Mechanics Review

This section presents a sketch of the fundamentals of quantum mechanics that are required for our future treatment. We will focus mostly on the abstract mathematical description of quantum mechanical systems. A quite comprehensive modern textbook that includes these topics is Nielsen and Chuang [**10**]. Another good reference for general quantum mechanics is [**11**]. Measurement theory in quantum mechanics is a vast topic - a good starting place may be the summary of [**12**]. A standard textbook of quantum optics is [**13**].

First, we comment on the Dirac notation [**10**] for vectors and linear functionals in a Hilbert space. A vector ψ in the space is written as a so-called *ket* $|\psi\rangle$, and the linear functional corresponding to taking the inner product with ψ is written as a so-called *bra* $\langle\psi|$. The inner product of two vectors ϕ and ψ then becomes in Dirac notation the bra(c)ket $\langle\phi|\psi\rangle$. A projection operator P_ψ onto the (normalized) vector ψ is written conveniently in Dirac notation as $P_\psi = |\psi\rangle\langle\psi|$. In quantum mechanics, the inner product is conventionally taken to be linear in the *second* argument and conjugate-linear in the first.

1.2.1. Quantum Mechanics Axioms

(1) Every physical system is associated with a Hilbert space $\mathcal{H}$, i.e. a complete normed space over the complex numbers with an inner product denoted $\langle\cdot|\cdot\rangle$. The system is completely specified by its *state vector* , denoted $|\psi\rangle$, which is a unit vector in the space, i.e. $\langle\psi|\psi\rangle= 1$.

(2) The time evolution of a closed quantum system is described by a unitary, i.e., inner product-preserving, linear operator, on $\mathcal{H}$ (via the Schrodinger equation):

$$|\psi(t)\rangle = U(t,0)|\psi(0)\rangle \tag{1.19}$$

(3) A *Measurement* $\mathcal{M}$ with a finite number M of outcomes $\{m\}_1^M$ on the system is described by a set of operators $\{M_m\}_1^M$ that satisfy

$$\sum_m M_m^\dagger M_m = I \tag{1.20}$$

where I is the identity operator on $\mathcal{H}$, and the $\dagger$ denotes the adjoint operation. The *probability* $p(m)$ of obtaining the outcome m when measurement $\mathcal{M}$ is performed on a system in the state $|\psi\rangle$ is

$$p(m) = \langle\psi|M_m^\dagger M_m|\psi\rangle = Tr[|\psi\rangle\langle\psi|M_m^\dagger M_m]. \tag{1.21}$$

Here Tr stands for operator trace. The state of the system after obtaining the outcome m is

$$|\psi_m\rangle = \frac{M_m|\psi\rangle}{\sqrt{\langle\psi|M_m^\dagger M_m|\psi\rangle}}. \tag{1.22}$$

Note that Eq. (1.20) ensures that the probabilities $p(m)$ sum to one and the denominator of Eq. (1.22) is simply a normalization.

(4) The state space of a composed quantum system with two components is the *tensor product* (or Kronecker product) $\mathcal{H} \otimes \mathcal{K}$ of the component state spaces $\mathcal{H}$ and $\mathcal{K}$. The extension to more than two component systems is obvious. In Dirac notation, the tensor product of $|\psi\rangle$ and $|\phi\rangle$ is denoted simply by $|\psi\rangle|\phi\rangle$ omitting the $\otimes$ sign. Recall that if $\{|e_i\rangle\}_1^m$ and $\{|f_j\rangle\}_1^n$ are bases for $\mathcal{H}$ and $\mathcal{K}$, then $\{|e_i\rangle|f_j\rangle\}_{(1,1)}^{(m,n)}$ is a basis for $\mathcal{H} \otimes \mathcal{K}$. Thus the dimension of the composite state space grows multiplicatively. This feature is characteristic of quantum mechanics. The operator space $\mathcal{L}(\mathcal{H} \otimes \mathcal{K}) = \mathcal{L}(\mathcal{H}) \otimes \mathcal{L}(\mathcal{K})$, and so the dimension multiplies here too. If $A_1, A_2, \cdots$ are operators on $\mathcal{H}_1, \mathcal{H}_2, \cdots$ respectively, the notation $A_1 A_2 \cdots$ is used conventionally to refer to the tensor product operator $A_1 \otimes A_2 \otimes \cdots$.

The above axiom (3) is more general than the special case of *von Neumann* (or *Projective*) Measurement that was first recognized historically. The projective measurements are those where $\{M_m\}$ are mutually orthogonal projection operators, i.e. $M_m^\dagger = M_m = M_m^2$, and $M_m M_{m'} = M_{m'} M_m = M_m \delta_{mm'}$. The physical *observables* like position, total energy etc. are associated with Hermitian operators $X = X^\dagger$. Every Hermitian operator has a (discrete or continuous) spectral decomposition:

$$X = \sum_x x P_x \tag{1.23}$$

where $\{x\}$ is the spectrum of X. The operators $\{P_x\}$ constitute a projective measurement and it is this measurement that is meant when one speaks in quantum mechanics of 'measuring an operator X'.

It can be in fact be shown that a general measurement of the form of axiom (3) can be realized on a system with state space $\mathcal{H}$ by adjoining an ancillary system with state space $\mathcal{K}$ of sufficiently large dimension and performing a von Neumann measurement on the composite system $\mathcal{H} \otimes \mathcal{K}$. This follows from Neumark's extension theorem (See [**11**] and [**14**]) and is in fact how the above definition of a general measurement suggests itself.

It is interesting that the general definition of measurement given above covers a large set of possible experimental procedures. It can be shown that two successive measurements $\{L_l\}$ and $\{M_m\}$ are equivalent (in terms of final state and probability distribution of outcomes) to a single measurement $\{M_m L_l\}$ ([**10**], p. 86). An experimenter in the lab can conceive many operational measurement procedures that seem to fall outside the purview of the measurement axiom. For example, given a quantum system, the experimenter may choose to make a measurement on one portion of the system, and depending on the outcome, make another measurement on the rest of the system and so on. This kind of procedure is referred to as an *Adaptive* measurement. However, it can be shown again that such procedures can also be reduced back to performing an equivalent measurement of the form of axiom (3) [**15**]. This fact is of importance to the sequel since we will be interested in using the measurement formalism to prove results that need to be valid under very general conditions. We also mention that in many situations of interest, the full description of a measurement provided by axiom (3) is not required and a less detailed description is sufficient. Such situations arise in quantum detection theory that is the subject of the Subsection 1.2.4.

1.2.2. Density Operators

We now discuss a useful generalization of the concept of state of a quantum system. Note that the quantum state $|\psi\rangle$ is equally well represented by the projection operator $\rho := |\psi\rangle\langle\psi|$. Indeed, the overall phase of $|\psi\rangle$ has no observable manifestation. Similarly, the probability distribution of measurement outcomes can be given in terms of ρ (the second equality in (1.21)). and the post-measurement state conditioned on outcome m is simply:

$$\rho_m = \frac{M_m \rho M_m^\dagger}{Tr[M_m \rho M_m^\dagger]}. \tag{1.24}$$

This observation allows us to compactly describe the system state when additional (classical) randomness is present. For example, our knowledge of the system may be limited to knowing that it is one of many pure states $|\psi_i\rangle$ with probability p_i. In this case, it is convenient to represent the system state by the *density operator* ρ:

$$\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|. \tag{1.25}$$

States whose density operator is not a one-dimensional projection are called *mixed* states. ρ in the above equation is said to be a *mixture* of the states $|\psi_i\rangle$. Those ρ that are of the form $|\psi\rangle\langle\psi|$ are referred to as *pure* states. The operator in (3.9)can be seen to have the following properties [**10**] which serve to define a general density operator:

$$\rho = \rho^{\dagger} \tag{1.26}$$

$$\rho \geq 0 \tag{1.27}$$

$$Tr[\rho] = 1. \tag{1.28}$$

Here the operator inequality $A \geq B$ means that $A - B$ is a positive (semi-definite) operator. In terms of the density operator formalism, measurement probabilites are given by

$$p(m) = Tr[M_m^{\dagger} M_m \rho] \tag{1.29}$$

and final states by (1.24). We will use the density operator formulation in the following except sometimes when dealing with pure states.

1.2.3. Quantum Optics

We quickly mention some facts from quantum optics that are needed in the sequel. The treatment is mainly to introduce standard notations and omits a lot of detail, which can be found in [**13**]. A single field mode has an infinite dimensional Hilbert space that is isomorphic to $\mathcal{L}^2(\mathbb{R})$. Thus, as in the case of a particle in one dimension, a pure state of the field can be given as a wave function $\psi(x)$ where x represents a quadrature component of the field. The quadrature components, denoted by operators a_1 and a_2, when suitably normalized, are analogous to the position and momentum operators and satisfy the commutation relation:

$$[a_1, a_2] = i/2 \tag{1.30}$$

It is useful to define the non-Hermitian operator a, called the *annihilation operator*, by

$$a = a_1 + ia_2. \tag{1.31}$$

The positive operator $N = a^\dagger a$ is called the photon *number operator* and has a discrete spectrum comprising the non-negative integers. Its eigenvectors span the Hilbert space and are called number states. The number state corresponding to the eigenvalue n is denoted $|n\rangle$, and so the spectral decomposition of N is:

$$N = \sum_{n=0}^{\infty} n|n\rangle\langle n|. \tag{1.32}$$

The eigenvalues of a are all the complex numbers $\alpha = \alpha_1 + i\alpha_2 \in \mathbb{C}$ with corresponding eigenvectors denoted $|\alpha\rangle$. These states are called *coherent states* and are of great importance. In the number state basis, the coherent state $|\alpha\rangle$ has the representation:

$$|\alpha\rangle = e^{-|\alpha|^2/2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \tag{1.33}$$

Note that a measurement of photon number N on the coherent state (1.33) yields a Poisson distribution of outcomes. The coherent states are usually represented as points on the two dimensional plane. However, the state space is not two-dimensional - Indeed the inner product $|\langle\alpha|\beta\rangle| \rightarrow 0$ as the Euclidean distance on the plane between $|\alpha\rangle$ and

$|\beta\rangle$ increases. The coherent states have great practical importance. The quantum state of a laser operating far above threshold is well-modeled by a coherent state [**13**]. Coherent states in a lossy channel, e.g., an optical fiber, are well-behaved and do not become quantum- correlated or *entangled* [**10**] with the environment. The product of the quantum fluctuations in the field components has the minimum possible value (as given by the Heisenberg inequality) for coherent states. As we mentioned, the coherent state is a good model for the states in optical communications, at least when no amplification is done. Even when amplifiers are used, the field state can be represented by a density operator that is a mixture of coherent states. Thus, the coherent state is fairly ubiquitous. Indeed, the 'α' in $\alpha\eta$ highlights the fact that the protocol works with coherent states in contrast to other quantum key generation protocols which ideally require exotic states like the photon number eigenstate $|1\rangle$ that are hard to produce.

1.2.4. Quantum detection theory: Introduction and the POM concept

We now study the modification of standard communication theory that arises when the states of a quantum system are used as signals. The associated detection theory is called Quantum Detection Theory. The aim in this section is to introduce the mathematical problem, mention its possible solution, and to highlight the features of the theory that makes key generation possible.

The situation is as follows: As in classical communication, the sender Alice wishes to communicate to the receiver Bob a random variable X taking values $x \in \mathcal{X}$, which we assume to be a finite set in our work. (If not, as in classical communication theory, we talk of an *estimation* problem rather than a detection problem). To do so, she *modulates* a quantum system as a function of the particular value x to generate a state with density operator $\rho_x \in \mathcal{H}$ with a priori probability p_x. This state is put onto a physical channel such as an optical fiber, and may suffer some noise and loss before it gets to Bob. In general, the action of the physical channel can be modeled as a map from the space of density operators to itself that satisfies some conditions. In the work to follow, however, we neglect channel noise, and as mentioned in the previous section, the coherent states that we use in our system transform very simply under loss and retain their coherent state character. Therefore, we shall not be concerned with the channel map. In either case, Bob's task at the receiver is to make a measurement on the states $\{\rho_x\}$ with the purpose of deciding which x was sent. The optimal way of doing this is the subject of quantum detection theory.

Before we formulate and give the solution of the general quantum detection problem, we remark that the general theory of quantum measurement that we gave earlier is not required in its entirety for the detection problem. Specifically, for a measurement $\mathcal{M}$, we are concerned only with the *probabilities* of the various outcomes m and not with the final states $M_m \rho M_m^\dagger$. This is because we are concerned only with the extraction of classical information (residing in the outcomes m of the measurement) from the system and not the final quantum state. Indeed, in optics, most measurements that we can do today, in the final analysis, correspond to the measurement of the photon number operator N of some or the other field mode using a photodiode, which destroys the post-measurement state in that mode.

Eq. (1.29) reveals that the set of operators $\{\Pi_m = M_m^\dagger M_m\}_1^M$ is sufficient to generate the probability distribution of outcomes for any state ρ. Such a set of operators can be seen to satisfy the properties:

$$\Pi_m \geq 0, \tag{1.34}$$

$$\sum_{m=0}^{M} \Pi_m = I. \tag{1.35}$$

and is called a *Positive Operator-Valued Measure* or POM and its elements are called POM elements. Also, any given POM $\{\Pi_m\}$ can be realized by a measurement $M_m = U\sqrt{\Pi_m}$ where $\sqrt{\cdot}$ denotes the positive square root of the positive operator Π_m and U is any unitary operator. Note that the earlier discussion of adaptive measurements etc. carries over to the POM case, so that any such measurement procedure has an equivalent POM associated with it. It is believed (which is highly plausible but not proved) that any conceivable measurement procedure in quantum mechanics is associated to a unique POM. In our discussion of quantum detection theory, therefore, we can confine our characterization of possible measurements to their associated POM's. POM's are discussed in more detail in [**10**]. The standard reference for quantum detection and estimation theory is the book by Helstrom [**14**].

1.2.5. M-ary Quantum Detection: Setup of the Mathematical Problem and its Solution

The general M-ary quantum detection (or hypothesis testing) problem is as follows. Given a set of M density operators $\{\rho_i\}_1^M$ with respective a priori probabilities $\{p_i\}_1^M$ and a cost function C_{ij} on $J \times J$, $J = \{1, \ldots, M\}$. The task is to find a POM $\{\Pi_j\}_1^M$ that maximizes (or minimizes) the average cost:

$$\overline{C} = \sum_{i,j=1}^{M} p_i C_{ij} Tr[\Pi_j \rho_i]. \tag{1.36}$$

Note that $Tr[\Pi_j \rho_i]$ is the conditional probability of obtaining outcome j given state ρ_i. We shall only be interested in the cost function $C_{ij} = 1 - \delta_{ij}$, in which case the average cost is the average probability of error in distinguishing the states $\{\rho_i\}_1^M$. A necessary and sufficient condition for operators $\{\hat{\Pi}_j\}_1^M$ to minimize the cost (1.36) was given by Yuen, Kennedy and Lax [**16**]. In the form given in [**17**] the theorem is:

YKL Theorem: A necessary and sufficient condition for the set $\{\hat{\Pi}_j\}_1^M$ to minimize the cost (1.36) is that there exists a Hermitian operator X satisfying

$$X - p_1 \rho_i \geq 0, \quad \forall i \in J \tag{1.37}$$

$$(X - p_1 \rho_i)\hat{\Pi}_i = 0 \ \ \forall i \in J. \tag{1.38}$$

The matrix X is the solution of the dual problem $\min_{X \in \mathcal{B}} Tr[X]$ subject to the conditions (23). Here $\mathcal{B}$ is the set of Hermitian operators on the space $\mathcal{H}$.

The YKL theorem is useful for checking if a proposed POM is optimal. Actually obtaining the optimal POM is a quite difficult task. In the case where the range of the operators $\{\rho_i\}$ taken together is finite-dimensional of dimension n, it can be solved numerically by methods of semidefinite programming [**17**]. The space of Hermitian operators on a n-dimensional complex vector space is a n^2-dimensional real vector space. Since there are M POM elements, we have a total dimension of Mn^2 with the $M + 1$ constraints

(1.34). The dimension of the space can be reduced in some cases where the states to be distinguished have some symmetries.

It may be helpful to explain qualitatively the difference between classical and quantum detection theory. Classical detection theory is contained in quantum detection theory as the special case when the density operators $\{\rho_x\}$ under consideration all commute, i.e. have a common set of eigenvectors. In this case, the optimal decision procedure is simply to make the von Neumann measurement defined by these eigenvectors and map each possible outcome to the most likely hypothesis. These two steps have as analogs in classical decision theory the making of an observation of the signal(i.e. obtaining a point in the observation space which in digital communication is a subset of $\mathbb{R}^n$) and partitioning the observation space into regions that map to the various hypotheses so as to minimize the error probability. In the POM formalism given above for the quantum case, these two steps are collapsed into one but there is no essential difference (Indeed, a "two-step" formulation of quantum detection can be given -See e.g.[**16**]. However, the problem reduces to finding an optimal POM as given above.)

The difference only arises when the $\{\rho_x\}$ do not commute. In this case, there is no single analog of the classical observation space on which all signals $\{\rho_X\}$ can be assumed consistently to live. This is a characteristic feature of quantum mechanics. In effect, each possible measurement creates its own observation space and the probability distributions of the outcomes of different observables for a particular state are not compatible with (i.e. cannot all be derived from) a probability distribution on a "universal" observation space. In the classical theory, an observation space that includes the values of all the degrees of freedom of the system serves as such a universal space.

1.2.6. Binary Signal Sets

In the case of a binary signal set $\{\rho_0, \rho_1\}$, with a priori probabilities $\{p_0, p_1\}$, Helstrom (See [**12**]) gave a formula for the optimal error probability $\overline{P_e}$

$$\overline{P_e} = \frac{1}{2} - \frac{1}{2}||p_0\rho_0 - p_1\rho_1||_1. \tag{1.39}$$

Here $||A||_1 = Tr[\sqrt{A^\dagger A}]$ is the operator trace-norm. We use this formula to give the optimum $\overline{P_e}$ for a quantum binary phase shift keying (BPSK) signal set $\{|\alpha\rangle, |-\alpha\rangle\}$ with equal a priori probabilities. Here $|\alpha\rangle$ and $|-\alpha\rangle$ are coherent states. Analogously to classical communication, we define the average energy $S = \langle\alpha|N|\alpha\rangle = |\alpha|^2$. Then Eq. (1.39) gives:

$$\overline{P_e}^{opt} \approx \frac{1}{4}e^{-4S} \tag{1.40}$$

Interestingly, this performance can be reached to within a factor of two by a concrete setup called the Kennedy receiver. As we will see, this signal set is used in our $\alpha\eta$ key generation protocol. Two other common suboptimal measurements called the homodyne and heterodyne measurements (similar to their rf counterparts) on this signal set yield the respective error probabilities:

$$\overline{P_e}^{hom} = \frac{1}{2}e^{-2S}, \quad \overline{P_e}^{het} = \frac{1}{2}e^{-S} \tag{1.41}$$

This example illustrates the difference in performance that results from using different measurements on the same signal set.

CHAPTER 2

Error Probability Bounds for M-ary Quantum Detection

2.1. Literature Review and Motivation

The problem of obtaining the optimum quantum detector is solved explicitly only for a few special cases. Ban et. al [**18**] derive the solution for a pure state set $|\psi_i\rangle$, where the vectors $|\psi_i\rangle$ form a cyclic set, i.e. they are generated by applying a cyclic group of unitary operators to a single vector of the set. Eldar and Forney [**19**] discuss the so-called 'Least Squares Measurement' (LSM), which consists of rank-one POM elements chosen to minimize the sum of the squared differences between the ith state vector and the ith measurement vector. This measurement is shown to be optimum for geometrically uniform (GU) state sets, i.e., a set of equally likely pure states that are generated from one of them by the action of an Abelian group of unitary operators. This optimality result has also been extended to the case of geometrically uniform *mixed* states generated by even some non-Abelian groups of operators under certain conditions [**20**]. After we introduce the $\alpha\eta$ protocol in Chapter 3, we will see that the states corresponding to each plaintext sequence $\mathbf{x}_n$ in a ciphertext-only attack are indeed a geometrically uniform set. Unfortunately, no explicit solution for the measurement operators is found in [**20**], but the problem is left as a convex semidefinite programming problem, albeit one of much smaller dimension than the original one. Therefore, the problem of analytical estimates of error probability is still open. Furthermore, the semidefinite programming problem deriving from it is one of dimension exponential in n, ruling out even the possibility of numerical computation for large n. Thus, there is good reason to develop new upper bounds on the error probability that yield analytical estimates even under some assumptions such as geometric uniformity or less specific ones such as linear independence. In addition, to the author's knowledge, there are no general *lower bounds* on the M-ary error probability in the literature. Since analytical lower bounds on Eve's error probability are important for proving *security* of a quantum protocol (as opposed to upper bounds which can prove the *insecurity* of a protocol), searching for such bounds is well motivated. In the rest of this chapter, we develop both a novel upper bound and a novel lower bound for the M-ary quantum detection problem. The lower bound is of general validity, while the upper bound is valid when the M density operators involved have linearly independent support.

2.2. Sequential Detection Upper Bound on M-ary Quantum Error Probability

2.2.1. Pure State Case

Consider first the case of a *linearly independent* (LI) pure state ensemble $\{\pi_m, \psi_m\}_1^M$. Here the $\{\pi_m\}_1^M$ represent a priori probabilities, and we dispense with Dirac notation, which we will only use when convenient. We derive an upper bound on the *optimal* average probability of error (which we will call the APE), $\overline{P}_e$, by estimating it for a particular detection method that may be called 'Sequential Detection'.

The general idea is the following:- We first make a two-valued projection measurement that 'decides' between hypothesis H_1 and the remaining hypotheses. If H_1 is detected, we declare $\hat{m} = 1$ and stop. If not, we make a second projection measurement that 'decides' between H_2 and the hypotheses $H_3, \ldots, H_M$, and continue in a similar fashion until one of the hypotheses is detected. If the $(M-1)$th such measurement does not yield $\hat{m} = M-1$, we declare $\hat{m} = M$. Note that the said measurements are only made conceptually for the purpose of bounding the APE and, if made in actuality, would necessitate holding intact the state after each measurement before making the succeeding one.

Let us make the above precise. Let the M-dimensional subspace spanned by the M LI states $\{\psi_m\}$ be denoted V. Let the $(M-k+1)$-dimensional space spanned by $\{\psi_k, \ldots, \psi_M\}$ be denoted $V_k, 1 \leq k \leq M$, so that $V = V_1 \supset \ldots \supset V_M$. All the inclusions are strict because of the linear independence condition. The Projection Valued Measurement (PVM) $\{P_i, P_{\bar{i}}\}$ that describes the i-th measurement is given by

$$P_{\bar{i}} = P_{V_{i+1}}; P_i = P_{V_{i+1}} - P_{V_{i+2}}. \tag{2.1}$$

Here P_{V_i} is the orthogonal projection operator onto the subspace V_i. In the form above, the i-th measurement PVM is defined on the space V_i and not the entire space, but this is permissible since the $(i-1)$-th measurement projects the state into V_i when the outcome $\bar{i}$ is obtained. Let us now write the sequential detection error probability $\overline{P}_e^{seq}$ as:

$$\overline{P}_e^{seq} = \sum_{m=1}^{M} \pi_m Pr[\text{Error}|\psi_m \text{sent}]. \tag{2.2}$$

From the way the $\{P_i\}$ are defined, and because of the linear independence of the states, we can see that if ψ_m is sent, one *never* obtains the results $i \in \{1, \ldots, (m-1)\}$ because the support space of each such P_i is orthogonal to ψ_m. Thus, when ψ_m is sent, an error occurs if and only if the result $\overline{m}$ is obtained in the $\{P_m, P_{\overline{m}}\}$ measurement. Since

$Pr[\text{Error}|\psi_m \text{sent}] = \langle \psi_m | P_{\overline{m}} | \psi_m \rangle = \| P_{\overline{m}} \psi_m \|^2$, we have

$$\overline{P}_e^{seq} = \sum_{m=1}^{M} \pi_m \| P_{\overline{m}} \psi_m \|^2, \tag{2.3}$$

which is the Sequential Detection upper bound for arbitrary linearly independent pure states.

2.2.2. Mixed State Case

The upper bound on the APE just derived can be extended to an analogous one valid for mixed states. Similar to the linear independence condition assumed in the pure state case, we make the following assumption regarding the M density operators $\{\rho_m\}_{m=1}^M$ corresponding to the M hypotheses. Define $V_m' := \text{support}(\rho_m) = \text{range}(\rho_m)$. If $\dim(V_m') = d_m$, we will assume that d_m is finite and that

$$\dim(V_1' + \ldots + V_M') = \sum_{m=1}^{M} d_m. \tag{2.4}$$

In other words, if we collect into a larger set the basis vectors of each V_m', the larger set is still linearly independent. A common situation in which this condition holds is when each ρ_m is a mixture of pure states $\psi_m^l, l = 1, \ldots L_m$ and the set of vectors $\bigcup_{m=1}^M \bigcup_{l=1}^{L_m} \{\psi_m^l\}$ is linearly independent. In particular, the condition holds when the $\{\psi_m^l\}$ are *distinct* coherent states.

When Eq (2.4) holds, the argument used to define the sequential measurement projectors $\{P_m, P_{\overline{m}}\}$ in the previous subsection goes through as before with V_m in Eq. (2.1) replaced by

$$V_m = V_m' + \ldots + V_M. \tag{2.5}$$

The probability of error is again given by

$$\overline{P}_e^{seq} = \sum_{m=1}^{M} \pi_m Pr[\text{Error } |\rho_m \text{ sent}]. \tag{2.6}$$

Since the support space of ρ_m is orthogonal to those of the projectors $\{P_i\}_{i=1}^{m-1}$, when ρ_m is sent, none of the outcomes $i \in \{1, \ldots, (m-1)\}$ can be obtained. Thus, we again have

$$\overline{P}_e^{seq} = \sum_{m=1}^{M} \pi_m \text{tr}(\rho_m P_{\overline{m}}). \tag{2.7}$$

If $\rho_m = \sum_{l=1}^{L_m} p_m^l |\psi_m^l\rangle\langle\psi_m^l|$, we get the following upper bound on the APE for mixed states:

$$\overline{P}_e^{seq} = \sum_{m=1}^{M} \pi_m \sum_{l=1}^{L_m} p_m^l \parallel P_{\overline{m}} \psi_m^l \parallel^2 \geq \overline{P}_e. \tag{2.8}$$

2.2.3. ϵ-Orthogonal States

To illustrate the above upper bounds, let us calculate them for what may be called ϵ-orthogonal states. In the pure state case, we mean by this that we have the uniform upper bound

$$0 \leq |\langle\psi_m|\psi_{m'}\rangle| \leq \epsilon \quad \forall m \neq m' \tag{2.9}$$

on the inner product between any two states. Intuitively, such an upper bound on the pairwise inner products should yield an upper bound on $\parallel P_{\overline{m}} \psi_m \parallel$ for each m, which will then also upper bound $\overline{P}_e^{seq}$ via Eq. (2.3). Indeed, such an upper bound can be derivedfrom the following lemma:

Lemma 2.2.1. *Let $\mathcal{S} = \{x, e_1, \ldots, e_n\}$ be a linearly independent set of unit vectors with the property that $|\langle s|s'\rangle| \leq \epsilon \ \forall s, s' \in \mathcal{S}$. Let $\mathcal{E}$ denote the subspace spanned by $\{e_1, \ldots, e_n\}$ and let E be the orthogonal projector onto this subspace. If $(n-1)\epsilon < 1$, then $\| Ex \| \leq \frac{\epsilon\sqrt{n}}{1-(n-1)\epsilon}$.*

Proof: First observe that

$$\| Ex \| = \max_{y\in\mathcal{E}, \|y\|=1} |\langle y|x\rangle|. \tag{2.10}$$

Now $y \in \mathcal{E}$ implies that $y = \sum_{i=1}^{n} y_i e_i$ for some vector of coefficients $\mathbf{y} := (y_1, \ldots, y_n)$ and $\| y \| = 1$ implies that $(\mathbf{y}, \mathbf{G}\mathbf{y}) = 1$. In the latter equation $(\cdot, \cdot)$ is the usual inner product on $\mathbb{C}^n$ (not to be confused with the inner product in the state space), and $\mathbf{G}$ is the matrix with entries $\mathbf{G}_{ij} = \langle e_i|e_j\rangle$. Now

$$\| Ex \| = \max_{y\in\mathcal{E}, \|y\|=1} |\langle y|x\rangle| \tag{2.11}$$

$$= \max_{\mathbf{y}\in\mathbb{C}^n, (\mathbf{y},\mathbf{G}\mathbf{y})=1} |\sum_{i=1}^{n} y_i \langle x|e_i\rangle| \tag{2.12}$$

$$\leq \max_{\mathbf{y}\in\mathbb{C}^n, (\mathbf{y},\mathbf{G}\mathbf{y})=1} \epsilon |\sum_{i=1}^{n} y_i| \tag{2.13}$$

$$= \max_{\mathbf{y}\in\mathbb{C}^n, (\mathbf{y},\mathbf{G}\mathbf{y})=1} \epsilon \| \mathbf{y} \|_1 \tag{2.14}$$

$$\leq \epsilon\sqrt{n} \max_{\mathbf{y}\in\mathbb{C}^n, (\mathbf{y},\mathbf{G}\mathbf{y})=1} \| \mathbf{y} \|_2 \tag{2.15}$$

$$= \frac{\epsilon\sqrt{n}}{\lambda_{\min}(\mathbf{G})}. \tag{2.16}$$

Here $\|\cdot\|_1$ and $\|\cdot\|_2$ refer respectively to the l_1 and l_2 vector norms in $\mathbb{C}^n$ and $\lambda_{\min}(\mathbf{G})$ is the smallest eigenvalue of $\mathbf{G}$. Inequality (2.13) is a consequence of the assumption of the theorem and the rest of the inequalities are standard matrix theory results (See, e.g., [**21**]). We know that $\mathbf{G}_{ii} = 1 \ \ \forall i$ and $|\mathbf{G}_{ij}| \le \epsilon \ \ \forall i \ne j$. Also, being a Gram matrix of linearly independent vectors, $\mathbf{G}$ is positive definite. Thus, according to the Gerschgorin Disk Theorem (See [**22**], p. 344), all the eigenvalues of $\mathbf{G}$ are located in the interval $[1-(n-1)\epsilon, 1+(n-1)\epsilon]$, and hence so is its minimum eigenvalue. If $(n-1)\epsilon < 1$, $1-(n-1)\epsilon > 0$ and the result follows.

Applying Lemma 2.1 with the appropriate identifications to Eq (2.3) we get, provided $(m-1)\epsilon < 1$, the bound

$$\overline{P}_e \le \overline{P}_e^{seq} \le \sum_{m=1}^{M-1} \frac{\pi_m \epsilon^2 m}{[1-(m-1)\epsilon]^2} \tag{2.17}$$

for the APE. If $\pi_m = 1/M$, we may use the integral bound

$$\frac{1}{M}\sum_{m=1}^{M-1} \frac{\epsilon^2 m}{[1-(m-1)\epsilon]^2} \le \frac{1}{M}\int_{x=1}^{M} \frac{\epsilon^2 x}{[1-(x-1)\epsilon]^2} \tag{2.18}$$

giving the analytical upper bound

$$\overline{P}_e \le \frac{\epsilon(1+\epsilon)}{M[1-(M-1)\epsilon]} + \frac{\ln[1-(M-1)\epsilon]}{M}. \tag{2.19}$$

As an application of this result, observe that if $\epsilon M \to 0$, $\overline{P}_e \le \epsilon + \ln(1+\epsilon)/M$. We will apply this result to $\alpha\eta$ in Chapter 4.

For the states ψ_m^l in Eq. (2.8), we may also assume an ϵ-orthogonality condition such as

$$0 \leq |\langle\psi_m^l|\psi_{m'}^{l'}\rangle| \leq \epsilon \text{ whenever } \delta(m,m')\delta(l,l') = 0. \tag{2.20}$$

Under this assumption, and assuming that each L_m of Eq (2.8) equals L for all m, and also that $[(M-1)L-1]\epsilon < 1$, one can again apply Lemma 2.1, which for the case of uniform a priori probabilities gives the upper bound

$$\overline{P}_e \leq \overline{P}_e^{seq} \leq \frac{\epsilon^2 L}{M}\sum_{m=1}^{M-1}\frac{m}{[1-(mL-1)\epsilon]^2}, \tag{2.21}$$

from which an integral bound yields

$$\overline{P}_e \leq \frac{1}{M}\{\frac{(1+\epsilon)L\epsilon(M-1)}{[1-(ML-1)\epsilon][1-(L-1)\epsilon]} + \ln[\frac{1-(ML-1)\epsilon}{1-(L-1)\epsilon}]\}. \tag{2.22}$$

2.3. Binary Detection Lower Bound on M-ary Quantum Error Probability

2.3.1. Binary Detection Lower Bound

As before, we are concerned with an ensemble $\{(\pi_m, \rho_m)\}_{m=1}^M$ of M quantum states corresponding to each of the M hypotheses. If $\{M_m\}$ denotes the *optimal* M-ary decision POM, we have for the optimal average probability of correct decision $\overline{P}_c$ (referred to as APC in the following):

$$\overline{P}_c = \sum_{m=1}^{M} \pi_m \mathrm{tr}(\rho_m M_m). \tag{2.23}$$

Let us define a state $\rho_{\overline{m}}$ as

$$\rho_{\overline{m}} := \frac{1}{1-\pi_m} \sum_{m' \neq m} \pi_{m'} \rho_{m'}. \tag{2.24}$$

Conceptually, this is the state obtained when we know that state ρ_m was not prepared, but we do not know which of the remaining states $m' \neq m$ was prepared. As such, it is analogous to the classical case of a *composite hypothesis* (See, e.g., [**23**], p. 86) consisting of $H_{m'}$ for $m' \neq m$. Now consider the situation where $\rho_{\overline{m}}$ is prepared with probability $(1 - \pi_m)$ and ρ_m is prepared with probability π_m. As far as the receiver is concerned, this situation is *physically identical* to the preparation of the original ensemble $\{(\pi_m, \rho_m)\}_{m=1}^M$. This is also true in the classical case, with density operators replaced by probability distributions on a given observation space. In both cases, what we mean by 'physically identical' is that the receiver sees the same probability distribution for any observation in the two cases.

Given this fact, suppose that the receiver desires to distinguish with minimum probability of error the two hypotheses of the ensemble $\{(\pi_m, \rho_m), (1-\pi_m, \rho_{\overline{m}})\}$. A *suboptimal* way of doing so is to make the optimum M-ary measurement $\{M_m\}$ and declare all outcomes $m' \neq m$ as $\overline{m}$. If she does so, her probability of correct decision P_m for the binary problem is given by

$$P_m := \pi_m \text{tr}(\rho_m M_m) + \sum_{m',m''\neq m} \pi_{m'} \text{tr}(\rho_{m'} M_{m''}) \leq \overline{P}^{bin}_{c(m)}, \tag{2.25}$$

where we have denoted her optimal probability of correct decision by $\overline{P}^{bin}_{c(m)}$. Incidentally, comparing the above equation to Eq (2.23) reveals that

$$\overline{P}_c \leq P_m \leq \overline{P}^{bin}_{c(m)}, \tag{2.26}$$

so that the quantity P_m and hence $\overline{P}^{bin}_{c(m)}$ are upper bounds on $\overline{P}_c$. Instead of using, e.g., $\min_m \overline{P}^{bin}_{c(m)}$ as an upper bound, we can attempt to improve on the bound (2.26) by evaluating $\sum_{m=1}^{M} P_m$ by summing over the M equations of the form (2.25) for each m. Doing so reveals that, in fact, $\sum_{m=1} P_m = 2\overline{P}_c + (M-2)$. Plugging this into Eq. (2.26), and writing everything in terms of APE's ($\overline{P}_c = 1 - \overline{P}_e$ etc.), we get the simple result

$$\overline{P}_e \geq \frac{1}{2} \sum_{m=1}^{M} \overline{P}^{bin}_{e(m)}. \tag{2.27}$$

Substituting the Helstrom formula Eq (1.39) into the RHS gives the following 'Binary Detection' lower bound on $\overline{P}_e$:

$$\overline{P}_e \geq \frac{1}{4} \sum_{m=1}^{M} [1- \parallel \pi_m \rho_m - (1-\pi_m)\rho_{\overline{m}} \parallel_1]. \tag{2.28}$$

In order to further evaluate Eq. (2.28), we can write

$$\| \, \pi_m \rho_m - (1-\pi_m)\rho_{\overline{m}} \, \|_1 = \| \sum_{m' \neq m} \{ \frac{\pi_m}{M-1} \rho_m - \pi_{m'} \rho_{m'} \} \, \|_1 \tag{2.29}$$

$$\leq \sum_{m' \neq m} \| \, \frac{\pi_m}{M-1} \rho_m - \pi_{m'} \rho_{m'} \, \|_1, \tag{2.30}$$

where the inequality follows from the convexity of the trace norm. We can further split up the density operators in the above equation to mixtures of pure states and re-apply convexity until we end up with trace norm terms involving the difference of pure state density operators. These can be evaluated in terms of the inner product of the corresponding states using the following formula:

$$\| \, |\psi_0\rangle\langle\psi_0| - \lambda|\psi_0\rangle\langle\psi_0| \, \|_1 = \; [(1+\lambda)^2 - 4\lambda|\langle\psi_0|\psi_1\rangle|^2]^{1/2}. \tag{2.31}$$

This technique will be applied in Chapter 4.

CHAPTER 3

Quantum Direct Encryption: The $\alpha\eta$ protocol

3.1. Random Ciphers – Quantitative Definition

The characterization of a general random cipher in Sec. 1.1 merely using Eq. (1.3) or (1.5) is perhaps not well-motivated. The reason for studying random ciphers is in fact the belief that they enhance the security of the cipher against various attacks. By bringing into focus the intuitive mechanism by which a random cipher may provide greater security than a nonrandom counterpart against known-plaintext attacks, we will propose one possible quantitative characterization of a general random cipher (or more exactly, a general random *stream* cipher. See below.). For a description of known-plaintext and other attacks on ciphers, together with the known results on their security, we refer the reader to Section 1.1.

We now discuss the intuitive mechanism of security enhancement in a random cipher. To this end, a schematic depiction of encryption and decryption with a random cipher is given in Fig. 3.1. For a binary alphabet $\mathcal{X} = \{0, 1\}$, let $\mathcal{X}^n = \{a_1, \ldots, a_N\}$ be the set of $N = 2^n$ possible plaintext n-sequences. Let k be a particular key value. One can view the key k as dividing the ciphertext space $\mathcal{Y}^n$ into N parts, denoted by the $\mathcal{A}^k_{a_j}, j \in \{1, \ldots, N\}$, in the figure. Encryption of plaintext a_j proceeds by first determining the relevant region $\mathcal{A}^k_{a_j}$ and randomly selecting (this is the function of the private randomizer) as ciphertext some $y \in \mathcal{A}^k_{a_j}$. The decryption condition Eq.(1.2) is satisfied by virtue of the regions

$\mathcal{A}_{a_j}^k$ being disjoint for a given k. Also shown in Fig. 3.1 is the situation where a different key value k' is used in the system. The associated partition of $\mathcal{Y}^n$ consists of the sets ${\mathcal{A}'}_{a_j}^k$ that are shown with shaded boundaries in Fig. 3.1. The *important point* here is that the respective partitions of the ciphertext space for the key values k and k' should be sufficiently 'intermixed'. More precisely, for any given plaintext a_j, and any observed ciphertext $\mathbf{y}_n$, we require that there exist sufficiently many key values k (and hence a sufficiently large probability of the set of possible keys corresponding to a given plaintext and observed ciphertext) for which $\mathbf{y}_n \in \mathcal{A}_{a_j}^k$. In other words, a given plaintext-ciphertext pair can be connected by many possible keys. This is the intuitive basis why random ciphers offer better quantitative security (as measured either by Eve's information on the key or her complexity in finding it.

While the above arguments hold for any type or random cipher whatsoever, we will restrict our scope to the so-called *stream ciphers.* Most ciphers in current use (which are all nonrandom), such as AES, are stream ciphers [**3**]. In a nonrandom stream cipher, the key K is first expanded using a deterministic function into a much longer sequence $(Z_1, \ldots, Z_n)$ called the *keystream* or *running key.* The defining property of a *stream cipher* is that the i-th ciphertext symbol y_i be a function of just the i-th keystream symbol z_i and the earlier and current plaintext symbols $x_1, \ldots, x_i$:

$$y_i = E^i(x_1, \ldots, x_i; z_i). \tag{3.1}$$

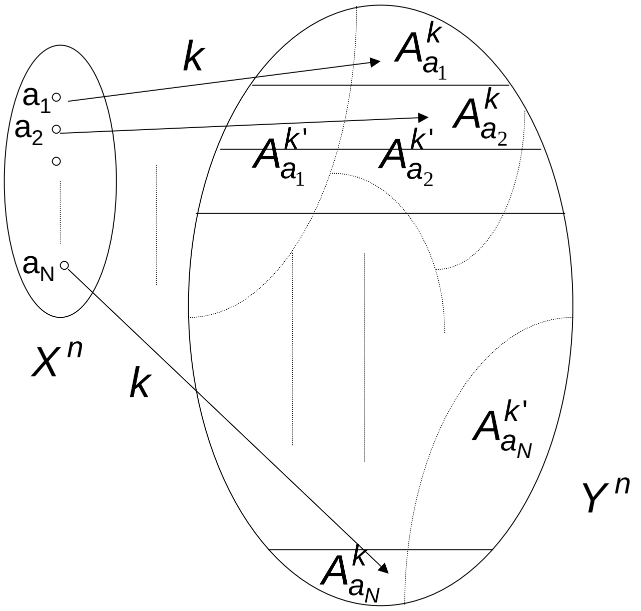

Figure 3.1. Schematic of a random cipher: The plaintexts a_i are carried, under the key k, into the corresponding regions $A^k_{a_i}$ of ciphertext space Y^n. The subsets of Y^n associated with a different key value k' are shown with curved boundaries.

It follows that decryption of the first i symbols of plaintext is possible from the first i symbols of ciphertext and the running key. A *synchronous* stream cipher is one for which

$$y_i = E^i(x_i; z_i). \tag{3.2}$$

Thus, the i-th ciphertext symbol depends only on the i-th plaintext symbol and the i-th keystream symbol, i.e., the cipher is memoryless. For our discussion of random ciphers, we will restrict ourselves for concreteness to the case of *random stream ciphers*, that are defined by:

$$y_i = E^i(x_1, \ldots, x_i; z_i; r_i). \tag{3.3}$$

Here, the $\{R_i\}$ are randomizers that may be assumed to be independent random variables (this is the case in $\alpha\eta$), but this is not necessary. In the following, a *random cipher* will always mean a *random stream cipher.*

For a nonrandom stream cipher given by Eq. (3.1), it is usually the case that given the plaintext vector $\mathbf{x}_i$ of length i and ciphertext symbol y_i, the value of the keystream z_i is uniquely determined. This is typically the case also in a random stream cipher *when the value r taken by the randomizer R_i is known.* In the absence of such knowledge, however, the different possible values taken by R_i will in general allow many different values of the keystream for the given plaintext vector and ciphertext symbol. The more such possibilities exist, the less information is obtained about the keystream and the more 'secure' the cipher is. Our quantitative definition of random cipher given below introduces a parameter Γ that provides one way of quantifying the different knowledge of the keystream obtained in the above two scenarios by the number of additional possible keystreams for a given pair of input data and corresponding ciphertext symbols.

Definition (Γ- *Random Cipher*) **:**

A Γ-Random Cipher is a random stream cipher of the form of Eq. (3.3) for which the following condition holds:

For every plaintext sequence, $\mathbf{x}_i$, for every i, for every ciphertext symbol y_i obtainable by encryption of $\mathbf{x}_i$, and for every value r of R_i,

$$(3.4) \quad |\{z_i | y_i = E^i(x_1, \ldots, x_i; z_i; r') \text{ for some } r'\}| - |\{z_i | y_i = E^i(x_1, \ldots, x_i; z_i; r)\}| \geq \Gamma.$$

The bars $|\cdot|$ indicate size of the enclosed set. For a nonrandom stream cipher, the keystream z_i is uniquely fixed by the plaintext vector $\mathbf{x}_i$ and the ciphertext symbol y_i. Therefore, if the randomizer in (3.4) is ignored so that it applies to a nonrandom cipher, a nonrandom cipher would have $\Gamma = 0$. Note that the sets whose sizes appear in the above equation, both for random ciphers and their nonrandom reductions, are constructed only on the basis of the i-th ciphertext symbol y_i, and not on the basis of the entire ciphertext sequence. Thus, the definition of Γ only gives the *number of possible keys per symbol of ciphertext* under known-plaintext attack, while the number of possible keys based on the entire ciphertext sequence (that is illustrated schematically by the overlap sets in Fig. 3.1) may be significantly less. In this sense, our definition has a restricted symbol by symbol scope but is easy to calculate with, similar to the independent particle approximation in many-body physics. It does not by itself determine the precise security of the cipher, but rather is the starting point of precise analysis, which is a difficult task just as correlations in interacting many-body systems are always difficult to deal with in a rigorous quantitative manner.

It is possible to satisfy the random cipher condition (1.5) with $\Gamma = 0$. This happens, e.g., when (3.4) holds for some ciphertext symbols with $\Gamma > 0$ but some others with $\Gamma = 0$, so the overall condition (3.4) is only satisfied for $\Gamma = 0$. A different measure of randomization Λ, bearing directly on (1.5), may be introduced which has the property that $\Lambda = 0$ is equivalent to a nonrandom cipher. For the case where the ciphertext alphabet is finite and for given $\mathbf{x}_i, z_i$ and r, let

$$(3.5) \quad \Lambda = |\{y_i | y_i = E^i(x_1, \cdots, x_i; z_i; r') \text{ for some } r'\}| - |\{y_i | y_i = E^i(x_1, \cdots, x_i; z_i; r)\}|.$$

Thus, condition (1.5) is equivalent to $\Lambda > 0$ for some $\mathbf{x}_i, z_i$ and r. It follows that $\Lambda = 0$ for all $(\mathbf{x}_i, z_i)$ is equivalent to the cipher being nonrandom. $\Lambda + 1$ is the number of possible output signal symbols corresponding to a given input symbol and running key value. Thus, the parameter Λ measures directly the degree of per symbol ciphertext randomization, while Γ measures the per symbol key redundancy. It is possible that a $\Gamma = 0$ random cipher is still useful due to the additional loads on Eve to record and store more information from her observation. On the other hand, for the *typical* case where z_i is in one-to-one correspondence with y_i for given $\mathbf{x}_i$ and r, $\Gamma > 0$ implies $\Lambda > 0$ for every $\mathbf{x}_i$ and z_i, which in turn implies that a cipher with $\Gamma > 0$ is random in the sense of (1.5). The following simple example serves to illustrate the above definitions:

Example (Random cipher)
Let $\mathcal{X} = \{0, 1\}$, $\mathcal{K} = \{k_0, k_1, k_2, k_3, k_4\}$ and $\mathcal{Y} = \{a, b, c, d, e\}$. Fig. 3.2 lists the possible ciphertexts for each plaintext and key pair.

For this cipher, one can easily verify that at least 2 key values connect every possible plaintext-ciphertext pair. In addition, every plaintext-key pair can lead to at least two different ciphertexts. In terms of the definitions given above, this cipher has $\Gamma = 1$ and $\Lambda = 1$.

x	k	y
0	k_0	a, b
1	k_0	c, d, e
0	k_1	c, d
1	k_1	e, a, b
0	k_2	e, a
1	k_2	b, c, d
0	k_3	b, c
1	k_3	d, e, a
0	k_4	d, e
1	k_4	a, b, c

Figure 3.2. Encryption table for a simple random cipher.

An inspection of the defining equation Eq. (3.4) for a random cipher (or Fig. 3.1) suggests how a random cipher may provide greater security against KPAs. For a given plaintext-ciphertext sequence pair, Eq.(3.4) suggests that one has some residual uncertainty on the value of the keystream $(Z_1, \ldots, Z_n)$, which does not exist for a corresponding nonrandom cipher. On the other hand, Eq.(3.4) refers only to the per-symbol uncertainty of the key stream calculated without regard to the ciphertext observed for the other symbols in the sequence. When such correlations are taken into account, the uncertainty on the keystream may be drastically reduced and we can give no general quantitative assertions of information-theoretic security. Note, however, that due to the randomization, the unicity distance n_1 of a random cipher under known-plaintext attacks can be expected to be bigger than that of any of its nonrandom reductions. Thus, the complexity-based security would be greater.

3.2. Quantum Random Ciphers

The known and possible advantages of a random classical cipher over a nonrandom one were discussed in the previous section. While it is possible to implement a random cipher classically using random numbers generated on Alice's side, this is not currently practical at high ($\sim$ Gbps) rates. As will become clear in the sequel, the quantum encryption protocol $\alpha\eta$ (Various implementations are described in [**24, 25, 26, 27, 28**] - The protocol in [**28**] is a variation on the original $\alpha\eta$ of [**24**]) effectively implements a random cipher from Eve's point of view for a given choice of her measurement, the difference from a classically random cipher being that it uses coherent-state quantum noise to perform the needed randomization. Before we describe $\alpha\eta$, we define some concepts that capture the relevant features of a quantum random cipher. As emphasized earlier, we will confine our attention to *stream* ciphers. First, we straightforwardly extend the usual stream cipher to one where the ciphertext is a quantum state. Our motivation for this definition is that, from the point of view of the legitimate users Alice and Bob, $\alpha\eta$ is a quantum stream cipher with negligible λ in the sense given below:

Definition *(λ-Quantum Stream Cipher (QSC))***:**
A quantum stream cipher is a cipher for which the following two conditions are satisfied:

A. The encryption map $e_k(\cdot)$ takes the n-symbol plaintext sequence $\mathbf{x}_n$ to a quantum state n-sequence ρ in the n-fold tensor product form:

$$\rho = e_k(\mathbf{x}_n) = \rho_1(x_1; z_1) \otimes \ldots \otimes \rho_n(x_1, \ldots, x_n; z_n), \tag{3.6}$$

and

B. Given the key k, there exists a measurement on the encrypted state sequence, that recovers each plaintext symbol x_i with probability $P_{dec} > 1 - \lambda$.

Here, as in Section 3.1, $(Z_1, \ldots, Z_n)$ is the keystream generated from the seed key K. A few comments will help clarify the definition. First, note that the tensor product form of the state in condition A retains for a quantum cipher the property of a classical cipher that one can generate the components in the n-sequence of states that constitute the output of a cipher one after the other in a time sequence. Note also that, analogous to a classical stream cipher, the i-th tensor component of ρ depends on just z_i and $(x_1, \ldots, x_i)$. Condition B is the generalized counterpart of the decryption condition Eq.(1.2) for a classical cipher – we now allow a small enough decryption error probability. Thus, the per-symbol error probability is bounded above by $\lambda < 1$.

We now want to bring the concept of classical *random* cipher defined in the previous section into the quantum setting. Our motivation in doing so is to show that, for an attacker making the same measurement on a mode-by-mode basis without knowledge of the key, $\alpha\eta$ reduces to an equivalent Γ-Random Cipher with significantly large Γ. Since the output of a quantum cipher is a quantum state and not a random variable, we will need to specify a POM $\{\Pi_{\mathbf{y}_n}\}$ whose measurement result $\mathbf{Y}_n$ supplies the classical ciphertext. Note that in this quantum situation different choices of measurement may result in radically different kinds of ciphertext. Note also that the user's and the attacker's measurements may be different. Our definition of a quantum random stream cipher below will apply relative to a chosen ciphertext $\mathbf{Y}_n$ defined by its associated POM. We will also assume that, from the eavesdropper's viewpoint, the same measurement is made on each of the n components of the cipher output. In other words, the POM defining the ciphertext $\mathbf{Y}_n$ is a tensor product of identical POMs $\{\pi_y\}$.

Definition ($(\Gamma, \lambda, \lambda', \{\pi_y\})$- *Quantum Random Stream Cipher (QRC)*)**:**
An $(\Gamma, \lambda, \lambda', \{\pi_y\})$ - quantum random stream cipher is a λ-quantum stream cipher such that for the ciphertext given by the result of the product POM $\{\Pi_{\mathbf{y}_n} = \bigotimes_{i=1}^{i=n} \pi_{y_i}\}$,

A. one has an Γ-random stream cipher satisfying Eq.(3.4), and

B. the probability of error per symbol P'_{dec} using the key *after* measurement is $P'_{dec} > 1 - \lambda'$.

Several comments are given to explain this definition:

1. While condition QRC-B above appears similar to the condition QSC-B for a quantum stream cipher, there is a crucial difference. In the latter, the decryption probability P_{dec} takes into account the possibility that the *quantum measurement* (as well as classical post-processing) made on the cipher state can depend on the key, i.e. it refers to Bob's rather than Eve's error probability. In QRC-B, we are considering the probability of error involved for Eve when she decrypts using a quantum measurement independent of the key followed by classical post-processing that is , in general, "collective" and depends on the key. Thus, the parameter λ' is related to the symbol error probability under this latter restriction while the parameter λ in QSC-B is tied to the symbol error probability for a quantum measurement allowed to depend on the key. We see that there are two measurements implicit in our definition of a QRC - one made by the user with the help of the key, and the other given by $\{\pi_y\}$ made by the attacker without

the key. See also Item 3 below. As we shall see, $\alpha\eta$ satisfies QRC-B with negligible λ' under a heterodyne or phase measurement attack by Eve.

2. Γ in QRC-A, as in Eq.(3.4), is a measure of the 'degree of intermixing' of the regions of ciphertext space corresponding to different key values on a symbol-by-symbol basis. If $\{\pi_y\}$ describes a discrete measurement, a Λ corrresponding to Eq.(3.5) can also be introduced.

3. Our stipulation that the same POM be measured on each of the components of the cipher output is tantamount to restricting the attacker to identical measurements on each tensor component followed by collective processing. We will call such an attack a *collective attack* in this paper (also in [**9**]). This definition is different from the usual collective attack in quantum cryptography [**29**]: in the latter, following the application of identical probes to each qubit/qumode, a joint quantum measurement on all the probes is allowed. In our case, there is no probe for Eve to set as we conceptually allow her a full copy of the quantum state. Doing so, we can upper bound her performance. (This is an important feature of our so-called KCQ approach to encryption and key generation. See [**1**] for discussion.) Thus, allowing a joint measurement, as also nonidentical measurements on each output component, will be called a joint attack.

4. In analogy with the classical random cipher definition Eq. (3.4), one may wonder why the private randomizers R_i used in that definition are missing from that of the quantum random cipher. Indeed, one may randomize the quantum state $\rho_i(x_1, \ldots, x_i; z_i)$ to $\rho_i(x_1, \ldots, x_i; z_i; r_i)$ using a private random variable with probability distribution p_{r_i}. However, since the value of R_i remains unknown to both user and attacker (Indeed, the user should not need to know R_i in order to decrypt or even to encrypt in the

case of $\alpha\eta$), one sees that all probability distributions of Bob's or Eve's measurements in this situation are given by the state $\rho'_i(x_1, \ldots, x_i; z_i) = \sum_{r_i} p_{r_i} \rho_i(x_1, \ldots, x_i; z_i; r_i)$, in which there is no explicit dependence on r_i. In particular, we mention here that exactly such quantum state randomization, called Deliberate Signal Randomization (DSR), has been proposed in the context of $\alpha\eta$ in [**1**] for the purposes of enhancing the information-theoretic security of $\alpha\eta$.

5. It is important to observe that the definitions given above both for classical and quantum random ciphers are not arbitrary ones, but rather the mathematical characterizations of very typical situations involving randomization in classical and quantum cryptosystems.

We present an example of a QRC in the next section: the $\alpha\eta$ cryptosystem.

3.3. The $\alpha\eta$ cryptosystem

3.3.1. Operation

We now describe the $\alpha\eta$ system and its operation as a quantum cipher:

(1) Alice and Bob share a secret key $\mathbf{K}_s$.

(2) Using a *key expansion function* $ENC(\centerdot)$, e.g., a linear feedback shift register or AES in stream cipher mode, the seed key $\mathbf{K}_s$ is expanded into a running key sequence that is chopped into n blocks: $\mathbf{K}_{mn} = ENC(\mathbf{K}_s) = (K_1, \ldots, K_{mn})$. Here, $m = \log_2(\mu)$, so that $Z_i \equiv (K_{(i-1)m+1}, \ldots, K_{im})$ can take μ values. The Z_i constitute the *keystream.*

(3) The encrypted state $e_{\mathbf{K}_s}(\mathbf{X}_n)$ of Eq.(3.6)is defined as follows. For each bit X_i of the plaintext sequence $\mathbf{X}_n = (X_1, \ldots, X_n)$, Alice transmits the *coherent state*

$$|\psi(X_i, Z_i)\rangle = |\alpha e^{i\theta(X_i, Z_i)}\rangle. \tag{3.7}$$

Here, $\alpha \in \mathbb{R}$ and $\theta(X_i, Z_i)$ takes values in the set $\{0, \pi/\mu, \ldots, (2\mu - 1)\pi/\mu\}$. The function θ taking the data bit and keystream symbol to the actual angle on the coherent state circle is called the *mapper*. In this paper, we choose $\theta(X_i, Z_i) = [Z_i/\mu + (X_i \oplus Pol(Z_i))]\pi$. $Pol(Z_i) = 0$ or 1 according to whether Z_i is even or odd. This distribution of possible states is shown in Fig. 3.3. Thus K_i can be thought of as choosing a 'basis' with the states representing bits 0 and 1 as its end points. In general, one has the freedom to vary the mapper in various ways for practical reasons. See, e.g, [**26**].

(4) In order to decrypt, Bob runs an identical ENC function on his copy of the seed key. For each i, knowing Z_i, he makes a quantum measurement to discriminate just the two states $|\psi(0, Z_i)\rangle$ and $|\psi(1, Z_i)\rangle$.

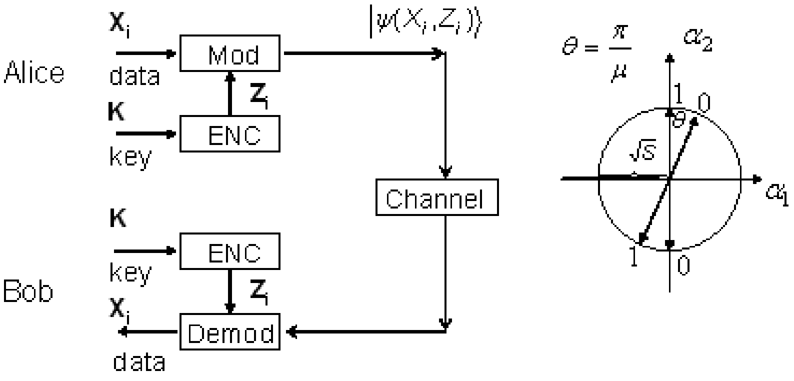

Figure 3.3. Left – Overall schematic of the $\alpha\eta$ encryption system. Right – Depiction of two of μ bases with interleaved logical bit mappings.

To decrypt in step (4) above, Bob, in general would need a phase reference. This is effectively provided by the use of Differential Phase Shift Keyed (DPSK) signals in the implementations of $\alpha\eta$. See [**25, 26, 27**] for details. Doing so does not compromise security as we still assume that Eve has a perfect copy of the transmitted state.

If the line transmittance between Alice and Bob is η, Bob receives a coherent state with energy ηS instead of $S \equiv |\alpha|^2$. The optimal quantum measurement Eq (1.39) for Bob has error probability

$$P_e^B \sim \frac{1}{4}\exp(-4\eta S). \tag{3.8}$$

It is thus apparent that $\alpha\eta$ is a λ-quantum cipher in the sense of Section 3.2 with $\lambda \sim \frac{1}{4}\exp(-4\eta S)$. For the $S \sim 4 \times 10^4$ of [**26**], over a distance of 80 km at a loss of 0.2 dB/km, we have $\eta S \sim 10^3$ photons. For this mesoscopic level, λ is $\sim \exp(-1000)$, which is completely negligible compared, say, to the standard acceptable BER limit of 10^{-9}, which arises from device imperfections, for an uncoded optical on-off keyed line.

Let us briefly indicate how this system may provide data security by considering an *individual attack* on each data bit X_i by Eve. Under such an attack, one only looks at the per-bit error probability ignoring correlations between the bits. Under this assumption, Eve, not knowing Z_i, is faced with the problem of distinguishing the density operators ρ^0 and ρ^1 where

$$\rho^b = \sum_{Z_i} \frac{1}{\mu} |\psi(b, Z_i)\rangle\langle\psi(b, Z_i)|. \tag{3.9}$$

For a fixed signal energy S, Eve's optimal error probability is numerically seen to go asymptotically to 1/2 as the number of bases $\mu \to \infty$ (See Fig. 1 of [**24**]). The intuitive reason for this is that increasing μ more closely interleaves the states on the circle representing bit 0 and bit 1, making them less distinguishable. Therefore, at least under such individual attacks on each component qumode [1] of the cipher output, $\alpha\eta$ offers any desired level of security determined by the relative values of S and μ. While we are not concerned in this paper with key generation, it may be observed that unambiguous state determination (USD) attacks on $\alpha\eta$ are totally ineffective due to the large number of 2μ states involved.

In our security analysis, Eve is always assumed to be at the transmitter so that $\eta = 1$ for her. Without knowing the key, however, her performance on the data is still poor as described in the above paragraph. Her attacks on the key are described in the following. We have assumed that the users can utilize the signal energy ηS to maintain a proper bit error rate without channel coding, despite possible interference from Eve. This does not place a stringent requirement on η itself as one can typically go around 80 km in fiber before the signal needs to be amplified. In case Eve's interference is too strong and causes error, it would be detected in a message authentication code which always goes with encryption. There is clearly no need to do separate intrusion detection in this direct encryption case, but it turns out there is also no need in the key generation regime [**9, 1**] which we do not discuss here.

[1] When referring to an optical field mode, we use the term *qumode* (for 'quantum mode', in analogy to 'qubit').

3.3.2. $\alpha\eta$ as a Random Cipher

We showed in the previous subsection that $\alpha\eta$ may be operated in a regime of S, η and μ where it is a λ-quantum cipher for $\lambda \sim 0$. We now show, that from Eve's point of view, under both a heterodyne and phase measurement attack, $\alpha\eta$ appears effectively as a quantum *random* cipher according to the characterization of Section 3.2. Note that the randomization in $\alpha\eta$ can also be effected in principle by using an additional classical random number generator. This is not required in $\alpha\eta$ as high-speed randomization is automatically provided by the coherent-state quantum noise.

To see the quantum random cipher characteristic of $\alpha\eta$, consider employing the following two measurements for obtaining $\{\pi_y\}$ in the quantum random cipher definition:

1) (Heterodyne measurement) $\pi_y = \frac{1}{\pi}|y\rangle\langle y|, y \in \mathbb{C}$.

2) (Canonical Phase measurement) $\pi_\theta = \frac{1}{2\pi}\sum_{n,n'=0}^{\infty} e^{i(n-n')\theta}|n\rangle\langle n'|, \theta \in [0, 2\pi)$.

To show that the conditions for a QRC are satisfied, let us first consider QRC-B. It may be shown [**1**] that the error probabilities λ' involved are respectively $\sim \frac{1}{2}e^{-S}$ and $\sim \frac{1}{2}e^{-2S}$ for the heterodyne and phase measurements.

Turning to QRC-A, let us estimate the value of Γ under heterodyne and phase measurement. For a signal energy S, the heterodyne measurement is Gaussian distributed around the transmitted amplitude with a standard deviation of $1/2$ for each quadrature while the phase measurement has an approximately Lorentzian distribution around the transmitted phase with standard deviation $\sim 1/\sqrt{S}$. If we assume that, given a certain transmitted amplitude/phase, the possible ciphertext values are uniformly distributed within a standard deviation on either side and ciphertext values outside this range are not reached (this will be called the *wedge approximation*), we get the following estimates

N_{het} and N_{phase} for the number of keystream values z_i covered by the quantum noise under heterodyne and phase measurements:

$$N_{het} = 2N_{phase} = \mu/(\pi\sqrt{S}). \tag{3.10}$$

If the value of the randomizer R is fixed (corresponding to rotation by a given angle within the wedge), Z_i is fixed by the plaintext and ciphertext. Thus we have according to Eq. (3.4) that

$$\Gamma_{het} = N_{het} - 1 \cong \mu/(\pi\sqrt{S}), \tag{3.11}$$

and that

$$\Gamma_{phase} \cong \Gamma_{het}/2 \cong \mu/(2\pi\sqrt{S}). \tag{3.12}$$

As expected, the Γ's of both measurements increase as the number of bases μ increases, and decrease with increasing signal energy S that corresponds to decreasing quantum noise. For example, using the experimental parameters in [**26**] of $S \sim 4 \times 10^4$ photons and $\mu \sim 2 \times 10^3$ has $\Gamma_{het} \sim 3$. These numbers can be directly related to bounds on the unicity distance against known-plaintext attacks with individual measurements and collective processing. See [**30**] or [**7**] for further details. In the rest of this thesis, we will be concerned with applying the results of Chapter 2 to *joint* attacks, which are more general than those that appear in the definition of a Quantum Random Cipher given here.

CHAPTER 4

Joint Attack Security Analysis of $\alpha\eta$: Results and Future Directions

A general mathematical model for $\alpha\eta$, or any other system based on the KCQ (Keyed Communication in Quantum Noise) principle (see [**1**] and [**9**] for detailed discussion of the KCQ principle in the context of both direct encryption and key generation) is the following: A pair of random variables $\mathbf{X}_n$ (the plaintext/data) and K (the secret key) together determine a quantum state $\rho^k_{\mathbf{x}_n}$ in the Hilbert space of the system. $\mathbf{X}_n$ and K are independent random variables with probability distributions $p(\mathbf{X}_n)$ and $p(K)$, the latter being assumed uniform, $p(k) = 2^{-|K|}\ \ \forall k$.

As for the case of classical private-key cryptography, one can consider the cases of known-plaintext attack (KPA) and ciphertext-only attack (CTA), implying respectively that $p(\mathbf{X}_n)$ is degenerate or uniform. In the case of KPA, Eve tries to obtain the key K with minimum error probability using a joint measurement attack. In the case of CTA, she can either attempt to get the data $\mathbf{X}_n$ with minimum error probability or she can attempt to obtain the K with minimum error probability. If the latter attack succeeds with high enough probability, she can use the obtained key to make the same measurement as Bob and undermine the system security by obtaining all future data. In this chapter, we set up the appropriate quantum detection problem for the above attacks, and indicate the kind of results obtainable by applying the upper and lower bounds developed in Chapter 2.

4.1. Key Security under Known-Plaintext Attack

The behavior of the minimum error probability $\overline{P}_e$ for attacking the key under KPA as a function of the data length n provides some understanding of how fast the cryptosystem leaks information on the key to Eve. We mentioned in Chapter 1 that a classical nonrandom cipher is broken with probability 1 under KPA at its nondegeneracy distance n_d, which is usually quite small. $\alpha\eta$, however, is a quantum random cipher, and as such, it is interesting to study how the attacker's error probability on the key varies with increasing n. In particular, it would be very significant practically if this probability remains bounded at a high enough level (i.e., an appreciable fraction of the maximum possible level $2^{-|K|}$) away from zero for arbitrary n. Unfortunately, we shall see that this is not the case in $\alpha\eta$.

Since a fixed $\mathbf{x}_n \in \mathcal{X}^n$ is known to Eve, she is faced with an M-ary detection problem between the $M = 2^{|K|}$ states $\{\rho^k_{\mathbf{x}_n}\}_{k=1}^{2^{|K|}}$ with a priori probabilities $p_k = 2^{-|K|}$. When the ENC box (see Fig. 3.3) in an $\alpha\eta$ system consists of a linear feedback shift register (LFSR), the $\rho^k_{\mathbf{x}_n}$ are pure states according to Eq. (3.7) and the pairwise inner products $|\langle\psi(\mathbf{x}_n, k)|\psi(\mathbf{x}_n, k')\rangle|$ (and hence the optimum error probability) are independent of the particular $\mathbf{x}_n \in \mathcal{X}^n$ under consideration.

When the ENC box is an LFSR, one can readily obtain simple upper and lower bounds on the pairwise inner product between any two signals ρ^k, where we may now suppress the plaintext subscript. Note that, for any plaintext length n, one may consider the LFSR as implementing a linear map from the $|K|$ possible seed keys to mn-bit keystreams $\mathbf{Z}_n$. Let us denote by $d_{\min}(n)$ and $d_{\max}(n)$ respectively the minimum and maximum non-zero (binary) Hamming weight of a keystream sequence of length n.

Using the notations of Section 3.3, for $n = |K|/m$, let us compare the keystream sequences $\mathbf{Z}_n^k$ and $\mathbf{Z}_n^{k'}$ corresponding to different keys k and k'. Since the bit length of $\mathbf{Z}_n$ is $|K|$ for this value of n and the LFSR used in the protocol is a $|K|$-stage one, at least one of the bits of $\mathbf{Z}_n^k$ is different from the corresponding one in $\mathbf{Z}_n^{k'}$. Thus, $d_{\min}(|K|/m) = 1$. Let us define $\epsilon_{\max}(n) = \max_{k \neq k'} |\langle \psi(\mathbf{x}_n, k) | \psi(\mathbf{x}_n, k') \rangle|$. From the form (3.7) of the states used, we can bound, for integers $l > 0$,

$$\epsilon_{\max}(l|K|/m) \leq e^{-2Sl \sin^2(\pi/\mu)} \simeq e^{-2Sl\pi^2/\mu^2}. \tag{4.1}$$

Using the above value for ϵ in Eq. (2.19) of Chapter 2, we get, for the n values above, under the condition that $\epsilon_{max}(l|K|/m)|K| < 1$ (this happens eventually as ϵ_{max} decays exponentially),

$$\overline{P}_e(n) \leq 1/M\{(M-1)\epsilon_{max}(n) - (1+\epsilon_{max}(n)) \ln[1-(M-1)\epsilon_{\max}(n)]\}. \tag{4.2}$$

Here, $M = 2^{|K|}$. In fact, the following approximate equation follows from the one above:

$$\overline{P}_e(n) \leq (1 - M^{-1})\epsilon^2_{\max}(n) + O(\epsilon^3_{\max}(n)). \tag{4.3}$$

Thus, as l becomes large, the upper bound goes to zero for any values of S, $|K|$ and $m = \log_2(\mu)$, and the system is broken under known-plaintext attack.

We now obtain a lower bound on $\overline{P}_e(n)$ from the Binary Detection lower bound Eq.(2.28). First, we note that $d_{\max}(n) \leq n$, and that a corresponding lower bound for $\epsilon_{\min}(n) = \min_{k \neq k'} |\langle\psi(\mathbf{x}_n, k)|\psi(\mathbf{x}_n, k')\rangle|$ is $\epsilon_{\min}(n) \geq e^{-2nS}$. Using convexity repeatedly in Eq.(2.28) and Eq.(2.31), we get

$$\overline{P}_e \geq \frac{1 - M^{-1}}{2} \epsilon^2_{\min}(n) \geq \frac{1 - M^{-1}}{2} e^{-4nS}. \tag{4.4}$$

This is a very weak bound due to the weakness of the bound on $d_{\max}(n)$. It should be possible to use properties of the LFSR output sequences to obtain better bounds on $d_{\max}(n)$ and $d_{\min}(n)$ and thereby to improve the above bounds for known-plaintext attack security.

4.2. Key Security under Ciphertext-Only Attack

Let us first set up the appropriate detection problem for attacking the key under a ciphertext only attack. If $\{M_k\}_{k=1}^{|K|}$ represents the optimal POM for this case, we may write the probability of correct decision as

$$\overline{P}_c = \sum_{\mathbf{x}_n, k} p_k p_{\mathbf{x}_n} \mathrm{tr}(M_k \rho^k_{\mathbf{x}_n}) = \sum_k p_k \mathrm{tr}(M_k (\sum_{\mathbf{x}_n} p_{\mathbf{x}_n} \rho^k_{\mathbf{x}_n}). \tag{4.5}$$

The problem is thus equivalent to a standard $M = 2^{|K|}$-ary detection between the operators $\rho_k := \sum_{\mathbf{x}_n} p_{\mathbf{x}_n} \rho^k_{\mathbf{x}_n}$ with a priori probabilities $p(k) = 2^{-|K|}$.

Let us attempt to apply the mixed state upper bound Eq. (2.22) of Section 2.2 to this case. To do so, we need to find a *uniform* upper bound on the pairwise inner products $|\langle\psi(\mathbf{x}_n, k)|\psi(\mathbf{x'}_n, k')\rangle|$. Here a complication arises when $k = k'$. It can happen that the two sequences $\mathbf{x}_n$ and $\mathbf{x'}_n$ differ by only one bit no matter how large n is. In that case, we would have $|\langle\psi(\mathbf{x}_n,k)|\psi(\mathbf{x'}_n,k')\rangle| = e^{-2S}$, which is a constant. Thus, with $L = 2^n$, the condition $[(M-1)L-1]\epsilon < 1$ needed for the validity of Eq. (2.22) would not hold and the upper bound is not applicable. In the absence of a good uniform upper bound on the inner product, we are forced to fall back upon the basic bound Eq.(2.8). According to this equation, we require estimates of the projections $\parallel P_{\overline{k}}\psi_k^{\mathbf{x}} \parallel$, where $P_{\overline{k}}$ is the projector onto the subspace spanned by $\psi(\mathbf{x},k')$ for $k' > k$ and for all $\mathbf{x}$. Following a closely parallel sequence of steps as in the case of ϵ-orthogonal signals, we find that a minimum eigenvalue estimate of a Gram matrix is again called for. Unfortunately, the Gerschgorin Theorem works in this case only under the condition

$$(1+\gamma)^n + 2^n(2^{|K|}-k-1)\epsilon(n) - 1 < 1, \tag{4.6}$$

where $\gamma = e^{-2S}$ and $\epsilon(n) = e^{-n\frac{Sm\pi^2}{2|K|\mu^2}}$. This is not satisfied except for very small values of n and S, and so our method fails to give good results. The binary detection lower bound also gives the weak result

$$\overline{P}_e \geq 2^{-n}\frac{(1+\Gamma)^n - 1}{2}, \tag{4.7}$$

for $\Gamma = e^{-4S}$.

4.3. Conclusion and Future Directions

We saw in the previous subsection that while $\alpha\eta$ can be proved to be insecure for large enough n under KPA for any values of the system parameters, no conclusion about its security or insecurity against CTA could yet be drawn. It seems that the reason for this failure are that good estimates of the pairwise inner products between the signal states of the $\alpha\eta$ system are not available. These are in turn tied to estimates of the weight distribution or, at the least, the minimum and maximum distance between codewords of the code induced by the LFSR. This direction of study seems most urgent for further results on $\alpha\eta$ security under joint attack. It is also advisable to look for methods that yield sharper eigenvalue estimates needed in applying our upper bounds than the completely general Gerschgorin Disk Theorem. The research directions indicated above are being pursued as are efforts to apply the lower and upper bounds derived in this thesis to other problems in quantum information.

References

[1] H. Yuen, KCQ: A new approach to quantum cryptography I. General principles and qumode key generation, quant-ph/0311061.

[2] J.L. Massey, Proc. IEEE, 76 (1988) 533-549.

[3] D.R. Stinson, *Cryptography: Theory and Practice*, Chapman and Hall/CRC, 3nd ed, 2006.

[4] C. Shannon, Bell Syst. Tech. J. 28 (1949) 656–715.

[5] T.M. Cover & J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[6] U. Maurer, IEEE Trans. IT, vol. 39, No. 3, 1993, pp. 733-742.

[7] R. Nair, H.P. Yuen, E. Corndorf, T. Eguchi, P. Kumar, Quantum Noise Randomized Ciphers, To appear in Phys. Rev. A.

[8] H.N. Jendahl, Y.J.B. Kuhn, J.L. Massey, Advances in Cryptology -EUROCRYPT '89, Lect. Notes in Comp. Science 434, 382-394, 1990, Springer-Verlag, Berlin.

[9] H.P. Yuen, R. Nair, E. Corndorf, G.S. Kanter, P. Kumar, quant-ph/0509091, To appear in *Quantum Information & Computation.*

[10] M. Nielsen, I. Chuang, *Quantum Computation and Quantum Information*, Cambridge Univ. Press, 2000.

[11] A. Peres, *Quantum theory: Concepts and Methods*, Kluwer Academic Pub., 1995.

[12] A. Holevo, *Statistical Structure of Quantum Theory*, Springer Verlag, 2001.

[13] M. Scully, M.S. Zubairy, *Quantum Optics*, Cambridge Univ. Press, 1997.

[14] C. Helstrom, *Quantum Detection and Estimation Theory*, Academic Press, 1976.

[15] P. Benioff, Decision procedures in Quantum mechanics, J. Math. Phys., 13 (1972) 908-915.

[16] H. Yuen, R. Kennedy, M. Lax, Optimum Testing of Multiple Hypotheses in Quantum Detection Theory, IEEE Trans. IT 21 (1975) 125-134.

[17] Y. Eldar, A. Megretski, G. Verghese, Designing optimal quantum detectors via semidefinite programming, IEEE Trans. IT 49 (2003) 1007-1012.

[18] M. Ban, K. Kurokawa, R. Momose, O. Hirota, Optimum measurements for discrimination among symmetric quantum states and parameter estimation, Int. J. Theor. Phys., vol. 36, pp. 1269-1288, 1997.

[19] Y. Eldar, G.D. Forney, On Quantum detection and the Square-Root Measurement, IEEE Trans. IT 47 (2001), pp. 858-872.

[20] Y. Eldar, A. Megretski, G. Verghese, Optimal Detection of mixed symmetric quantum states, IEEE Trans. IT 50 (2004) pp. 1198-1207.

[21] R. Bhatia, *Matrix Analysis*, Springer Verlag, New York, 1997.

[22] R. Horn, C. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

[23] H.L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, Wiley Inter-Science, 2001.

[24] G. Barbosa, E. Corndorf, P. Kumar, H. Yuen, Phys. Rev. Lett. 90 (2003) 227901.

[25] E. Corndorf, G. Barbosa, C. Liang, H. Yuen, P. Kumar, Opt. Lett. 28, 2040-2042, 2003.

[26] E. Corndorf, C. Liang, G.S. Kanter, P. Kumar, and H.P. Yuen, Phys. Rev. A 71, pp. 062326, 2005.

[27] C. Liang, G.S. Kanter, E. Corndorf, and P. Kumar, Photonics Tech. Lett. 17, pp. 1573-1575, 2005.

[28] O. Hirota, M. Sohma, M. Fuse, and K. Kato, Phys. Rev. A. 72 (2005) 022335; quant-ph/0507043.

[29] N. Gisin, G. Ribordy, W. Tittel and H. Zbinden, Rev. Mod. Phys. 74, 145 (2002).

[30] T. Eguchi, M.S. Thesis, Northwestern University, Jun 2006.