VARIATIONAL QUANTUM HYPOTHESIS TESTING

Mahadevan Subramanian, Undergraduate at Department of Physics, IIT Bombay *

Guide: Professor Sai Vinjanampathy, Department of Physics, IIT Bombay

December 2022

Contents

*email: mahadevan_s@iitb.ac.in, roll no: 190260027

ABSTRACT

Distinguishing between two quantum maps is a fundamental task in quantum mechanics which can be thought of as a hypothesis testing task. In this report, a novel variational quantum algorithm is presented which constructs the optimal probe and the optimal measurement to distinguish two quantum map. The obtained state aims at saturating the Holevo-Helstrom bound and is bipartite in nature. The maps are only acted on one part of the state and since it is bipartite it aims at being the state which is used to find the diamond distance between two quantum maps. To this end our novel algorithm also is capable of diamond distance estimation. We also apply this algorithm to the task of quantum illumination demonstrating a variational quantum algorithm using purely Gaussian states.

1 Introduction

Hypothesis testing is of great relevance in classical statistics with the main task revolving around finding out whether a certain observed quantity belongs to one statistic or the other. The key principle revolves around introducing two hypotheses and defining the condition that either accepts or rejects one of the hypotheses [58]. This method naturally results in possible errors due to incorrect classification. Naturally one would wish to minimize these errors as much as possible. The general notion is that there is a null hypothesis H_0 and some alternate hypothesis H_1 and a false negative is if we claim H_0 when it was truly H_1 and a false positive is if we claim H_1 when it was truly H_0. The error probabilities of a false negative and false positive are denoted as β and α respectively. Depending on the aims of our testing scenario, we may either wish to minimize both these errors which is the symmetric error testing case, or we may wish to minimize one of them as long as the other is bounded which is the asymmetric error testing case.
Similarly one may wish to be able to distinguish between two quantum states. Here the hypotheses can be whether one received some state ρ_0 or some state ρ_1. The task now would be to carry out measurements on these states to tell them apart with as much accuracy as possible. These measurements will be two outcome measurements which correspond to the two hypotheses. This task has been studied in various limits for the single-shot case [27, 30] and the asymptotics with many copies [3, 50, 28, 42, 63]. These bring about natural interpretations for various distance measures which can be defined between two states such as the trace distance or relative entropy which itself speak about error bounds in the symmetric error testing and asymmetric error testing cases. We explore these relations in section 3. These bring us to certain bounds which limit us in how well we can tell apart two states. An important interpretation we obtain here is that certain measures such as the trace distance, are infact nothing but a solution of an optimization problem, the optimization being over finding the best possible POVM measurement scheme [17].
Now suppose we were to generalize this problem to distinguishing between two quantum maps $\mathcal{E}_0$ and $\mathcal{E}_1$ instead of two quantum states. In this scenario we would have to send some initial state ρ to the quantum map and then perform measurements to distinguish between the state $\rho_0 = \mathcal{E}_0(\rho)$ and $\rho_1 = \mathcal{E}_1(\rho)$. In this case, in addition to finding the optimal distinguishing measurement, we need to also find the optimal probe. A natural bound provided here is that of the diamond distance [2] between the two quantum maps. Interestingly, the diamond distance itself is calculated as the maximum possible trace norm between the states obtained by acting the map on a bipartite state where the map only acts on one part of it. This means that the optimal way of distinguishing two quantum maps naturally requires and entangled probe where the map is acted on one part of the probe and the other part is left unaffected.
This exact idea of acting the probe on only one part of the state can be found in quantum illumination [44, 60, 61, 56, 20, 40, 35]. In quantum illumination, a two mode probe is created which generates photons that are entangled. The task is to distinguish between the presence and absence of a very weakly reflective beamsplitter in a bright thermal bath. This means that the detector either receives purely thermal light or it receives thermal light mixed with a small part of the signal that reflected. It has been shown that making use of an entangled probe will outperform using just a signal without an idler. This is because a joint measurement scheme is able to extract more information despite the entanglement breaking down since even if the photon has reflected it has lost a lot of its entanglement. There are similar hypothesis testing tasks in a similar context such as quantum reading which tries to identify classically stored data using quantum light [55] and quantum ranging which performs a ranging radar task as a multiary hypothesis test using quantum light [68].
In this report we try to build on these ideas to provide a novel shallow-depth variational quantum algorithm which can be used for both finding the optimal probe and optimal measurement scheme for distinguishing quantum maps. The description of the algorithm can be found in section 4 and it has been applied to the task of quantum illumination with the results shown in section 5. Our algorithm is designed for NISQ (noisy intermediate scale quantum) devices [7] given that it doesn't have a large depth and also shows robustness to errors.

2 Variational quantum algorithms

The variational method in quantum mechanics [59] tries to find the ground state energy of a Hamiltonian $\hat{H} : \mathcal{H} \to \mathcal{H}$ by treating the task as an optimization problem.

$$E_0 = \inf_{|\psi\rangle \in \mathcal{H}} \frac{\langle\psi|\hat{H}|\psi\rangle}{\langle\psi|\psi\rangle} \tag{1}$$

Now suppose we were to parameterize this state as $|\psi(\boldsymbol{\theta})\rangle$ by preparing it using some unitary, we could classically optimize $\boldsymbol{\theta}$ to get $\boldsymbol{\theta}_0 = \text{argmax}_{\boldsymbol{\theta}}(\langle\psi|\hat{H}|\psi\rangle)$ (assuming the state is normalized). This is the underlying principle for the operation of the variational quantum algorithm [15]. These are specifically designed for NISQ devices [7]. The NISQ era described quantum computers with hundreds of noisy qubits which are not error corrected. In this regime, VQAs have been able to offer interesting results. A VQA workflow as described in [7] can be boiled down to the following components

1. The objective function O that encodes the problem to be solved.
2. A parametrized quantum circuit which has some set of parameters $\boldsymbol{\theta}$ which are to be tuned in a way to minimize O.
3. A measurement scheme that is used to evaluate O by measuring expectation values $\langle H \rangle$ and do basis changes.
4. A classical optimizer which minimizes O by variationally updating $\boldsymbol{\theta}$.

The parametrized quantum circuit is defined using a circuit ansatz which can be defined in accordance to the system for which the VQA is being applied. The *hardware-efficient* ansatz [34] aims at providing a low depth ansatz which is constructed with the constraints of the kind of connectivity between the qubits in the given processor.
The training of a VQA is an NP-Hard task [8], which means that in general the classical optimization task is pretty hard. In addition to this there is the issue of barren plateaus that form when using large depth circuits [46]. The issues of barren plateaus can be avoided using low depth circuits of depth $\mathcal{O}(\log n)$ or using local observables [14]. For any variational quantum algorithm, expressibility of the ansatz is of great importance. In [62] the relation between external driving and disorder to the expressibility of parametrized analog circuits is studied with ways of still maintinging quantum supremacy presented.
Finally there are applications in quantum sensing for variational quantum algorithms as well [47]. This has been explored with variational quantum Ramsey-interferometry [36, 45]. These provide an interesting structure of VQAs where both the state preparation and measurement parts of the circuit are made variational. This gives a physical interpretation in the case of quantum sensing as preparation of both the optimal probe and optimal measurement scheme. The algorithm we use for hypothesis testing is loosely based on this structure since it also prepares an optimal probe along with optimal measurements.
Some other interesting uses of variational quantum algorithms are the estimation of quantum Fisher information [4] and estimation of trace distance and fidelity [17]. In the case of [17], the optimization mainly picks out a measurement scheme for distinguishing the two states between which the trace distance needs to be found. This is also a principle we make use of in our algorithm for hypothesis testing.

3 Quantum hypothesis testing

This section will provide a brief review of the theoretical limits of quantum hypothesis testing.

3.1 Statistical hypothesis testing

Statistical hypothesis testing in essence is the task of distinguishing between two probability distributions. There are two major philosophical interpretations for this task, one developed by Fisher [31] and the other by Neyman-Pearson [58]. While differing in how they look at their outcomes, they are in principle equivalent in practice [41]. To briefly describe the two approaches, they can be simplified as follows:

- Fisher's approach: the method was called a significance test with respect to a null hypothesis H_0 with no alternate hypotheses. An exact level of significance is defined as how well the observed statistic matches the null hypothesis. On the basis of this level of significance the hypothesis conclusions are drawn only if the result is significant enough.
- Neyman-Pearson's approach: the method was called decision theory and sets up two statistical hypothesis H_0 and H_1. Here a similar level of significance is calculated but this time a threshold is defined as well. This

threshold is used to decide if a hypothesis is to be accepted or rejected. Since a decision between the two hypotheses is always made no matter what, there will be certain error probabilities due to false negatives and false positives.

Our approach will follow the hypothesis testing defined by Neyman-Pearson. This lets us define type-I error probability α and type-II error probability β. Here we can see how the decision which rejects H_0 can be adjusted based on which

	H_0 is true	H_1 is true
H_0 accepted	✓	Type II error (β)
H_0 rejected	Type I error (α)	✓

Table 1: Description of decision outcomes

kind of error we wish to minimize. To understand what a decision could be like, let us imagine two probability distributions $f_0(x) = \mathcal{N}(\mu = 0, \sigma = 0.5)$ and $f_1(x) = \mathcal{N}(\mu = 1, \sigma = 0.5)$. Here we recieve the value of x and our hypotheses H_0 and H_1 are whether x was sampled from $f_0(x)$ or $f_1(x)$. We could define the decision to be that H_0 is accepted if $x < 0.5$ and this gives us some value for α and β.
This brings about a fundamental measure for distance between two probability distributions which is the Kullback-Leibler divergence [39].

Definition 3.1 (Kullback-Leibler divergence)**.** The Kullback-Leibler divergence or relative entropy for two discrete probability distributions which are defined over the same probability space $\mathcal{S}$ is given as follows.

$$D_{KL}(P\|Q) = \sum_{x\in\mathcal{S}} P(x)\log\left(\frac{P(x)}{Q(x)}\right) \tag{2}$$

More generally we need P and Q to be probability measures on a measurable space $\mathcal{S}$ and P must be absolutely continuous with respect to Q. This lets us define the following for Kullback-Leibler divergence.

$$D_{KL}(P\|Q) = \int_{\mathcal{S}} \log\left(\frac{P(dx)}{Q(dx)}\right) P(dx) \tag{3}$$

Here $P(dx)/Q(dx) = r(x)$ which is well defined on S and is known as the Radon-Nikodym derivative of P with respect to Q. This quantity is well defined due to the P being absolutely continuous with respect to Q.

The importance of the Kullback-Leibler divergence is in how it tells us the distance between two probability distributions which in a sense also tells us how well they can be discriminated using hypothesis testing.

3.2 Distance measures for quantum states

In this subsection we will define the different distance measures used for quantum states and the relations between them.

Definition 3.2 (Trace distance)**.** The trace distance between density matrices $\rho_0 : \mathcal{H} \to \mathcal{H}$ and $\rho_1 : \mathcal{H} \to \mathcal{H}$ acting on Hilbert space $\mathcal{H}$ is defined as follows.

$$T(\rho_0, \rho_1) = \frac{1}{2}\|\rho_0 - \rho_1\|_1 = \frac{1}{2}\sum_i |\lambda_i| \tag{4}$$

Here $\|\|_1$ is the trace norm. λ_i are the eigenvalues of $\rho_0 - \rho_1$. This can also be written in a variational form.

$$T(\rho_0, \rho_1) = \sup_{P\leq\mathbb{I}}(\text{Tr}(P(\rho_0 - \rho_1))) \tag{5}$$

Here P is a positive operator with $P \leq \mathbb{I}$ hence can be thought of a POVM element as well.

The expression in equation 5 will be of prime importance in our variational algorithm for quantum hypothesis testing. The proof of equation 5 can be found in [48].

