

Designing quantum experiments with a genetic algorithm

Rosanna Nichols,¹ Lana Mineh,^{2,3} Jesús Rubio,⁴ Jonathan C. F. Matthews,⁵ and Paul A. Knott^{1,*}

¹*Centre for the Mathematics and Theoretical Physics of Quantum Non-Equilibrium Systems (CQNE), School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, UK.*

²*Quantum Engineering Centre for Doctoral Training, University of Bristol, Bristol, UK.*

³*Now at: School of Mathematics, University of Bristol, UK.*

⁴*Department of Physics and Astronomy, University of Sussex, Brighton, BN1 9QH, UK.*

⁵*Quantum Engineering Technology Labs, H. H. Wills Physics Laboratory and Department of Electrical & Electronic Engineering, University of Bristol, BS8 1FD, UK.*

(Dated: September 23, 2019)

We introduce a genetic algorithm that designs quantum optics experiments for engineering quantum states with specific properties. Our algorithm is powerful and flexible, and can easily be modified to find methods of engineering states for a range of applications. Here we focus on quantum metrology. First, we consider the noise-free case, and use the algorithm to find quantum states with a large quantum Fisher information (QFI). We find methods, which only involve experimental elements that are available with current or near-future technology, for engineering quantum states with up to a 100-fold improvement over the best classical state, and a 20-fold improvement over the optimal Gaussian state. Such states are a superposition of the vacuum with a large number of photons (around 80), and can hence be seen as Schrödinger-cat-like states. We then apply the two most dominant noise sources in our setting – photon loss and imperfect heralding – and use the algorithm to find quantum states that still improve over the optimal Gaussian state with realistic levels of noise. This will open up experimental and technological work in using exotic non-Gaussian states for quantum-enhanced phase measurements. Finally, we use the Bayesian mean square error to look beyond the regime of validity of the QFI, finding quantum states with precision enhancements over the alternatives even when the experiment operates in the regime of limited data.

PACS numbers:

Engineering quantum states with specific properties plays a part in all experiments and technologies in quantum physics. But designing optimal experiments to engineer such states can be challenging, in part due to the counter-intuitive nature of the theory. This has prompted a number of recent works allocating the task of experiment-design to artificial intelligence and machine learning [1–5]. Many of the computer-designed experiments have surpassed the best human-designed alternatives.

Building on this approach, we introduce a genetic algorithm, named AdaQuantum [79], that designs quantum optics experiments to produce quantum states with the required properties. The algorithm is flexible and modifiable, so that researchers with a range of requirements and available quantum-optics equipment can use it to design and optimise their experiments, as well as facilitating theoretical research in quantum state engineering and experimental design. In addition, the algorithm has a powerful computational engine, so that it efficiently and effectively searches for the optimal experimental designs in a given setting, and can allow for truncations of the Hilbert space up to 170 photons, enabling even more exotic states to be found. A selection of the experiments designed by AdaQuantum are shown schematically in Fig. 5.

To run AdaQuantum we must specify two things: first, we specify the *toolbox* of equipment that is available to construct our experiment, or more specifically, which quantum states, operations, and measurements we would like AdaQuantum to optimise over. Second, we must specify a fitness function, which allows us to quantify *how good* the quantum state outputted by AdaQuantum is for our purposes. In principle, any fitness function that takes as input a quantum state, and outputs a real number, can be used. Given this, AdaQuantum then performs an automated search to find arrangements and parameter settings of the experiential equipment that produce a state that maximises (or minimises) the fitness function. The flexibility built into AdaQuantum means that different fitness functions can easily be implemented, such as the fidelity with useful quantum states such as Schrödinger-cat states or states used in optical quantum computation [6]; or measures of entanglement, coherence, non-classicality, and so on. A flow chart illustrating the overall structure and usage of our algorithm is shown in Fig. 1.

As a demonstration of the great potential of our algorithm, in this work we focus on fitness functions relevant to the field of quantum metrology, and we will see how its application reveals important findings that will help to design better metrological protocols. We first consider noise-free experiments, and use as our fitness function the quantum Fisher information (QFI) of the (pure) output state. The QFI allows us to quantify how precisely a given state can measure a phase shift in an interfer-

*Contact email address: Paul.Knott@Nottingham.ac.uk

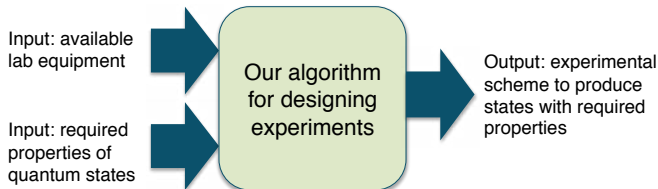


FIG. 1: A flow chart illustrating the overall structure and usage of our algorithm, AdaQuantum, for designing quantum experiments.

ometer – a valuable fitness function for optical quantum metrology [7, 8]. Next, we apply the two most dominant noise sources in our quantum optics setting – photon loss at the output, and imperfect heralding measurements – and use the mixed state QFI as our fitness function.

The third fitness function that we optimise is the Bayesian mean square error (BMSE). Although the QFI plays a crucial role in quantum metrology, using the theory associated with it requires a number of assumptions that are not always fulfilled in experiments. For example, it often assumes that a large number of repetitions has taken place and that certain prior information was available [9–12]. The BMSE, on the other hand, factors in and allows us to control both of these explicitly, and thus gives a reliable measure of the phase-measuring capability of a given state in realistic experimental settings [13].

Our main results obtained by AdaQuantum are threefold: First, in the noiseless case we find a number of methods of producing quantum states with QFIs up to 20 times larger than the optimal Gaussian state, which amounts to a 100-fold improvement over the best classical state. We show that these states can be thought of as Schrödinger-cat-like states, as they contain superpositions of the vacuum with a large numbers of photons (around 80). Second, we find methods of producing states that can still beat the optimal Gaussian state, even when realistic levels of the dominant noise sources (photon loss in the heralding measurements and in the final state) are included. Finally, using the BMSE for the problem of measuring a phase shift in an interferometer, we find states that beat both the optimal Gaussian state and the coherent state when the experiment operates with realistic prior knowledge and a reasonable number of experimental repetitions. This will enable experiments to create exotic states with enhanced phase-measuring capabilities.

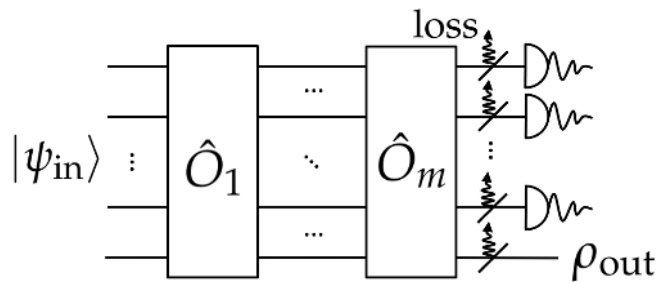


FIG. 2: The generic blueprint for the quantum optics experiments that we use an algorithm to optimise. An input state, $|\psi\rangle_{\text{in}}$, is acted on by a series of operators, \hat{O}_i , before heralding measurements are performed to produce an output state, ρ_{out} .

I. USING A GENETIC ALGORITHM TO DESIGN EXPERIMENTS

A. Optical quantum state engineering

The quantum optics experiments we design follow the blueprint depicted in Fig. 2. In this scheme, we start with an N -mode state, $|\psi\rangle_{\text{in}}$, which consists of independent one- and two-mode states. This state is acted on by up to m operators in sequence, which each act on one or two modes, with the two-mode operations serving to mix and entangle the modes. Finally, in $N - 1$ of the output modes heralding measurements are performed, modelled by POVMs. When each of these heralding measurements are simultaneously successful, the state in the N^{th} mode is the output state, which we then apply our fitness function to. With appropriate choices of experimental elements (which we introduce below), many of the state engineering schemes in the literature fit into our blueprint illustrated in Fig. 2 [14–18].

Although N is flexible, here we concentrate on $N = 2$ modes because having more modes is more difficult for both experiments and for our computer simulations (later we outline future work to overcome the computational challenges and enable $N > 2$ modes). Experimentally, having more modes substantially lowers the overall probability for producing the desired output state, and adds significant additional noise.

To design an experiment using the blueprint in Fig. 2, we need a toolbox of states, operations and measurements – the full list of these elements that we consider is given in Table I. Our focus is on producing experimental designs that are feasible with current or near future technology, and hence our toolbox only contains such elements. Other elements can easily be added to this list, and the list can be customised to match the capabilities of a particular laboratory.

Here we only introduce the most important details of the toolbox; more details can be found in Appendix A. Firstly, the input states we include are the single-mode squeezed vacuum $|\zeta\rangle_i$; the two-mode squeezed vacuum

States	$ n\rangle, \alpha\rangle, \zeta\rangle_i, \zeta\rangle_{ij}$
Operations	$\hat{D}(\alpha), \hat{S}_i(\zeta), \hat{S}_{ij}(\zeta), \exp(i\hat{n}\phi), \hat{U}_{ij}(T)$
Measurements	$ n\rangle\langle n , \text{Bucket}, \text{Multiplex}, x_\lambda\rangle\langle x_\lambda $

TABLE I: The states, operations and measurements that make up our toolbox of experimentally feasible elements. Here, i and j are arbitrary modes. The identity operator is also included in all runs of the algorithm.

