Machine Learning for Quantum Phase Estimation

Nelson Filipe Ferreira de Almeida Costa
nelson.filipe.costa@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

October 2018

Abstract

Metrology is the science of the measurement and, as such, encapsulates a wide range of aspects, from defining the units of measurement and realizing them in practice, to the fundamental limits that can be achieved in the precise estimation of parameters. These limits are set by the underlying nature of the measurement scheme. For schemes with independent classical measurement inputs, this limit is defined in metrology by the Standard Quantum Limit (SQL), while for schemes exploring quantum enhancing techniques at the input level, the limit of the measurement precision is set by the Heisenberg Limit (HL). As the employment of quantum enhanced techniques in measurement schemes is still limited by the current level of technology, it is important to develop alternatives to overcome the SQL. Thus, the objective of this work was to devise a quantum phase estimation scheme capable of beating this limit, resorting exclusively to machine learning techniques with a non-entangled input. For this task, two algorithms were studied: *Differential Evolution* and *Particle Swarm Optimization*. Under their optimal configurations, both algorithms were applied to an adaptive phase estimation scheme associated to different configurations of the Mach-Zehnder interferometer. Under the ideal and the noisy interferometer configuration the algorithms were able to go beyond the SQL. However for the noisy interferometer with photon loss this was only verified for a certain number of photons. These shortcomings, nevertheless, could be simply overcome by allowing for a larger amount of time for the learning process of the algorithms.

Keywords: Metrology, Quantum Phase Estimation, Machine Learning, Differential Evolution, Particle Swarm Optimization

1. Introduction

Metrology is the science of measurement and encapsulates a wide range of aspects, from defining the units of measurement and realising them in pratice, to the fundamental limits that can be achieved in the precise estimation of parameters [1, 2, 3]. Since these limits are set by the underlying theory of Nature, quantum mechanics, research on the foundations of this theory can lead to important developments in science and technology, such as the search for gravitational wave detection [4], in atomic magnetometry [5, 6] and atomic clock synchronization [7].

Common to any measurement process is the preparation of a probe in a desired initial state, followed by its interaction with the physical systems that encases the unknown parameter to be estimated and, finally, the measurement of the probe to infer the value of the unknown parameter [8]. However, regardless of its scale and due to its physical nature, this measurement process is often plagued by systematic or statistical errors [9]. The source of the former stems from the physical realization of the measurement, either from imperfect calibration of the measurement instruments or from an altogether poorly conceived measurement scheme. While the source of the latter can be either accidental, resulting from an insufficient control of the measurement scheme, or fundamental, deriving from the nature of the physical measurement and whether or not it explores the laws of quantum mechanics [10].

Fortunately, the impact of the statistical errors, regardless of their origin, can be minimized by repeating the measurement process and averaging the resulting outcomes, as a consequence of the central limit theorem [11]. In metrology, for N independent measurements, the central limit theorem leads to a scaling in error of $1/\sqrt{N}$ and is refered to as the Standard Quantum Limit (SQL) [12]. However, this limit is associated to experiments which do not fully exploit the quantum nature of the system under investigation. Consequently, the SQL is not a fundamental bound as it can be surpassed by quantum enhanced strategies. Nonetheless, these non-classical strategies, due to their quantum nature, are also bounded by the Heisenberg uncertainty relations, which pose the ultimate limit to the precision to which any measurement can be performed. This ultimate limit bounds the error of N averaged independent results to a deviation scaling proportionally to $1/N$, which in metrology is refered to as the Heisenberg Limit (HL) [12].

Naturally, with the developlment of technology and the growing ability to prepare and control quantum resources, the exploration of these quantum enhancing techniques is a often employed strategy for quantum metrology [13]. Unfortunately, one main consequence of the quantum nature of the measurement is the quantum back action, which states that the extraction of information from a system can give rise to a feedback effect in which the sys-

tem configuration after the estimation is collapsed by the measurement outcome [14]. As most applications benefit from conserving the quantum nature of the physical state of the system after the estimation process, it is important to search for non-demolishing quantum measurement techniques [15, 16].

Therefore, instead of focusing on employing quantum enhancing techniques at the input level of the measurement scheme, by for example seeking quantum entangled states of a probe of multiple particles, a recent and more experimentally feasible approach is gaining relevance: machine learning [17]. This approach seeks to exploit the advantages of machine learning algorithms to manipulate the output results of the measurement scheme to drive it towards estimations with precisions exceeding the limit set by SQL [18].

Through the manipulation of the available quantum resources and the exploitation of the ever growing field of machine learning, metrology is a field ripe for improvement [18, 19, 20, 21]. Whether by replacing the often all too ilusive mathematical model of the quantum measurement dynamics, or by further developing the precision of the already existing quantum estimation scheme. This is possible since machine learning enables the treatment of the whole physical system as a black box, thus overcoming the shortcomings associated with quantum metrology [22, 23].

The employment of machine learning algorithms to the task of quantum parameter estimation has already seen promising results, namely in the task of quantum phase estimation [24, 25, 26, 27, 28, 29]. However, machine learning algorithms in metrology keeps being exclusively applied to measurement schemes that are already exploiting resources of quantum nature. It is, therefore, desirable to explore their application to classical forms of input.

2. Adaptive Phase Estimation Scheme

To perform the estimation of an unknown phase, an adaptive scheme can be employed with the help of a Mach-Zenhder interferometer [30]. The Mach-Zenhder interferometer, depicted in Figure 1, consists of two input ports, two output ports, two beam splitters and two phase shifters, one whose phase is known and which can be controlled, $\tilde{\phi}$, and other whose phase value is unknown and the object of the estimation, ϕ.

Under this estimation scheme, a known input set of N photons, either in the state $|0\rangle$ or in the state $|1\rangle$, is sequentially injected into the interferometer and its output measurement outcome used to update the controllable phase shifter in preparation for the next input particle. After the N^{th} photon is injected to the interferometer and its outcome measured, the value of the updated controllable phase shifter is considered to be the estimated value of the unknown phase shifter.

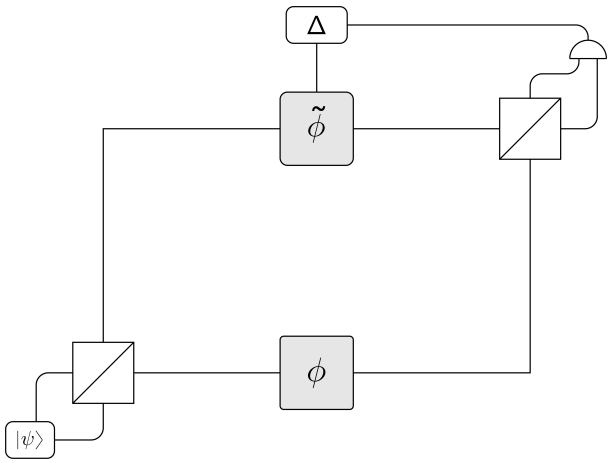

Figure 1: Adaptive quantum phase estimation scheme. An N photon input bundle is injected to the Mach-Zehnder interferometer one photon at a time through one of its two input ports. Inside the interferometer there are two phase shifters, one whose value is unknown, phase shifter ϕ, an on whose value can be controlled, phase shifter $\tilde{\phi}$. After each photon is measured at one of the two output ports of the interferometer, the controllable phase shifter uses the measurement outcome to adjust its value according to some policy Δ. After the N^{th} photon is injected into the interferometer and its final state measured, the estimated value of the unknown phase shifter ϕ is inferred from the value of the controllable phase shifter $\tilde{\phi}$.

The policy used to update the controllable shifter is, therefore, the key ingredient of all adaptive phase estimation schemes. Representing the policy by an N-dimensional vector Δ, the update rule for the controllable phase shifter to be used in this work is:

$$\tilde{\phi}_m = \tilde{\phi}_{m-1} + (-1)^{x_m}\Delta_m \tag{1}$$

To evaluate the performance of a policy it must be considered its ability to arrive to a correct estimation despite the value of the unknown phase shifter ϕ. Hence, a common choice for assessing the performance of a policy is the following equation:

$$S = \left| \sum_{k=1}^{K} \frac{e^{i\theta_k}}{K} \right| \tag{2}$$

Where $\theta_k = \tilde{\phi}_k - \phi_k$ and $K = N^2$. Note that the value K represents the number of different values of the unknown phase shifter considered in the learning task imposed to the machine learning algorithms. Hence, it can be viewed as the total number of training instances used in the estimation scheme and this particular choice for its value is one which, when employed in parallel with Equation (2), was shown to guarantee the convergence of the algorithms [29].

However, as both the SQL and the HL are expressed in the form of a measurement imprecision, the performance criteria of Equation (2) must too be expressed in the form of a imprecision value. Fortunately, this imprecision value

can be quantified by the Holevo variance, which is specially suited for periodically bounded parameters, as:

$$V_H = S^{-2} - 1 = \left(\left| \sum_{k=1}^{K} \frac{e^{i\theta_k}}{K} \right| \right)^{-2} - 1 \tag{3}$$

Finally, for the designed estimation process to bear resemblance to a real life estimation experiment, the interferometer must be able to accommodate noise in its application and reading of the controllable phase shifter, as well as entertain a photon loss probability, in which case the algorithms are instructed to do nothing regarding the update of the controllable phase shifter.

3. Machine Learning Algorithms

As discussed in the previous sections, the objective of this work is to devise a policy that successfuly finds a correct estimation of an unknown interferometric phase based on a given input of photons. For this task two machine learning algorithms will be employed. Stemming from genetic and evolutionary algorithms, the first algorithm to be explored will be the Differential Evolution (DE) [31]. In addition, and inspired by social behaviour simulations, the second algorithm to be explored will be the Particle Swarm Optimization (PSO) [32].

3.1. Differential Evolution

In order to successfully employ the DE algorithm to the task for which it was designed, it is important to first study its performance with its different possible controllable parameter configurations and find which one delivers the most promising results. There are two parameters driving the learning process of the DE algorithm: the crossover parameter C and the amplification parameter F.

To evaluate the performance of the algorithm there are basically two criteria that must be verified: its ability to converge to a solution and, more importantly, its ability to converge to a valid solution. To verify the convergence of the solutions, it is important to remember that at each iteration of the algorithm there will be different candidate solutions, with each population representing a candidate solution. As the algorithm iterates, the candidate solutions should move closer to each other, finally converging to a final solution. Thus, one way of inferring the convergence of the algorithm is by calculating the deviation from the average value of each population for each entry of their candidate solution and then average these values in what must be considered the convergence value L of the algorithm:

$$L = \sum_{j}^{N} \frac{1}{N} \left(\sum_{i}^{P} \frac{\bar{x}_j - x_{i,j}}{P} \right) \tag{4}$$

Where $\bar{x}_i$ corresponds to the average value of entry i over all the candidate solutions, whereas $x_{j,i}$ corresponds to the entry i of the candidate solution vector j. Lower values of L occur when all the candidate solutions are relatively close to each other, indicating that the algorithm has converged to a solution. On the other hand, larger values of L indicate that the algorithm was not able to converge to a solution within the given number of iterations.

To study the convergence of the algorithm and its ability to converge to a valid solution, the control tests will be done only for varying values of the controllable parameters of interest, F and C, while keeping all the other values constant. Therefore, the tests were conducted for $N = 10$ photons, $P = 20$ populations, $G = 50$ generations and 1000 training instances.

The first parameter control test to be made is exclusively concerned with the ability of the algoritm to converge to a solution, regardless of its validity. The convergence values obtained for each controllable parameter configuration are displayed in Table 1.

F \ C	0	0.2	0.4	0.6	0.8	1
0	1.01659	**0.06956**	**0.02612**	**0.02235**	**0.00189**	**0.00096**
0.2	0.93340	0.37261	**0.03258**	**0.01161**	**0.02037**	**0.00936**
0.4	1.08966	0.74905	0.72306	**0.02352**	**0.02019**	**0.02686**
0.6	1.16473	0.83644	0.77643	0.52262	**0.02506**	**0.01930**
0.8	1.21830	1.13563	0.93326	1.09012	1.01326	**0.06910**
1	1.31005	1.24603	1.24334	1.30810	1.13585	1.27975

Table 1: Convergence values $L(F, C)$ for the Differential Evolution algorithm. Each value on this table corresponds to the convergence average of the different solutions found by the DE algorithm for the corresponding set of controllable parameters. Values in bold represent configurations where convergence was achieved. Convergence was considered only for values of $L \leq 0.1256$, which corresponds to a maximum dispersion of approximately 2% of the entire 2π search space of each member of the different candidate solutions.

Inspecting the results obtained and considering that convergence was only obtained for the values in bold, it is possible to see a pattern appearing among the different combinations. That is, for all combinations where the value of the amplification constant F is strictly smaller than the crossover constant C the algorithm is able to converge within the given number of generations. This leads to the first conclusion of this study: $F < C$ guarantees convergence of the DE algorithm.

This newfound rule can be understood from a physical interpretation of the process. Larger values for the crossover parameter C encourage the different populations to experience more member configurations, thus covering a bigger area of the entire search space. This increased exploration of the search space, in consequence, leads to a more exhaustive pursuit of the best policy solution. However, by increasing the amplification constant F, the step of this search process is also increased, which may lead to overlooking possibly desirable candidate solutions and not allowing the algorithm to ultimately converge. Hence, smaller values of F ensure a more careful examination of the search space at each iteration of the algorithm.

Having arrived at the convergence values of the algorithm, it is also important to verify if the converged so-

lution is indeed a valid solution. After all, there is no interest in converging to a solution if it is not a correct one. To do so, one must refer to the already mentioned Holevo variance (3). Recalling that lower values of the Holevo variance correspond to solutions with an higher degree of accuracy, that is, solutions which lead to a very close approximation of the actual value that is being estimated.

Keeping exactly the same configuration setup as before, the performance results obtained are displayed on Table 2

F \ C	0	0.2	0.4	0.6	0.8	1
0	−0.37349	**-1.20117**	**-0.78445**	**-0.77582**	**-0.70829**	**-0.41960**
0.2	5.21114	−1.62661	**-1.37373**	**-1.44006**	**-1.02438**	**-1.09483**
0.4	−0.38969	−0.23975	0.61811	**-1.56134**	**-1.29756**	**-1.56106**
0.6	0.31307	−1.18959	−1.15885	−1.36995	**-1.83210**	**-1.54534**
0.8	0.63256	−1.03954	0.36938	−0.66454	0.87628	**-1.67842**
1	−0.17861	1.24603	2.55647	0.61314	4.23639	0.08642

Table 2: Performance values $V_H(F, C)$ for the Differential Evolution algorithm. Each value on this table corresponds to the Holevo variance for the corresponding controllable parameters. Values in bold represent the configurations which have achieved convergence in Table 1. Lower variance values are obtained by solutions that lead to more precise estimations of the unknown parameter that is being measured.

Regarding now the different performances obtained for the different parameter combinations and remembering the already discussed results and consequent conclusion, it is easy to verify that the parameter configurations that previously lead the algorithm to converge to a solution, do indeed lead to a valid solution if $F > 0$. Hence, the previously found rule can now be made more specific: $F < C, \forall F > 0$ guarantees the convergence of the DE algorithm to a valid solution.

A more careful inspection of the obtained results in Table 2, shows that increasing values of the crossover parameter, C, leads to poorer results. Complementing this information with the results obtained in Table 1, this can be understood as a greedy exploitation of the environment. The same results are also found for lower values of the amplification parameter F. Greedy exploitation occurs when the algorithm is mostly focused in the immediate reward, disregarding a more thourough exploration of the surrounding search space and, hence, getting stuck at the first local minima it finds.

In addition, another remark can and should be made regarding the apparently good performances obtained even when the rule $F < C,\ \forall F > 0$ is not verified. These are misleading results and can be quickly disregarded, considering that as the algorithm is told to stop regardless of having converged to a solution after $G = 50$ generations, the resulting averaged solution among all the $P = 20$ populations may haphazardly coincide with a valid one. These situations must, nevertheless, be regarded as chanceful events and not be considered as possible valid solutions, as the algorithm was not able to indeed converge to any solution.

In order to have a more sensible understanding of the general landscape of the parameter configurations, a more thourough control test was repeated three times, following the same configurations as before and its results averaged and displayed on Table 3. Note that this time, the iterating step between parameters was reduced in order to avoid overlooking possible candidate parameter configurations.

F \ C	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.2	-	−1.40030	−1.48014	−1.54717	−1.40994	−1.10353	−1.19144	−1.54393	−1.05422
0.3	-	-	−1.69780	−1.65036	−1.48013	−1.30289	−1.40975	−1.55556	−1.12456
0.4	-	-	-	−1.75715	−1.71944	−1.44125	−1.47032	−1.46008	−1.45009
0.5	-	-	-	-	−1.54267	−1.64228	−1.71262	−1.49899	−1.48686
0.6	-	-	-	-	-	−1.62500	−1.73507	−1.52755	−1.59668
0.7	-	-	-	-	-	-	−1.86063	−1.64474	−1.46033
0.8	-	-	-	-	-	-	-	−1.27129	−1.22001
0.9	-	-	-	-	-	-	-	-	−0.77590
1	-	-	-	-	-	-	-	-	-

Table 3: Performance results $V_H(F, C)$ for the Differential Evolution algorithm. Each value on this table corresponds to the averaged performance obtained for each corresponding parameter configuration after three runs of the DE algorithm. Only parameter configurations following the rule $F < C,\ \forall F > 0$ were considered for a more efficient employment of the computational time.

Observing the results in Table 3 it is possible to confirm the previously made suppositions. Therefore, bigger values of F, while still considering only configurations where $F < C, \forall F > 0$, lead to better results. Ensuring that the step size at each iteration of the algorithm is not too big as to overstep a possible valid solution, it should not be too small either to avoid getting stuck in a local minimum. This explains why larger values of F, while still in the interval $F < C$, lead to better results in detriment of a faster convergence. In fact, low values of F lead to a very quick convergence of the algorithm, with it usually getting stuck in local minima.

Hence, considering the study made, the parameter configuration that leads to optimal results, even if it may take a larger number of iterations, and which will be used for the ultimate quantum phase estimation task will be the amplification constant set to $F = 0.7$ and the crossover constant set to $C = 0.8$.

3.2. Particle Swarm Optimization

Once again, in order to successfully employ the PSO algorithm to the quantum phase estimation task it is necessary to find which controllable parameters are able to deliver the best results. Previously, for the DE algorithm there were only two parameters to consider, the amplification and the crossover constant. For the PSO algorithm, however, there are four parameters of interest. Two of these parameters, the α and β constants, are mostly concerned with the ability of the algorithm to converge to a valid solution, while the other two, the w and v_{max} constants, regulate the speed at which the parameter will converge to a solution.

As before, the first step consists of studying the overall ability of the algorithm to converge to a solution for the different possible parameter configurations. In this case, as only the parameters α and β have a direct impact on the convergence process, the other two controllable parameters can be set to constant values for this first test, as well as all the already previosuly fixed parameters. Here the same fixed parameters as before will be kept. Thus the tests will be conducted for $N = 10$ photons, for $P = 20$ populations, $G = 50$ generations and $k = 1000$ training instances. The two additional controllable parameters for the PSO algorithm will be set to $w = 0.8$ and to $v_{max} = 0.2$ for the first test.

The convergence value L of the algorithm can be calculated through the same formula as before, Equation 4. Keeping this formula in mind, the results for the different parameter configurations of α and β are shown in Table 4.

β \ α	0	0.2	0.4	0.6	0.8	1
0	1.32312	1.32920	1.31393	1.33413	1.40445	1.31051
0.2	**0.12386**	**0.10720**	0.21930	0.23794	0.47360	0.45795
0.4	**0.10139**	**0.09982**	**0.09738**	0.17423	0.23284	0.42804
0.6	**0.07710**	**0.07881**	**0.08076**	**0.09016**	0.12964	0.24098
0.8	**0.08206**	**0.08206**	**0.10719**	**0.07340**	**0.08058**	0.15889
1	**0.08349**	**0.09757**	**0.10079**	**0.08873**	**0.07076**	**0.09328**

Table 4: Convergence values $L(\alpha, \beta)$ for the Particle Swarm Optimization algorithm. Each value on this table corresponds to the convergence average of the different solutions found by the DE algorithm for the corresponding set of controllable parameters. Values in bold represent configurations where convergence was achieved. Convergence was considered only for values of $L \leq 0.1256$, which corresponds to a maximum dispersion of approximately 2% of the entire 2π search space of each member of the different candidate solutions.

Considering the results obtained, it is possible to highlight a few patterns among the different configurations analyzed. First and foremost, it is possible to see that as long as the β value is bigger than zero the algorithm is able to converge to a solution, even if it may take more than $G = 50$ iterations. This can be intuitively understood by remembering that the β parameter defines the desirability of each individual in following the best found solution of the entire group of candidate solutions until that point. Ergo, as long as this value is bigger than zero, each individual will feel the urge to move towards the best globally found solution.

Secondly, for β values equal or larger than α the algorithm is always able to converge to a solution within the stipulated number of iterations. However, for β values much larger than α the algorithm will converge within very few iterations, possibly getting stuck in local minima and overlooking possible valid solutions. These first conclusions can be made into the following rules: $\beta > 0$ guarantees convergence of the algorithm and $\beta \geq \alpha$ guarantees convergence of the algorithm within the stipulated number of iterations.

These recently found rules, appear to show a pattern among collectives of individuals in search for a common goal. From the rule $\beta > 0$, and as was already mentioned, it is possible to verify that as long as each individual candidate solution has knowledge of the best solution found by its collective group of candidate solutions it is able to converge to desired policy. However, the rule $\beta \geq \alpha$ provides a more interesting observation among collective cooperative systems. This rule mathematically shows that while each particle lends more value to the collective knowledge and interest of the group, defined by parameter β, against its own individual impulses, defined by parameter α, it is always able to arrive at a valid solution for the cooperative search problem.

Having arrived at the conditions that ensure the convergence of the algorithm within the designated number of iterations, it is now possible to study which of these parameter configurations lead to valid solutions. To that end, and once more referring to Equation 3, a second test was performed regarding the performance of the algorithm. The results obtained are displayed in Table 5.

β \ α	0	0.2	0.4	0.6	0.8	1
0	1.30823	2.37173	0.70764	1.57271	1.75400	6.05629
0.2	**-1.54746**	**-1.36506**	−1.66229	−1.45299	−0.83828	−1.02809
0.4	**-1.40162**	**-1.55983**	**-1.52946**	−1.68306	−1.60050	1.21259
0.6	**-1.62291**	**-1.51998**	**-1.93241**	**-1.36500**	−1.66050	−1.54079
0.8	**-1.72272**	**-1.45393**	**-1.56148**	**-1.58284**	**-1.83612**	−1.21324
1	**-1.68694**	**-1.32900**	**-1.39564**	**-1.67255**	**-1.59088**	**-1.48256**

Table 5: Performance values $V_H(\alpha, \beta)$ for the Particle Swarm Optimization algorithm. Each value on this table corresponds to the Holevo variance for the corresponding controllable parameters. Values in bold represent the configurations which have achieved convergence in Table 4. Lower variance values are obtained by solutions that lead to more precise estimations of the unknown parameter that is being measured.

Regarding these values it is possible to state that all the converged solutions found by the algorithm are valid ones, as all result in very low Holevo variance values. To determine which one of these solutions attains the optimal results, the same test was repeated three times and its results averaged. Notice that, once again, there is a smaller increment between each parameter step in order to avoid overlooking any possibly better parameter configuration. The results obtained are shown in Figure 6.

β \ α	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0	-	-	-	-	-	-	-	-	-	-	-
0.1	−1.09416	−1.52975	-	-	-	-	-	-	-	-	-
0.2	−1.38981	−1.36721	−1.41636	-	-	-	-	-	-	-	-
0.3	−1.40107	−1.48233	−1.49761	−1.51265	-	-	-	-	-	-	-
0.4	−1.46069	−1.61997	−1.58691	−1.63963	−1.53599	-	-	-	-	-	-
0.5	−1.40841	−1.55627	−1.32771	−1.66059	−1.57489	−1.32981	-	-	-	-	-
0.6	−1.45473	−1.52685	−1.58137	−1.54599	−1.68291	−1.69163	−1.53121	-	-	-	-
0.7	−1.42202	−1.56280	−1.66484	−1.51098	−1.64009	−1.64734	−1.55591	−1.50283	-	-	-
0.8	−1.68987	−1.56790	−1.55538	−1.57038	−1.65150	−1.40194	−1.57077	−1.57445	−1.79330	-	-
0.9	−1.45009	−1.58862	−1.50124	−1.68416	−1.66751	−1.67140	−1.66556	−1.52001	−1.60796	−1.74155	-
1	−1.62750	−1.32900	−1.35927	−1.40775	−1.55569	−1.61148	−1.60453	−1.55369	−1.55450	−1.60948	−1.57246

Table 6: Performance results $V_H(\alpha, \beta)$ for the Particle Swarm Optimization algorithm. Each value on this table corresponds to the averaged performance obtained for each corresponding parameter configuration after three runs of the PSO algorithm. Only parameter configurations following the rule $\beta \geq \alpha$ were considered for a more efficient employment of the computational time.

Considering now the results in Table 6, it is possible to determine which parameter configuration delivers better results. Considering only the quality of the solutions found by the different configurations, there does not seem to exist a big fluctuation between the results. However, despite the very similar results, there is one which stands out, even if only slightly, and it is the pair $\alpha = 0.8$ and $\beta = 0.8$. For this reason, the first two of the four controllable parameters of the Particle Swarm Optimization algorithm to be adopted for the estimation of the quantum phase are $\alpha = 0.8$ and $\beta = 0.8$.

Having arrived at the optimal values for the controllable parameters α and β it is still necessary to study how the remaining two parameters, the update weight w and the maximum velocity v_{max}, influence the convergence time and performance of the algorithm. It is important to note that as these two parameters have a bigger influence on the convergence time, it is important to choose a configuration that ensures that the algorithm is able to converge in the given number of iterations. Hence, setting the α and β parameters to a constant value the convergence values for the remaining two parameters are organized according to Table 7.

v_{max} \ w	0.2	0.4	0.6	0.8	1
0.1	0.18387	0.15023	**0.08428**	**0.08831**	**0.09111**
0.2	**0.10215**	**0.08578**	**0.11030**	**0.08927**	**0.06185**
0.3	**0.09191**	**0.10147**	**0.08468**	**0.12393**	**0.09591**
0.4	0.21586	**0.08683**	0.13114	**0.09963**	0.13237
0.5	0.22722	0.27548	**0.08743**	0.14339	0.16289

Table 7: Convergence values $L(v_{max}, w)$ for the Particle Swarm Optimization algorithm. Each value on this table corresponds to the convergence average of the different solutions found by the PSO algorithm for the corresponding set of controllable parameters. Values in bold represent configurations where convergence was achieved. Convergence was considered only for values of $L \leq 0.1256$, which corresponds to a maximum dispersion of approximately 2% of the entire 2π search space of each member of the different candidate solutions.

Even though not all parameter configurations in Table 7 were able to converge according to the $L \leq 0.1256$ criteria in the given number of iterations, it is still possible to see that they were nevertheless close to achieving the desired value. Yet, to confirm that the solutions to which these configurations have converged are close to a correct solution, their performance was, as previously done, studied according to equation 3 and displayed in Table 8.

v_{max} \ w	0.2	0.4	0.6	0.8	1
0.1	−1.12405	−1.42792	**-1.66680**	**-1.15179**	**-1.35982**
0.2	**-1.42288**	**-1.39072**	**-1.66791**	**-1.56242**	**-1.63574**
0.3	**-1.39536**	**-1.42527**	**-1.81183**	**-1.41887**	**-1.65919**
0.4	−1.42395	**-1.51831**	−1.80198	**-1.91076**	−1.43842
0.5	−1.31452	−1.32775	**-1.71329**	−1.61740	−1.40301

Table 8: Performance values $V_H(v_{max}, w)$ for the Particle Swarm Optimization algorithm. Each value on this table corresponds to the Holevo variance for the corresponding controllable parameters. Values in bold represent the configurations which have achieved convergence in Table 7. Lower variance values are obtained by solutions that lead to more precise estimations of the unknown parameter that is being measured.

Inspecting the values in Table 8 it is possible to confirm that, despite some fluctuatuion, all parameter configurations are able to converge to an acceptable solution. This further confirms the hypothesis that both the update weight w and the maximum velocity v_{max} parameter do not directly influence the ability of the algorithm to converge to a correct solution, although playing an important role in ensuring that the algorithm does indeed converge to a solution. As the range of each of these two parameters was bounded by small values for the maximum velocity, this second statement is less evident, yet a reduced step is ultimately necessary to ensure that the different solutions do not swing around a desired solution without converging.

Therefore, compared to the study made for the parameter values of *alpha* and *beta*, for the w and v_{max} parameters it is possible to have a more flexible choice of optimal configurations. Indeed, there is more than one set of configurations which could be considered and that would ensure a convergence of the algorithm to a correct solution. Yet, to be coherent, the same configuration used for testing the parameters α and β will be employed for the final results. Therefore, the final optimal configuration of the different controllable parameters for the quantum phase estimation task are $\alpha = 0.8$, $\beta = 0.8$, $w = 0.8$ and $v_{max} = 0.2$.

4. Results

Having performed a thorough analysis on the performance of both algorithms response to the different parameter configurations, it is finally possible to apply them to the quantum phase estimation task. The optimal parameters obtained in the previous chapter are organized

on Table 9.

F	C	α	β	w	v_{max}
0.7	0.8	0.8	0.8	0.8	0.2

Table 9: Optimal controllable parameters for the machine learning algorithms.

To better study the performance of both algorithms in estimating the value of the unknown phase of the Mach-Zehnder interferomer, the analysis will be broken into three parts with increasing degree of complexity. Firstly, it will only be considered an utopic interferometer where any source of noise and photon loss will be neglected. Secondly, it will be considered an interferometer already experiencing a gaussian noise but still neglecting photon loss. Thirdly and finally, the interferometer will be experiencing both a gaussian noise and a photonic loss probability. The fixed parameters considered for all the following experimental setups are grouped in Table 10.

P	G	K
$20 + 2 \cdot int(N/10) - 1$	100	$10N^2$

Table 10: Fixed parameters for the machine learning algorithms.

As the number of photons N being injected into the interferometer increases, so must the number of populations P and the number of training instances K in Table 10 to accomodate the increasing complexity of the search space, which as mentioned in Chapter 2 scales linearly with N. Note that $int(\cdot)$ represents the integer part of the number. Note as well that the number of generations was changed to $G = 100$. This adjustment was made to accomodate the possible longer number of iterations required for the convergence of the algorithms due to the increased complexity associated to an increasing number of photons.

For the following tests, the performance of the algorithms will be studied for a different number of photons, ranging from 10 to 25[1]. In addition, and to further optimize the computational time, an heuristic will be implemented to the convergence process of the algorithms. When the converging index for each position of all the P candidate solutions is below a threshold value of 2% of their entire 2π search space, the estimation process will be concluded at that current iteration and convergence will be assumed.

4.1. Ideal Interferometer

The first experimental setup consists on an utopic interferometer where all noise and photon loss are neglected. Applying the estimation scheme studied to this scenario it was possible to arrive at the results in Figure 2.

Observing the results in Figure 2 it is immediately possible to see that both algorithms were indeed able to arrive at estimation precisions superior to the SQL as was the objective of this work. However, on a more careful inspection, it is also possibe to note that the increased performance curve for the increasing number of photons seems to break at $N \geq 20$. Intuitively, the estimation precision of the measurement process should increase with the number of photons, following both the SQL and HL curves, as the interferometer would have more information to evaluate the value of the unknown phase shifter. Yet, this is not a malfunction of the algorithms, but a consequence of the time available. Unfortunately, the increasing number of photons also leads directly to an increased computational time and requires a larger number of generations to converge to a final policy capable of guiding the measurement scheme to a more precise estimation. Thus, the $G = 100$ generations are rendered insufficient to guarantee convergence for $N \geq 20$ photons.

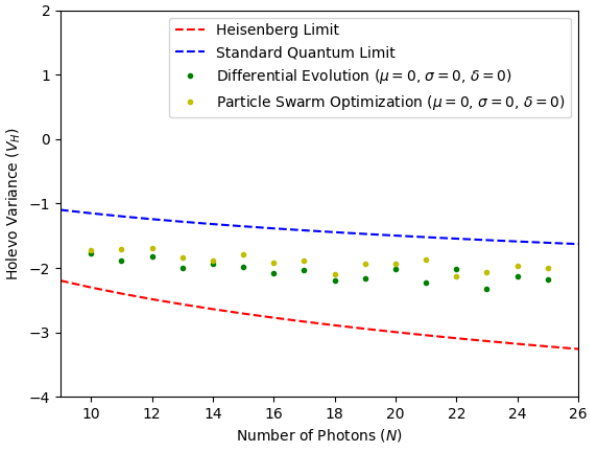

Figure 2: Final results for the ideal interferometer. Both the results obtained for the clean interferometer with the DE and the PSO algorithms obtained measurement precisions beyond the SQL. However, the performance of both algorithms seems to slightly break the slope it was following for $N > 20$ photons. This, however, can be explained by the limited number of iterations, $G = 100$, for the convergence of the estimation process, as for increasing N the algorithms require more iterations to converge.

At this point, it could be objected that a larger number of generations would be sufficient to overcome this convergence problem and it would indeed be correct. However, it also must be taken in consideration that with the current configuration for each data point in Figure 2 the algorithms were already taking more than five days before arriving to an estimation policy. A further optimization of the complexity of the code may lead to faster computational times and, in consequence, allow for a larger number of generations to be considered for each estimation process.

Additionally, another remark might be made regarding the individual performance of each algorithm. Considering only the number of photons used, it is possible to notice

[1]The maximum number of photons for which the study was made was limited by the computational time. Note that for $N > 20$ photons the algorithm was already taking more than five days for each point in each of the figures that follow.

that the DE algorithm is always able to arrive at better results than the PSO algorithm, with the only exception at $N = 22$. However, before arriving at a definite conclusion between the adaptability of each algorithm for the present estimation task, it would be needed to see how they scale with larger numbers of photons. It is indeed possible, that for a larger N one of the algorithms might break. But again, for this supplementary study to be made possible a further optimization of the code, or even access to more powerful computing reosurces, would be required.

4.2. Interferometer with Noise

Having obtained satisfying results for the ideal interferometer, it is time to study the resilience of the algorithms to noise in the interferometer. Denoting the Gaussian noise average value by μ, its corresponding standard deviation by σ and the photonic loss probability by δ, the algorithms will at this point be tested for increasing amounts of Gaussian noise at the controllable phase shifter. As the controllable phase shifter can only assume values in a periodic and bounded interval, $\tilde{\phi} \in [-\pi, \pi]$, it is sufficient to fix the average value of the Gaussian noise at $\mu = 0$ and only increase its standard deviation value. Therefore, the increasing noise in the interferometer will be studied for $\sigma = \{0.2, 0.4, 0.8\}$. The results obtained under these conditions are shown in Figure 3.

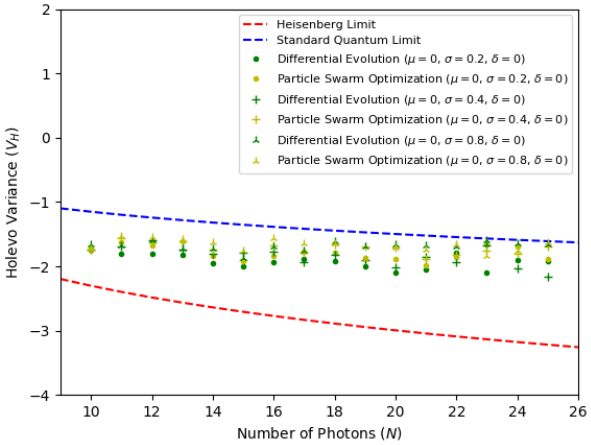

Figure 3: Final results for the noisy interferometer. Both the results obtained for the interferometer with different values of noise with the DE and the PSO algorithms obtained measurement precisions beyond the SQL. However, the performance of both algorihtm clearly breaks the slope it was following for $N > 20$ photons. This, however, can be explained by the limited number of iterations, $G = 100$, for the convergence of the estimation process, as for increasing N the algorithms require more iterations to converge. It is also possible to see that increasing values of noise in the interferometer lead to poorer results in the precision of the estimation process.

Once again, the results obtained in Figure 3 were able to attain precision values superior to the theoretical limit of the SQL. From the figure it is also possible to see that although both algorithms are able to exceed the SQL, as the noise fluctuations increase the precision of the algorithms at arriving to a precise estimation of unknown phase shifter diminishes. Even if is an unfortunate result, it was nevertheless expected that the precision would decrease when in presence of noise.

Considering only the worse case scenario where there was a noise deviation of $\sigma = 0.8$, which implies a fluctuation of around 12% of the entire possible range of values for the controllable phase shifter $\tilde{\phi} \in [-\pi, \pi]$, and the degree of precision obtained despite it, it is possible to evaluate the robustness of the algorithms to a more realistic experimental configuration of the quantum phase estimation task. It is indeed possible to see that the implemented algorithms are capable of arriving at valid policy solutions in spite of noises in the order of magnitude of at least 10% on the entire value interval of the controllable phase shifter in the interferometer.

In addition, it is also possible to notice the already observed convergence issues in Figure 2 for $N \geq 20$ photons. This effect is, in fact, even more notorious in Figure 3 due the more chaotic dispersion of the performance values after this threshold. Likewise, as in the ideal interferometric configuration, the DE algorithm is still consistently able to devise better policy schemes than the PSO algorithm for the quantum phase estimation task.

4.3. Interferometer with Noise and Photon Loss

So far, the algorithms have achieved promising results in both the clean and the noisy interferometric configuration. However, besides their resilience to noise, it is important to study their performance under the possibility of photonic loss in the interferometer. This final test attemps to recreate the most realistic realization of a real life experiment of the estimation process that is being described in this work. Hence, it will determine the likelihood of the present methodology being successfully applied to current metrology challenges. Analogue to the noise scenario, the algorithm will be tested against an increasing probability of photonic loss occurence according to $\delta = \{2\%, 4\%\}$. Note that the noise parameters are left constant, $\mu = 0$ and $\sigma = 0.4$, as the focus of interest now falls exclusively in the ability of the algorithms to entertain photon loss. With these interferometric configurations, the results obtained are displayed in Figure 4.

On a first inspection, the results obtained this time in Figure 4 are not as satisfactory as the ones obtained in the two previous scenarios. It is, nevertheless, still possible to verify that both algorithms are able to surpass the SQL, despite the noise and photonic loss in the interferometer, for the first number of photons considered. However, as N increases the algorithms are not able to follow the performance curve imposed by the two physical limits, SQL and HL, within the number of generations $G = 100$ available for their convergence. Indeed, their performance seems to break at $N \geq 17$.

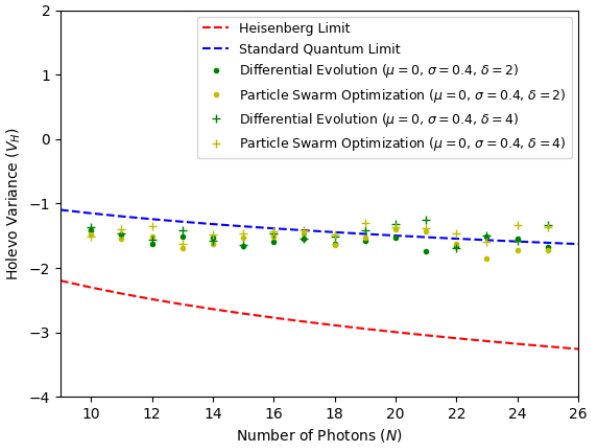

Figure 4: Final results for the noisy interferometer with photon loss. Both the results obtained for the interferometer with different values of photon loss percentage with the DE and the PSO algorithms obtained measurement precisions beyond the SQL up to $N = 19$ photons. The break in performance that occurs for $N \geq 20$ photons can be explained by the limited number of iterations, $G = 100$, for the convergence of the estimation process, as for increasing N the algorithms require more iterations to converge. It is also possible to see that the increasing values of the photon loss percentage in the interferometer lead to poorer results in the precision of the estimation process.

Whereas in Figure 2 and Figure 3 the performance breaking point was only visible for $N \geq 20$ photons, the increased complexity associated with the experimental configuration of Figure 4, requires an even larger number of generations to guarantee that both algorithms are able to converge to a valid policy. As the number of generations remained the same, $G = 100$, throughout all three different experiments, it is natural that the current configuration arrives faster at a breaking point in its performance.

Nevertheless, as both the DE and PSO algorithms are always able to exceed the SQL for the first set of input photons, there is no reason to believe that this shortcoming would not be overcome by allowing a larger number of generations for the algorithms to explore the entire search space of the problem. This would require, however, either a further optimization of the code or a more powerful computer to handle the learning process of the algorithms.

5. Conclusions

The objective of this work was to devise a quantum phase estimation scheme capable of beating the Standard Quantum Limit without requiring any quantum enchancing techniques in the photonic input, resorting exclusively to machine learning techniques with a non-entangled input. For this task, two algorithms were studied and the controllable parameters that promised the best results analyzed. Under their optimal configurations, both algorithms were applied to the adaptive phase estimation scheme associated to the Mach-Zehnder interferometer. This task was then performed under three different configurations, corresponding to an increasing resemblance to a real life experience.

Under this thorough and careful study, it was possible to arrive at very satisfying results regarding the quantum phase estimation task in quantum metrology, as shown in Figure 5.

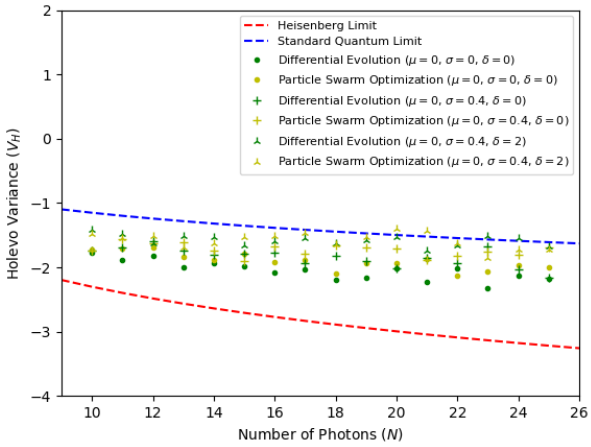

Figure 5: Final results for the different interferometer configurations. The comparison of the performances obtained for the different interferometric configurations show that the ideal interferometer is able to attain better precision values than the interferometer with only noise and the interferometer with noise alongside with photon loss. It is also possible to see that the occurance of photon loss further decreases the precision values of estimation process. The break in performance approximately seen at $N > 20$ can be explained by the limited number of iterations, $G = 100$, for the convergence of the estimation process, as for increasing N the algorithms require more iterations to converge.

It is indeed possible to see in Figure 5 that the main objective of this entire study was accomplished. Both of the implemented algorithms were able to exceed the SQL, the theoretical limit deemed unbeatible for classical inputs. Futhermore, with a slightly better performance throughout all different experimental configurations by the DE algorithm when compared to the PSO algorithm, it is also possible to benchmark both approaches regarding the problem of quantum phase estimation in favour of the former algorithm.

Nevertheless, and despite the positive results, it was also possible to see in this work where the current methodology may also be improved to guarantee even better results.

- Optimize the code to allow for shorter runtimes at each iteration of both algorithms. This would allow the same amount of time to be employed by larger number of iterations for each algorithm to converge to a valid solution and, therefore, significantly improve their performance for increasing number of input particles.

- Parallely, allow the code to run on more a more powerful computer. Similarly to the previous reasoning, this would allow for a better exploitation of the available time resources which posed the ultimate barrier to the attainable precision for the increasing number of input particles in this work.

References

[1] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum-enhanced measurements: beating the standard quantum limit. *Science*, 306(5700):1330–1336, 2004.

[2] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Physical review letters*, 96(1):010401, 2006.

[3] Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Advances in quantum metrology. *Nature photonics*, 5(4):222, 2011.

[4] J Abadie, BP Abbott, R Abbott, TD Abbott, M Abernathy, C Adams, R Adhikari, C Affeldt, B Allen, GS Allen, et al. A gravitational wave observatory operating beyond the quantum shot-noise limit. *Nature Physics*, 7(12):962, 2011.

[5] S Danilin, AV Lebedev, A Vepsäläinen, GB Lesovik, G Blatter, and GS Paraoanu. Quantum-enhanced magnetometry by phase estimation algorithms with a single artificial atom. *npj Quantum Information*, 4(1):29, 2018.

[6] Jonatan Bohr Brask, Rafael Chaves, and J Kołodyński. Improved quantum magnetometry beyond the standard quantum limit. *Physical Review X*, 5(3):031010, 2015.

[7] Johannes Borregaard and Anders S Sørensen. Near-heisenberg-limited atomic clocks in the presence of decoherence. *Physical review letters*, 111(9):090801, 2013.

[8] Masanao Ozawa. Realization of measurement and the standard quantum limit. In *Squeezed and Nonclassical Light*, pages 263–286. Springer, 1989.

[9] Calyampudi Radhakrishna Rao, Calyampudi Radhakrishna Rao, Mathematischer Statistiker, Calyampudi Radhakrishna Rao, and Calyampudi Radhakrishna Rao. *Linear statistical inference and its applications*, volume 2. Wiley New York, 1973.

[10] Yi Peng and Heng Fan. Information theoretical analysis of quantum metrology. *arXiv preprint arXiv:1801.07071*, 2018.

[11] Harald Cramér. *Mathematical methods of statistics (PMS-9)*, volume 9. Princeton university press, 2016.

[12] Géza Tóth and Iagoba Apellaniz. Quantum metrology from a quantum information science perspective. *Journal of Physics A: Mathematical and Theoretical*, 47(42):424006, 2014.

[13] Artur Widera, Olaf Mandel, Markus Greiner, Susanne Kreim, Theodor W Hänsch, and Immanuel Bloch. Entanglement interferometry for precision measurement of atomic scattering properties. *Physical review letters*, 92(16):160406, 2004.

[14] HP Robertson. Hp robertson, phys. rev. 34, 163 (1929). *Phys. Rev.*, 34:163, 1929.

[15] Vladimir B Braginsky, Vladimir Borisovich Braginsky, and Farid Ya Khalili. *Quantum measurement*. Cambridge University Press, 1995.

[16] Vladimir B Braginsky, Yuri I Vorontsov, and Kip S Thorne. Quantum nondemolition measurements. *Science*, 209(4456):547–557, 1980.

[17] Ryszard S Michalski, Jaime G Carbonell, and Tom M Mitchell. *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.

[18] Alessandro Lumino, Emanuele Polino, Adil S Rab, Giorgio Milani, Nicolò Spagnolo, Nathan Wiebe, and Fabio Sciarrino. Experimental phase estimation enhanced by machine learning. *arXiv preprint arXiv:1712.07570*, 2017.

[19] Nathan Wiebe and Chris Granade. Efficient bayesian phase estimation. *Physical review letters*, 117(1):010503, 2016.

[20] Marin Bukov, Alexandre GR Day, Dries Sels, Phillip Weinberg, Anatoli Polkovnikov, and Pankaj Mehta. Reinforcement learning in different phases of quantum control. *arXiv preprint arXiv:1705.00565*, 2017.

[21] Murphy Yuezhen Niu, Sergio Boixo, Vadim Smelyanskiy, and Hartmut Neven. Universal quantum control through deep reinforcement learning. *arXiv preprint arXiv:1803.01857*, 2018.

[22] Pavel Sekatski, Michalis Skotiniotis, Janek Kołodyński, and Wolfgang Dür. Quantum metrology with full and fast quantum control. *Quantum*, 1:27, 2017.

[23] Hideo Mabuchi and Navin Khaneja. Principles and applications of control in quantum systems. *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal*, 15(15):647–667, 2005.

[24] Alexander Hentschel and Barry C Sanders. Machine learning for precise quantum measurement. *Physical review letters*, 104(6):063603, 2010.

[25] Barry C Sanders and Alexander Hentschel. An efficient algorithm for optimizing adaptive quantum metrology processes. In *International Quantum Electronics Conference*, page I184. Optical Society of America, 2011.

[26] Neil B Lovett, Cécile Crosnier, Martí Perarnau-Llobet, and Barry C Sanders. Differential evolution for many-particle adaptive quantum metrology. *Physical review letters*, 110(22):220501, 2013.

[27] Pantita Palittpongarnpim, Peter Wittek, and Barry C Sanders. Single-shot adaptive measurement for quantum-enhanced metrology. In *Quantum Communications and Quantum Imaging XIV*, volume 9980, page 99800H. International Society for Optics and Photonics, 2016.

[28] Pantita Palittapongarnpim, Peter Wittek, and Barry C Sanders. Controlling adaptive quantum phase estimation with scalable reinforcement learning. In *24th European Symposium on Artificial Neural Networks, Bruges, April 27–29, 2016*, pages 327–332, 2016.

[29] Pantita Palittapongarnpim, Peter Wittek, Ehsan Zahedinejad, Shakib Vedaie, and Barry C Sanders. Learning in quantum control: High-dimensional global optimization for noisy quantum dynamics. *Neurocomputing*, 268:116–126, 2017.

[30] HM Wiseman and RB Killip. Adaptive single-shot phase measurements: The full quantum theory. *Physical Review A*, 57(3):2169, 1998.

[31] Rainer Storn and Kenneth Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

[32] Russell Eberhart and James Kennedy. A new optimizer using particle swarm theory. In *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on*, pages 39–43. IEEE, 1995.