Definition 3.3 (Fidelity)**.** The fidelity between two density matrices $\rho_0 : \mathcal{H} \to \mathcal{H}$ and $\rho_1 : \mathcal{H} \to \mathcal{H}$ acting on Hilbert space $\mathcal{H}$ is defined as follows.

$$F(\rho_0, \rho_1) = \left(\text{Tr}\sqrt{\sqrt{\rho_0}\rho_1\sqrt{\rho_0}}\right)^2 \tag{6}$$

Fidelity is qualitatively equivalent to the trace distance as is evident from the following inequalities [48].

$$1 - \sqrt{F(\rho_0, \rho_1)} \leq T(\rho_0, \rho_1) \leq \sqrt{1 - F(\rho_0, \rho_1)} \tag{7}$$

Definition 3.4 (Relative entropy and variance)**.** The quantum relative entropy of density matrix $\rho_0 : \mathcal{H} \to \mathcal{H}$ with respect to density matrix $\rho_1 : \mathcal{H} \to \mathcal{H}$ acting on Hilbert space $\mathcal{H}$ is defined as follows.

$$D(\rho_0 \| \rho_1) = \text{Tr}(\rho_0(\log(\rho_0) - \log(\rho_1))) \tag{8}$$

The quantum relative variance of $\rho_0 : \mathcal{H} \to \mathcal{H}$ with respect to density matrix $\rho_1 : \mathcal{H} \to \mathcal{H}$ acting on Hilbert space $\mathcal{H}$ is defined as follows.

$$V(\rho_0 \| \rho_1) = \text{Tr}[\rho_0(\ln(\rho_0) - \ln(\rho_1))^2] - D(\rho_0 || \rho_1) \tag{9}$$

Another important inequality is the quantum Pinsker's inequality given as follows [29].

$$D(\rho_0, \rho_1) \geq \frac{2}{\ln 2} T(\rho_0, \rho_1)^2 \tag{10}$$

The quantities of relative entropy and variance have importance in the assymetric error setting for hypothesis testing.

3.3 Hypothesis testing in the quantum world

Let us have two quantum states ρ_0 and ρ_1 Quantum hypothesis testing is the task of identifying whether the given state to us ρ_0 or ρ_1 and make a decision based on this. The hypothesis H_0 corresponds to having received ρ_0 and H_1 for ρ_1. One can immediately see the link to statistical hypothesis testing since the density matrix formalism is closely related to probability distributions. Our decision here would simply be based on the result of applying a measurement on the state that is given to us and using these results to make a decision.
This would require using a binary result POVM $\{\Pi_0, \Pi_1\}$ where $\Pi_0 + \Phi_1 = 1$. This gives us the following table of probabilities.

	H_0 is true	H_1 is true
H_0 accepted	$\text{Tr}(\Pi_0 \rho_0)$	$\beta = \text{Tr}(\Pi_0 \rho_1)$
H_0 rejected	$\alpha = \text{Tr}(\Pi_1 \rho_0)$	$\text{Tr}(\Pi_1 \rho_1)$

Table 2: Probability of decision outcomes using a POVM $\{\Pi_0, \Pi_1\}$

Given that we wish to discriminate between two states, the error probability will respect certain bounds based on the distance between ρ_0 and ρ_1. If we wish to optimize both α and β simultaneously by picking an appropriate POVM, we run into the Holevo-Helstrom bound [27, 30].

Theorem 3.1 (Holevo-Helstrom bound)**.** Given the task of discriminating two states ρ_0 and ρ_1 from a Hilbert space $\mathcal{H}$ using a two-outcome POVM $\{\Pi, \mathbb{I} - \Pi\}$ we have the following bound for the error probability.

$$P_{\text{err}} \geq \frac{1}{2}(1 - \|\rho_0 - \rho_1\|_1) \tag{11}$$

The optimal POVM which gives rise to this error probability is defined as following Jordan decomposition.

$$\Pi_{\text{optimal}} = (\rho_0 - \rho_1)_+ = \sum_{\lambda_i > 0} |i\rangle \langle i| \tag{12}$$

Here λ_i are the eigenvalues of $\rho_0 - \rho_1$ and $|i\rangle$ are the corresponding eigenvectors.

Now in the same way if we were to have n copies input, we are essentially discriminating between $\rho_0^{\otimes n}$ and $\rho_1^{\otimes n}$. While equation 11 holds true, one needs to find how the value $\|\rho_0^{\otimes n} - \rho_1^{\otimes n}\|$ behaves. This is where we make use of the Chernoff bound [3, 50]

Theorem 3.2 (Quantum Chernoff bound)**.** Given the task of discriminating the two states $\rho_0^{\otimes n}$ and $\rho_1^{\otimes n}$ from the Hilbert space $\mathcal{H}^{\otimes n}$ using a two-outcome POVM $\{\Pi_n, \mathbb{I} - \Pi_n\}$ we have the following bound for the error probability.

$$\liminf_{n \to \infty} \frac{1}{n} \log(P_{\text{err},n}) \geq \inf_{0 \leq s \leq 1} \log \text{Tr}(\rho_0^{1-s} \rho_1^s) \tag{13}$$

For notation we define the following quantity $\kappa(\rho_0, \rho_1)$ as the Chernoff bound.

$$\kappa(\rho_0, \rho_1) = \inf_{0 \leq s \leq 1} \log \text{Tr}(\rho_0^{1-s} \rho_1^s) \tag{14}$$

A weaker but related bound is the quantum Bhattacharya bound which substitutes $s = 1/2$.

$$\kappa_B(\rho_0, \rho_1) = \log \text{Tr}(\rho_0^{1/2} \rho_1^{1/2}) \tag{15}$$

The proof of the above found can be found in [50]. While both the Chernoff bound and Holevo-Helstrom bound look at symmetric hypothesis testing (Bayesian approach) where we wish to minimize both α and β simultaneously, one can also choose to optimize one while making sure the other is bounded the asymmetric error setting. Here we define $\alpha_n(\Pi_n) = \text{Tr}(\rho_0^{\otimes n}(\mathbb{I} - \Pi_n))$ and $\beta_n(\Pi_n) = \text{Tr}(\rho_1^{\otimes n}\Pi_n)$ for a sequence of POVM measurements $\{\Pi_n, \mathbb{I} - \Pi_n\}$. The quantum Stein's lemma shows that the relative entropy is a critical jump point in the asymptotic behavior of asymmetric hypothesis testing.

- For an arbitrary $R \leq D(\rho_0\|\rho_1)$, there exists a sequence of POVMs $\{\Pi_n, \mathbb{I} - \Pi_n\}$ such that the following holds [28].
$$\liminf_{n\to\infty} \frac{-1}{n} \log(\beta_n(\Pi_n)) \geq R \text{ and } \lim_{n\to\infty} \alpha_n(\Pi_n) = 0 \tag{16}$$
- Conversely [51] if a sequence of POVMs $\{\Pi_n, \mathbb{I} - \Pi_n\}$ such that
$$\liminf_{n\to\infty} \frac{-1}{n} \log(\beta_n(\Pi_n)) > D(\rho_0\|\rho_1) \tag{17}$$
then $\lim_{n\to\infty} \alpha_n(\Pi_n) = 1$

To understand what happens when the asymptotic leading term in n is exactly $D(\rho_0\|\rho_1)$ for β_n we note theorem 2 in [42].

Theorem 3.3. With null hypothesis of $\rho_0^{\otimes n}$ and alternate hypothesis of $\rho_1^{\otimes n}$ we denote error probabilities of $\alpha_n(\Pi_n)$ and $\beta_n(\Pi_n)$ for sequence of measurements $\{\Pi_n, \mathbb{I} - \Pi_n\}$, we have the following:

- *Achievability:* for any $E_2 \in \mathbb{R}$ and $f(n) \in o(\sqrt{n})$, there exists a sequence of measurements $\{\Pi_n, \mathbb{I} - \Pi_n\}$ such that
$$\beta_n(\Pi_n) \leq \exp\{-(nD(\rho_0\|\rho_1) + E_2\sqrt{n} + f(n))\} \tag{18}$$
$$\limsup_{n\to\infty} \alpha_n(\Pi_n) \leq \Phi\left(\frac{E_2}{\sqrt{V(\rho_0\|\rho_1)}}\right) \tag{19}$$
- *Optimality:* if there is a sequence of measurements $\{\Pi_n, \mathbb{I} - \Pi_n\}$ such that
$$\beta_n(\Pi_n) \leq \exp\{-(nD(\rho_0||\rho_1) + E_2\sqrt{n} + f(n))\} \tag{20}$$
then the following holds for any given $E_2 \in \mathbb{R}$ and $f(n) \in o(\sqrt{n})$
$$\limsup_{n\to\infty} \alpha_n(\Pi_n) \leq \Phi\left(\frac{E_2}{\sqrt{V(\rho_0\|\rho_1)}}\right) \tag{21}$$

Here $\Phi : \mathbb{R} \to \mathbb{R}$ is the Gauss error function.

The proof of this theorem is the main result of [42]. The importance of the above theorem is mainly to do with the scaling this gives us for the type I and II error at the exact transition point which is given by the quantum Stein's lemma. Looking at equations 20 and 21 we see that there is some sequence of measurements which allows us to make $\beta_n \to 0$ while keeping α_n bounded by some non-zero quantity. What is important here is the rate at which β_n goes to zero is mainly dictated by $D(\rho_0\|\rho_1)$.
An additional result on symmetric error testing with n copies can be found in [49] which uses the fidelity between the states.
$$\frac{1 - \sqrt{1 - \mathcal{F}(\rho_0, \rho_1)^{2n}}}{2} \leq P_{err}(\rho_0, \rho_1) \leq \frac{\mathcal{F}(\rho_0, \rho_1)^n}{2} \tag{22}$$
An important point so far is that we are considering the hypotheses H_0 and H_1 to be receiving a particular state ρ_0 or ρ_1. This can be re-framed to instead defining the hypotheses H_0 and H_1 as whether a CPTP map $\mathcal{E}_0$ or $\mathcal{E}_1$ has been applied on some ρ. This simply means we write $\rho_0 = \mathcal{E}_0(\rho)$ and $\rho_1 = \mathcal{E}_1(\rho)$ and do a state discrimination between these two states.

4 A variational quantum algorithm to distinguish quantum maps

In this section we will describe a novel algorithm which is capable of producing an optimal probe and an optimal measurement as well for the task of discriminating between two quantum maps in the single shot regime.

4.1 What is the cost function?

Since we have a variational quantum algorithm, we must first understand what an appropriate cost function would be which should also be efficient to evaluate. Here we will frame the problem at hand we wish to solve.
Inputs: CPTP maps $\mathcal{E}_0$ and $\mathcal{E}_1$ taking density matrices on Hilbert space $\mathcal{H}$ to density matrices acting on Hilbert space $\mathcal{H}$
Outputs: Density matrix $\rho : \mathcal{H} \to \mathcal{H}$ such that $\rho = \text{argmin}_\rho P_{err}(\mathcal{E}_0(\rho), \mathcal{E}_1(\rho))$ where $P_{err}(\mathcal{E}_0(\rho), \mathcal{E}_1(\rho))$ is the error probability in distinguishing between $\mathcal{E}_0(\rho)$ and $\mathcal{E}_1(\rho)$ using a two outcome POVM.
Using the results from the symmetric error testing case with only one copy of the states (equation 11), we can see that the following holds

$$\underset{\rho}{\text{argmin}}\, P_{err}(\mathcal{E}_0(\rho), \mathcal{E}_1(\rho)) = \underset{\rho}{\text{argmax}}\, T(\mathcal{E}_0(\rho), \mathcal{E}_1(\rho)) \tag{23}$$

Now if one were to somehow conduct the test with a large number of copies, that would make the LHS of the above equation simply become $\text{argmax}_\rho\, \kappa(\mathcal{E}_0(\rho), \mathcal{E}_1(\rho))$. The asymmetric error setting will make the LHS to be $\text{argmax}_\rho\, D(\mathcal{E}_0(\rho), \mathcal{E}_1(\rho))$.
An important thing here is that all of these quantities are not easy to calculate simply by measurements. All of these require having the density matrix of the state being available to us, a task which requires quantum tomography which itself scales exponentially [18]. Estimation of trace distance is probably hard even for quantum computers given that judging whether a trace distance is small or large itself belongs to the complexity class QSZK-complete [65, 1]. There exist various methods to estimate the quantum relative entropy. Some of these are Taylor approximations [16] or solving a semi-definite programming problem to attain the entropy value efficiently [21]. However these do not scale well with system size due to exponential scaling of the Hilbert space dimension.
However there are ways of estimating trace distance using a variational quantum algorithm as shown in [17] and [1]. The main clue lies in the definition of trace distance as can be seen in equation 5. We need to variationally optimize the POVM P to obtain the trace distance. To do this using a unitary operation, we must embed the POVM into the unitary operator. For this the Naimark extension can be used [66].

Theorem 4.1 (Naimark extension)**.** For any POVM $\{\Gamma_i\}_{i\in O}$ acting on a system S, there exists a unitary U_{PS} (acting on a probe system P and the system S) and an orthonormal basis $\{|i\rangle_P\}_{i\in O}$ such that

$$\text{Tr}\left((|i\rangle\langle i| \otimes I_S) U_{PS} (|i\rangle\langle i| \otimes \rho_S) U_{PS}^\dagger\right) = \text{Tr}(\Gamma_i \rho_S) \tag{24}$$

As pointed out in [66], the two-outcome POVM $\{\Gamma, \mathbb{I} - \Gamma\}$ can be encoded in the following unitary with the probe system being a qubit.

$$U_{PS} = \mathbb{I}_P \otimes (\sqrt{\Gamma})_S + i(\sigma_Y)_P \otimes (\sqrt{\mathbb{I} - \Gamma})_S \tag{25}$$

Let us define a parameterized unitary $V(\phi)$ which acts over both the probe and the system. We define the following quantity as an estimate of trace distance,

$$T_{\boldsymbol{\phi}}(\rho_0, \rho_1) = |p_0 - p_1| \tag{26}$$

$$p_i = \text{Tr}\left((|0\rangle\langle 0| \otimes I) V(\boldsymbol{\phi}) (|0\rangle\langle 0| \otimes \rho_i) V(\boldsymbol{\phi})^\dagger\right) \quad i \in \{0, 1\} \tag{27}$$

On combining theorem 4.1 and equation 5, we get that $\forall \boldsymbol{\phi}(T_{\boldsymbol{\phi}}(\rho_0, \rho_1) \leq T(\rho_0, \rho_1))$. This is the main crux of using a variational algorithm for estimating trace distance in [17]. As an extension to this, we define the following cost function for our algorithm where we use an additional qubit as the probe subsystem.

$$T^{\text{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) = |p_0(\boldsymbol{\theta}, \boldsymbol{\phi}) - p_1(\boldsymbol{\theta}, \boldsymbol{\phi})| \tag{28}$$

$$p_i(\boldsymbol{\theta}, \boldsymbol{\phi}) = \text{Tr}\left((|0\rangle\langle 0| \otimes I) V(\boldsymbol{\phi}) (|0\rangle\langle 0| \otimes \mathcal{E}_i(\rho(\boldsymbol{\theta}))) V(\boldsymbol{\phi})^\dagger\right) \quad i \in \{0, 1\} \tag{29}$$

Our optimization procedure will have to optimize both $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ for obtaining the best possible state preparation and measurement. Let us define $\boldsymbol{\theta}_0$ and $\boldsymbol{\phi}_0$ as follows

$$\boldsymbol{\theta}_0 = \underset{\boldsymbol{\theta}}{\text{argmax}}(\|\mathcal{E}_1(\rho_{\boldsymbol{\theta}}) - \mathcal{E}_0(\rho_{\boldsymbol{\theta}})\|_1) \tag{30}$$

$$\boldsymbol{\phi}_0 = \underset{\boldsymbol{\phi}}{\text{argmax}}(TD_{\boldsymbol{\phi}}(\boldsymbol{\theta}_0)) \tag{31}$$

Here $\boldsymbol{\theta}_0$ optimizes toward the state that saturates the Holevo-Helstrom bound [30, 27]. As per the definition of the optimization problem of trace distance estimation [17], $\boldsymbol{\phi}_0$ represents the best parameters to estimate the trace distance for $\rho_{\boldsymbol{\theta}_0}$. Now let us define the parameters obtained by a complete optimization as follows

$$\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\phi}} = \underset{\boldsymbol{\theta}, \boldsymbol{\phi}}{\text{argmax}}(\text{Tr}_{\boldsymbol{\phi}}(\mathcal{E}_1(\rho_{\boldsymbol{\theta}})) - \text{Tr}_{\boldsymbol{\phi}}(\mathcal{E}_0(\rho_{\boldsymbol{\theta}}))) \tag{32}$$

Clearly $T^{\text{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) \leq \|\mathcal{E}_1(\rho_{\boldsymbol{\theta}}) - \mathcal{E}_0(\rho_{\boldsymbol{\theta}})\|_1/2$. Our task now would be to verify if $\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\phi}} \equiv \boldsymbol{\theta}_0, \boldsymbol{\phi}_0$, to see whether the global optimization reaches a meaningful result.

Claim 4.1.1. Under the following assumption

$$\forall \boldsymbol{\theta} \exists \boldsymbol{\phi} \left(T^{\text{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta}) = T(\mathcal{E}_1(\rho_{\boldsymbol{\theta}}), \mathcal{E}_0(\rho_{\boldsymbol{\theta}})) \right)$$

We can claim the following equivalence

$$\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\phi}} \equiv \boldsymbol{\theta}_0, \boldsymbol{\phi}_0$$

Proof. From the assumption we have taken, it is quite clear that $TD_{\boldsymbol{\phi}_0}(\boldsymbol{\theta}_0) = \|\mathcal{E}_1(\rho_{\boldsymbol{\theta}_0}) - \mathcal{E}_0(\rho_{\boldsymbol{\theta}_0})\|_1$. Along with this, since the parameters $\tilde{\theta}, \tilde{\phi}$ are from an optimization of $T^{\text{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$, we must have the following inequality hold

$$TD_{\tilde{\boldsymbol{\phi}}}(\tilde{\boldsymbol{\theta}}) \geq TD_{\boldsymbol{\phi}_0}(\boldsymbol{\theta}_0)$$

Now from the assumption that we have taken we can make the following claim

$$TD_{\tilde{\boldsymbol{\phi}}}(\tilde{\boldsymbol{\theta}}) = \|\mathcal{E}_1(\rho_{\tilde{\boldsymbol{\theta}}}) - \mathcal{E}_0(\rho_{\tilde{\boldsymbol{\theta}}})\|_1$$

If this doesn't hold, there will exist some $\boldsymbol{\phi}$ which gives the exact trace distance and the TD function always is less than the trace bound hence resulting in a contradiction. Hence we have the following hold

$$\|\mathcal{E}_1(\rho_{\tilde{\boldsymbol{\theta}}}) - \mathcal{E}_0(\rho_{\tilde{\boldsymbol{\theta}}})\|_1 \geq \|\mathcal{E}_1(\rho_{\boldsymbol{\theta}_0}) - \mathcal{E}_0(\rho_{\boldsymbol{\theta}_0})\|_1$$

Let us assume that $\|\mathcal{E}_1(\rho_{\tilde{\boldsymbol{\theta}}}) - \mathcal{E}_0(\rho_{\tilde{\boldsymbol{\theta}}})\|_1 > \|\mathcal{E}_1(\rho_{\boldsymbol{\theta}_0}) - \mathcal{E}_0(\rho_{\boldsymbol{\theta}_0})\|_1$. This would contradict the fact that $\boldsymbol{\theta}_0$ is a parameter that saturates the trace distance. Hence we finally get the following hold.

$$\|\mathcal{E}_1(\rho_{\tilde{\boldsymbol{\theta}}}) - \mathcal{E}_0(\rho_{\tilde{\boldsymbol{\theta}}})\|_1 = \|\mathcal{E}_1(\rho_{\boldsymbol{\theta}_0}) - \mathcal{E}_0(\rho_{\boldsymbol{\theta}_0})\|_1$$

Hence both these parameters saturate the Holevo bound and also they have the perfect trace distance estimators, hence proving their equivalence. □

While the assumption in the above claim requires $V(\boldsymbol{\phi})$ to be able to reach the optimal POVM's Naimark extension for all $\rho_{\boldsymbol{\theta}}$, this does show that the optimization procedure is sound and produces meaningful results.
In essence we will have to optimize our estimated trace distance since the real trace distance is not as easy to calculate but this optimization will end up optimizing the true trace distance as well as the estimate of trace distance toward the true value. This will be reflected in our results which we showcase in the sections that follow.

4.2 Description of the algorithm

Algorithm 1: Variational quantum hypothesis testing

Take input as $\mathcal{E}_0$ and $\mathcal{E}_1$ which are CPTP maps from density matrices in $\mathcal{H}$ to density matrices in $\mathcal{H}$.
Initialize $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ which are parameters for $U(\boldsymbol{\theta})$ and $V(\boldsymbol{\phi})$.
Define convergence condition for cost function $T^{\text{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$.
while $T^{\text{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$ *has not converged* **do**
  for $\boldsymbol{\theta}_i = \boldsymbol{\theta} + \Delta\boldsymbol{\theta}_i$ *and* $\boldsymbol{\phi}_i = \boldsymbol{\phi} + \Delta\boldsymbol{\phi}_i$ *for some set of* $\Delta\boldsymbol{\theta}_i$ *and* $\Delta\boldsymbol{\phi}_i$ **do**
    Run circuit with parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_i$ with CPTP map as $\mathcal{E}_0$
    Using measurements on q_0 obtain value of p_0.
    Run circuit with parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_i$ with CPTP map as $\mathcal{E}_1$
    Using measurements on q_0 obtain value of p_1.
    Assign $TD_{\boldsymbol{\phi}_i}(\boldsymbol{\theta}_i) = |p_0 - p_1|$
  end
  Using $TD_{\boldsymbol{\phi}_i}(\boldsymbol{\theta}_i)$ update $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ with a classical optimizer.
  Check convergence of $T^{\text{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$.
end
return final values of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$.

The structure of the algorithm takes inspiration in structure from quantum variational optimization for Ramsey-interferometry [36] which has been implemented experimentally as well [45]. The circuit consists of N atomic spins and is exclusively operated in the Dicke basis [24]. Ramsey interferometry is the task of estimating a small phase shift applied on an atomic spin. The optimal state for estimating the value of the phase shift applied on N atomic spins is

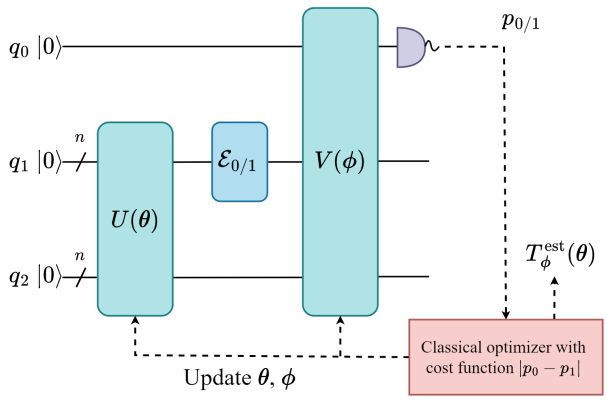

Figure 1: Quantum circuit for diamond norm estimation. The state preparation is over the two n qubit registers q_1 and q_2. q_0 is used for measurements that are used to estimate the trace distance between the two states. The map is only applied over register q_1 and the optimization hence aims to prepare the state which gives the value of diamond distance as the trace distance between the states.

the GHZ state [9] which also reaches the Heisenberg limit [53, 25]. However, as pointed out in [**empty citation**], the optimal state if we assume ϕ to have a prior width $\delta\phi$ is actually not the GHZ but is an entangled state.
In such a situation the capability of a variationally optimized state can be used to construct the appropriate the correct state. However obtaining information from the state requires a measurement and here that is included by what is called as the decoding layer. This is essentially just a parametrized unitary before measurements of each spin is done hence can give rise to joint measurements.
Our circuit does a very similar thing with trying to optimize a state preparation and a Neimark extension of a POVM simultaneously. From the task we accomplish, we obtain both a state preparation and POVM which are optimized for distinguishing the maps $\mathcal{E}_0$ and $\mathcal{E}_1$. Since our aim is to approach the states that saturate the Holevo-Helstrom bound, this minimizes error probability in the single-shot regime.

4.3 An algorithm for diamond distance estimation

Diamond distance is a metric used to characterize how much different two quantum maps are from each other. Naturally this is deeply linked with the task of distinguishing two quantum maps.

Definition 4.1 (Diamond norm & distance)**.** Given $\Phi : M_n(\mathbb{C}) \to M_m(\mathbb{C})$ which is a linear map and $X \in M_{n^2}(\mathbb{C})$ where $M_n(\mathbb{C})$ is the set of $n \times n$ complex valued matrices.

$$\|\Phi\|_\diamond := \max_{X;\|X\|_1 \leq 1} \|(\Phi \otimes \mathbb{I}_n)X\|_1, \tag{33}$$

Here $\|A\|_1 = \text{Tr}\sqrt{A^\dagger A}$. Using the above equation we define the diamond distance for two CPTP maps Φ_1 and Φ_2 as follows for density matrices ρ

$$d_\diamond(\Phi_1, \Phi_2) = \|\Phi_1 - \Phi_2\|_\diamond := \max_\rho \|(\Phi_1 \otimes \mathbb{I}_n)\rho - (\Phi_2 \otimes \mathbb{I}_n)\rho\|_1, \tag{34}$$

Clearly equation 34 highlights the exact optimization problem we need to solve to estimate the diamond distance between two states. The diamond distance also has a lot of important physical interpretations as a distance measure.

Lemma 4.2 (Properties of diamond norm)**.** The following properties hold for a quantum gate $T : \mathbf{L}(\mathcal{N}) \to \mathbf{L}(\mathcal{M})$.

1. $\|T\|_\diamond = \|T \otimes I_\mathcal{G}\|_1 \geq \|T\|_1$ where $\dim \mathcal{G} \geq \dim \mathcal{N}$
2. $\|T \circ \rho\|_\diamond \leq \|T\|_\diamond \|\rho\|_1$
3. $\|T \circ R\|_\diamond \leq \|T\|_\diamond \|R\|_\diamond$
4. $\|T \otimes R\|_\diamond = \|T\|_\diamond \|R\|_\diamond$
5. $\|T\|_\diamond = 1$ if it is physically allowed.

All of the above properties are proven in [2]. The important result we will be using is property 1 of lemma 4.2 which directly relates the diamond norm to the 1-norm. The definition of the 1-norm for an operator T is as follows.

$$\|T\|_1 = \sup_{X \neq 0} \frac{\|T(X)\|_1}{\|X\|_1} \tag{35}$$

What is important with the diamond distance is that it comments on how the different states we obtain from the two maps are even when we allow entangled states to enter them. This is important since we can obtain more information from a quantum map if we send in a bipartite state to it by keeping one part unaffected and act the map on the other part. This is exactly why $\|T\|_\diamond \geq \|T\|_1$.
This tells us a very interesting result by simply extending the Holevo-Helstrom bound.

Theorem 4.3 (Holevo-Helstrom bound for quantum maps)**.** The error probability in trying to distinguish two CPTP maps $\mathcal{E}_0 : M_n(\mathcal{C}) \to M_m(\mathcal{C})$ and $\mathcal{E}_1$ is bounded as follows.

$$P_{err} \geq \frac{1}{2}(1 - \|\mathcal{E}_0 - \mathcal{E}_1\|_\diamond) \tag{36}$$

The state used as probe to saturate the bound must satisfy the following

$$\rho = \underset{\rho}{\operatorname{argmax}}(\|(\mathcal{E}_0 \otimes I_n)(\rho) - (\mathcal{E}_1 \otimes I_n)(\rho)\|_1) \tag{37}$$

The measurement must be the Helstrom POVM $\{P, \mathbb{I} - P\}$ such that

$$P = \underset{0 \leq P \leq \mathbb{I}}{\operatorname{argmax}}(\text{Tr}(P((\mathcal{E}_0 \otimes I_n)(\rho) - (\mathcal{E}_1 \otimes I_n)(\rho)))) \tag{38}$$

Proof. We know that there must be some state $\rho \in M_{n^2}(\mathbb{C})$ such that

$$(\|(\mathcal{E}_0 \otimes I_n)(\rho) - (\mathcal{E}_1 \otimes I_n)(\rho)\|_1 = \|\mathcal{E}_0 - \mathcal{E}_1\|_\diamond$$

This follows from the variational definition of diamond norm in equation 34. Now we essentially just follow the Holevo-Helstrom bound for distinguishing the two states $(\mathcal{E}_0 \otimes I_n)(\rho)$ and $(\mathcal{E}_1 \otimes I_n)(\rho)$.
To prove that this is the true minimum in the sense that entangling with an even larger dimension state will not produce more distinguishability, we impose lemma 4.2, property 1. This says that $\|T \otimes I_\mathcal{G}\| = \|T \otimes I_\mathcal{N}\|$ as long as $\dim \mathcal{G} \geq \dim \mathcal{N}$. This shows that we do not gain any more distinguishability on moving to a higher dimension showing that the minimum error probability is captured truly by the diamond norm. □

This shows that the diamond norm has a fundamental physical interpretation in the form of minimum error probability.

Algorithm 2: Variational quantum algorithm to estimate diamond distance

Take input as $\mathcal{E}_0$ and $\mathcal{E}_1$ which are CPTP maps from density matrices in $\mathcal{H}$ to density matrices in $\mathcal{H}$.
Initialize $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ which are parameters for $U(\boldsymbol{\theta})$ and $V(\boldsymbol{\phi})$.
Define convergence condition for cost function $T^{\text{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$.
while $T^{\text{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$ *has not converged* **do**
  for $\boldsymbol{\theta}_i = \boldsymbol{\theta} + \Delta\boldsymbol{\theta}_i$ *and* $\boldsymbol{\phi}_i = \boldsymbol{\phi} + \Delta\boldsymbol{\phi}_i$ *for some set of* $\Delta\boldsymbol{\theta}_i$ *and* $\Delta\boldsymbol{\phi}_i$ **do**
    Run circuit with parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_i$ with CPTP map as $\mathcal{E}_0 \otimes \mathbb{I}_n$
    Using measurements on q_0 obtain value of p_0.
    Run circuit with parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_i$ with CPTP map as $\mathcal{E}_1 \otimes \mathbb{I}_n$
    Using measurements on q_0 obtain value of p_1.
    Assign $TD_{\boldsymbol{\phi}_i}(\boldsymbol{\theta}_i) = |p_0 - p_1|$
  end
  Using $TD_{\boldsymbol{\phi}_i}(\boldsymbol{\theta}_i)$ update $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ with a classical optimizer.
  Check convergence of $T^{\text{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$.
end
Assign estimated diamond norm $= T^{\text{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$ **return** final values of $\boldsymbol{\theta}$, $\boldsymbol{\phi}$ and estimated diamond norm.

This is pretty much equivalent to algorithm 1 with the exception that we feed in a map of form $\mathcal{E}_0 \otimes \mathbb{I}_n$ instead. The importance of this algorithm is that it gives a method to find the diamond distance between two CPTP maps which is possible with low depth circuits as well. There are existing ways to find the diamond distance in $\mathcal{O}(poly(\dim \mathcal{H}))$ since it is a convex optimization problem [5] but this clearly will grow exponentially if we consider n qubits. Our algorithm is clearly scalable for such situations and can produce results using a Hardware efficient ansatz.

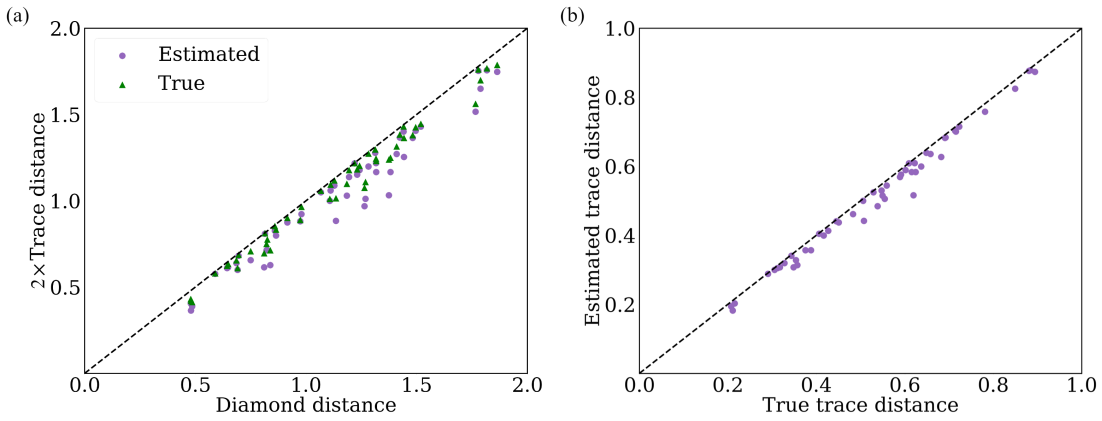

Figure 2: Estimation for diamond distance between the identity map and an arbitrary 2 qubit unitary map. (a) Shows how well the value of estimated $T_{\boldsymbol{\phi}}^{\text{est}}(\boldsymbol{\theta})$ and the true trace distance $T(\mathcal{E}_0(\rho_{\boldsymbol{\theta}}), \mathcal{E}_1(\rho_{\boldsymbol{\theta}}))$ for the optimized state matches against the analytical value. (b) Shows how well the estimated trace distance matches against the true trace distance for the state $\rho_{\boldsymbol{\theta}}$

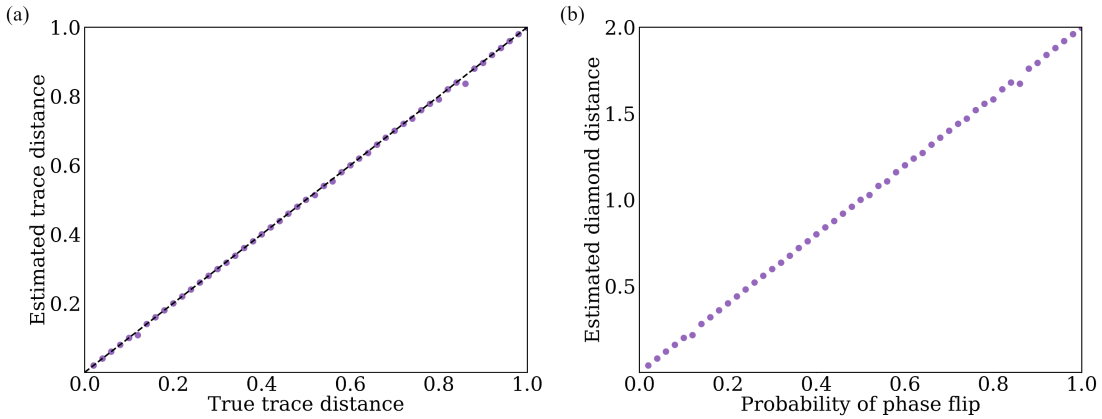

Figure 3: Estimation for diamond distance between the identity map and a phase flip map. (a) Shows how well the estimated trace distance matches against the true trace distance for the state $\rho_{\boldsymbol{\theta}}$. (b) Shows the result of the estimated diamond norm against the probability of phase flip.

Figures 2 and3 show results of simulations for estimating of diamond norm. The diamond distance between a unitary map and the identity map is the diameter of the circle which is able to contain all the eigenvalues of the unitary operation [6]. For Kraus maps such as the phase flip map of form $\mathcal{E}(\rho) = (1-p)\rho + pZ\rho Z$ have a diamond distance of $2p$ from the identity map [64].
The simulations were carried out using QuTip [33, 32]. The circuit had 5 qubits with one of them being used in the trace distance estimation subroutine and the other 4 being used for state preparation after which the quantum map is applied on the first 2 of the 4.

5 Variational quantum illumination

In this section we describe how we apply algorithm 1 in a modified setting to the task of quantum illumination.

5.1 Quantum illumination

The task of quantum illumination presents a quantum hypothesis testing problem with the following hypotheses on the presence of an object probed by a single photon.