(TMSV) $|\zeta\rangle_{ij}$, the coherent state $|\alpha\rangle$, and Fock states $|n\rangle$; the parameters z , α and n are constrained by what is possible experimentally.

Next are the operators, of which the most important is the beam splitter $\hat{U}_{ij}(T)$, where T is the probability of transmission, which serves to mix and entangle the two modes, enabling more exotic and useful states to be produced when part of the entangled state is measured. The other operators we use are the displacement operator $\hat{D}(\alpha)$, the phase shift $\exp(i\hat{n}\phi)$, and the single- and two- mode squeezing operators, given by $\hat{S}_i(\zeta)$ and $\hat{S}_{ij}(\zeta)$ respectively.

The first measurement we include is the *photon number resolving detection* (PNRD), given by $|n\rangle\langle n|$. However, due to the difficulty of implementing a PNRD we include two easier measurements, the *bucket detector* [19] and the *multiplex detector* [20]. The latter is achieved by separating the mode you want to measure into several modes via spatial- or time-multiplexing [20–22], followed by a bucket detection performed on each of these separated modes. Finally we include the homodyne detector, which is the projection onto a line in phase space, characterised by the projector $|x_\lambda\rangle\langle x_\lambda|$.

We focus on the two most dominant noise sources in optics experiments taking the form of Fig. 2: i) imperfect photon detectors, which can be modelled as photon loss prior to detection, and ii) photon loss on the output state. Noise is also present in the initial state preparation and in implementing the operations, but we do not include these as they are typically smaller than the detection and output-state loss (though future work will also incorporate these). See Appendix A 6 for details of how we simulate noise.

B. Our genetic algorithm

We now introduce and describe our algorithm, AdaQuantum, that designs quantum optics experiments. To run our algorithm, the user first has to specify which input states, operators and measurements they would like the algorithm to search over. This is done via the user interface shown in Fig. 9. As well as selecting the toolbox, the user can specify the ranges of the parameters, the loss rates of the detectors, the loss rate on the final state, and a number of other details that we will introduce below.

Once a user has specified their toolbox, they next tell

the algorithm what kind of quantum state they require by specifying a fitness function. The fitness function must take as input a quantum state, and output a real number that we wish to maximise (or minimise). In section II we introduce three different fitness functions that quantify the performance of the state for measuring a phase shift, but in principle any fitness function that can be written in the above-mentioned form can be incorporated into the algorithm (more on this below).

Next, we ask the question: *How can we best arrange the states, operators and measurements, so that the final quantum state maximises the fitness function?* This is a search problem; to perform this search, we use a genetic algorithm, which is a powerful and flexible global search metaheuristic based on evolution by natural selection.

To use a genetic algorithm for our problem, we must first encode each potential experiment into a *genome*. Our genome is a vector containing a mixture of integers and real numbers. For the most part, the integers in the genome encode *which* quantum optics elements will make up the experiment and the *arrangement* of these elements. The real numbers then encode the *parameters* of the different elements, such as the transmission probability of a beam splitter or the magnitude and phase of a coherent state. Given any (valid) genome, AdaQuantum then simulates the corresponding quantum optics experiment, determines what quantum state will be outputted, and calculates the fitness value by applying the chosen fitness function to the output state. See Fig. 5 for examples of different experiments, each of which corresponds to a different genome.

The task we then set the genetic algorithm is to search for a genome that maximises the fitness value. Genetic algorithms start by creating a collection of genomes, which together are known as the *population*. Next, the fitness function for each genome in the population is evaluated. The “fittest” genomes – i.e. the genomes with the largest fitness value – are then selected, and a new population of genomes is generated by mixing some of the genomes together (*crossover*) and by modifying (*mutating*) others. This next population should, in principle, be comprised of genomes that are “fitter”. This process repeats through a number of *generations*, until it is unlikely that any more generations will result in improvements. At this stage, if the algorithm has been designed appropriately, the fittest genomes will encode optimised solutions. A flow chart of a genetic algorithm is given in Fig. 3, and a more detailed description of how genetic algorithms work can be found in Appendix B.

C. Using a 3-stage algorithm

The slowest part of running the genetic algorithm is the simulation of each quantum optics experiment, and this must be done a large number of times in order to complete a thorough search of the search space (here the search space is the space of all possible quantum optics

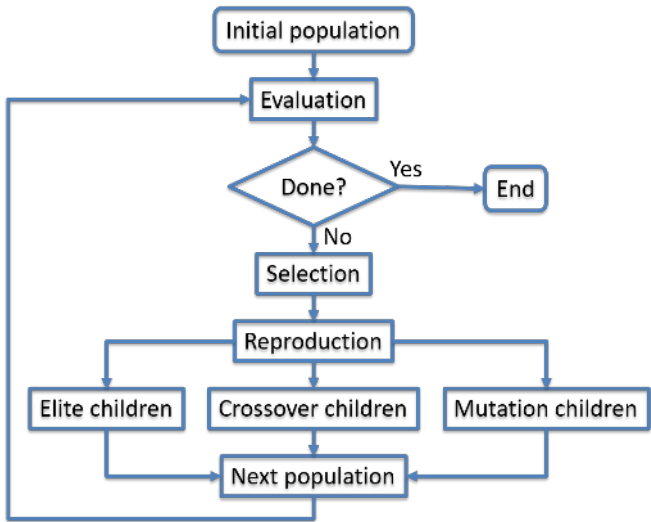


FIG. 3: Flowchart of a genetic algorithm – the various steps are described in the main text.

experiments that can be constructed with the selected toolbox). The dominant parameter that determines the speed of the simulation is the *truncation* of the Hilbert space: we have to truncate each state $|\psi\rangle$ (and correspondingly the operators and measurements), such that $|\psi\rangle = \sum_{i=0}^{\infty} c_i|i\rangle$ becomes $|\psi\rangle \approx \sum_{i=0}^T c_i|i\rangle$, where T is the truncation. The larger the value of T the more accurate the simulation will be, but the time taken for each simulation increases (approximately) exponentially as we increase T .

However, while the “best” quantum states we found required a large value of T in order for the simulation to be accurate (e.g. see Fig. 6), we also found that for a significant proportion of the possible experiments an approximate simulation, using a small value of T , can still provide valuable information to guide the search. To exploit this fact, and overcome the challenge of requiring a large T for accuracy but a small T for speed, we settled upon the following 3-stage algorithm:

Stage 1: A large number (around 10^7 for our main algorithm runs) of random genomes are created and evaluated using simulations with a small T , e.g. $T = 30$. A collection of genomes with the best fitness values are selected for the next stage. While this first stage is only approximate due to the small value of T , it still rules out many ineffective experiments and gives the genetic algorithm in the next stage a much stronger starting population.

Stage 2: We run Matlab’s inbuilt genetic algorithm [23] with a medium-sized population (around 10^5 for our main algorithm runs). Here T is larger (e.g. $T = 80$) and the genetic algorithm runs for a set number of generations, usually 10.

Stage 3: In the final stage, the simulation is accurate but slow. In this stage, the fitness function will first simulate the circuit specified by the input genome at a very low truncation (e.g. $T = 20$), and then repeat this, increasing T on each iteration (in steps of 10), until both the mean number of photons in the final state and the fitness value converge (indicating that increasing T does not change our results), or until the maximum value of T is reached (where the maximum value of T is specified by the user, labeled “Max. Truncation” in Fig. 9). This ensures the results are reliable and accurate, while running much faster than if we had chosen the maximum value of T for each genome. Here the population is smaller (around 10^4 for our main algorithm runs). By the end of this stage (which usually runs for around 40 generations), the algorithm should converge to a genome corresponding to a quantum optics experiment that produces the desired quantum state.

We compared our 3-stage genetic algorithm to Matlab’s built-in standard genetic algorithm, pattern search, swarm, and simulated annealing algorithms [23], and our algorithm performed significantly better than all. Our tests also showed that removing any one of the 3 stages described above gave sub-optimal results. In Appendix C we overview the genetic algorithm settings (such as the selection, mutation and crossover functions), and give details of the hyperparameters we use.

We took a number of steps to speed up our algorithm [80]. Most importantly, we utilised the algorithm outlined in [24], and available as a Matlab function at [25], which implements our operations by calculating $\exp(tA)b$ without needing to explicitly form $\exp(tA)$, where t is a scalar, A is a matrix and b is a vector. This produces a dramatic speed up, and allows us to simulate experiments with a value of T as large as 170 photons (in two modes) in the order of seconds. This is a significantly larger value of T than comparable algorithms in the literature [1, 26]. As we will see next, this allows us to generate, among other things, quantum states comprised of a superposition of the vacuum with 80 photons.

D. Designing AdaQuantum for flexibility

A key design focus when constructing AdaQuantum was flexibility, so that researchers with diverse requirements can use it to achieve their goals. The first flexible aspect is the toolbox selection, which is done using the user interface in Fig. 9. Once the user has selected the states, operators, and measurements, AdaQuantum automatically designs a genome that reflects this choice. Therefore, different runs of the algorithm with different toolbox selections will have different genome structures – both the genome length and the positions of the integers and real numbers will change. The number of modes in the experiment can also be modified, as can the number of operations prior to measurement (see Fig. 2), and

again AdaQuantum adapts to this. For example, when we select 4-modes the algorithm automatically knows that the input state can contain at most two 2-mode states, and that we will require three heralding measurements in order to produce a single mode output state.

In addition, when the user chooses a loss-free experiment with either number resolving or homodyne detectors, then the output state will be a pure state, whereas if there is loss (on either the final state or the measurements), or if the measurements have uncertain outcomes (i.e. when using the on/off or multiplex detector), then the output will be mixed. These features allow a huge range of experiments to be searched over, with applications for both experimentalists looking for lab-ready experiments, and theorists looking to push the limits of what is possible in hypothetical experiments.

Another key part of the flexibility is that AdaQuantum has been designed so that elements can be easily added, removed or modified from the toolbox, and similarly for the fitness functions. AdaQuantum is available on GitHub [27], which includes a User Guide that explains how to modify the toolbox and fitness functions. Modifying the toolbox will allow experimental elements that we haven't yet considered to be included, increasing the applicability of the algorithm. Perhaps more importantly, adding more fitness functions will allow AdaQuantum to be used to find quantum states for a wide range of applications. As long as a fitness function can be programmed in a form that takes as input a density matrix, and outputs a real number, then it can be incorporated into the algorithm.

Once a fitness function or toolbox element is added/modified, AdaQuantum automatically updates the user interface (Fig. 9), and again the genome changes accordingly. Therefore, with simple modifications AdaQuantum can, for example, search for non-Gaussian states [28, 29], states for multi-parameter estimation [30–33], or states with certain non-classical properties [34–36]. In this paper we focus on fitness functions for quantum-metrology applications (see next section), and in a spin-off project we have already designed fitness functions – and then used AdaQuantum to design experiments – to produce states with a high fidelity to a range of target states, such as cat states [37].

II. RESULTS

In this section, we present the results of running the algorithm with several different fitness functions and toolboxes, which are all relevant to the field of quantum metrology. In Appendix C we give a detailed discussion of the numerous steps and choices involved in running AdaQuantum to obtain these results.

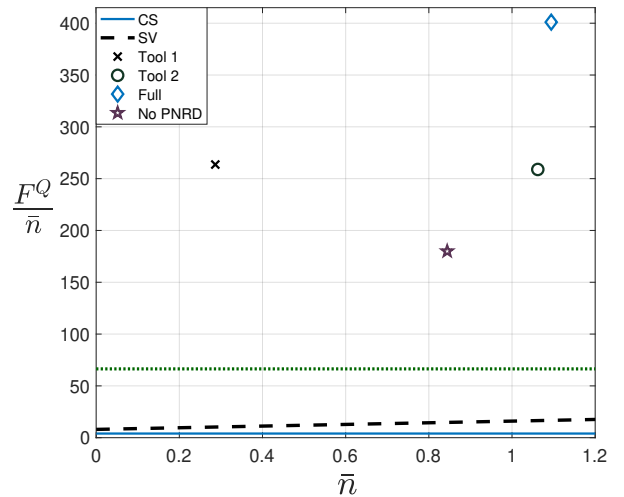


FIG. 4: Noiseless results: The QFI (for phase estimation), scaled by the mean photon number, for each of the states produced by our algorithm, compared to the squeezed vacuum state (SV) and coherent state (CS). For comparison, the value of F^Q/\bar{n} for the coherent state is 4. See Table. II for details of all the states. The toolboxes *Full*, *Tool 1*, *Tool 2*, and *No PNRD* are described in the main text. We see large improvements over the CS and SV, even for the experimentally-viable toolboxes *Tool 1*, *Tool 2*, and *No PNRD*. We also compare our results to the squeezed vacuum state with the largest value of squeezing so far produced in experiments, which, to the best of our knowledge, is the 15dB squeezed state in Ref. [38]. For this state, $F^Q/\bar{n} = 66.4$, which is illustrated as the horizontal dotted green line.

A. Fitness function 1: Pure state QFI

For our first fitness function we consider only noise-free experiments, and we consider the quantum Fisher information (QFI) of the (pure) output state $|\psi\rangle$ for measuring an unknown phase ϕ . This is given by [7]

$$F^Q = 4(\langle\psi'_\phi|\psi'_\phi\rangle - |\langle\psi'_\phi|\psi_\phi\rangle|^2). \quad (1)$$

Here $|\psi_\phi\rangle = \exp(i\hat{n}\phi)|\psi\rangle$ is our state after the phase shift ϕ is applied, and $|\psi'_\phi\rangle \equiv \frac{\partial}{\partial\phi}|\psi_\phi\rangle$. The QFI is useful because it can be used to tell us the precision with which our state $|\psi\rangle$ can measure the phase shift ϕ . This is found using the quantum Cramér-Rao bound (CRB) [7]:

$$\delta\phi \geq \frac{1}{\sqrt{\mu F^Q}}, \quad (2)$$

where $\delta\phi$ is said precision, and μ is the number of repetitions of the experiment.

Note that $|\psi'_\phi\rangle = i\hat{n}|\psi_\phi\rangle$. We can therefore write Eq. (1) independently of ϕ , for example

$$|\langle\psi'_\phi|\psi_\phi\rangle|^2 = |\langle\psi_\phi|\hat{n}|\psi_\phi\rangle|^2 \quad (3)$$

$$= |\langle\psi|\hat{n}|\psi\rangle|^2 \quad (4)$$

In this way, we can calculate the QFI directly from the output state of our circuit, $|\psi\rangle$, without needing to perform any differentiation, which improves the performance of our algorithm.

The actual fitness function we use here is the QFI scaled by \bar{n} (the mean photon number of the output state), F^Q/\bar{n} . If we just used the un-scaled QFI, then the algorithm would just try to make bigger and bigger states, as even a classical state has an unbounded QFI in the large photon-number limit. Such an optimisation would soon produce states that require Hilbert-space truncations that are both too large to simulate, and require unrealistic experimental equipment. Using F^Q/\bar{n} overcomes this problem. (Note that it would also be possible to include a restraint in the fitness function that penalizes the deviation of \bar{n} from a given number.)

We also use F^Q/\bar{n} to compare our results with alternatives in the literature. While a large QFI does amount to a high precision, it does not give a fair comparison when different states have different intensities (where we define intensity as the mean number of photons in the state). To compare different states the key quantity of interest is the precision that a given state can estimate the phase shift ($\delta\phi$), but in an attempt to keep the comparison fair we choose to fix the total intensity that each state is allowed to use in the estimation procedure. We label the total intensity (the ‘resources’) R , which is given by $R = \bar{n}\mu$, where μ is the number of times the state is sent through the phase shift. By taking R as a constant, the CRB in Eq. (2) can be written as

$$\delta\phi \geq \sqrt{\frac{\bar{n}}{RF^Q}} \propto \sqrt{\frac{\bar{n}}{F^Q}}. \quad (5)$$

We see that by fixing R , we can use F^Q/\bar{n} to compare the performance of different states.

To judge the success of our results, we will compare the states we produce to a coherent state, which is the optimal *classical* state, and to a squeezed vacuum state, which is the optimal Gaussian state when there is no noise [39, 40]. Fig. 4 shows the comparison of these states with our results for the noiseless experiments found by AdaQuantum. Here we chose to display the squeezed vacuum and coherent states by plotting \bar{n} against F^Q/\bar{n} . But this is not the only way to compare these states with our results. As an alternative, we can compare the states found by AdaQuantum against a squeezed vacuum state with the largest value of squeezing so far produced in experiments, which, to the best of our knowledge, is the 15dB squeezed state in Ref. [38]. For this state, $F^Q/\bar{n} = 66.4$ (calculated using [41], with $|\zeta| = 1.73$), which we include in Fig. 4 as the horizontal dotted green line.

In Fig. 4 a different experimental scheme has been found for each of the four different toolboxes we have tested. The most expansive toolbox we studied is named ‘Full’ in the figure. This includes all of the toolbox elements listed in Table I (and described in Section IA). Note that, for this toolbox, squeezing operators are allowed on arbitrary states. In this toolbox, the strength

of the squeezing $|\zeta|$ is limited to 1.4 in both the squeezing operators and squeezed states; $|\alpha|$ is limited to 5 in both initial coherent states and displacement operators; and number states are limited to $n = 5$. The limit to the number of photons resolved by PNRDs is $n = 10$. The best experiment devised by the algorithm, detailed in Table II, utilises the squeezing operators and heralds on the detection of 10 photons. The resulting state has a F^Q/\bar{n} approximately 20 times higher than the corresponding squeezed vacuum state, and 100 times higher than the coherent state (and note that the squeezed vacuum state has a larger QFI than both the NOON state and the Holland and Burnett state [41–43]). This state even beats the largest squeezed vacuum so far produced in experiments (to the best of our knowledge) [38] by a factor of 6 (though if we had allowed this much squeezing in our runs of AdaQuantum, as opposed to limiting the squeezing to $|\zeta| = 1.4$, then our results would likely be improved significantly).

However, acting with squeezing operations on arbitrary states is extremely challenging at present, so we will move on to toolboxes ‘Tool 1’ and ‘Tool 2’. These are both the same as ‘Full’ but with the squeezing operations removed. They are very similar to each other, the only difference being that ‘Tool 2’ has the photon detection limited at $n = 10$, as in ‘Full’, and ‘Tool 1’ has the photon detection limited to $n = 6$. Running the algorithm with these toolboxes produced states with F^Q/\bar{n} still far higher than the squeezed vacuum and coherent state. Finally, we remove the ideal number measurements to form the toolbox ‘No PNRD’, which instead includes bucket detectors and multiplex detection consisting of 16 bucket detectors, for up to 6 photon heralding. The resulting experiment found by AdaQuantum uses a heralding measurement on the vacuum $|0\rangle\langle 0|$ and produces a state with F^Q/\bar{n} still significantly higher than the squeezed vacuum. The states found by AdaQuantum for all these toolboxes are detailed in Table II, and schematics of a selection of the experiments designed by AdaQuantum are given in Fig. 5.

Why do these states give such large improvements over the alternatives? This can be revealed by looking at the number-distribution of the states. Here we focus on the state labeled ‘Tool 2’, and leave such an analysis of the other states for future work. By writing the state as

$$|\psi\rangle = \sum_{n=0}^{\infty} c_n |n\rangle, \quad (6)$$

in Fig. 6 we plot the log of $|c_n|^2$ against n [81]. The figure reveals that this state has a large contribution of the vacuum, and a small but significant contribution of a huge number of photons, around 60-100. Such a state can be seen as a Schrödinger-cat-like state: the human eye can directly detect less than 100 photons [44], so, arguably, this state can be seen as a superposition of a macroscopic state with the vacuum. More importantly for metrology, such a state has a large QFI because it has a large variance (with respect to the encoding Hamiltonian), whilst

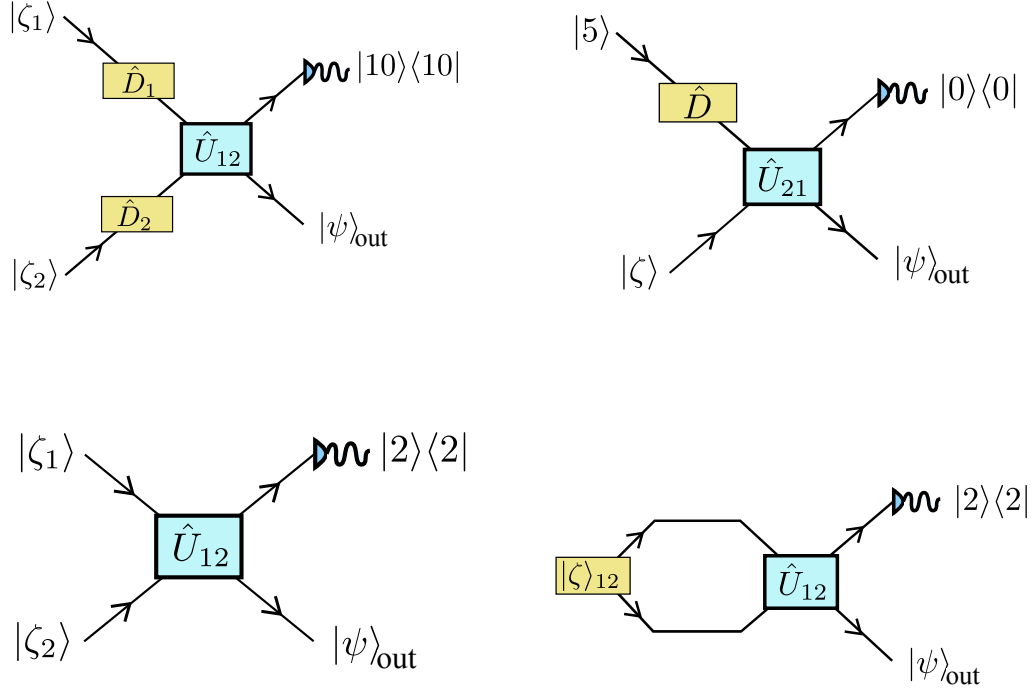


FIG. 5: A selection of the experiments designed by AdaQuantum. Top left and top right: These designs produce states that optimise the QFI, as introduced in Section II A. The performance of these states is shown in Fig. 4, and the details of the parameters are given in Table. II. The top left design uses the toolbox *Tool 1*, whereas top right uses toolbox *No PNRD*. Bottom left: A design to produce a state that optimises the QFI, after loss is applied to the measurement and final state, as introduced in Section II B. The performance of this state is shown in Fig. 7, where each loss rate requires different parameters. Bottom right: A design to produce a state that optimises the Bayesian mean square error (BMSE) with 1 repetition ($\mu = 1$), as introduced in Section II C. The performance of this state is shown in Fig. 8, and the details of the parameters are given in Table. IV.

Toolbox	ψ_{in}	\mathcal{O}_1	\mathcal{O}_2	\mathcal{O}_3	POVM	$p(\%)$
Tool 1	$ \zeta_1 = 1.39e^{i2.50}, \zeta_2 = 0.34e^{i5.64}\rangle$	$\hat{D}_2(\alpha = 2.49e^{i5.92})$	$\hat{D}_1(\alpha = 1.66e^{i6.11})$	$\hat{U}_{12}(T = 0.30)$	$ n = 10\rangle\langle n = 10 $	1.19
Tool 2	$ \zeta_1 = 1.40e^{i2.35}, \zeta_2 = 0.31e^{i5.44}\rangle$	$\hat{D}_2(\alpha = 1.97e^{i2.72})$	$\hat{D}_1(\alpha = 2.34e^{i5.98})$	$\hat{U}_{21}(T = 0.18)$	$ n = 6\rangle\langle n = 6 $	1.72
Full	$ 0, \zeta = 1.32e^{i0.06}\rangle$	$\hat{S}_{12}(\zeta = 0.88e^{i4.73})$	$\hat{D}_1(\alpha = 3.20e^{i6.28})$	$\hat{S}_2(\zeta = 0.19e^{i6.25})$	$ n = 10\rangle\langle n = 10 $	3.05
No PNRD	$ n = 5, \zeta = 1.40e^{i6.09}\rangle$	$\hat{D}_1(\alpha = 2.30e^{i4.62})$	$\hat{U}_{21}(T = 0.66)$	—	$ n = 0\rangle\langle n = 0 $	8.60

TABLE II: Toolboxes used and circuits produced for the results presented in Fig. 4 (schematics of a selection of the experiments designed by AdaQuantum are given in Fig. 5). The contents of each of the toolboxes are described in the text, and $p(\%)$ refers to the heralding success probability as a percentage. To clarify the notation used here, the state found with toolbox *Tool 1*, for example, is engineered by first creating a pair of squeezed vacuum states (with the parameters given in the table). Next, displacement operators act on both modes. A beam splitter is then applied, which entangles the two modes, and finally a 10 photon heralding measurement is performed. The probability of this heralding measurement being successful is 1.19%. The algorithm's flexibility means that we could easily re-run it to search for states with a higher heralding probability, if this was desired.

retaining a small mean number of photons. This state can therefore be seen as a realistic version of the so-called ON state, which has already attracted much attention in metrology [45, 46]. This state is also useful for quantum computing with continuous variables [47].

B. Fitness function 2: Mixed-state QFI

For our next fitness function, we study the effect of the two most dominant noise sources in the quantum optics experiments we are considering: photon loss at the output and imperfect heralding measurements. After

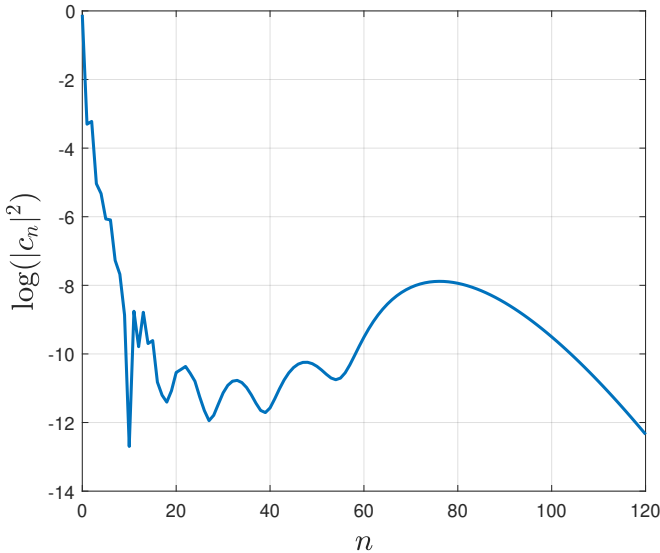


FIG. 6: Writing the state generated by ‘Tool 2’ as $|\psi\rangle = \sum_{n=0}^{\infty} c_n |n\rangle$, we plot the log of $|c_n|^2$ against n . We see that this state is a superposition of the vacuum with a large number of photons, and hence can be seen as a Schrödinger-cat-like state. To see how this state is made, see Table. II.

this noise is applied the resulting state will be a mixed state ρ . We then apply the mixed-state QFI to this state (again scaled by the mean photon number in the state, \bar{n}). To calculate the mixed-state QFI, we first need to apply the phase shift to ρ , giving ρ_ϕ , and then find the eigenvalues and eigenvectors of this state by writing it as $\rho_\phi = \sum_m q_m |\psi_m\rangle\langle\psi_m|$. We then use the form of the mixed-state QFI developed in [48]:

$$F^Q = \sum_i q_i F_i^Q - \sum_{i \neq j} \frac{8q_i q_j}{q_i + q_j} |\langle\psi'_i|\psi_j\rangle|^2 \quad (7)$$

where F_i^Q is the pure state QFI of eigenstate $|\psi_i\rangle$. We use a similar technique as with the pure-state QFI to write the mixed-state QFI so that it does not involve differentiation. It should be noted that we deal with pure states, in the form of vectors, until the measurement stage, when we switch to the density matrix formalism, as this allows us to use the matrix exponential technique discussed above and in [24].

In the noiseless case, we saw that AdaQuantum found quantum states with dramatic improvements over the squeezed vacuum and coherent state. Noise will inevitably diminish such improvements, but up to now it has been an open question whether it was possible to improve over the squeezed vacuum and coherent state *at all* when realistic experimental equipment and noise are factored in. Note that the squeezed vacuum and coherent state do not require heralding, so they escape the effects of imperfect heralding.

As discussed above, imperfect heralding can be modeled by applying loss to the state prior to heralding. For

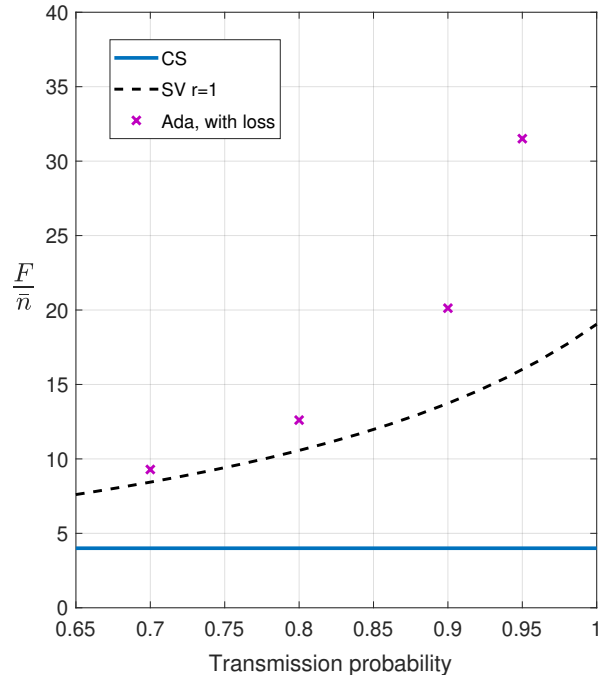


FIG. 7: We plot F^Q/\bar{n} of the states found by AdaQuantum against the transmission probability ($1 - \text{loss rate}$) of both the loss on the output state, and the loss in the heralding measurements. All states found by AdaQuantum take the form $|\psi\rangle = \mathcal{N}\langle 2|\hat{U}_{12}|\zeta_1, \zeta_2\rangle$, where \mathcal{N} is the normalisation, i.e. these states are formed by sending two squeezed vacuum states through a beam splitter, followed by 2-photon heralding. Note that, despite the fact that the states found by AdaQuantum have the same form for all loss rates, we optimised AdaQuantum for each loss rate separately. I.e. we first fixed the loss rate, then ran AdaQuantum to find states to maximise the QFI for that rate.

simplicity, we fix the loss to be the same for both the heralding measurements and on the output state. Unlike the noiseless case, here we are specifically focused on experiments that can be realistically performed without specialised equipment. Therefore, the toolbox we use here does not use squeezing operations and has the following limits to the parameters: $n = 4$ for Fock states; $|\alpha| = 5$ for coherent states and displacement operator; $|\zeta| = 1$ for squeezed states; and photon number resolving up to $n = 6$.

The results produced for varying values of loss are shown in Fig. 7. The value of F^Q/\bar{n} for each of these results is higher than the squeezed vacuum at the same loss rate – this opens up the possibility for experiments to create these exotic non-Gaussian states, and use them to beat the squeezed vacuum state in phase estimation. The heralding probabilities to produce the desired output states are quite reasonable (around 10-20%) and so the experiments proposed here can be carried out experimentally in reasonable time. Note that this experiment requires a 2-photon heralding measurement (which can

be performed with multiplexed bucket-detectors) and single mode squeezed states, which can all be achieved with current or near-future technology (for example see [20–22, 49, 50]).

C. Fitness function 3: Beyond the QFI

The third fitness function that we optimise here is the Bayesian mean square error (BMSE). Despite the importance of the QFI as a method of quantifying the metrological performance of different states, in general its usefulness depends on the possibility of recasting the problem at hand in the language of the local approach to estimation theory [9, 11–13, 51]. In particular, approaching the CRB in equation (2) typically requires either having certain prior knowledge and repeating the experiment a large amount of times [9, 10], or just having a large amount of prior information [11, 12]. But in realistic scenarios the number of repetitions of the experiment can be small, and the formalism based on the QFI does not take into account the possibility of having a moderate amount of prior information [13]. The BMSE factors in both of these, and hence gives a reliable measure of the phase-measuring capability of a given state in a realistic experimental setting.

To use the BMSE, unlike when we use the QFI, we need to specify what measurement scheme will be used to extract the information about the phase shift from the probe state. For the results in this paper each run of AdaQuantum produces a *single-mode* state $|\psi\rangle$, which cannot be used on its own to estimate a phase shift because in experiments we can only access the information about relative phase shifts. One approach we can use is to take a pair of such states, $|\psi\rangle \otimes |\psi\rangle$, and then encode a difference of phase shifts as $|\psi_\theta\rangle = \exp[-i(\hat{n}_1 - \hat{n}_2)\theta/2]|\psi\rangle \otimes |\psi\rangle$. A measurement can then be performed on both modes, which for simplicity we take to be ideal and given by the POVM elements $\{|M\rangle\langle M|\}$. Examples of measurements are given below.

The probability of finding the outcome M given the unknown value of the phase is $p(M|\theta) = |\langle M|\psi_\theta\rangle|^2$, and if the previous preparation and measurement stages are repeated μ times and the outcomes $\mathbf{M} = (M_1, \dots, M_\mu)$ are recorded, then their probability is $p(\mathbf{M}|\theta) = \prod_i p(M_i|\theta)$. Furthermore, imagine that we know in advance that the phase can only be found within an interval of width $\pi/12$ centred around zero, a state of knowledge that can be represented by the prior probability $p(\theta) = 12/\pi$, for $\theta \in [-\pi/24, \pi/24]$, and zero otherwise. We can then combine all these pieces of information using Bayes' theorem, which gives us the posterior probability $p(\theta|\mathbf{M}) = p(\theta)p(\mathbf{M}|\theta)/p(\mathbf{M})$, where $p(\mathbf{M})$ is the normalisation factor.

At this point we can introduce the BMSE as $\bar{\epsilon}_{\text{mse}} =$

$\int d\mathbf{M} p(\mathbf{M}) \epsilon(\mathbf{M})$, where

$$\epsilon(\mathbf{M}) = \left\{ \int d\theta p(\theta|\mathbf{M}) \theta^2 - \left[\int d\theta p(\theta|\mathbf{M}) \theta \right]^2 \right\} \quad (8)$$

is the variance of the posterior probability. We note that since the posterior $p(\theta|\mathbf{M})$ contains the information about the phase, the BMSE can be understood as an optimal way of quantifying the quality of this information on average for a given POVM. In particular, $\bar{\epsilon}_{\text{mse}}$ is a measure of uncertainty. Further details and the justification of this framework can be found in [9, 11, 13, 52].

The task given to AdaQuantum is, in this case, to find single-mode states $|\psi\rangle$ that minimise $\bar{\epsilon}_{\text{mse}}$ for a given number of repetitions and measurement scheme. Since the calculation of the BMSE is more demanding than finding the QFI (see, e.g., [9]), we have chosen a narrow prior to simplify the calculation of the integrals in equation (8). However, this does not imply that we could have used the QFI instead, since it was argued in [13] that the prior width needs to be 0.1 or smaller if the QFI is to be a suitable figure of merit in a similar scheme, and here the width is $\pi/12 \approx 0.3$. Thus our choice corresponds to the regime of intermediate prior knowledge where the BMSE is useful [13]. In addition, note that, unlike for the QFI, we do not need to rescale the BMSE by the average photon-number because the BMSE is already a direct measure of the estimation performance. To keep the comparison fair, all states in this section contain an average photon number of 1.

We will optimise the BMSE using two different strategies, both using the same toolbox as for the mixed-state QFI above. First we focus on a specific and practically-motivated POVM: counting photons after the action of a 50:50 beam splitter, and we set the algorithm to optimise the BMSE for $\mu = 4$, $\mu = 8$ and $\mu = 12$ repetitions. This search produces a state that takes the form $|\psi\rangle = \mathcal{N} \langle n | \hat{U}_{12} \hat{S}_{12} | 0, 0 \rangle$ for $\mu = 4$ and $\mu = 12$, where \mathcal{N} is the normalisation, while for $\mu = 8$ we find $|\psi\rangle = \mathcal{N} \langle n | \hat{P}_1 \hat{U}_{12} \hat{S}_{12} | 0, 0 \rangle$, where P_1 is a phase shift in the first mode. Table IV provides the numerical parameters that generate these states. The uncertainty associated with two copies of these probes for each number of trials has been represented in Fig. 8.a (individual points), and we have also included the BMSE of both an interferometer with a squeezed vacuum per port (dashed line) and the configuration that mixes a coherent state and the vacuum with a 50:50 beam splitter (solid line) [82]. This figure shows that the states found by AdaQuantum perform better than these two benchmarks, which constitutes a clear demonstration that AdaQuantum is able to optimise a Bayesian figure of merit beyond the regime where the QFI is useful.

More concretely, we can quantify this improvement by introducing the quantity

$$I_r = \frac{\bar{\epsilon}_r - \bar{\epsilon}_{\text{ada}}}{\bar{\epsilon}_r}, \quad (9)$$

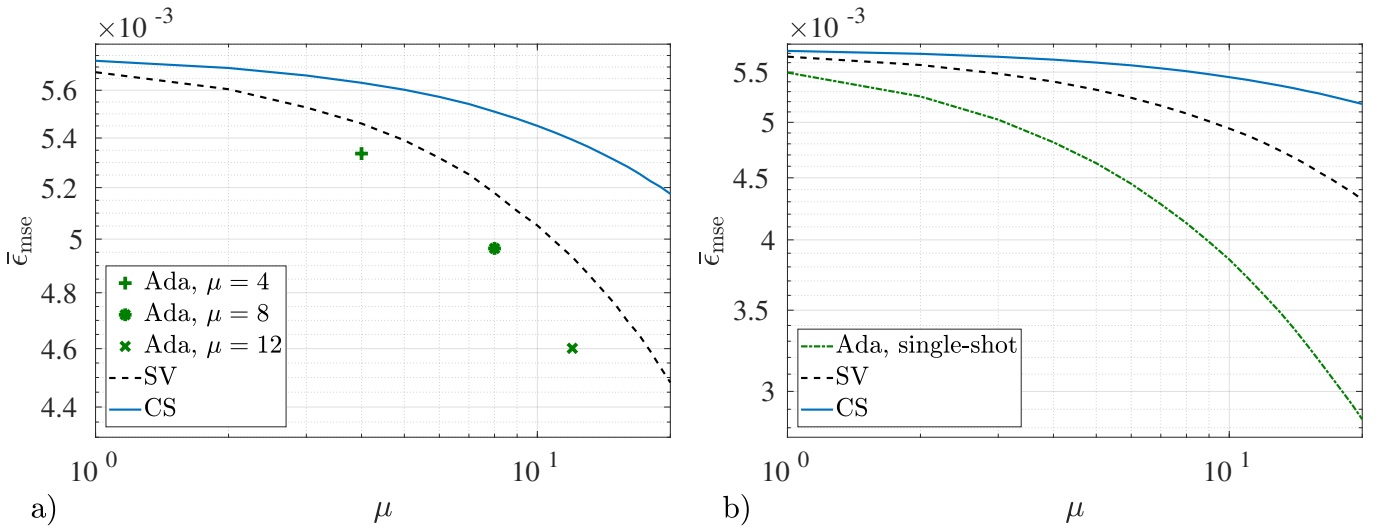


FIG. 8: Bayesian mean square error as a function of the number of repetitions for a) a coherent state mixed with the vacuum (solid line), two squeezed vacuums (dashed line) and states found by our genetic algorithm, AdaQuantum, for $\mu = 4$ (plus sign), $\mu = 8$ (asterisk) and $\mu = 12$ (cross) repetitions. Here the measurement scheme is based on counting photons after the action of a 50:50 beam splitter. In b) we again consider the coherent state and squeezed state, but now measured by their respective optimal single-shot POVMs. The state found by AdaQuantum (dash-dot line) is then based on the bound in equation (10), which already takes into account the optimal single-shot POVM for the given state. All the configurations are based on a two-mode interferometer with 1 photon on average, and where the phase shift can be found in an interval of width $\pi/12$, centred around zero.

AdaQuantum's relative enhancement					
50:50 beam splitter & photon counting				Single-shot POVM	
Ref.	$I_r(\mu = 4)$	$I_r(\mu = 8)$	$I_r(\mu = 12)$	Ref.	$I_r(\mu = 1)$
SV	0.02	0.04	0.07	SV	0.03
CS	0.05	0.10	0.15	CS	0.04

TABLE III: Improvement factor as defined in equation (9) to quantify the enhancement of the states found by AdaQuantum with respect to two squeezed vacuums (SV) and a coherent state mixed with the vacuum (CS). The details of the experimental configuration are those indicated in Fig. 8 and in the main text.

where $\bar{\epsilon}_r$ is the BMSE of any of the two reference states that we are employing and a positive I_r indicates that there has been an improvement. Its calculation, whose results are summarised in table III, shows an enhancement of between 2% and 7% with respect to the squeezed vacuum, and between 5% and 15% with respect to the coherent state.

The second strategy is to perform an analytical optimisation over all possible POVMs first, and then set AdaQuantum to find experiments that optimise this strategy. (Note that during AdaQuantum's search, the optimal measurement scheme for estimating the parameters has been selected – AdaQuantum just optimises over the state-engineering part of the experiment, i.e. the arrangement of elements in Fig. 2). Following [13], first we

recall that the single-shot BMSE satisfies [53–55]

$$\bar{\epsilon}_{\text{mse}}(\mu = 1) \geq \int d\theta p(\theta) \theta^2 - \text{Tr}(\bar{\rho}S), \quad (10)$$

where $S\rho + \rho S = 2\bar{\rho}$, $\rho = \int d\theta p(\theta) |\psi_\theta\rangle\langle\psi_\theta|$ and $\bar{\rho} = \int d\theta p(\theta) \theta |\psi_\theta\rangle\langle\psi_\theta|$. This bound can always be saturated when the measurement scheme is given by the projections onto the eigenstates of S [53–55], where S is an operator that only depends on the transformed state and the prior. Therefore, we can set the algorithm to search for states that minimise $\bar{\epsilon}_{\text{mse}}(\mu = 1, |M\rangle\langle M| = |s\rangle\langle s|)$, where $\{|s\rangle\}$ are the eigenvectors of S . Here we find another state with the form $|\psi\rangle = \mathcal{N}\langle n|\hat{U}_{12}\hat{S}_{12}|0,0\rangle$ but with different parameters (see table IV).

As table III shows, the state found by AdaQuantum for the optimal single-shot measurement is 3% better than two squeezed vacuums measured by their correspondent single-shot POVM, and 4% better than the coherent states. In addition, we note that using this measurement scheme in a sequence of repeated experiments is an appropriate strategy when we cannot or we do not wish to correlate different trials [13]. The performance of this state for the first 20 repetitions of the scheme, which has been represented in Fig. 8.b, shows that the state found by AdaQuantum using the optimal single-shot POVM is better than the benchmarks even when the number of repetitions grows. To summarise, we can say that the combination of AdaQuantum and the methodology introduced in [13] provides a robust method to find practical probe states with a strong performance for those systems that operate in the regime of limited data.

Setting	ψ_{in}	\mathcal{O}_1	\mathcal{O}_2	\mathcal{O}_3	POVM
BMSE, $\mu = 8$	$ 0, 0\rangle$	$\hat{S}_{12}(\zeta = 0.89 e^{i0.031})$	$\hat{U}_{12}(T = 0.69)$	$e^{i\hat{n}_1 0.32}$	$ n = 4\rangle\langle n = 4 $
BMSE, $\mu = 4, 12$	$ 0, 0\rangle$	$\hat{S}_{12}(\zeta = 0.91 e^{i0.040})$	$\hat{U}_{12}(T = 0.66)$	—	$ n = 6\rangle\langle n = 6 $
BMSE, $\mu = 1$	$ 0, 0\rangle$	$\hat{S}_{12}(\zeta = 0.95 e^{i6.1})$	$\hat{U}_{12}(T = 0.72)$	—	$ n = 2\rangle\langle n = 2 $

TABLE IV: Details of the circuits produced by AdaQuantum using the Bayesian framework. The first two are for a photon counting measurement after the beam splitter and that the last one is for the optimal single-shot POVM.

III. DISCUSSION AND CONCLUSIONS

This paper is not the first to apply techniques from machine learning to the task of designing quantum experiments [1–5] – we give a full comparison to these alternatives in Appendix D.

The results in the final section on the BMSE may have important general consequences for quantum metrology. The formal solution to the optimisation problem posed by a general quantum estimation scheme has been known for some time (see [54]). However, finding an analytical form of this solution for specific problems is challenging and generally not possible, which explains why we usually rely on bounds such as the CRB. And while in a sense the limits generated by the latter can be seen as fundamental [12], it can be argued that this is only useful when the bound can be safely applied to a problem in practice, which is not always the case (because realistic experiments put limits on the number of experiment repetitions and the prior knowledge available) [9, 10]. Therefore, the fact that our algorithm is able to find useful metrology protocols with more general fitting functions such as the BMSE or its single-shot optimum in equation (10) allows us to see it as a promising route to design quantum experiments using more reliable and general figures of merit, including, for instance, other cost functions different from the square error, or even multi-parameter systems [33, 56].

One clear direction for future work is to extend our runs of the algorithm to more than 2 optical modes. We attempted to do this in this project, but the additional simulation time required for > 2 optical modes meant that our global search was not effective. For 2 optical modes, most of the results presented here were generated by running the algorithm for 96 hours on 16 cores of the University of Nottingham’s High Performance Computing facility. Running for such times allowed us to run the genetic algorithm with very large populations (in the order of thousands or tens of thousands). When moving to 3 modes, the exponential slow-down in computing time required to simulate this larger Hilbert space meant that such large populations were not possible. To overcome this, one main focus of our future work will be to significantly enhance the global search, by, among other things, exploring a range of metaheuristic search methods, and by improving the way the genome is encoded. We are confident that this will allow effective searches for quan-

tum experiments in *at least* 3 modes, with the potential of finding quantum states with significant enhancements.

A further direction is to extend the fitness functions optimised by the algorithm. Ongoing work is searching for experiments to produce a range of specific states (such as squeezed cat states [57] and GKP states [6]). Beyond this, there are a broad range of fitness functions that could easily be incorporated, including searching for non-Gaussian states, highly entangled states, and states with a negative Wigner function (our GitHub page [27] includes a user guide that explains how new fitness functions can be added).

Another interesting question for future work is to understand *why* the experiments we presented here perform so well in maximising their respective fitness functions. This might be especially revealing in the states that are robust to noise: why do these specific experimental arrangements found by AdaQuantum create states that still perform well with photon loss?

In conclusion, we have introduced a genetic algorithm for designing quantum optics experiments. We demonstrated the flexibility of our algorithm by optimising three different fitness functions, and in all three the algorithm found improvements over the commonly used alternatives. Perhaps most notably, the algorithm found a realistic method of producing a Schrödinger-cat-like state comprised of a superposition of the vacuum with a large number of photons (around 80), which displays substantial phase-measuring improvements over the competing strategies. We emphasise here that our algorithm can in principle be used to design experiments to produce optical states with *any* desired properties, so long as these properties can be quantified and hence optimised over. Our algorithm, together with related work in the literature [1–5], highlights the power of utilising methods from artificial intelligence and machine learning to design and optimise quantum experiments.

Acknowledgements: We acknowledge helpful discussions with Gerardo Adesso, Jacob Dunningham, Ryuji Takagi, Ender Özcan, Chris Wade and Guillaume Thekkadath. JR acknowledges support from the South East Physics Network (SEPnet). LM was supported by the Bristol Quantum Engineering Centre for Doctoral Training, EPSRC Grant No. EP/L015730/1. JCFM acknowledges support from an EPSRC Early Careers Fellowship (EP/M024385/1) and the EPSRC UK Quan-

-
- [1] P. A. Knott, *New J. Phys.* **18**, 073033 (2016).
- [2] M. Krenn, M. Malik, R. Fickler, R. Lapkiewicz, and A. Zeilinger, *Phys. Rev. Lett.* **116**, 090405 (2016).
- [3] A. A. Melnikov, H. P. Nautrup, M. Krenn, V. Dunjko, M. Tiersch, A. Zeilinger, and H. J. Briegel, *Proc. Natl. Acad. Sci. p.* 201714936 (2018).
- [4] J. M. Arrazola, T. R. Bromley, J. Izaac, C. R. Myers, K. Brádler, and N. Killoran, arXiv preprint arXiv:1807.10781 (2018).
- [5] K. K. Sabapathy, H. Qi, J. Izaac, and C. Weedbrook, arXiv preprint arXiv:1809.04680 (2018).
- [6] D. Gottesman, A. Kitaev, and J. Preskill, *Phys. Rev. A* **64**, 012310 (2001).
- [7] S. L. Braunstein and C. M. Caves, *Phys. Rev. Lett.* **72**, 3439 (1994).
- [8] C. M. Caves, *Phys. Rev. D* **23**, 1693 (1981).
- [9] J. Rubio, P. Knott, and J. Dunningham, *J. Phys. Commun.* **2**, 015027 (2018).
- [10] D. W. Berry, M. J. W. Hall, M. Zwierz, and H. M. Wiseman, *Phys. Rev. A* **86**, 053813 (2012).
- [11] R. Demkowicz-Dobrzański, M. Jarzyna, and J. Kołodyński, *Progress in Optics* **60**, 345 (2015).
- [12] J. F. Haase, A. Smirne, J. Kołodyński, R. Demkowicz-Dobrzański, and S. F. Huelga, *Precision limits in quantum metrology with open quantum systems*, arXiv:1807.11882 (2018).
- [13] J. Rubio and J. Dunningham, *Quantum metrology in the presence of limited data*, arXiv: 1810.12857 (2018).
- [14] T. Gerrits, S. Glancy, T. S. Clement, B. Calkins, A. E. Lita, A. J. Miller, A. L. Migdall, S. W. Nam, R. P. Mirin, and E. Knill, *Phys. Rev. A* **82**, 031802 (2010).
- [15] T. J. Bartley, G. Donati, J. B. Spring, X.-M. Jin, M. Barbieri, A. Datta, B. J. Smith, and I. A. Walmsley, *Phys. Rev. A* **86**, 043820 (2012).
- [16] A. Ourjoumtsev, H. Jeong, R. Tualle-Brouri, and P. Grangier, *Nature* **448**, 784 (2007).
- [17] K. Huang, H. Le Jeannic, J. Ruaudel, V. B. Verma, M. D. Shaw, F. Marsili, S. W. Nam, E. Wu, H. Zeng, Y.-C. Jeong, et al., *Phys. Rev. Lett.* **115**, 023602 (2015).
- [18] J. Etesse, M. Bouillard, B. Kanseri, and R. Tualle-Brouri, *Phys. Rev. Lett.* **114**, 193602 (2015).
- [19] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, *Appl. Opt.* **35**, 1956 (1996).
- [20] J. C. F. Matthews, X.-Q. Zhou, H. Cable, P. J. Shadbolt, D. J. Saunders, G. A. Durkin, G. J. Pryde, and J. L. O'Brien, *npj Quantum Information* **2**, 16023 (2016).
- [21] G.-Y. Xiang, B. L. Higgins, D. Berry, H. M. Wiseman, and G. Pryde, *Nature Photon.* **5**, 43 (2011).
- [22] D. Achilles, C. Silberhorn, C. Sliwa, K. Banaszek, I. A. Walmsley, M. J. Fitch, B. C. Jacobs, T. B. Pittman, and J. D. Franson, *J. Mod. Opt.* **51**, 1499 (2004).
- [23] T. M. Inc., *Global optimization toolbox: User's guide (r2018a)*, <https://uk.mathworks.com/help/gads/index.html>. Last accessed 2018-07-17.
- [24] A. H. Al-Mohy and N. J. Higham, *SIAM journal on scientific computing* **33**, 488 (2011).
- [25] A. H. Al-Mohy and N. J. Higham, *expmv MATLAB code*, <https://github.com/higham/expmv>. Last accessed 2018-09-08.
- [26] N. Killoran, J. Izaac, N. Quesada, V. Bergholm, M. Amy, and C. Weedbrook, arXiv preprint arXiv:1804.03159 (2018).
- [27] *AdaQuantum on GitHub*, <https://github.com/paulk444/AdaQuantum> (2019).
- [28] I. Straka, L. Lachman, J. Hloušek, M. Miková, M. Mičuda, M. Ježek, and R. Filip, *npj Quantum Information* **4**, 4 (2018).
- [29] R. Takagi and Q. Zhuang, arXiv preprint arXiv:1804.04669 (2018).
- [30] P. C. Humphreys, M. Barbieri, A. Datta, and I. A. Walmsley, *Phys. Rev. Lett.* **111**, 070403 (2013), URL <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.111.070403>.
- [31] M. D. Vidrighin, G. Donati, M. G. Genoni, X.-M. Jin, W. S. Kolthammer, M. S. Kim, A. Datta, M. Barbieri, and I. A. Walmsley, *Nat. Commun.* **5** (2014), URL <https://www.nature.com/articles/ncomms4532>.
- [32] M. Szczykulska, T. Baumgratz, and A. Datta, *Adv. Phys. X* **1**, 621 (2016), URL <http://www.tandfonline.com/doi/full/10.1080/23746149.2016.1230476>.
- [33] T. J. Proctor, P. A. Knott, and J. A. Dunningham, *Phys. Rev. Lett.* **120**, 080501 (2018).
- [34] C. Lee, S. Ferrari, W. H. Pernice, and C. Rockstuhl, *Physical Review A* **94**, 053844 (2016).
- [35] H. Zhu, *Physical review letters* **117**, 120404 (2016).
- [36] J. Sperling, A. Eckstein, W. Clements, M. Moore, J. Renema, W. Kolthammer, S. W. Nam, A. Lita, T. Gerrits, I. Walmsley, et al., *Physical Review A* **96**, 013804 (2017).
- [37] L. O'Driscoll, R. Nichols, and P. Knott, arXiv preprint arXiv:1812.03183 (2018).
- [38] H. Vahlbruch, M. Mehmet, K. Danzmann, and R. Schnabel, *Physical review letters* **117**, 110801 (2016).
- [39] A. Monras, *Phys. Rev. A* **73**, 033821 (2006), URL <https://link.aps.org/doi/10.1103/PhysRevA.73.033821>.
- [40] A. De Pasquale, P. Facchi, G. Florio, V. Giovannetti, K. Matsuoka, and K. Yuasa, *Phys. Rev. A* **92**, 042115 (2015).
- [41] J. Sahota and N. Quesada, *Phys. Rev. A* **91**, 013808 (2015).
- [42] H. Lee, P. Kok, and J. P. Dowling, *J. Mod. Opt.* **49**, 2325 (2002).
- [43] M. Holland and K. Burnett, *Phys. Rev. Lett.* **71**, 1355 (1993).
- [44] F. Rieke and D. A. Baylor, *Rev. Mod. Phys.* **70**, 1027 (1998).
- [45] A. Rivas and A. Luis, *New J. Phys.* **14**, 093052 (2012).
- [46] P. Knott, T. Proctor, A. Hayes, J. Ralph, P. Kok, and J. Dunningham, *Phys. Rev. A* **94**, 062312 (2016).
- [47] K. K. Sabapathy and C. Weedbrook, *Phys. Rev. A* **97**, 062315 (2018).
- [48] Y. M. Zhang, X. W. Li, W. Yang, and G. R. Jin, *Phys. Rev. A* **88**, 043832 (2013).

- [49] P. C. Humphreys, B. J. Metcalf, T. Gerrits, T. Hiemstra, A. E. Lita, J. Nunn, S. W. Nam, A. Datta, W. S. Kolthammer, and I. A. Walmsley, *New Journal of Physics* **17**, 103044 (2015).
- [50] I. Afek, O. Ambar, and Y. Silberberg, *Science* **328**, 879 (2010).
- [51] M. G. A. Paris, *Int. J. Quant. Inf.* **07**, 125 (2009).
- [52] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003), ISBN 0521592712 hardback.
- [53] S. Personick, *IEEE Transactions on Information Theory* **17**, 240 (1971).
- [54] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
- [55] K. Macieszczak, M. Fraas, and R. Demkowicz-Dobrzański, *New J. Phys.* **16**, 113002 (2014).
- [56] P. C. Humphreys, M. Barbieri, A. Datta, and I. A. Walmsley, *Phys. Rev. Lett.* **111**, 070403 (2013).
- [57] P. Knott, T. Proctor, A. Hayes, J. Cooling, and J. Dunningham, *Phys. Rev. A* **93**, 033859 (2016).
- [58] P. G. Kwiat, K. Mattle, H. Weinfurter, A. Zeilinger, A. V. Sergienko, and Y. Shih, *Phys. Rev. Lett.* **75**, 4337 (1995).
- [59] A. Lamas-Linares, J. C. Howell, and D. Bouwmeester, *Nature* **412**, 887 (2001).
- [60] C. Simon and D. Bouwmeester, *Phys. Rev. Lett.* **91**, 053601 (2003).
- [61] J. Sperling, W. Vogel, and G. S. Agarwal, *Phys. Rev. A* **85**, 023820 (2012).
- [62] S. M. Barnett and P. M. Radmore, *Methods in theoretical quantum optics*, vol. 15 (Oxford University Press, 2002).
- [63] Y. Miwa, J.-i. Yoshikawa, N. Iwata, M. Endo, P. Marek, R. Filip, P. van Loock, and A. Furusawa, *Phys. Rev. Lett.* **113**, 013601 (2014).
- [64] N. C. Harris, Y. Ma, J. Mower, T. Baehr-Jones, D. Englund, M. Hochberg, and C. Galland, *Opt. Express* **22**, 10487 (2014).
- [65] T. Inc., *Thorlabs online catalogue - optical beam splitters*, https://www.thorlabs.com/navigation.cfm?guide_id=18. Last accessed 2018-07-16.
- [66] P. A. Knott, Ph.D. thesis, University of Leeds (2015).
- [67] M. Nielsen, I. Chuang, and L. Grover, *American J. Phys.* **70**, 558 (2002).
- [68] M. O. Scully and W. E. Lamb, Jr, *Phys. Rev.* **179**, 368 (1969).
- [69] P. Kok, W. J. Munro, K. Nemoto, T. C. Ralph, J. P. Dowling, and G. J. Milburn, *Rev. Mod. Phys.* **79**, 135 (2007).
- [70] P. Kok and S. L. Braunstein, *Phys. Rev. A* **61**, 042304 (2000).
- [71] D. E. Goldberg, Goldberg,–[USA]: Addison-Wesley (1989).
- [72] A. R. Conn, N. I. M. Gould, and P. Toint, *SIAM Journal on Numerical Analysis* **28**, 545 (1991).
- [73] *Blog post with an introduction to genetic algorithms, using the medium of LEGO®.*, <https://quantarei.wordpress.com/2018/08/08/using-a-genetic-algorithm-to-design-quantum-experiments/> (2018).
- [74] H. Barnum, H. J. Bernstein, and L. Spector, *Journal of Physics A: Mathematical and General* **33**, 8047 (2000).
- [75] G. Hornby, A. Globus, D. Linden, and J. Lohn, in *Space 2006* (American Institute of Aeronautics and Astronautics, 2006), p. 7242.
- [76] T. Geijtenbeek, M. Van De Panne, and A. F. Van Der Stappen, *ACM Transactions on Graphics (TOG)* **32**, 206 (2013).
- [77] B. Evans and S. P. Walton, *Applied Mathematical Modelling* **52**, 215 (2017).
- [78] K. Deep and M. Thakur, *Applied mathematics and Computation* **193**, 211 (2007).
- [79] The algorithm, AdaQuantum, is named after Ada Lovelace, the world’s first computer programmer, and resident of Nottingham, where our own algorithm was born.
- [80] Among other techniques, we used Matlab’s sparse matrix feature, and used analytical calculations to speed up a number of steps.
- [81] To rule out the possibility that the large QFI could be due to numerical inaccuracies, we double-checked the calculation leading to Fig. 6 analytically. This can be done by writing the displaced-squeezed states in the number basis in terms of creation and annihilation operators, propagating these creation and annihilation operators through a beam splitter, before finally acting on the LHS with $\langle 6 | \otimes \mathbb{I}$ and normalising the resulting single-mode state.
- [82] In [13] it was argued that if the prior probability is centred around zero and the POVM is based on mixing the light beams and counting photons, then for some probes it is necessary to also shift the phase of the second arm after the unknown parameter has been encoded in order to achieve the optimal BMSE. The calculations presented in this section have already taken this into account as prescribed by [13].

Appendix A: Optical quantum state engineering

1. Number states and number measurements

The first input state we include is the number state, or Fock state, $|n\rangle$. Typically, an experiment would be restricted to number states of low n , as generating a state with a large number of photons is technically challenging, so we restrict our simulations to have at most $n = 5$ (though this can easily be modified). As an example, a single photon state can be produced using spontaneous parametric down-conversion [58–60]. In this process, a photon of a particular frequency may be spontaneously converted into two identical photons, each with half of the original frequency. A heralding measurement is then used to detect one photon in this pair, which signals the existence of the other (this can be extended for $n > 1$).

The first measurements we include is measuring the number of photons in a mode – *photon number resolving detection* (PNRD) – which is implemented with the POVM element $|n\rangle\langle n|$. However, it can be difficult to precisely measure the number of photons so we also include two more-easily-implementable measurements. The first of these is a *bucket detector* (aka *on/off*, *threshold*, or *click* detector). This measures whether at least one photon is absorbed by the detector (it ‘clicks’), but it is unable to determine the exact number of photons. It is represented by the POVM, $\{|0\rangle\langle 0|, \mathbb{I} - |0\rangle\langle 0|\}$, where \mathbb{I} is the identity. These are commonly found in laboratories, for example in the form of avalanche photodiodes [19].

If exact PNRDs are not available, an approximate photon number measurement can be implemented by using several bucket detectors in a *multiplex detector*. This is achieved by separating the mode you want to measure (the target mode) into several modes via spatial- or time-multiplexing. The former can be implemented via a series of beam splitters, as in [21], or by optical fibre splitters, as in [20]. The latter can be implemented using a series of interferometers with paths of varying length, as in [22]. A bucket detection is then performed on each of these separated modes. A perfect PNRD is obtained in the infinite limit of multiplexed bucket detectors, but otherwise there will be a non-zero probability that more than one photon enters the same bucket detector in a given run, resulting in uncertainty in the exact number of photons entering the detector.

We use the following POVM, from [20], to simulate this multiplex detection, which takes into account all possible coincidence detection patterns. Here, r is the number of bucket detection events and c is the actual number of photons in the target mode. The weights $w_r(c)$ are non-negative and satisfy $\sum_r w_r(c) = 1$. These rely on the number of bucket detectors, d , and on $S(c, r)$, which is the Stirling number of the second kind, which counts the number of ways of partitioning c objects into r non-empty subsets. More details on constructing this POVM are found in the Supplemental Material of [20], which uses methods from [61] to compute the weights. The POVM elements E_r are given by

$$E_r = \sum_{c \geq r}^{\infty} w_r(c) |c\rangle\langle c|, \quad w_r(c) = \frac{d! S(c, r)}{(d-r)! d^c}. \quad (\text{A1})$$

In our simulations we use a multiplex detector consisting of either five [21] or sixteen [20] bucket detectors, but again this can be easily altered (the choices to use *five* and *sixteen* bucket detectors wasn’t due to any fundamental significance to these numbers, but rather because they have been used in experiments [20, 21]).

2. