Cracking Quantum Key Distribution: Basis Estimation and Optimal Measurements

Matthew Thill*, Sam Dolinar*, and Dariush Divsalar*

ABSTRACT. — In this report, we examine vulnerabilities in quantum key distribution (QKD) schemes that may arise in practice. In particular, we assume a probability bias in the distribution of states and bases used in a BB84-like protocol, and show how this can allow an eavesdropper to estimate the states used in the protocol and to correspondingly design a positive-operator valued measure (POVM) to estimate secret key bits for use in an intercept-relay attack. We quantify the error in the state estimation and the probability of correct key-bit detection with respect to the probability bias.

I. Introduction

Today's world sees a constant demand for large amounts of sensitive data, which must be communicated securely and protected from malicious eavesdroppers. An optimal approach to encrypting data sent over a classical channel is to use a one-time pad (OTP), originally discovered by Miller in 1882 [1], patented by Gilbert Vernam in 1919 [2] and proven information-theoretically secure by Shannon in 1949 [3]. The OTP requires a secret key of bit-length equal to that of the message communicated. Thus, there is an inherent demand for efficient protocols to establish secret key bits between sender and receiver with which an eavesdropper shares no mutual information.

Quantum key distribution (QKD) (thoroughly reviewed in [4]) exploits the principles of quantum science to produce these secret bits. Two parties who share access to a quantum channel, and typically an authenticated classical channel, attempt to establish a secret key by exchanging, sharing, and measuring quantum states transmitted over the channel. Any quantum correlations that are lost during the communication translate directly to information leaked to the environment (or an eavesdropper), allowing the

*Communications Architectures and Research Section

The research described in this publication was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

two parties "Alice" and "Bob" to detect the eavesdropper "Eve" if she gains too much information about their established secret key. The authenticated classical channel allows Alice and Bob to estimate this information loss. Furthermore, the *no-cloning theorem* of quantum mechanics [5] ensures that any copy (or partial copy) that Eve attempts to make of Alice and Bob's shared state will perturb or damage the original, again risking exposing Eve's presence. In short, an effective eavesdropper needs to be able to ascertain as much information as possible with few measurements.

The most popular QKD protocols, such as BB84 [6] (see Section II), involve Alice and Bob transmitting and receiving one of several known states over the quantum channel, and performing quantum measurements catered to these states. If Eve does not know these states, she needs a practical method to learn them using as few measurements as possible in order to have any hope of learning Alice and Bob's secret key bits without being detected. The process by which Eve learns Alice and Bob's states is a version of *quantum tomography* [7]. Past research has explored how to learn a quantum state with a small number of measurements by exploiting techniques from fields such as compressive sensing [8]. Learning the states is only the first step for Eve, who must use this knowledge to inform how to best eavesdrop on the quantum channel and perform measurements to learn Alice and Bob's secret key bits. We will consider how Eve can perform these tasks by exploiting imperfections, which could easily arise in practice in Alice and Bob's QKD scheme. In particular, we will quantify the accuracy with which Eve can estimate Alice and Bob's state as a function of the number of measurements she is able to make, and how she can construct an optimal eavesdropping POVM based on this estimate. An important point to mention, however, is that the more measurements Eve makes, the more she perturbs Alice and Bob's shared state and increases her likelihood of being detected. For the time being, we will not consider this in our derivations, though realistically, it would affect how much information Eve can ascertain about Alice and Bob's secret key.

II. The BB84 Protocol

The first practical QKD protocol was introduced by Bennett and Brassard in 1984 [6], and as such is referred to as "BB84." In this method, Alice establishes a secure key with Bob by transmitting photons that she prepares in one of two orthonormal polarization bases: either the horizontal/vertical polarizations, $\{|H\rangle, |V\rangle\}$, or the rotated polarizations of ± 45 degrees, $\{|+45\rangle, |-45\rangle\}$. Here, we use the bra-ket notation [9] with the convention that "$|\psi\rangle$" represents a column vector corresponding to a pure state ψ, and "$\langle\psi|$" is its conjugate transpose (a row vector). Explicitly, $|\pm 45\rangle = \frac{1}{\sqrt{2}}(|H\rangle \pm |V\rangle)$. Alice randomly selects one of these two bases as well as a bit: 0 (corresponding to $|H\rangle$ or $|+45\rangle$) or 1 (corresponding to $|V\rangle$ or $|-45\rangle$). Bob randomly chooses one of the two bases in which to measure the received photon, and interprets the corresponding bit as a 0 or a 1. If Alice and Bob used the same basis and the photon is unperturbed, Bob

will have correctly interpreted Alice's bit. If they have used opposite bases, Bob has a 50% chance of having interpreted Alice's bit correctly. Alice and Bob use an authenticated classical channel to determine when their bases were compatible, a process called "sifting," and in doing so establish a secret key.

An eavesdropper, Eve, seeks to learn as much about the secret key bits as possible. There are a number of attacks she can perform, the simplest of which is an intercept-resend attack [4], in which she chooses one of Alice and Bob's bases to measure in and relays the resulting state to Bob. With luck, Alice and Bob will have both used the same basis as Eve, in which case she will know their established bit. Alice and Bob can choose to share some of their bits over the classical channel to confirm that they match and to perform *information reconciliation* to correct bit-errors, such as with the cascade protocol [10], and *privacy amplification* [11] to minimize information leaked to Eve, though this will shrink the size of their secret key. If Eve has perturbed the photons too much, Alice and Bob may choose to abort the protocol altogether, so it behooves Eve to remain undetected.

In the following sections, we consider a somewhat more complicated regime than BB84 in which Alice and Bob have not restricted themselves to using only the $\{|H\rangle, |V\rangle\}$ and $\{|+45\rangle, |-45\rangle\}$ bases. We assume the role of an eavesdropper, Eve, with no a priori knowledge of the states or the bases in which Alice and Bob are communicating, and we examine how quickly Eve can learn these states using just simple projective measurements provided that Alice and Bob have some small bias in the distribution of their transmitted states. We begin with the single-basis case in Section III, where Alice sends bits to Bob only encoded in the '0' or '1' element of one orthonormal basis. We then extend our results in Section IV to a two-basis communication scheme such as BB84. In Section IV-A, we show how Eve can further design a positive-operator valued measure (POVM) to estimate which bit Alice has transmitted with as high a probability of correct detection as possible.

The BB84 protocol as described here is an example of a *prepare-and-measure* QKD scheme where Alice prepares a state on which Bob performs a measurement. It is worth noting that there are alternative ways to implement these protocols using entangled photon sources, such as BBM [12], the entanglement-based version of BB84. Our results can be easily adapted to these entanglement-based alternatives, but for the sake of simplicity, we will pose our discussion in the context of prepare-and-measure protocols.

III. Single Basis Estimation

We first consider the following simple scenario: Alice selects an element from the orthonormal basis $\mathcal{A} = \{|a_0\rangle, |a_1\rangle\} \in \mathbb{C}^{2\times 1}$, which she sends to Bob over a noiseless channel. She selects element $|a_i\rangle$ with probability p_i, $i = 0, 1$, and she repeats this process many times. This gives rise to Alice's expected density operator of $\rho =$

$p_0|a_0\rangle\langle a_0| + p_1|a_1\rangle\langle a_1|$, which is simply the diagonal matrix $\rho = \begin{bmatrix} p_0 & 0 \\ 0 & p_1 \end{bmatrix}$ when expressed with respect to the basis $\mathcal{A}$.

An eavesdropper, Eve, would like to determine Alice's basis $\mathcal{A}$ and the probabilities of its two elements, perhaps with an eye toward performing intercept/resend attacks. We note that the probabilities p_0 and p_1 are simply the singular values of ρ, and provided they are not exactly equal to each other (which is the case if Alice lacks a perfectly unbiased random number generator), their associated singular vectors are unique, and equal to the basis elements $|a_0\rangle$ and $|a_1\rangle$. In this case, Eve needs only to determine the mixed state ρ with respect to *some* basis, and perform a singular value decomposition. In the case where p_0 and p_1 are each exactly equal to $\frac{1}{2}$, then ρ is the identity operator, and Eve will not be able to learn the basis elements $|a_0\rangle$ and $|a_1\rangle$ using standard independent projective measurements, as we consider in the next section.

If the distribution is biased so that $p_0 \neq p_1$, then Eve has a hope of recovering the basis elements and their probabilities. To do this, she chooses her own orthonormal basis $\mathcal{E} = \{|e_0\rangle, |e_1\rangle\}$ with respect to which she will express ρ. Note that in this basis, ρ will have the matrix form $\begin{bmatrix} \langle e_0|\rho|e_0\rangle & \langle e_0|\rho|e_1\rangle \\ \langle e_1|\rho|e_0\rangle & \langle e_1|\rho|e_1\rangle \end{bmatrix} = \begin{bmatrix} |e_0\rangle & |e_1\rangle \end{bmatrix}^H \rho \begin{bmatrix} |e_0\rangle & |e_1\rangle \end{bmatrix}$. Since this matrix must be Hermitian with trace equal to 1, we may express it in the form $\begin{bmatrix} x & y+iz \\ y-iz & 1-x \end{bmatrix}$, with x, y, and z real. Eve is now tasked with estimating these three parameters. It is well-known that a density operator can be determined with three types of projective measurements on the Bloch (or Poincaré) Sphere [7, 13]. In the following section, we examine how Eve can optimize her protocol by using an arbitrary number of projective measurement types to estimate x, y, and z. Since Eve would like to avoid being detected, we focus on the tradeoff between the estimation error and the total number of measurements performed.

A. Estimation Using Projective Measurements

We first consider the regime in which Eve performs a series of projective measurements to estimate x, y, and z. Each measurement will amount to Eve intercepting a state from Alice and projecting it onto some orthonormal basis, which without loss of generality can be expressed in the form $\{\mathbf{U}_i|e_0\rangle, \mathbf{U}_i|e_1\rangle\}$, $i = 1, ..., N_T$. Here, each $\mathbf{U}_i$ is a unitary matrix and N_T is the number of different types of projective measurements that Eve chooses between. Eve performs n_i measurements with basis type i and records the empirical fraction X_i of instances in which she measures $\mathbf{U}_i|e_0\rangle$. We can express this vector as a linear combination of the original basis elements in $\mathcal{E}$, writing $\mathbf{U}_i|e_0\rangle = \alpha_i|e_0\rangle + \beta_i|e_1\rangle$, where $|\alpha_i|^2 + |\beta_i|^2 = 1$. In this form, we may express the expected value

of X_i as

$$\begin{aligned} E[X_i \mid \rho, \mathcal{E}, \mathbf{U}_i] &= \langle e_0|\mathbf{U}_i^H \rho \mathbf{U}_i|e_0\rangle \\ &= |\alpha_i|^2\langle e_0|\rho|e_0\rangle + \alpha_i^*\beta_i\langle e_0|\rho|e_1\rangle + \beta_i^*\alpha_i\langle e_1|\rho|e_0\rangle + |\beta_i|^2\langle e_1|\rho|e_1\rangle. \quad (1) \end{aligned}$$

Using the relations $\langle e_0|\rho|e_0\rangle = x$, $\langle e_1|\rho|e_1\rangle = 1 - x$, $\langle e_0|\rho|e_1\rangle = y + iz$, and $\langle e_1|\rho|e_0\rangle = y - iz$, we can rewrite all these equations in the matrix form

$$\begin{aligned} &\begin{bmatrix} E[X_1 \mid \rho, \mathcal{E}, \mathbf{U}_1] + |\alpha_1|^2 - 1 \\ \vdots \\ E[X_{N_T} \mid \rho, \mathcal{E}, \mathbf{U}_{N_T}] + |\alpha_{N_T}|^2 - 1 \end{bmatrix} \\ &= \begin{bmatrix} |\alpha_1|^2 - |\beta_1|^2 & 2\,\mathrm{Re}(\alpha_1^*\beta_1) & -2\,\mathrm{Im}(\alpha_1^*\beta_1) \\ \vdots & \vdots & \vdots \\ |\alpha_{N_T}|^2 - |\beta_{N_T}|^2 & 2\,\mathrm{Re}(\alpha_{N_T}^*\beta_{N_T}) & -2\,\mathrm{Im}(\alpha_{N_T}^*\beta_{N_T}) \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \quad (2) \end{aligned}$$

We will express this compactly as $\mathbf{w} = \mathbf{A}\mathbf{v}$, where

$$\mathbf{w} := \begin{bmatrix} E[X_1 \mid \rho, \mathcal{E}, \mathbf{U}_1] + |\alpha_1|^2 - 1 \\ \vdots \\ E[X_{N_T} \mid \rho, \mathcal{E}, \mathbf{U}_{N_T}] + |\alpha_{N_T}|^2 - 1 \end{bmatrix}, \quad \mathbf{v} := \begin{bmatrix} x \\ y \\ z \end{bmatrix},$$

$$\mathbf{A} := \begin{bmatrix} |\alpha_1|^2 - |\beta_1|^2 & 2\,\mathrm{Re}(\alpha_1^*\beta_1) & -2\,\mathrm{Im}(\alpha_1^*\beta_1) \\ \vdots & \vdots & \vdots \\ |\alpha_{N_T}|^2 - |\beta_{N_T}|^2 & 2\,\mathrm{Re}(\alpha_{N_T}^*\beta_{N_T}) & -2\,\mathrm{Im}(\alpha_{N_T}^*\beta_{N_T}) \end{bmatrix}.$$

At this point, it becomes clear that we need at least $N_T = 3$ distinct measurement types to unambiguously recover $\mathbf{v}$, but we would like to quantify the error in our estimation with respect to both N_T and the number of projective measurements n_i performed in each basis. To this end, let us fix a constant number of measurements per basis: $n_i = M$ for $i = 1, ..., N_T$. In this case, each X_i corresponds to the fraction of the M measurements whose outcome is the '0' basis element. If we let $\hat{\mathbf{w}}$ denote the vector $\left[X_1 + |\alpha_1|^2 - 1, \quad \ldots, \quad X_{N_T} + |\alpha_{N_T}|^2 - 1\right]^T$, an estimate for $\mathbf{w}$, then we could estimate $\mathbf{v}$ as

$$\hat{\mathbf{v}} = \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} = \mathbf{A}^+\hat{\mathbf{w}},$$

where $\mathbf{A}^+$ denotes the Moore-Penrose pseudoinverse of $\mathbf{A}$. An estimate for the density operator ρ is then reconstructed as $\hat{\rho} = \begin{bmatrix} \hat{x} & \hat{y} + i\hat{z} \\ \hat{y} - i\hat{z} & 1 - \hat{x} \end{bmatrix}$. As previously mentioned, the estimated basis element probabilities $\hat{p}_0$ and $\hat{p}_1$ are the eigenvalues of $\hat{\rho}$, which can be expressed in terms of $\hat{x}$, $\hat{y}$, and $\hat{z}$ as

$$\hat{p}_0, \hat{p}_1 = \frac{1}{2}\left(1 \pm \sqrt{1 - 4[\hat{x}(1-\hat{x}) - \hat{y}^2 - \hat{z}^2]}\right). \quad (3)$$

The estimated basis elements are the normalized eigenvectors associated with $\hat{p}_0$ and $\hat{p}_1$, which are

$$|\hat{a}_k\rangle = \begin{bmatrix} -(\hat{y} + i\hat{z}) \\ \hat{x} - \hat{p}_k \end{bmatrix} / \sqrt{\hat{y}^2 + \hat{z}^2 + (\hat{x} - \hat{p}_k)^2}, \qquad k = 0, 1. \tag{4}$$

The following theorem characterizes the first and second order statistics of $\hat{\mathbf{v}}$:

Theorem 1. *For $N_T \geq 3$, the estimator $\hat{\mathbf{v}} = \mathbf{A}^+\hat{\mathbf{w}}$ is an unbiased estimator for $\mathbf{v}$ (that is, $E[\hat{\mathbf{v}} \mid \rho, \mathcal{E}, \{\mathbf{U}_i\}] = \mathbf{v}$). If the basis $\mathcal{E}$ is chosen uniformly at random on the complex unit sphere, then the expected covariance matrix of $\hat{\mathbf{v}}$ conditioned on $\{\alpha_i\}$, $\{\beta_i\}$, and ρ is $\mathbf{R}_{\rho,\{\alpha_i\},\{\beta_i\}} = \frac{1}{M}\left(\frac{1}{6} + \frac{1}{3}p_0p_1\right)\left(\mathbf{A}^T\mathbf{A}\right)^{-1}$, where M is the number of projective measurements per basis type.*

Proof. The fact that $\hat{\mathbf{v}}$ is unbiased follows from the fact that $\hat{\mathbf{w}}$ is clearly an unbiased estimator for $\mathbf{w}$, so

$$E[\hat{\mathbf{v}} \mid \rho, \mathcal{E}, \{\mathbf{U}_i\}] = \mathbf{A}^+ E[\hat{\mathbf{w}} \mid \rho, \mathcal{E}, \{\mathbf{U}_i\}] = \mathbf{A}^+\mathbf{w} = \mathbf{v}.$$

By requiring that $\mathcal{E}$ be uniformly randomly chosen on the complex unit sphere, we in particular demand that $|e_0\rangle \in \mathbb{C}^{2\times 1}$ have a distribution equivalent to selecting the real and imaginary components as a random real vector $\begin{bmatrix} \mathrm{Re}\left(|e_0\rangle\right) \\ \mathrm{Im}\left(|e_0\rangle\right) \end{bmatrix} \in \mathbb{R}^{4\times 1}$ by selecting a vector $\mathbf{n}$ according to the multivariate normal distribution $\mathbf{n} \sim \mathcal{N}(\mathbf{0}_{4\times 1}, \mathbf{I}_{4\times 4})$, and then setting $\begin{bmatrix} \mathrm{Re}\left(|e_0\rangle\right) \\ \mathrm{Im}\left(|e_0\rangle\right) \end{bmatrix} = \dfrac{\mathbf{n}}{||\mathbf{n}||_2}$.

Note that the X_i are independent binomial random variables scaled by M, with the distribution $M \cdot X_i \sim B\left(M, \langle e_0|\mathbf{U}_i^H\rho\mathbf{U}_i|e_0\rangle\right)$.

By decomposing $\mathbf{U}_i|e_0\rangle$ in the basis $\mathcal{A}$, we can write

$$\langle e_0|\mathbf{U}_i^H\rho\mathbf{U}_i|e_0\rangle = p_1 + (p_0 - p_1)\left|\langle a_0|\mathbf{U}_i|e_0\rangle\right|^2. \tag{5}$$

The statistics of the squared inner product can be derived by defining $|a_{0,i}\rangle := \mathbf{U}_i^H|a_0\rangle$, and for convenience, expressing this as a real unit vector $\mathbf{a}_{0,i} := \begin{bmatrix} \mathrm{Re}\left(|a_{0,i}\rangle\right) \\ \mathrm{Im}\left(|a_{0,i}\rangle\right) \end{bmatrix}$. We also express $|e_0\rangle$ as the real vector $\mathbf{e}_0 := \begin{bmatrix} \mathrm{Re}\left(|e_0\rangle\right) \\ \mathrm{Im}\left(|e_0\rangle\right) \end{bmatrix}$. We then note that

$$\begin{aligned} \left|\langle a_0|\mathbf{U}_i|e_0\rangle\right|^2 = \left|\langle a_{0,i}|e_0\rangle\right|^2 &= \mathbf{e}_0^T\left(\mathbf{a}_{0,i}\mathbf{a}_{0,i}^T + \mathbf{M}\mathbf{a}_{0,i}\mathbf{a}_{0,i}^T\mathbf{M}^T\right)\mathbf{e}_0 &(6)\\ &= \mathbf{e}_0^T\mathbf{\Phi}\mathbf{e}_0, &(7)\end{aligned}$$

where $\mathbf{M} := \begin{bmatrix} \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} \\ -\mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} \end{bmatrix}$ and $\mathbf{\Phi} := \mathbf{a}_{0,i}\mathbf{a}_{0,i}^T + \mathbf{M}\mathbf{a}_{0,i}\mathbf{a}_{0,i}^T\mathbf{M}^T$ is the projection onto the space spanned by $\{\mathbf{a}_{0,i}, \mathbf{M}\mathbf{a}_{0,i}\}$, which is quickly verified to be a 2-dimensional orthonormal set.

Since $\mathbf{e}_0 \in \mathbb{R}^{4\times 1}$ is uniformly distributed on the unit sphere, then $\mathbf{e}_0^T \mathbf{\Phi} \mathbf{e}_0$ will be distributed as $\frac{Z_1^2+Z_2^2}{Z_1^2+Z_2^2+Z_3^2+Z_4^2}$, where $Z_i \sim \mathcal{N}(0,1)$ for each i. Since $Z_1^2 + Z_2^2 \sim \chi^2(2)$ and $Z_3^2 + Z_4^2 \sim \chi^2(2)$, we have that $\mathbf{e}_0^T \mathbf{\Phi} \mathbf{e}_0 \sim \text{Beta}(1,1)$, which has mean $1/2$ and variance $1/12$, making its first two moments

$$E[\mathbf{e}_0^T \mathbf{\Phi} \mathbf{e}_0] = \frac{1}{2}, \tag{8}$$

$$E\left[\left(\mathbf{e}_0^T \mathbf{\Phi} \mathbf{e}_0\right)^2\right] = \frac{1}{12} + \left(\frac{1}{2}\right)^2 = \frac{1}{3}. \tag{9}$$

Now, returning to the statistics of X_i, we have that

$$E[X_i \mid \rho, |e_0\rangle, \mathbf{U}_i] = p_1 + (p_0 - p_1)\mathbf{e}_0^T \mathbf{\Phi} \mathbf{e}_0, \tag{10}$$

$$\text{var}\left[X_i \mid \rho, |e_0\rangle, \mathbf{U}_i\right] = \frac{1}{M}\left(\langle e_0|\mathbf{U}_i^H \rho \mathbf{U}_i|e_0\rangle \left(1 - \langle e_0|\mathbf{U}_i^H \rho \mathbf{U}_i|e_0\rangle\right)\right) \tag{11}$$

$$= \frac{1}{M}\left(p_1 + (p_0 - p_1)\mathbf{e}_0^T \mathbf{\Phi} \mathbf{e}_0\right)\left(p_0 + (p_1 - p_0)\mathbf{e}_0^T \mathbf{\Phi} \mathbf{e}_0\right) \tag{12}$$

$$= \frac{1}{M}\left(p_0 p_1 + (p_0 - p_1)^2 \mathbf{e}_0^T \mathbf{\Phi} \mathbf{e}_0 - (p_0 - p_1)^2 \left(\mathbf{e}_0^T \mathbf{\Phi} \mathbf{e}_0\right)^2\right). \tag{13}$$

Averaging these over our uniform distribution on the basis $\mathcal{E}$, we obtain

$$E[X_i \mid \rho, \alpha_i, \beta_i] = p_1 + (p_0 - p_1)/2 \tag{14}$$

$$= \frac{1}{2}, \tag{15}$$

$$\text{var}\left[X_i \mid \rho, \alpha_i, \beta_i\right] = \frac{1}{M}\left(p_0 p_1 + (p_0 - p_1)^2/2 - (p_0 - p_1)^2/3\right) \tag{16}$$

$$= \frac{1}{M}\left(p_0 p_1 + (p_0 - p_1)^2/6\right) \tag{17}$$

$$= \frac{1}{M}\left(\frac{1}{6} + \frac{1}{3} p_0 p_1\right). \tag{18}$$

Both of these are independent of α_i and β_i, so to be concise we may just consider expectations conditioned on ρ.

We can now derive the conditional expected covariance matrix as

$$\begin{aligned}
\mathbf{R}_{\rho,\{\alpha_i\},\{\beta_i\}} &= E\left[(\hat{\mathbf{v}} - \mathbf{v})(\hat{\mathbf{v}} - \mathbf{v})^T \mid \rho, \{\alpha_i\}, \{\beta_i\}\right] \\
&= \mathbf{A}^+ E\left[(\hat{\mathbf{w}} - \mathbf{w})(\hat{\mathbf{w}} - \mathbf{w})^H \mid \rho, \{\alpha_i\}, \{\beta_i\}\right] \left(\mathbf{A}^+\right)^T \\
&= \mathbf{A}^+ E\left[(\mathbf{X} - E[\mathbf{X} \mid \rho])(\mathbf{X} - E[\mathbf{X} \mid \rho])^T \mid \rho\right] \left(\mathbf{A}^+\right)^T,
\end{aligned}$$

where $\mathbf{X} := \begin{bmatrix} X_1, & ..., & X_{N_T} \end{bmatrix}^T$. Since the X_i are independent,
$E\left[(\mathbf{X} - E[\mathbf{X} \mid \rho])(\mathbf{X} - E[\mathbf{X} \mid \rho])^T \mid \rho\right]$ will be a diagonal matrix, with diagonal entries given by $\text{var}\left[X_i \mid \rho\right]$. It follows that

$$\begin{aligned}
\mathbf{R}_{\rho,\{\alpha_i\},\{\beta_i\}} &= \frac{1}{M}\left(\frac{1}{6} + \frac{1}{3} p_0 p_1\right) \mathbf{A}^+ \left(\mathbf{A}^+\right)^T \\
&= \frac{1}{M}\left(\frac{1}{6} + \frac{1}{3} p_0 p_1\right) \left(\mathbf{A}^T \mathbf{A}\right)^{-1}.
\end{aligned} \tag{19}$$

Corollary 1. *The conditional expected covariance matrix satisfies*

$$\frac{1}{6M}\left(\mathbf{A}^T\mathbf{A}\right)^{-1} \preccurlyeq \mathbf{R}_{\rho,\{\alpha_i\},\{\beta_i\}} \preccurlyeq \frac{1}{4M}\left(\mathbf{A}^T\mathbf{A}\right)^{-1},$$

where the lower bound is achieved if $p_0 = 0$ or 1 and the upper bound is achieved when $p_0 = p_1 = \frac{1}{2}$. In particular, if ρ is chosen according to a distribution such that $E[p_0] = E[p_1] = \frac{1}{2}$, then if we average over ρ, the expected covariance matrix is

$$\mathbf{R}_{\{\alpha_i\},\{\beta_i\}} = \frac{1}{4M}\left(1 - \frac{4}{3}\mathrm{var}[p_0]\right)\left(\mathbf{A}^T\mathbf{A}\right)^{-1}.$$

Proof. This follows directly from the fact that $(\mathbf{A}^T\mathbf{A})^{-1}$ is positive semidefinite, and the fact that $\frac{1}{6} \leq \left(\frac{1}{6} + \frac{1}{3}p_0p_1\right) \leq \frac{1}{4}$ as p_0 varies over the interval $[0, 1]$ and $p_1 = 1 - p_0$. The lower bound is achieved when $p_0 = 0$ or 1 and the upper bound is achieved when $p_0 = 1/2$.

In general, we will be considering scenarios in which p_0 and p_1 are both nearly $1/2$, and in this case, we should expect the conditional covariance matrix to nearly achieve its upper bound:

$$\mathbf{R}_{\rho,\{\alpha_i\},\{\beta_i\}} \approx \frac{1}{4M}\left(\mathbf{A}^T\mathbf{A}\right)^{-1} \text{ for } p_0 \approx p_1 \approx \frac{1}{2}. \tag{20}$$

B. Estimation Error

We can analyze the error in our estimate for the density operator ρ by noting that

$$\hat{\rho} - \rho = \begin{bmatrix} \Delta x & \Delta y + i\Delta z \\ \Delta y - i\Delta z & -\Delta x \end{bmatrix}, \tag{21}$$

where $\Delta x = \hat{x} - x$, $\Delta y = \hat{y} - y$, and $\Delta z = \hat{z} - z$. Let us consider the trace norm (nuclear norm) of our estimation error,

$$||\hat{\rho} - \rho||_1 := \mathrm{Tr}\left(\sqrt{(\hat{\rho} - \rho)^H(\hat{\rho} - \rho)}\right) = 2\sqrt{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2} = 2||\hat{\mathbf{v}} - \mathbf{v}||_2. \tag{22}$$

Corollary 2. *The expected value of the estimation error trace norm is*

$$E(||\hat{\rho} - \rho||_1^2 \mid \rho, \{\alpha_i\}, \{\beta_i\}) = \frac{4}{M}\left(\frac{1}{6} + \frac{1}{3}p_0p_1\right) Tr\left((\mathbf{A}^T\mathbf{A})^{-1}\right),$$

which can take any value in the interval $\left[\frac{2}{3}\frac{1}{M} Tr\left((\mathbf{A}^T\mathbf{A})^{-1}\right), \frac{1}{M} Tr\left((\mathbf{A}^T\mathbf{A})^{-1}\right)\right]$, and is roughly equal to its upper bound when $p_0 \approx p_1 \approx \frac{1}{2}$.

Proof. We can use Theorem 1 to find the expected error as

$$
\begin{aligned}
&E(||\hat{\rho}-\rho||_1^2 \mid \rho,\{\alpha_i\},\{\beta_i\}) \\
&\quad = 4\left(\mathrm{var}(\hat{x} \mid \rho,\{\alpha_i\},\{\beta_i\}) + \mathrm{var}(\hat{y} \mid \rho,\{\alpha_i\},\{\beta_i\}) + \mathrm{var}(\hat{z} \mid \rho,\{\alpha_i\},\{\beta_i\})\right) && (23)\\
&\quad = \frac{4}{M}\left(\frac{1}{6}+\frac{1}{3}p_0p_1\right)\mathrm{Tr}\left((\mathbf{A}^T\mathbf{A})^{-1}\right) && (24)\\
&\quad \approx \frac{1}{M}\mathrm{Tr}\left((\mathbf{A}^T\mathbf{A})^{-1}\right) \quad \left(\text{for } p_0 \approx p_1 \approx \frac{1}{2}\right). && (25)
\end{aligned}
$$

The upper and lower bounds on the expected error come from varying p_0 between 0 and 1 (keeping $p_0 + p_1 = 1$) in Equation (24).

In this form, it becomes apparent that the best strategy to minimize Eve's estimation error of ρ is for her to minimize $\mathrm{Tr}\left((\mathbf{A}^T\mathbf{A})^{-1}\right)$. To this end, we state the following fact:

Lemma 1. *For any choice of the N_T unitary matrices $\mathbf{U}_i$ used to construct the matrix $\mathbf{A}$, we have $Tr\left(\mathbf{A}^T\mathbf{A}\right) = N_T$.*

Proof. This follows from essentially a direct calculation. Considering each pair (α_i, β_i) associated to the matrix $\mathbf{U}_i$, we can write $\alpha_i = |\alpha_i|e^{i\arg(\alpha_i)}$ and $\beta_i = |\beta_i|e^{i\arg(\beta_i)}$. Noting that $|\alpha_i|^2 + |\beta_i|^2 = 1$, and defining $\varphi_i := \arg(\beta_i) - \arg(\alpha_i)$, we can express $\mathbf{A}$ in the form

$$
\mathbf{A} = \begin{bmatrix}
2|\alpha_1|^2 - 1 & 2|\alpha_1|\sqrt{1-|\alpha_1|^2}\cos\varphi_1 & -2|\alpha_1|\sqrt{1-|\alpha_1|^2}\sin\varphi_1 \\
\vdots & \vdots & \vdots \\
2|\alpha_{N_T}|^2 - 1 & 2|\alpha_{N_T}|\sqrt{1-|\alpha_{N_T}|^2}\cos\varphi_{N_T} & -2|\alpha_{N_T}|\sqrt{1-|\alpha_{N_T}|^2}\sin\varphi_{N_T}
\end{bmatrix}. \tag{26}
$$

It follows that the diagonal elements of $\mathbf{A}^T\mathbf{A}$ are

$$
\begin{aligned}
&\mathrm{diag}(\mathbf{A}^T\mathbf{A}) \\
&\quad = \left(\sum_i (2|\alpha_i|^2-1)^2,\ \sum_i 4|\alpha_i|^2(1-|\alpha_i|^2)\cos^2\varphi_i,\ \sum_i 4|\alpha_i|^2(1-|\alpha_i|^2)\sin^2\varphi_i\right).
\end{aligned}
$$

We can thus compute

$$
\mathrm{Tr}\left(\mathbf{A}^T\mathbf{A}\right) = \sum_i (2|\alpha_i|^2-1)^2 + 4|\alpha_i|^2(1-|\alpha_i|^2) = \sum_i 1 = N_T. \tag{27}
$$

As a result, we can derive a lower bound on $\mathrm{Tr}\left((\mathbf{A}^T\mathbf{A})^{-1}\right)$, which we state in the following theorem:

Theorem 2. *For any choice of the N_T unitary rotations $\{\mathbf{U}_i\}$, the resulting matrix $\mathbf{A}$ satisfies $Tr\left((\mathbf{A}^T\mathbf{A})^{-1}\right) \geq \frac{9}{N_T}$. As a result, we find that for p_0 and p_1 close to $\frac{1}{2}$, $E(||\hat{\rho}-\rho||_1^2 \mid \rho,\{\alpha_i\},\{\beta_i\})$ is approximately lower bounded by $\frac{9}{M\cdot N_T}$.*

Proof. This can be concisely shown with the Cauchy-Schwartz inequality applied to the Frobenius inner product and norm: Setting $\mathbf{M} = \sqrt{\mathbf{A}^T\mathbf{A}}$, we have

$$\left|\langle \mathbf{M}, \mathbf{M}^{-1}\rangle_F\right| \leq ||\mathbf{M}||_F \cdot ||\mathbf{M}^{-1}||_F, \tag{28}$$

where $\langle \mathbf{B}, \mathbf{C}\rangle_F = \mathrm{Tr}\left(\mathbf{B}^H\mathbf{C}\right)$ and $||\mathbf{B}||_F = \sqrt{\langle \mathbf{B}, \mathbf{B}\rangle_F}$. Since $\mathbf{M}$ is real and symmetric, we have that $\left|\langle \mathbf{M}, \mathbf{M}^{-1}\rangle_F\right| = \mathrm{Tr}\left(\mathbf{I}_{3\times 3}\right) = 3$, $||\mathbf{M}||_F = \sqrt{\mathrm{Tr}\left(\mathbf{A}^T\mathbf{A}\right)} = \sqrt{N_T}$, and $||\mathbf{M}||_F = \sqrt{\mathrm{Tr}\left((\mathbf{A}^T\mathbf{A})^{-1}\right)}$, yielding the inequality. The rest follows from Equation (25).

Since it is convenient to express our error bounds in terms of the matrix $\mathbf{A}$, whose components are functions of the sets $\{\alpha_i\}$ and $\{\beta_i\}$, we briefly describe how to design the measurement matrices $\mathbf{U}_i$, which will give rise to $\mathbf{A}$. Assuming the reference basis $\mathcal{E} = \{|\mathbf{e}_0\rangle, |\mathbf{e}_1\rangle\}$ has been selected, we may simply set

$$\mathbf{U}_i = \alpha_i|\mathbf{e}_0\rangle\langle\mathbf{e}_0| + e^{i\gamma_i}\beta_i^*|\mathbf{e}_0\rangle\langle\mathbf{e}_1| + \beta_i|\mathbf{e}_1\rangle\langle\mathbf{e}_0| - e^{i\gamma_i}\alpha_i^*|\mathbf{e}_1\rangle\langle\mathbf{e}_1|, \tag{29}$$

for any $\gamma_i \in [0, 2\pi)$.

C. Minimizing Error in $\hat{\rho}$

We now discuss a practical construction of a set of projective measurements that comes close to the lower bound in Theorem 2. We first define the equally spaced angles

$$\varphi_k := \frac{2\pi(k-1)}{N_T},\ k = 1, ..., N_T. \tag{30}$$

To define our projective measurement types, it is enough to specify α_k and β_k for each $k = 1, ..., N_T$. Notice that, from Equation (26), we may without loss of generality take the α_k to be real. Once they are fixed, β_k is completely determined by α_k and φ_k. To this end, notice that if we set

$$\begin{aligned}\alpha_k &:= \pm\sqrt{\frac{2+\sqrt{2}}{4}},\\ \beta_k &:= \left(\sqrt{1-\alpha_k^2}\right)e^{i\varphi_k},\ k = 1, ..., N_T,\end{aligned} \tag{31}$$

then the expression for the matrix $\mathbf{A}$ in Equation (26) reduces to the much simpler form

$$\mathbf{A} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & \cos\varphi_1 & -\sin\varphi_1 \\ \vdots & \vdots & \vdots \\ 1 & \cos\varphi_{N_T} & -\sin\varphi_{N_T}\end{bmatrix}. \tag{32}$$

Then considering the matrix $\mathbf{A}^T\mathbf{A}$, the off-diagonal terms will be either $\frac{1}{2}\sum_k \cos\varphi_k$, $-\frac{1}{2}\sum_k \sin\varphi_k$, or $-\frac{1}{2}\sum_k \sin\varphi_k\cos\varphi_k = -\frac{1}{4}\sum_k \sin(2\varphi_k)$. Since the φ_k are the arguments of the $(N_T)^{th}$ roots of unity, we have

$$\sum_k \cos\varphi_k = \sum_k \sin\varphi_k = 0.$$

If N_T is odd, then the ordered set $(2\varphi_k)_{k=1}^{N_T}$ is a permutation of the angles $\{\varphi_k\}$ modulo 2π. If N_T is even, then $(2\varphi_k)_{k=1}^{N_T}$ contains each of the arguments of the $(N_T/2)^{th}$ roots of unity with multiplicity 2. In either case, we have

$$\sum_k \sin(2\varphi_k) = 0.$$

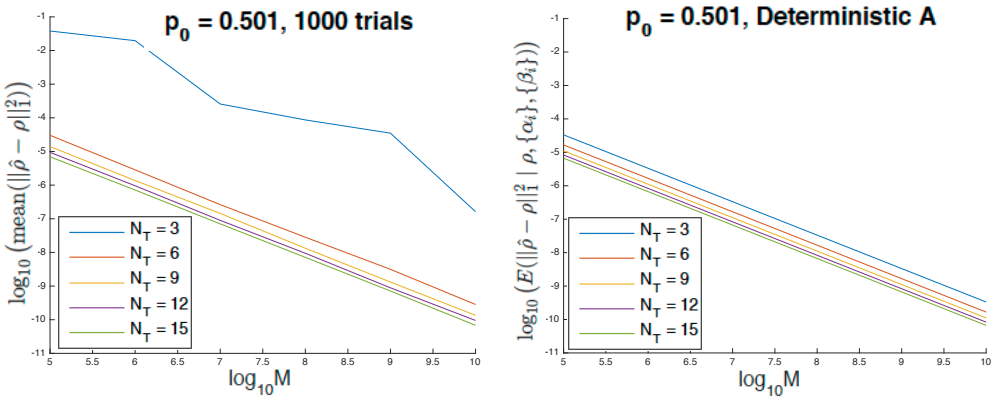
This establishes that $\mathbf{A}^T\mathbf{A}$ is a diagonal matrix, and a quick calculation shows that its diagonal terms are $\frac{1}{2}\sum_k 1 = \frac{N_T}{2}$, $\frac{1}{2}\sum_k \cos^2\varphi_k = \frac{N_T}{4}$, and $\frac{1}{2}\sum_k \sin^2\varphi_k = \frac{N_T}{4}$. As a result, we have in this case

$$\mathrm{Tr}\left((\mathbf{A}^T\mathbf{A})^{-1}\right) = \frac{2}{N_T} + \frac{4}{N_T} + \frac{4}{N_T} = \frac{10}{N_T},$$

which is within $\frac{1}{N_T}$ of the lower bound in Theorem 2.

Remark: We may also set $\alpha_k = \pm\sqrt{\frac{2-\sqrt{2}}{4}}$ in Equation (31) for each k, which will reverse the sign of the first column of $\mathbf{A}$, but will keep $\mathbf{A^T A}$ equal to the same diagonal matrix.

In Figure 1, we fix $p_0 = .501$ and compare the expected estimation error using the deterministic $\mathbf{A}$ described above to the empirical estimation error obtained by randomly choosing our measurement matrices. In the deterministic construction, we choose the reference basis $\mathcal{E} = \{|\mathbf{e}_0\rangle, |\mathbf{e}_1\rangle\}$ uniformly at random and then build the measurement matrices $\mathbf{U}_i$ using Equation (29), setting γ_i equal to 0 for each i. As we can see, there is a noticeable drop in the expected estimation error using our construction for a small number of measurement types (particularly $N_T = 3$) compared to the mean estimation error averaged over 1000 randomly chosen sets of measurement matrices $\{\mathbf{U}_i\}_{i=1}^{N_T}$.



(a) **(b)**

Figure 1. Convergence of the estimate $\hat{\rho}$ to the true density operator ρ with respect to the number of samples M for each of N_T different measurement types. The basis probability bias is fixed to $p_0 = .501$. (a) Each of the N_T measurement-defining unitary matrices $\mathbf{U}_i$ is chosen randomly, to estimate $\hat{\rho}$. This process is repeated 1000 times and the mean squared trace norm error is plotted. (b) The unitary measurement matrices are chosen deterministically according to the method in Section III-C, and the expected squared trace norm error is plotted.

In Figure 2, we instead fix the number of samples M per measurement type, and plot the convergence of our estimation error with respect to p_0 for our deterministic construction of $\mathbf{A}$. We see that when N_T is small, adding extra measurement types yields a significant drop in expected estimation error (as seen when increasing N_T from 3 to 6).

IV. Multiple Basis Estimation

We now shift our attention to the case where Alice randomly communicates to Bob in one of two bases, as in the BB84 protocol, and examine how Eve can learn about Alice's bases with a series of projective measurements. Let the first basis be $\mathcal{A} = \{|a_0\rangle, |a_1\rangle\}$, and without loss of generality the second basis can be expressed as $\mathcal{B} = \{\mathbf{U}|a_0\rangle, \mathbf{U}|a_1\rangle\}$, where $\mathbf{U}$ is a unitary transformation. Alice transmits an element from basis $\mathcal{A}$ with probability q_a and an element from $\mathcal{B}$ with probability $q_b = 1 - q_a$. For each basis, we assume Alice transmits the '0' element with probability p_0 and the '1' element with probability $p_1 = 1 - p_0$. This yields the mixed density operator

$$\Psi := q_a \rho_A + q_b \mathbf{U} \rho_A \mathbf{U}^H, \tag{33}$$

where $\rho_A := p_0 |a_0\rangle\langle a_0| + p_1 |a_1\rangle\langle a_1|$. Since the bases $\mathcal{A}$ and $\mathcal{B}$ live in 2-dimensional Hilbert spaces, we can express ρ_A and Ψ in the form

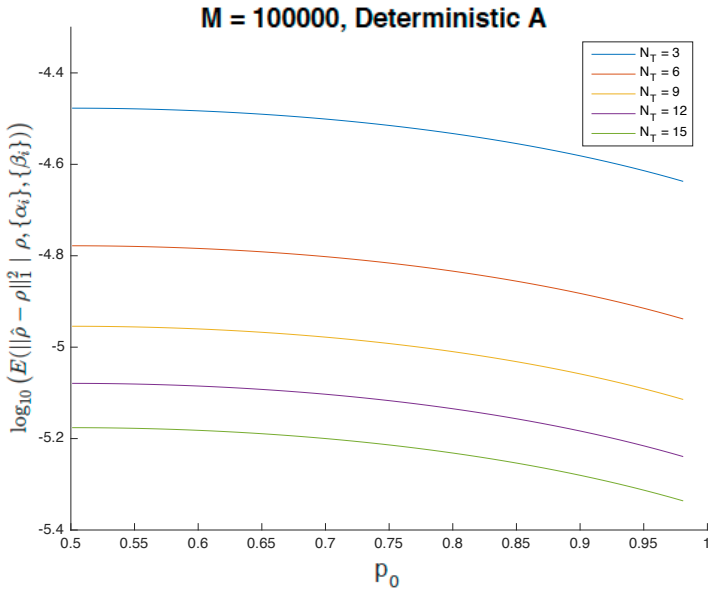


Figure 2. Convergence of the estimate $\hat{\rho}$ to the true density operator ρ with respect to the probability bias, indicated by p_0, for the deterministic measurement construction of Section III-C. The number of samples M per measurement type is fixed to 100,000, and the expected squared trace norm error is plotted for varying numbers N_T of measurement types.

$$\rho_A = \begin{bmatrix} x & w \\ w^* & 1-x \end{bmatrix}, \tag{34}$$

$$\Psi = \begin{bmatrix} \tilde{x} & \tilde{w} \\ \tilde{w}^* & 1-\tilde{x} \end{bmatrix}, \tag{35}$$

where x and $\tilde{x}$ are real, and w and $\tilde{w}$ are complex. For convenience, we will define $\lambda := \frac{q_a}{q_b}$ so that we may write $\Psi = q_b[\lambda\rho_A + \mathbf{U}\rho_A\mathbf{U}^H]$. We may also express $\mathbf{U}$ in the the form

$$\mathbf{U} = \begin{bmatrix} a & b \\ -e^{i\varphi}b^* & e^{i\varphi}a^* \end{bmatrix}, \tag{36}$$

with a and b complex numbers satisfying $|a|^2 + |b|^2 = 1$. This allows us to express the relationship between the parameters (x, w) and $(\tilde{x}, \tilde{w})$ from Equation (33) in the form of the homogeneous equation

$$\frac{1}{q_b}\begin{bmatrix} \tilde{x} \\ \tilde{w} \\ \tilde{w}^* \end{bmatrix} = \mathbf{M} \cdot \begin{bmatrix} x \\ w \\ w^* \end{bmatrix} + \mathbf{c}, \tag{37}$$

where

$$\mathbf{M} = \begin{bmatrix} |a|^2 - |b|^2 + \lambda & ab^* & a^*b \\ -2e^{-i\varphi}ab & e^{-i\varphi}a^2 + \lambda & -e^{i\varphi}b^2 \\ -2e^{i\varphi}a^*b^* & -e^{i\varphi}b^{*2} & e^{i\varphi}a^{*2} + \lambda \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} |b|^2 \\ e^{-i\varphi}ab \\ e^{i\varphi}a^*b^* \end{bmatrix}. \tag{38}$$

We are now equipped to prove the following theorem:

Theorem 3. *Assume Eve has knowledge of the probability bias $\lambda = \frac{q_a}{q_b}$ and of the basis transformation $\mathbf{U}$. Then Eve can recover the the operator ρ_A from Ψ provided that either $\lambda \neq 1$ or $|a|\cos(\theta - \varphi/2) \neq 0$, where a and φ appear in $\mathbf{U}$ as in Equation (36) and θ is the argument of a.*

Proof. Eve can recover x and w, the defining parameters of ρ_A, from $\tilde{x}$ and $\tilde{w}$ (which define Ψ) from Equation (37) provided that the matrix $\mathbf{M}$ is invertible. We can compute the determinant of $\mathbf{M}$ to be

$$\det(\mathbf{M}) = \lambda^3 + \beta(\lambda^2 + \lambda) + 1, \tag{39}$$

where $\beta = 4|a|^2\cos^2(\theta - \varphi/2) - 1$ and $a = |a|e^{i\theta}$. Note that $\lambda > 0$ (by assumption) and $-1 \leq \beta \leq 3$. $\mathbf{M}$ will be noninvertible only when its determinant is zero, which corresponds to when $\beta = \frac{-(\lambda^3+1)}{\lambda^2+\lambda}$. We claim that $\frac{-(\lambda^3+1)}{\lambda^2+\lambda} \leq -1$ for $\lambda > 0$. This is equivalent to the function $f(\lambda) = \lambda^3 - \lambda^2 - \lambda + 1$ being nonnegative for $\lambda > 0$. Examining the derivative $f'(\lambda)$ reveals a single local minimum for $f(\lambda)$ in this domain at $\lambda = 1$, at which point $f(1) = 0$. Thus, $f(\lambda) \geq 0$ (and $\frac{-(\lambda^3+1)}{\lambda^2+\lambda} \leq -1$) for $\lambda > 0$. Based on the

range in which β lies, we now see that $\mathbf{M}$ can only be noninvertible when $\beta = -1$, which corresponds to when $|a|\cos(\theta - \varphi/2) = 0$. In this case, the determinant of $\mathbf{M}$ becomes exactly the function $f(\lambda)$, and as we just discussed, it will only be zero if $\lambda = 1$.

Theorem 3 implies, in particular, that Eve can recover ρ_A from Ψ as long as the probabilities of Alice selecting basis $\mathcal{A}$ and $\mathcal{B}$ are not exactly $1/2$, provided she knows the ratio $\frac{q_a}{q_b}$ and the unitary relationship $\mathbf{U}$ between the two bases. This will be the case if Eve has some knowledge of a bias in Alice's random number generator, and perhaps knows that the two bases are related by an angular rotation of 45 degrees. Eve can use the techniques discussed in Section III to estimate Ψ with projective measurements, and to retrieve the basis $\mathcal{A}$ from ρ_A.

A. Optimal Probability of Detection with a POVM

Ultimately, Eve would like to design a protocol to estimate the actual information bits ('0' or '1') communicated by Alice to Bob. Ideally, she could use this POVM in an attack such as intercept-resend and relay the resulting state to Bob in a relatively unperturbed form, or transmit an alternative state that would maximize her likelihood of remaining undetected. To this end, we will write out explicitly the density operators corresponding to whether a '0' or '1' was transmitted in either basis:

$$\psi_i := q_a|a_i\rangle\langle a_i| + q_b\mathbf{U}|a_i\rangle\langle a_i|\mathbf{U}^H,\ i = 0, 1. \tag{40}$$

Eve would like to construct a POVM $\{F_0, F_1\}$ to estimate which bit is sent. The probability that Eve estimates a '0' given the state ψ_0 was transmitted is then $\mathrm{Tr}\,(\psi_0 F_0)$, and the total probability of correct bit-estimation is

$$P_c = p_0 \mathrm{Tr}\,(\psi_0 F_0) + p_1 \mathrm{Tr}\,(\psi_1 F_1)\,. \tag{41}$$

It behooves Eve to maximize this over all choices of POVMs. $\{F_0, F_1\}$.

In two-dimensions, a POVM can be expressed in the form

$$F_0 = \lambda_1 \mathbf{f}_1\mathbf{f}_1^H + \lambda_2 \mathbf{f}_2\mathbf{f}_2^H, \tag{42}$$

$$F_1 = (1-\lambda_1)\mathbf{f}_1\mathbf{f}_1^H + (1-\lambda_2)\mathbf{f}_2\mathbf{f}_2^H, \tag{43}$$

where $\{\mathbf{f}_1, \mathbf{f}_2\} \subset \mathbb{C}^2$ is an orthonormal basis and $0 \le \lambda_i \le 1$ for $i = 1, 2$. For notational convenience, we will write in vector form $\mathbf{a}_i := |a_i\rangle$ for $i = 0, 1$.

Lemma 2. *Given a POVM $\{F_0, F_1\}$ expressed as in Equations (42) and (43), we have*

$$Tr(\psi_0 F_0) = (\lambda_1 - \lambda_2)(q_a|\mathbf{f}_1^H\mathbf{a}_0|^2 + q_b|\mathbf{f}_1^H\mathbf{U}\mathbf{a}_0|^2) + \lambda_2, \tag{44}$$

$$Tr(\psi_1 F_1) = Tr(\psi_0 F_0) + 1 - \lambda_1 - \lambda_2. \tag{45}$$

As a result, the probability of correct bit estimation is $P_c = Tr((\psi_0 - p_1\mathbf{I}_2)F_0) + p_1$, where $\mathbf{I}_2$ is the 2×2 identity matrix.

Proof. Equation (44) can be derived algebraically:

$$\begin{aligned}
\mathrm{Tr}\left(\psi_0 F_0\right) &= q_a \mathbf{a}_0^H F_0 \mathbf{a}_0 + q_b \mathbf{a}_0^H \mathbf{U}^H F_0 \mathbf{U} \mathbf{a}_0 \\
&= q_a \lambda_1 |\mathbf{f}_1^H \mathbf{a}_0|^2 + q_a \lambda_2 |\mathbf{f}_2^H \mathbf{a}_0|^2 + q_b \lambda_1 |\mathbf{f}_1^H \mathbf{U} \mathbf{a}_0|^2 + q_b \lambda_2 |\mathbf{f}_2^H \mathbf{U} \mathbf{a}_0|^2 \\
&= q_a \lambda_1 |\mathbf{f}_1^H \mathbf{a}_0|^2 + q_a \lambda_2 (||\mathbf{a}_0||_2^2 - |\mathbf{f}_1^H \mathbf{a}_0|^2) + q_b \lambda_1 |\mathbf{f}_1^H \mathbf{U} \mathbf{a}_0|^2 \\
&\qquad + q_b \lambda_2 (||\mathbf{U}\mathbf{a}_0||_2^2 - |\mathbf{f}_1^H \mathbf{U} \mathbf{a}_0|^2) \\
&= (\lambda_1 - \lambda_2)(q_a |\mathbf{f}_1^H \mathbf{a}_0|^2 + q_b |\mathbf{f}_1^H \mathbf{U} \mathbf{a}_0|^2) + \lambda_2,
\end{aligned}$$

where the second equality follows from Equation (42), the third equality follows from the fact that $\{\mathbf{f}_1, \mathbf{f}_2\}$ is an orthonormal basis, and the fourth equality follows from the fact that $||\mathbf{a}_0||_2^2 = ||\mathbf{U}\mathbf{a}_0||_2^2 = 1$. Mimicking this argument using ψ_1 and F_1, and starting with Equation (43), we get

$$\begin{aligned}
\mathrm{Tr}\left(\psi_1 F_1\right) &= (\lambda_2 - \lambda_1)(q_a |\mathbf{f}_1^H \mathbf{a}_1|^2 + q_b |\mathbf{f}_1^H \mathbf{U} \mathbf{a}_1|^2) + (1 - \lambda_2) \\
&= (\lambda_2 - \lambda_1)(q_a (1 - |\mathbf{f}_1^H \mathbf{a}_0|^2) + q_b (1 - |\mathbf{f}_1^H \mathbf{U} \mathbf{a}_0|^2)) + (1 - \lambda_2) \\
&= (\lambda_2 - \lambda_1) + (\lambda_1 - \lambda_2)(q_a |\mathbf{f}_1^H \mathbf{a}_0|^2 + q_b |\mathbf{f}_1^H \mathbf{U} \mathbf{a}_0|^2) + (1 - \lambda_2) \\
&= \mathrm{Tr}\left(\psi_0 F_0\right) + 1 - \lambda_1 - \lambda_2.
\end{aligned}$$

Now we obtain the final expression for P_c as follows:

$$\begin{aligned}
P_c &= p_0 \mathrm{Tr}\left(\psi_0 F_0\right) + p_1 \mathrm{Tr}\left(\psi_1 F_1\right) \\
&= p_0 \mathrm{Tr}\left(\psi_0 F_0\right) + p_1 (\mathrm{Tr}\left(\psi_0 F_0\right) + 1 - \lambda_1 - \lambda_2) \\
&= \mathrm{Tr}\left(\psi_0 F_0\right) + p_1 (1 - \lambda_1 - \lambda_2) \\
&= \mathrm{Tr}\left(\psi_0 F_0\right) + p_1 (1 - \mathrm{Tr}\left(F_0\right)) \\
&= \mathrm{Tr}\left((\psi_0 - p_1 \mathbf{I}_2) F_0\right) + p_1.
\end{aligned}$$

We are now in a position to find the optimal probability of correct detection for a POVM:

Theorem 4. *The optimal probability of correct detection maximized over all POVMs $\{F_0, F_1\}$ is given by $P_c = \max\{p_0, p_1, ||\psi_0||_2, ||\psi_1||_2\}$, where $||\psi_i||_2 = \max_{||\mathbf{f}||_2 = 1} \mathbf{f}^H \psi_i \mathbf{f}$ is the maximum singular value of ψ_i. These are equal for ψ_0 and ψ_1 by construction.*

Proof. For a given POVM $\{F_0, F_1\}$, Lemma 2 gives us

$$\begin{aligned}
P_c &= \mathrm{Tr}\left((\psi_0 - p_1 \mathbf{I}_2) F_0\right) + p_1 \\
&= \mathrm{Tr}\left(\psi_0 F_0\right) - p_1 \mathrm{Tr}\left(F_0\right) + p_1 \\
&= (\lambda_1 - \lambda_2)(q_a |\mathbf{f}_1^H \mathbf{a}_0|^2 + q_b |\mathbf{f}_1^H \mathbf{U} \mathbf{a}_0|^2) + \lambda_2 - p_1(\lambda_1 + \lambda_2) + p_1.
\end{aligned}$$

We consider maximizing P_c by varying λ_1 and λ_2 for a fixed $\mathbf{f}_1$, and to this end we set

$$\chi := q_a |\mathbf{f}_1^H \mathbf{a}_0|^2 + q_b |\mathbf{f}_1^H \mathbf{U} \mathbf{a}_0|^2,$$

so that

$$\begin{aligned} P_c &= (\lambda_1 - \lambda_2)\chi + \lambda_2 - p_1(\lambda_1 + \lambda_2) + p_1 \\ &= \lambda_1(\chi - p_1) + \lambda_2(p_0 - \chi). \end{aligned}$$

In this form, we see that it is optimal to choose

$$\lambda_1 = \begin{cases} 0, & \chi < p_1 \\ 1, & \chi \geq p_1 \end{cases}, \tag{46}$$

$$\lambda_2 = \begin{cases} 0, & \chi > p_0 \\ 1, & \chi \leq p_0 \end{cases}. \tag{47}$$

For these choices of λ_1 and λ_2, we see that

- If $p_0 > p_1$: $\quad P_c = \begin{cases} 1-\chi, & \chi < p_1 \\ p_0, & p_1 \leq \chi \leq p_0 \\ \chi, & \chi > p_0 \end{cases}.$
- If $p_1 > p_0$: $\quad P_c = \begin{cases} 1-\chi, & \chi < p_0 \\ p_1, & p_0 \leq \chi \leq p_1 \\ \chi, & \chi > p_1 \end{cases}.$

We quickly verify that the optimal probability of correct detection for a fixed $\mathbf{f}_1$ is

$$P_c = \max\{1-\chi,\ \max(p_0, p_1),\ \chi\}.$$

Rewriting χ as $\mathbf{f}_1^H \psi_0 \mathbf{f}_1$, and $1-\chi$ as $\mathbf{f}_1^H(\mathbf{I}_2 - \psi_0)\mathbf{f}_1 = \mathbf{f}_1^H \psi_1 \mathbf{f}_1$, we see that the optimal POVM will select $\mathbf{f}_1$ to maximize $\max\{\mathbf{f}_1^H \psi_0 \mathbf{f}_1,\ \mathbf{f}_1^H \psi_1 \mathbf{f}_1\}$, which will lead to a probability of correct detection of $P_c = \max\{p_0, p_1, ||\psi_0||_2, ||\psi_1||_2\}$ as in the theorem statement.

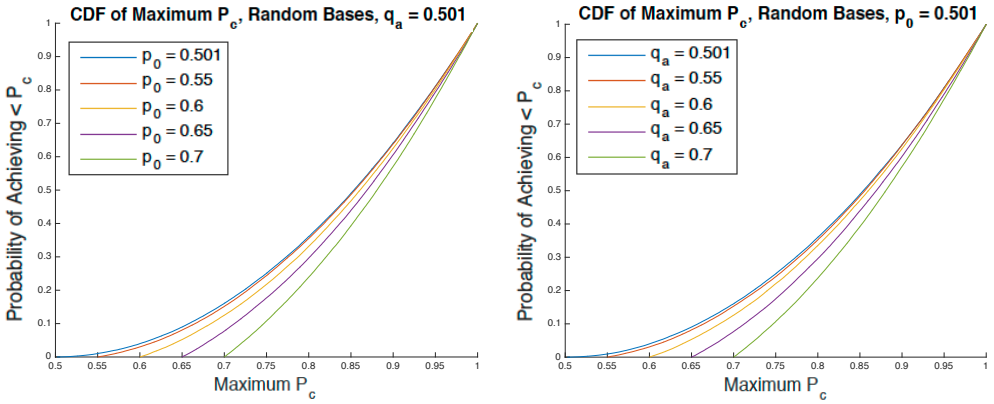
The final comment that $||\psi_0||_2$ and $||\psi_1||_2$ are equal can be seen as follows: For any choice of orthonormal basis $\{\mathbf{f}_1, \mathbf{f}_2\}$, we have

$$\begin{aligned} \mathbf{f}_1^H \psi_0 \mathbf{f}_1 &= q_a |\mathbf{f}_1^H \mathbf{a}_0|^2 + q_b |\mathbf{f}_1^H \mathbf{U}\mathbf{a}_0|^2 \\ &= q_a(1 - |\mathbf{f}_1^H \mathbf{a}_1|^2) + q_b(1 - |\mathbf{f}_1^H \mathbf{U}\mathbf{a}_1|^2) \\ &= q_a |\mathbf{f}_2^H \mathbf{a}_1|^2 + q_b |\mathbf{f}_2^H \mathbf{U}\mathbf{a}_1|^2 \\ &= \mathbf{f}_2^H \psi_1 \mathbf{f}_2, \end{aligned}$$

where the second equality follows from the fact that $\{\mathbf{a}_0,\ \mathbf{a}_1\}$ and $\{\mathbf{U}\mathbf{a}_0,\ \mathbf{U}\mathbf{a}_1\}$ are orthonormal bases, and the third equality follows from $\{\mathbf{f}_1,\ \mathbf{f}_2\}$ being orthonormal. Thus, the maximal singular values of ψ_0 and ψ_1 agree.

Figure 3 shows the distribution of values of the maximum achievable P_c for various fixed p_0 and q_a, obtained by selecting the orthonormal bases $\mathcal{A}$ and $\mathcal{B}$ uniformly at random.

We see clearly that for $p_0 > q_a$, the maximum value of P_c ranges between p_0 and 1, and likewise, for $q_a > p_0$ it will range between q_a and 1. Interchanging the values of p_0 and q_a seems to leave the distributions almost unchanged, though in Tables 1 and 2 we notice that the mean maximum achievable value of P_c is empirically higher in our simulations when fixing p_0 to 0.501 and varying q_a than when fixing q_a to the same value and varying p_0.



(a) **(b)**

Figure 3. The empirical cumulative distribution functions of the maximum achievable P_c, obtained by fixing p_0 and q_a and selecting the two bases $\mathcal{A}$ and $\mathcal{B}$ uniformly at random. (a) The basis probability q_a is set to .501, and the cumulative distribution function (cdf) is plotted for several p_0. (b) The bit probability p_0 is set to .501, and the cdf is plotted for several q_a.

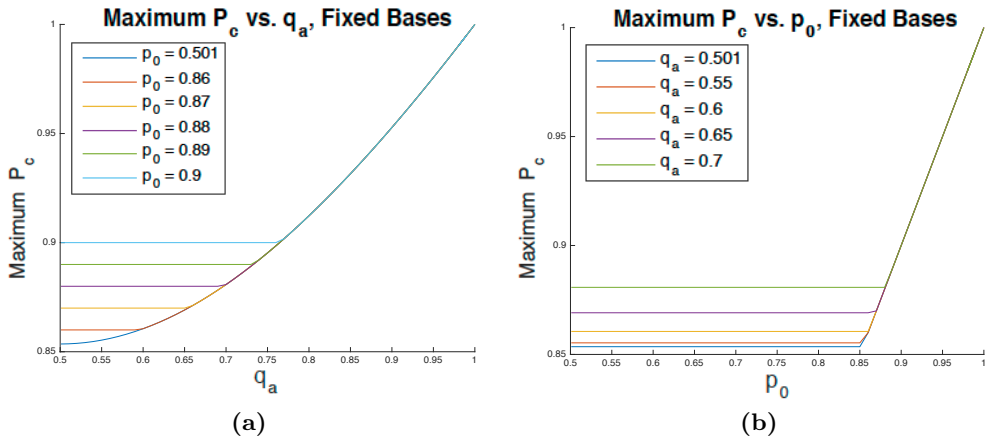
$p_0 = 0.501$	$p_0 = 0.55$	$p_0 = 0.6$	$p_0 = 0.65$	$p_0 = 0.7$
0.83262	0.83332	0.83453	0.8378	0.84411

Table 1. Empirical mean values of maximum achievable P_c for $q_a = 0.501$, averaged over 100,000 trials.

$q_a = 0.501$	$q_a = 0.55$	$q_a = 0.6$	$q_a = 0.65$	$q_a = 0.7$
0.8333	0.83637	0.84361	0.85605	0.8711

Table 2. Empirical mean values of maximum achievable P_c for $p_0 = 0.501$, averaged over 100,000 trials.

In Figure 4, we plot the maximum achievable P_c when the basis $\mathcal{A}$ is fixed to be the standard basis and $\mathcal{B}$ is a rotation of the standard basis by $\pi/4$. This scenario models a typical set of bases that might be used in a QKD scheme. In Figure 4(a), we can see clearly regions in which the maximum P_c is equal to p_0 (for low q_a) and in which it grows as the maximum singular value of the ψ_i (for high q_a). In the regime of Figure 4(b), we see that the maximum P_c is governed by the singular values of the ψ_i for low p_0 (which are determined by the fixed value of q_a for each curve), and reaches a point at which it grows linearly, where it is equal to p_0.



(a) **(b)**

Figure 4. The maximum probability of correct bit estimation P_c with a POVM, where basis $\mathcal{A}$ is the standard basis and $\mathcal{B}$ is the same with a rotation of $\pi/4$. In (a), P_c is plotted with respect to p_0 for several different q_a, and in (b) it is plotted with respect to q_a for several different choices of p_0.

V. Conclusion and Future Directions

In this report, we have evaluated how effectively an eavesdropper can exploit imperfections in a BB84-like protocol for QKD. Specifically, we focused on bias in a random number generator used by Alice in selecting between two orthonormal bases and between the basis elements therein. We showed how Eve can design a system of projective measurements to achieve a low-error estimate of the density operator of Alice's transmitted states, and derived bounds and an approximation for the expected covariance of her estimate conditioned on the choice of measurements. This led to an approximate lower bound on the trace norm of this error in terms of the number of types of projective measurements and the number of samples for each type. We designed a specific set of projective measurements that would enable Eve to achieve this bound. Finally, we gave conditions under which Eve can learn both bases in the BB84 protocol, and derived the optimal probability of correct key-bit estimation with a POVM.

Our analysis served mostly to identify vulnerabilities in a standard "prepare and measure" scheme between Alice and Bob, and there is some work required to frame it in the context of other QKD approaches. Future investigation could also entail incorporating other effects, such as noisy state transmissions or imperfect basis rotations, into our error calculations. The problem of extracting multiple sets of orthonormal states from a statistical mixture is interesting in its own right. It would be useful to derive conditions under which more than two bases could be estimated, and approaches Eve can take to overcome imperfect knowledge of the relationship between two bases (for instance, if the rotation matrix $\mathbf{U}$ is unknown in Section IV).

Finally, it remains to quantify how easily Eve's ability to estimate these states can be overcome through methods employed by Alice and Bob, such as privacy amplification and decoy states [14, 15], and the tradeoff between Eve's estimation error and her probability of detection by Alice and Bob using these methods.

References

[1] F. Miller. *Telegraphic Code to Insure Privacy and Secrecy in the Transmission of Telegrams*. C.M. Cornwell, 1882.

[2] G. Vernam. Secret signaling system: US1310719a. *United States Patent Office*, July 1919.

[3] C. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28(4):656–715, 1949.

[4] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev. The security of practical quantum key distribution. *Reviews of Modern Physics*, 81(3), September 2009.

[5] W. Wootters and W. Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, 1982.

[6] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, 175:8, 1984.

[7] J. B. Altepeter, D. F. V. James, and P. G. Kwiat. Quantum state estimation. *Lecture Notes in Physics, Springer, Berlin*, 2004.

[8] D. Gross, Y.-K. Liu, S. T. Flammia, S. Becker, and J. Eisert. Quantum state tomography via compressed sensing. *Physical Review Letters*, 105(15), 2010.

[9] P. A. M. Dirac. A new notation for quantum mechanics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 35(3):416–418, 1939.

[10] G. Brassard and L. Salvail. Secret key reconciliation by public discussion. *Advances in Cryptography: Eurocrypt 93*, pages 410–423, 1993.

[11] C. H. Bennett, G. Brassard, C. Crépeau, and U. Maurer. Generalized privacy amplification. *IEEE Transactions on Information Theory*, 41(6):1915–1923, November 1995.

[12] C. H. Bennett, G. Brassard, and N. D. Mermin. Quantum cryptography without Bell's theorem. *Physical Review Letters*, 68(5), February 1992.

[13] J. B. Altepeter, E. R. Jeffrey, and P. G. Kwiat. Photonic state tomography. *Advances In Atomic, Molecular, and Optical Physics*, 52(105), 2005.

[14] W.-Y. Hwang. Quantum key distribution with high loss: Toward global secure communication. *Physical Review Letters*, 91(5):057901, August 2003.

[15] H.-K. Lo, X. Ma, and K. Chen. Decoy state quantum key distribution. *Physical Review Letters*, 94:230503, 2005.