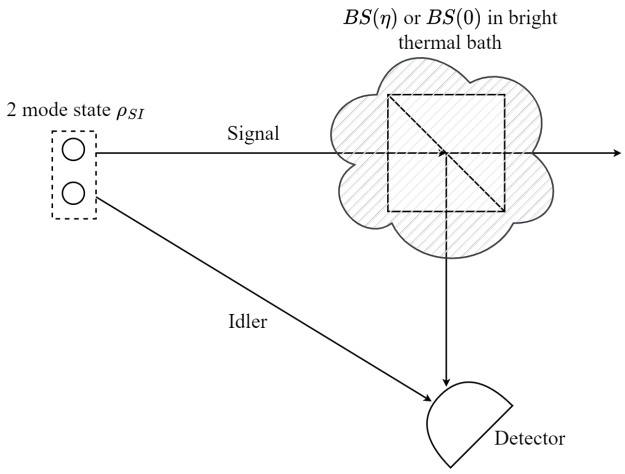

Figure 4: A figure representing the process of quantum illumination. The probe is a 2 mode state with a signal and idler component. The idler is unchanged and the signal goes to probe a beam splitter in a bright thermal bath. Most of the signal will be transmitted away in both cases. The detector receives a 2 mode state over which it does joint measurements to obtain information about the presence or absence of the beam splitter.

- H_0: The object is absent hence the detector only receives photons from background radiation.
- H_1: The object is present and behaves like a beamsplitter with some reflectivity η is present and the detector receives some of the probing photons as well as background radiation.

As it turns out, probing the object using photons that are entangled gives an advantage in the Chernoff bound as shown in [44]. The importance of this result is that the advantage offered by entanglement holds true even in the regime of high background radiation and low reflectivity. This advantage is explored in gaussian states that are maximally entangled in [61]. The state is of the following form

$$|\psi\rangle_{SI} = \sum_{n=0}^{\infty} \sqrt{\frac{N_S^n}{(N_S+1)^{n+1}}} |n\rangle_S |n\rangle_I \tag{39}$$

Here N_S is the average photon number per mode. The performance is quantified using the quantum Bhattacharya bound in [61]. The following bounds hold true for the case of using M copies of either a coherent state with average signal photon number of N_S or the state descried in 39. The reflectivity $\eta \ll 1$ and the background photon radiation is assumed to be $N_B \gg 1$.

$$P_{err,QI} \leq e^{-M\eta^2 N_S/N_B}/2 \quad P_{err,CS} \leq e^{-M\eta^2 N_S/4N_B}/2 \tag{40}$$

Here $P_{err,QI}$ is the error probability while using the entangled state and $P_{err,CS}$ uses unentangled coherent states. The above result clearly shows that we have atleast a factor of 4 improvement in the performance of an entangled state compared to unentangled states.
If we take a closer look at the hypotheses H_0 and H_1, we can define the following maps as the operation.

$$\mathcal{E}_\eta(\rho_{SI}) = \mathrm{Tr}_S(U_\eta(\rho_B \otimes \rho_{SI})U_\eta^\dagger) \tag{41}$$

$$U_\eta = \exp(i \sin^{-1}(\eta)(a_s^\dagger a_b - a_s a_b^\dagger)) \otimes \mathbb{I}_I \tag{42}$$

Here a_b, a_s are the annihilation operators for the background radiation and the signal mode. We can clearly see that the two maps we wish to distinguish between is $\mathcal{E}_0$ and $\mathcal{E}_\eta$. In this case where η is close to zero, this tasks is interchangeable with quantum sensing. The task of quantum sensing is fundamentally a parameter estimation task hence the quantum Fisher information is a very relevant measure here [43, 53]. In a general estimation task we would perform a set of measurements on a parameterized state to retrieve information on the parameters.
Due to the Cramér-Rao bound, we have a fundamental limit for the precision of an unbiased estimator $\hat{\eta}$.

$$\Delta\hat{\eta}^2 \geq \frac{1}{MH} \tag{43}$$

This tells us yet another way to measure the performance of the state. We can write the following expression for the quantum Fisher information.

$$H = 2 \sum_{mn, \lambda_m + \lambda_n \neq 0} \frac{|\langle\phi_m|(\partial_\eta \rho_\eta)|_{\eta=0}|\phi_n\rangle|^2}{\lambda_m + \lambda_n} \tag{44}$$

$$H \approx 8\frac{1 - F(\rho_{\eta=0}, \rho_{\eta=\delta})}{\delta^2} \tag{45}$$

Here $\rho_\eta = \mathcal{E}_\eta(\rho)$ and $\rho_{\eta=0} = \sum_n \lambda_n |\phi_n\rangle\langle\phi_n|$. The formula in equation 44 is exact and the formula in equation 45 is an easier to calculate approximation. The importance of this approximation is how it relates to error probability scaling with Fidelity as highlighted in equation 22. This is mainly explored in [49] where it can be shown that the error probability also approximately scales with the quantum Fisher information.

$$\frac{1}{4}e^{-MH\eta^2/4} \lesssim P_{err}(\rho_{\eta=0}, \rho_\eta) \lesssim \frac{1}{2}e^{-MH\eta^2/8} \tag{46}$$

In [56], it is highlighted that the quantum Fisher information of the TMSV as a probe exceeds that of the Coherent state, but the ratio by which this exceeds begins to approach 1 as the sinal power is increased. In [40], the prospect of using N-photon entangled states is explored and they can outperform in certain cases. From the point of view of all the different cost functions we have, such as the trace distance, the Chernoff bound, the quantum Fisher information and also the relative entropy, we can have very differing families of states which are optimal for this porblem in different interpretations.
The optimal strategy for asymmetric error testing is proven to be using n copies of the two modes squeezed vacuum (TMSV) as the probe [19]. At the same time we have a result for the quantum Chernoff bound [10].

Theorem 5.1 (Optimal probe for quantum illumination)**.** In the limit where $\eta \to 0$ (η being object reflectivity), the two-mode probe for quantum illumination that maximizes the quantum Chernoff bound subject to having a constraint on signal photon number is the TMSV state.

The proof of the above statement can be found in [10]. Hence the TMSV state represents a fundamental limit for two-mode states when we have a constraint on the signal photon number. There are ways to improve error probabilities using photon added or photon subtracted TMSV as shown in [20] but the conclusion that TMSV is optimal when the constraint is on N_S still holds. Keeping all these metrics in mind, we progress toward applying our variational algorithm to this specific problem and see how well our answers stack up against the performance of the two mode squeezed vacuum.

5.2 A primer for Gaussian quantum optics

In this subsection, we will introduce the basics of Gaussian quantum optics [11] along with certain important properties that we make use of. The system is composed of n bosonic harmonic oscillators, each with their own annihilation and creation operators $\hat{a}_k, \hat{a}_k^\dagger$ for $k = 1, \ldots, n$. We will have the commutation relations $[\hat{a}_j, \hat{a}_k] = [\hat{a}_j^\dagger, \hat{a}_k^\dagger] = 0$ and $[\hat{a}_j, \hat{a}_k^\dagger] = \delta_{jk}$. We now will move into a system of units where $\hbar = 1$. We define the following Hermitian position and momentum like operations for each mode.

$$\hat{x}_k = \frac{1}{\sqrt{2}}(\hat{a}_k + \hat{a}_k^\dagger) \quad \hat{p}_k = \frac{1}{i\sqrt{2}}(\hat{a}_k - \hat{a}_k^\dagger) \tag{47}$$

It is clear that $[\hat{x}_j, \hat{p}_k] = \delta_{jk}$. We now define a vector which has each of these operations as a component in the following equation.

$$\hat{r} = \begin{pmatrix} \hat{x}_1 \\ \hat{p}_1 \\ \vdots \\ \hat{x}_n \\ \hat{p}_n \end{pmatrix} \tag{48}$$

This allows us to define the commutation relations as follows.

$$[\hat{r}_j, \hat{r}_k] = i\Omega_{jk} \tag{49}$$

$$\Omega = \bigoplus_{k=1}^{2n} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \mathbb{I}_n \otimes \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{50}$$

The notation of $A \oplus B$ is the direct sum between the matrices A and B hence $A \oplus B = \textbf{diag}(A, B)$. We can then define a covariance matrix between components of $\hat{r}$ as follows.

$$\sigma_{jk} = \frac{1}{2}\langle\{\hat{r}_j, \hat{r}_k\}\rangle - \langle\hat{r}_j\rangle\langle\hat{r}_k\rangle \tag{51}$$

The importance of defining σ and $\hat{r}$ is that there is a certain family of n mode bosonic states which can fully characterized by using only σ and $\bar{r}$ where $\bar{r} = \langle \hat{r} \rangle$. More specifically this occurs if the Hamiltonian used to evolve the state is linear or bilinear in $\hat{a}_k$ and $\hat{a}_k^\dagger$ [57]. The canonical relations force the following uncertainty relation to hold.

$$\sigma + i\frac{\Omega}{2} \geq 0 \tag{52}$$

We can more quantitatively define a Gaussian state by writing a characteristic function of the state [52].

Definition 5.1 (Gaussian state (quantum optics))**.** A state ρ describing a n mode bosonic system of Harmonic oscillators is said to be Gaussian if the characteristic function $\chi[\rho] : \mathbb{R}^{2n} \to \mathbb{C}$ of ρ is Gaussian.

$$\chi[\rho](\Lambda) = \text{Tr}(\rho \exp(-i\Lambda^T \Omega \hat{r})) \tag{53}$$

This must be possible to write as follows.

$$\chi[\rho](\Lambda) = \exp(-\frac{1}{2}\Lambda^T \Omega \sigma \Omega^T \Lambda - i\Lambda^T \Omega \bar{r}) \tag{54}$$

The operation $\exp(-i\Lambda^T \Omega \hat{r})$ is a representation of a displacement in the phase space by the vector Λ which has $2n$ real components. We can equivalently express the state using the Wigner function $W : \mathbb{R}^{2n} \to \mathbb{R}$ which is the Fourier transform of the Characteristic function hence contains all the information in ρ.

$$W(r) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^{2n}} \exp(-ir^T \Omega s)\chi(s)ds \tag{55}$$

An important property of the Gaussian function is that it is a real-valued function and for Gaussian states it is non negative everywhere [13]. Also the Wigner function is normalized.

$$\int_{\mathbb{R}^{2n}} W(r)dr = 1 \tag{56}$$

For Gaussian states the Wigner function takes on the following form.

$$W(r) = \frac{1}{(2\pi)^n \sqrt{\det(\sigma)}} \exp\left(-\frac{1}{2}(r - \bar{r})^T \sigma^{-1}(r - \bar{r})\right) \tag{57}$$

We can see that this is pretty much a Gaussian centred at $\bar{r}$ with covariance σ and so is clearly an apt description for our Gaussian state. An important property of the Wigner function is that the trace of the product of two Gaussian states with covariance matrices σ_1, σ_2 and displacement vectors r_1, r_2 is the integral of the product of their Wigner functions ($\Sigma = \sigma_1 + \sigma_2$).

$$\text{Tr}(\rho_1 \rho_2) = \frac{1}{\sqrt{\det(\Sigma)}} \exp(-\frac{1}{2}(r_1 - r_2)^T \Sigma^{-1}(r_1 - r_2))) \tag{58}$$

Definition 5.2 (Gaussian unitary)**.** A unitary function which takes Gaussian states to Gaussian states can be characterized as a symplectic transform which transforms σ and $\bar{r}$ as follows.

$$\sigma \to \mathbf{F}\sigma\mathbf{F}^T \quad \bar{r} \to \mathbf{F}\bar{r} + \mathbf{d} \tag{59}$$

Here $\mathbf{d} \in \mathbb{R}^{2n}$ is a displacement and the matrix $\mathbf{F} \in \mathbb{R}^{2n \times 2n}$ must satisfy the following condition

$$\mathbf{F}\Omega\mathbf{F}^T = \Omega \tag{60}$$

Hence all unitary operations which are generated from a Hamiltonian that is at most quadratic in $\hat{a}_k$ and $\hat{a}_k^\dagger$ can be characterized into phase, displacements, squeezing and beamsplitters. A detailed list of these operations is summarized in the following table. More generally, we can also introduce positive maps which map Gaussian states to Gaussian states. We will now use this to propose an algorithm which can do hypothesis testing while working with Gaussian states.

Name	Operation	Symplectic transform $(\mathbf{F}, \mathbf{d})$
Displacement	$\hat{D}(\alpha) = \bigotimes_{k=1}^{n} \exp(\alpha \hat{a}_k^\dagger - \alpha^* \hat{a})$	$\mathbf{F} = \mathbb{I}_{2n}, \mathbf{d} = \sqrt{2} \begin{pmatrix} \mathrm{Re}(\alpha_1) \\ \mathrm{Im}(\alpha_1) \\ \vdots \\ \mathrm{Re}(\alpha_n) \\ \mathrm{Im}(\alpha_n) \end{pmatrix}$
Beam splitters	$BS(\eta) = \exp\left(\sin^{-1}(\eta)(a_1 a_2^\dagger - a_1^\dagger a_2)\right)$	$\mathbf{F} = \begin{pmatrix} \sqrt{\eta}\mathbb{I}_2 & \sqrt{1-\eta}\mathbb{I}_2 \\ -\sqrt{1-\eta}\mathbb{I}_2 & \sqrt{\eta}\mathbb{I}_2 \end{pmatrix}, \mathbf{d} = 0$
Single mode squeezing	$S_1(\zeta) = \exp\left(\frac{1}{2}\left(\zeta^* \hat{a}^2 - \zeta(\hat{a}^\dagger)^2\right)\right)$	$\mathbf{F} = \cosh(\lvert\zeta\rvert)\mathbb{I}_2 - \sinh(\lvert\zeta\rvert) S_{\arg(\zeta)}, \mathbf{d} = 0$
Two mode squeezing	$S_2(\zeta) = \exp(\zeta^* \hat{a}_1 \hat{a}_2 - \zeta \hat{a}_1^\dagger \hat{a}_2^\dagger)$	$\mathbf{F} = \begin{pmatrix} \cosh(\lvert\eta\rvert)\mathbb{I}_2 & -\sinh(\lvert\eta\rvert) S_{\arg(\eta)} \\ -\sinh(\lvert\eta\rvert) S_{\arg(\eta)} & \cosh(\lvert\eta\rvert)\mathbb{I}_2 \end{pmatrix}, \mathbf{d} = 0$

5.3 A Gaussian variational quantum algorithm

Gaussian states have a wide array of uses in quantum computing, more specifically in the use of continuous variable quantum computing [22]. Here we will propose a variational algorithm for hypothesis testing using Gaussian modes. The algorithm deals with three modes which we label as q_0, q_S and q_I.

Algorithm 3: Gaussian state variational quantum hypothesis testing with two modes.

Take input as $\mathcal{E}_0$ and $\mathcal{E}_1$ which are maps from single mode Gaussian states to single mode Gaussian states.
Initialize $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ which are parameters for $U(\boldsymbol{\theta})$ and $V(\boldsymbol{\phi})$.
Define convergence condition for cost function $T^{\mathrm{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$.
while $T^{\mathrm{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$ *has not converged* **do**
  for $\boldsymbol{\theta}_i = \boldsymbol{\theta} + \Delta\boldsymbol{\theta}_i$ *and* $\boldsymbol{\phi}_i = \boldsymbol{\phi} + \Delta\boldsymbol{\phi}_i$ *for some set of* $\Delta\boldsymbol{\theta}_i$ *and* $\Delta\boldsymbol{\phi}_i$ **do**
    Run circuit with parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_i$ having applied $\mathcal{E}_0$ on q_S
    Using vacuum measurements on q_0 obtain value of p_0 as $\langle|0\rangle\langle 0|\rangle_0$
    Run circuit with parameters $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_i$ having applied $\mathcal{E}_1$ on q_S
    Using vacuum measurements on q_0 obtain value of p_1 as $\langle|0\rangle\langle 0|\rangle_0$
    Assign $TD_{\boldsymbol{\phi}_i}(\boldsymbol{\theta}_i) = |p_0 - p_1|$
  end
  Using $TD_{\boldsymbol{\phi}_i}(\boldsymbol{\theta}_i)$ update $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ with a classical optimizer.
  Check convergence of $T^{\mathrm{est}}_{\boldsymbol{\phi}}(\boldsymbol{\theta})$.
end
return final values of $\boldsymbol{\theta}$, $\boldsymbol{\phi}$.

This algorithm aims at preparing a state which consists of two modes q_S (signal) and q_I (idler) and optimizing the state to prepare the optimal probe for distinguishing map $\mathcal{E}_0$ and $\mathcal{E}_1$ which is applied on the signal mode q_S. The circuit as shown in figure 5 is essentially the same as the circuit in figure 1 and using the previous results of diamond norm it should be clear that a 2 mode state will be able to reach the bound in equation 36 for a single mode map.
The algorithm described for estimating diamond norm can be extended to any possible setup since the main working principle is using the ancillary qubit q_0 in algorithm 2 as a substitute for measurements on the actual qubit registers q_1 and q_2. We could in principle substitute q_1 and q_2 by bosonic modes and the algorithm will carry out with the same principle since the Naimark extension example given in equation 25. However, it is largely unphysical to construct a physical setup which is capable of entangling a qubit to two bosonic modes and moreover the operations will no longer remain Gaussian.
Instead we propose using an additional mode since experimental realizations with multi-mode bosonic systems have been demonstrated such as the 100 mode device which reached an average of 45 photons [67]. This additional mode is either measured over the vacuum state to use for the trace distance estimation protocol.
To examine the case where we do vacuum state overlap, it is important to note that no two Gaussian states are orthogonal to each other. The overlap can be made as small as needed, but true orthogonality is impossible and hence we cannot reach the true canonical Naimark extension [54] if we choose to use only Gaussian operations. Neverthless, we can get sufficiently close for the task of estimation. The isometry of the canonical Naimark extension in this case is given as follows.

$$V = \mathbb{I}_{q_0} \otimes (\sqrt{\Gamma})_{SI} + i(-|0\rangle\langle 1|_{q_0} + |1\rangle\langle 0|_{q_0} + \mathbb{I} - |0\rangle\langle 0| - |1\rangle\langle 1|) \otimes (\sqrt{\mathbb{I} - \Gamma})_{SI} \tag{61}$$

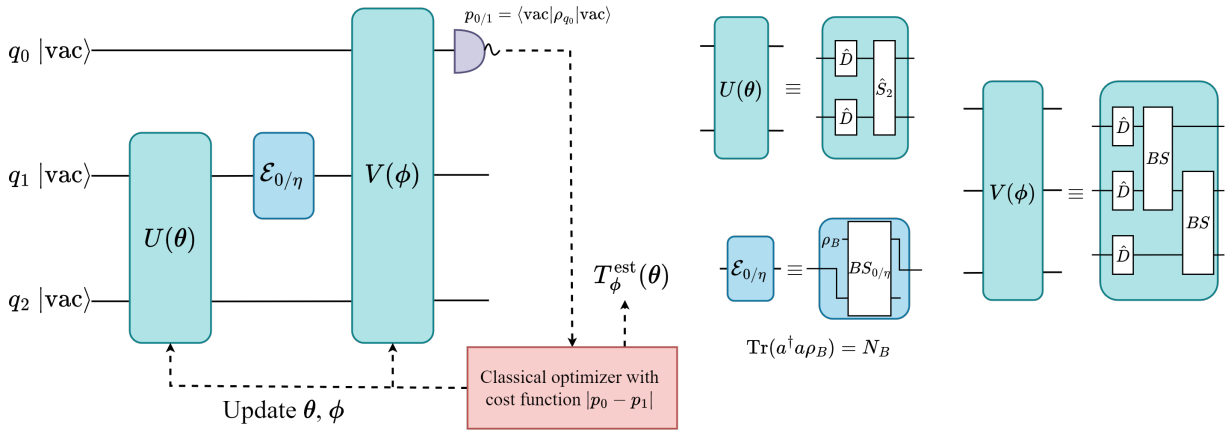

Figure 5: Description of the quantum circuit for variational quantum illumination. The state is prepared over the signal and idler mode using an ansatz which consists of displacements and two-mode squeezing. The ansatz shown for the measurement section consists of displacements and beam splitters. The simulations contain controlled phase gates for in the measurement section as well. The readout can be either measure the probability it is vacuum or the expectation value of the parity operator.

On the other hand, we can frame this as trying to perform a measurement on some n mode state by entangling it to a 1 mode system. This can be written as follows.

$$\text{Tr}(V^\dagger(P_0 \otimes \mathbb{I}_n)V(|0\rangle\langle 0| \otimes \rho)) = \text{Tr}(P\rho) \tag{62}$$

Here P_0 is a a projection which is in $L(\mathcal{H})$, the set of linear operators from $\mathcal{H}$ to $\mathcal{H}$ and $\mathcal{H}$ is the Hilbert space of a single mode of light. P is a projection in $L(\mathcal{H}^{\otimes n})$ and ρ is a density operator in $D(\mathcal{H}^{\otimes n})$ (set of density operators in $L(\mathcal{H}^{\otimes n})$) and $\mathbb{I}_n$ is the identity operator in $L(\mathcal{H}^{\otimes n})$. Since we want the above equation to hold for all ρ, we can rewrite it as follows.

$$V^\dagger(P_0 \otimes \mathbb{I}_n)V = \mathbb{I}_1 \otimes P \tag{63}$$

Here we are applying a transformation from one linear operator to another which means that as long as the norm of both $P_0 \otimes \mathbb{I}_n$ and $\mathbb{I}_1 \otimes P$ are equal, we can find a V. We can construct a $V = (\text{SWAP}_{1,2} \otimes \mathbb{I}_{n-1})(\mathbb{I}_1 \otimes V')$ where it performs a swap between the first two modes and then does some unitary V' only on the subsystem of n modes. This transforms the measurement from the space of the first mode to the space of the n modes.

$$(\mathbb{I}_1 \otimes V')^\dagger(\text{SWAP}_{1,2} \otimes \mathbb{I}_{n-1})^\dagger(P_0 \otimes \mathbb{I}_n)(\text{SWAP}_{1,2} \otimes \mathbb{I}_{n-1})(\mathbb{I}_1 \otimes V') = \mathbb{I}_1 \otimes V'^\dagger(P_0 \otimes \mathbb{I}_{n-1})V'^\dagger \tag{64}$$

This shows that we can construct any projection of form $P = V'^\dagger(P_0 \otimes \mathbb{I}_{n-1})V'^\dagger$ where V' is Gaussian since the swap operation between two modes can be trivially represented as a passive Gaussian operation. This recipe shows us that while we may not be able to construct the canonical Naimark extension of the optimal POVM, we can construct a Naimark extension that performs a POVM on the n mode subspace using a single-mode ancillary measurement.

Hence we can estimate trace distance overlap with vacuum which is also physically realizable as a measurement using a photon counter for measurement where the two outcomes are either zero photon number or non-zero [22]. The simulations were carried out using the Strawberry Fields [12, 38] and The Walrus [26] python packages.

5.4 Results from Fock simulations

Additionally instead of using q_0 as a bosonic mode, one could use a qubit in its place and allow the use of non-Gaussian operations. This can be simulated in the Fock space. For this the state preparation part is the same as the Gaussian circuit's state preparation. In the part where the measurement is optimized operations such as a controlled two-mode squeezing and controlled beam-splitters where the control is given to the qubit q_0.
These simulations are included in this report for the sake of completeness to show that there are alternative ways of applying the algorithm developed here.

6 Conclusion

The novel algorithm presented in this report shows capabilities in offering ways to distinguish quantum maps and has been applied to the task of quantum illumination and shows promising results. The models used here can be extended

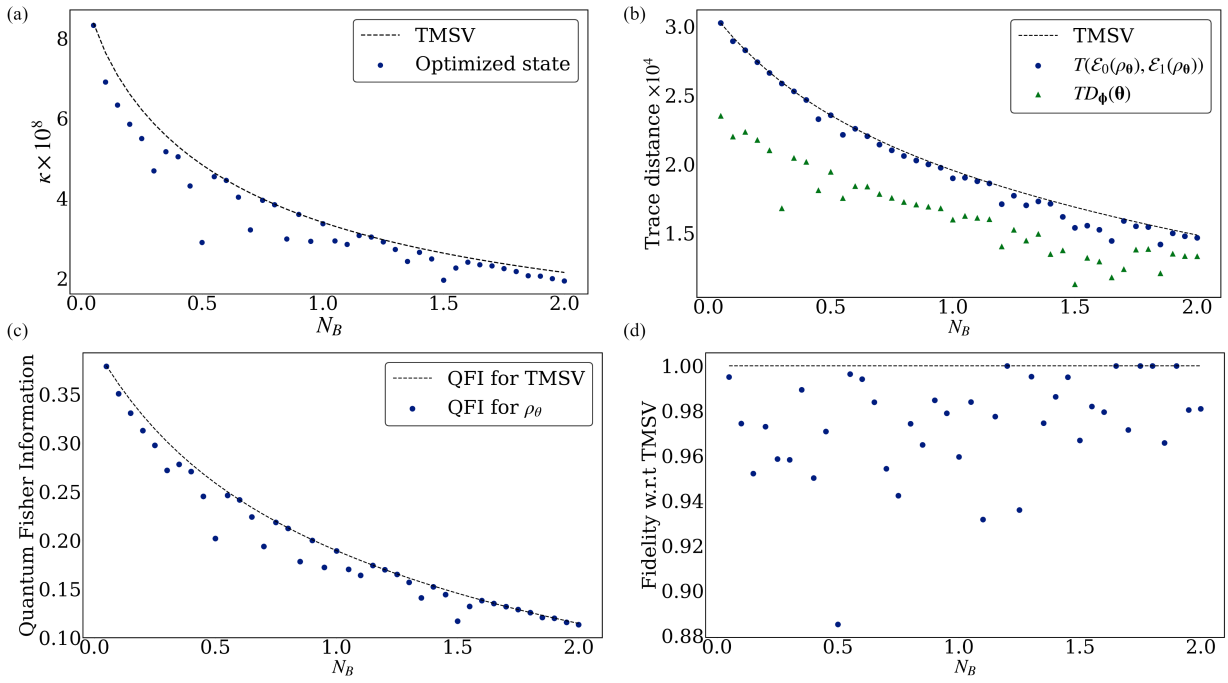

Figure 6: Simulation results using Gaussian backend with vacuum measurement. (a) Compares chernoff bound of the optimized state to TMSV, (b) compares trace distance (blue dots are true trace distance and green triangles are estimated trace distance), (c) compares QFI and (d) shows fidelity of the optimized state to TMSV.

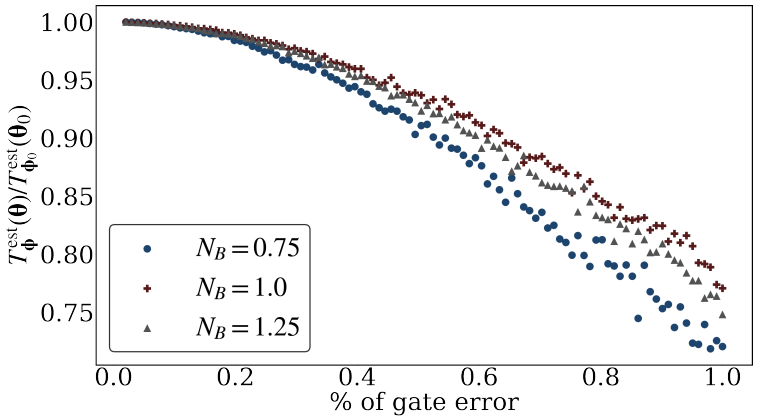

Figure 7: Noise robustness of variational quantum illumination. The quantity on y axis is the ratio of $T_{\boldsymbol{\phi}}^{\text{ve}}(\boldsymbol{\theta})$ to the value it takes with zero gate error. These are results that have been optimized with the constraint that $N_S = 0.1$ following which the optimized results were then simulated with Gaussian noise for each parameter to test robustness. The values plotted are the average after 500 runs with the Gaussian noise with percentage of gate error being the variance.

for tasks such as quantum reading [55] and quantum ranging [68] as well and more generally for larger scale quantum map discrimination on qubits. While the states we obtain in our optimization closely match the results of the two mode squeezed vacuum, an interesting thing to observe is that there is a clear manifold of optimality here of which the TMSV is only one part.

There were alternate optimizations done where 2 mode squeezing resources were not given to the state preparation but 1 mode squeezing resources were. In these it was able to prepare states which have exactly the same performance as the TMSV for the Helstrom bound but fell short in the Chernoff bound and more interestingly were very different from the TMSV. These states also were putting more photons into the idler than into the signal since the signal photons had an equality constraint in them but the idler photons did not have any constraints to follow. This was an

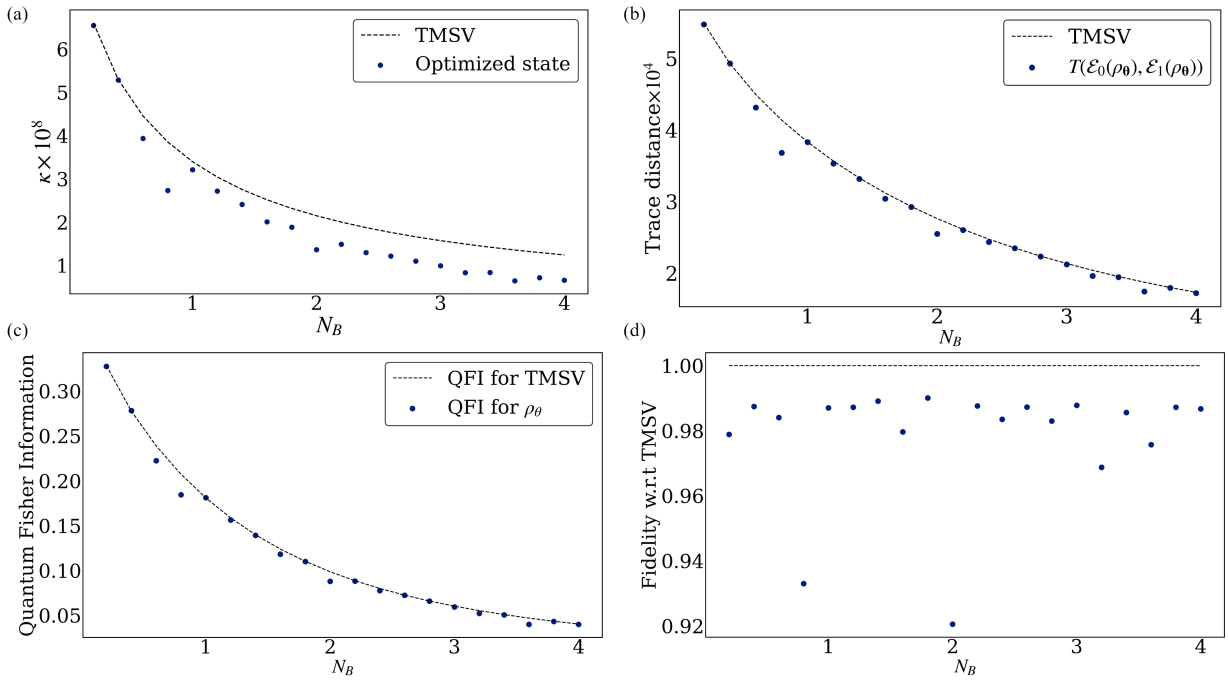

Figure 8: Simulation results of variational optimization for the task of quantum illumination which prepares the state in two bosonic modes (simulated in the Fock space) and uses an ancillary qubit for obtaining measurement probabilities on the two modes to use to optimize the estimate for trace distance. (a) Shows how the optimized state's Chernoff bound κ compares to that of the TMSV state, (b) compares the true trace distance, (c) compares the quantum Fisher information and finally (d) shows the fidelity of the prepared state to TMSV.

interesting consequence since the circuit could have prepared an exact TMSV but clearly chose not to which means the optimization landscape may have a lot of rich information.
It must be noted that all these simulations were done for extremely low depth circuits which could have had much more layers. This is especially relevant in the trace distance estimation section since we limit ourselves to one displacement followed by a controlled phase gate and then beam splitters. An important study to conduct on this would be the expressibility since there could be lots of uses for expressible ansatzes in continuous variable quantum computing [37] and extending this to study quantum sensing like tasks for Gaussian states [23].

Acknowledgements

I would like to thank Prof. Sai Vinjanampathy for guiding me throughout this project in navigating the various involved topics and offering very valuable feedback and discussions throughout this project.

References

[1] Rochisha Agarwal et al. *Estimating distinguishability measures on quantum computers*. 2021. DOI: 10.48550/ARXIV.2108.08406. URL: https://arxiv.org/abs/2108.08406.

[2] Dorit Aharonov, Alexei Kitaev, and Noam Nisan. "Quantum circuits with mixed states". In: *Proceedings of the thirtieth annual ACM symposium on Theory of computing*. 1998, pp. 20–30.

[3] K. M. R. Audenaert et al. "Discriminating States: The Quantum Chernoff Bound". In: *Phys. Rev. Lett.* 98 (16 Apr. 2007), p. 160501. DOI: 10.1103/PhysRevLett.98.160501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.98.160501.

[4] Jacob L. Beckey et al. "Variational quantum algorithm for estimating the quantum Fisher information". In: *Phys. Rev. Research* 4 (1 Feb. 2022), p. 013083. DOI: 10.1103/PhysRevResearch.4.013083. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.4.013083.

[5] Avraham Ben-Aroya and Amnon Ta-Shma. "On the complexity of approximating the diamond norm". In: *arXiv preprint arXiv:0902.3397* (2009).

[6] Giuliano Benenti and Giuliano Strini. "Computing the distance between quantum channels: usefulness of the Fano representation". In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 43.21 (2010), p. 215508.

[7] Kishor Bharti et al. "Noisy intermediate-scale quantum algorithms". In: *Reviews of Modern Physics* 94.1 (2022), p. 015004.

[8] Lennart Bittel and Martin Kliesch. "Training variational quantum algorithms is np-hard". In: *Physical Review Letters* 127.12 (2021), p. 120502.

[9] J. J . Bollinger et al. "Optimal frequency measurements with maximally correlated states". In: *Phys. Rev. A* 54 (6 Dec. 1996), R4649–R4652. DOI: 10.1103/PhysRevA.54.R4649. URL: https://link.aps.org/doi/10.1103/PhysRevA.54.R4649.

[10] Mark Bradshaw et al. "Optimal probes for continuous-variable quantum illumination". In: *Phys. Rev. A* 103 (6 June 2021), p. 062413. DOI: 10.1103/PhysRevA.103.062413. URL: https://link.aps.org/doi/10.1103/PhysRevA.103.062413.

[11] Jonatan Bohr Brask. *Gaussian states and operations – a quick reference*. 2021. DOI: 10.48550/ARXIV.2102.05748. URL: https://arxiv.org/abs/2102.05748.

[12] Thomas R Bromley et al. "Applications of near-term photonic quantum computers: software and algorithms". In: *Quantum Science and Technology* 5.3 (2020), p. 034010.

[13] William B Case. "Wigner functions and Weyl transforms for pedestrians". In: *American Journal of Physics* 76.10 (2008), pp. 937–946.

[14] M. Cerezo et al. "Cost function dependent barren plateaus in shallow parametrized quantum circuits". In: *Nature Communications* 12.1 (Mar. 2021). DOI: 10.1038/s41467-021-21728-w. URL: https://doi.org/10.1038%5C%2Fs41467-021-21728-w.

[15] Marco Cerezo et al. "Variational quantum algorithms". In: *Nature Reviews Physics* 3.9 (2021), pp. 625–644.

[16] Pietro Cerone. "Estimation of Relative Entropy Using Novel Taylor-Like Representations". In: *RGMIA research report collection* 5.4 (June 2002). URL: https://vuir.vu.edu.au/17790/.

[17] Ranyiliu Chen et al. "Variational quantum algorithms for trace distance and fidelity estimation". In: *Quantum Science and Technology* 7.1 (2021), p. 015019.

[18] G Mauro D'Ariano, Matteo GA Paris, and Massimiliano F Sacchi. "Quantum tomography". In: *Advances in Imaging and Electron Physics* 128 (2003), pp. 206–309.

[19] Giacomo De Palma and Johannes Borregaard. "Minimum error probability of quantum illumination". In: *Phys. Rev. A* 98 (1 July 2018), p. 012101. DOI: 10.1103/PhysRevA.98.012101. URL: https://link.aps.org/doi/10.1103/PhysRevA.98.012101.

[20] Longfei Fan and M. Suhail Zubairy. "Quantum illumination using non-Gaussian states generated by photon subtraction and photon addition". In: *Phys. Rev. A* 98 (1 July 2018), p. 012319. DOI: 10.1103/PhysRevA.98.012319. URL: https://link.aps.org/doi/10.1103/PhysRevA.98.012319.

[21] Hamza Fawzi and Omar Fawzi. "Efficient optimization of the quantum relative entropy". In: *Journal of Physics A: Mathematical and Theoretical* 51.15 (Mar. 2018), p. 154003. DOI: 10.1088/1751-8121/aab285. URL: https://doi.org/10.1088/1751-8121/aab285.

[22] Alessandro Ferraro, Stefano Olivares, and Matteo GA Paris. "Gaussian states in continuous variable quantum information". In: *arXiv preprint quant-ph/0503237* (2005).

[23] Yang Gao and Hwang Lee. "Bounds on quantum multiple-parameter estimation with Gaussian state". In: *The European Physical Journal D* 68.11 (2014), pp. 1–7.

[24] Barry M Garraway. "The Dicke model in quantum optics: Dicke model revisited". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369.1939 (2011), pp. 1137–1155.

[25] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. "Quantum Metrology". In: *Phys. Rev. Lett.* 96 (1 Jan. 2006), p. 010401. DOI: 10.1103/PhysRevLett.96.010401. URL: https://link.aps.org/doi/10.1103/PhysRevLett.96.010401.

[26] Brajesh Gupt, Josh Izaac, and Nicolás Quesada. "The Walrus: a library for the calculation of hafnians, Hermite polynomials and Gaussian boson sampling". In: *Journal of Open Source Software* 4.44 (2019), p. 1705.

[27] Carl W. Helstrom. "Quantum detection and estimation theory". In: *Journal of Statistical Physics* 1.2 (June 1969), pp. 231–252. ISSN: 1572-9613. DOI: 10.1007/BF01007479. URL: https://doi.org/10.1007/BF01007479.

[28] Fumio Hiai and Dénes Petz. "The proper formula for relative entropy and its asymptotics in quantum probability". In: *Communications in mathematical physics* 143.1 (1991), pp. 99–114.

[29] Osamu Hirota. *Application of quantum Pinsker inequality to quantum communications*. 2020. DOI: 10.48550/ARXIV.2005.04553. URL: https://arxiv.org/abs/2005.04553.

[30] Alexander Semenovich Holevo. "Bounds for the quantity of information transmitted by a quantum communication channel". In: *Problemy Peredachi Informatsii* 9.3 (1973), pp. 3–11.

[31] Carl J Huberty. "Historical origins of statistical testing practices: The treatment of Fisher versus Neyman-Pearson views in textbooks". In: *The Journal of Experimental Education* 61.4 (1993), pp. 317–333.

[32] J.R. Johansson, P.D. Nation, and Franco Nori. "QuTiP 2: A Python framework for the dynamics of open quantum systems". In: *Computer Physics Communications* 184.4 (2013), pp. 1234–1240. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2012.11.019. URL: https://www.sciencedirect.com/science/article/pii/S0010465512003955.

[33] J.R. Johansson, P.D. Nation, and Franco Nori. "QuTiP: An open-source Python framework for the dynamics of open quantum systems". In: *Computer Physics Communications* 183.8 (2012), pp. 1760–1772. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2012.02.021. URL: https://www.sciencedirect.com/science/article/pii/S0010465512000835.

[34] Abhinav Kandala et al. "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets". In: *Nature* 549.7671 (Sept. 2017), pp. 242–246. ISSN: 1476-4687. DOI: 10.1038/nature23879. URL: http://dx.doi.org/10.1038/nature23879.

[35] Athena Karsa et al. "Quantum illumination with a generic Gaussian source". In: *Phys. Rev. Research* 2 (2 June 2020), p. 023414. DOI: 10.1103/PhysRevResearch.2.023414. URL: https://link.aps.org/doi/10.1103/PhysRevResearch.2.023414.

[36] Raphael Kaubruegger et al. "Quantum Variational Optimization of Ramsey Interferometry and Atomic Clocks". In: *Phys. Rev. X* 11 (4 Dec. 2021), p. 041045. DOI: 10.1103/PhysRevX.11.041045. URL: https://link.aps.org/doi/10.1103/PhysRevX.11.041045.

[37] Nathan Killoran et al. "Continuous-variable quantum neural networks". In: *Physical Review Research* 1.3 (2019), p. 033063.

[38] Nathan Killoran et al. "Strawberry fields: A software platform for photonic quantum computing". In: *Quantum* 3 (2019), p. 129.

[39] S. Kullback and R. A. Leibler. "On Information and Sufficiency". In: *The Annals of Mathematical Statistics* 22.1 (1951), pp. 79–86. DOI: 10.1214/aoms/1177729694. URL: https://doi.org/10.1214/aoms/1177729694.

[40] Su-Yong Lee, Yong Sup Ihn, and Zaeill Kim. "Quantum illumination via quantum-enhanced sensing". In: *Phys. Rev. A* 103 (1 Jan. 2021), p. 012411. DOI: 10.1103/PhysRevA.103.012411. URL: https://link.aps.org/doi/10.1103/PhysRevA.103.012411.

[41] Erich L Lehmann. "The Fisher, Neyman-Pearson theories of testing hypotheses: one theory or two?" In: *Journal of the American statistical Association* 88.424 (1993), pp. 1242–1249.

[42] Ke Li. "Second-order asymptotics for quantum hypothesis testing". In: *The Annals of Statistics* 42.1 (Feb. 2014). ISSN: 0090-5364. DOI: 10.1214/13-aos1185. URL: http://dx.doi.org/10.1214/13-AOS1185.

[43] Jing Liu et al. "Quantum Fisher information matrix and multiparameter estimation". In: *Journal of Physics A: Mathematical and Theoretical* 53.2 (Dec. 2019), p. 023001. DOI: 10.1088/1751-8121/ab5d4d. URL: https://doi.org/10.1088%5C%2F1751-8121%5C%2Fab5d4d.

[44] Seth Lloyd. "Enhanced Sensitivity of Photodetection via Quantum Illumination". In: *Science* 321.5895 (2008), pp. 1463–1465. DOI: 10.1126/science.1160627. eprint: https://www.science.org/doi/pdf/10.1126/science.1160627. URL: https://www.science.org/doi/abs/10.1126/science.1160627.

[45] Christian D Marciniak et al. "Optimal metrology with programmable quantum sensors". In: *Nature* 603.7902 (2022), pp. 604–609.

[46] Jarrod R McClean et al. "Barren plateaus in quantum neural network training landscapes". In: *Nature communications* 9.1 (2018), pp. 1–6.

[47] Johannes Jakob Meyer, Johannes Borregaard, and Jens Eisert. "A variational toolbox for quantum multi-parameter estimation". In: *npj Quantum Information* 7.1 (2021), pp. 1–5.

[48] Michael A Nielsen and Isaac Chuang. *Quantum computation and quantum information*. 2002.

[49] Changsuk Noh, Changhyoup Lee, and Su-Yong Lee. *Quantum illumination with non-Gaussian states: Bounds on the minimum error probability using quantum Fisher information*. 2021. DOI: 10.48550/ARXIV.2110.06891. URL: https://arxiv.org/abs/2110.06891.

[50] Michael Nussbaum and Arleta Szkoła. "The Chernoff lower bound for symmetric quantum hypothesis testing". In: *The Annals of Statistics* 37.2 (2009), pp. 1040–1057. DOI: 10.1214/08-AOS593. URL: https://doi.org/10.1214/08-AOS593.

[51] T. Ogawa and H. Nagaoka. "Strong converse and Stein's lemma in quantum hypothesis testing". In: *IEEE Transactions on Information Theory* 46.7 (2000), pp. 2428–2433. DOI: 10.1109/18.887855.

[52] Stefano Olivares. "Quantum optics in the phase space". In: *The European Physical Journal Special Topics* 203.1 (2012), pp. 3–24.

[53] Matteo GA Paris. "Quantum estimation for quantum technology". In: *International Journal of Quantum Information* 7.supp01 (2009), pp. 125–137.

[54] Matteo GA Paris. "The modern tools of quantum mechanics". In: *The European Physical Journal Special Topics* 203.1 (2012), pp. 61–86.

[55] Stefano Pirandola. "Quantum Reading of a Classical Digital Memory". In: *Phys. Rev. Lett.* 106 (9 Mar. 2011), p. 090504. DOI: 10.1103/PhysRevLett.106.090504. URL: https://link.aps.org/doi/10.1103/PhysRevLett.106.090504.

[56] M. Sanz et al. "Quantum Estimation Methods for Quantum Illumination". In: *Phys. Rev. Lett.* 118 (7 Feb. 2017), p. 070803. DOI: 10.1103/PhysRevLett.118.070803. URL: https://link.aps.org/doi/10.1103/PhysRevLett.118.070803.

[57] Bonny L Schumaker. "Quantum mechanical pure states with Gaussian wave functions". In: *Physics Reports* 135.6 (1986), pp. 317–408.

[58] Clayton Scott and Robert Nowak. "A Neyman-Pearson approach to statistical learning". In: *IEEE Transactions on Information Theory* 51.11 (2005), pp. 3806–3819.

[59] Ramamurti Shankar. *Principles of quantum mechanics*. Springer Science & Business Media, 2012.

[60] Jeffrey H Shapiro and Seth Lloyd. "Quantum illumination versus coherent-state target detection". In: *New Journal of Physics* 11.6 (June 2009), p. 063045. ISSN: 1367-2630. DOI: 10.1088/1367-2630/11/6/063045. URL: http://dx.doi.org/10.1088/1367-2630/11/6/063045.

[61] Si-Hui Tan et al. "Quantum Illumination with Gaussian States". In: *Physical Review Letters* 101.25 (Dec. 2008). ISSN: 1079-7114. DOI: 10.1103/physrevlett.101.253601. URL: http://dx.doi.org/10.1103/PhysRevLett.101.253601.

[62] Jirawat Tangpanitanon et al. "Expressibility and trainability of parametrized analog quantum systems for machine learning applications". In: *Physical Review Research* 2.4 (Dec. 2020). DOI: 10.1103/physrevresearch.2.043364. URL: https://doi.org/10.1103%5C%2Fphysrevresearch.2.043364.

[63] Marco Tomamichel. *Quantum Information Processing with Finite Resources*. Springer International Publishing, 2016. DOI: 10.1007/978-3-319-21891-5. URL: https://doi.org/10.1007%5C%2F978-3-319-21891-5.

[64] John Watrous. "Distinguishing quantum operations having few Kraus operators". In: *arXiv preprint arXiv:0710.0902* (2007).

[65] John Watrous. "Quantum computational complexity". In: *arXiv preprint arXiv:0804.3401* (2008).

[66] Mark M Wilde. "Sequential decoding of a general classical-quantum channel". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 469.2157 (2013), p. 20130259.

[67] Han-Sen Zhong et al. "Quantum computational advantage using photons". In: *Science* 370.6523 (2020), pp. 1460–1463.

[68] Quntao Zhuang. "Quantum Ranging with Gaussian Entanglement". In: *Phys. Rev. Lett.* 126 (24 June 2021), p. 240501. DOI: 10.1103/PhysRevLett.126.240501. URL: https://link.aps.org/doi/10.1103/PhysRevLett.126.240501.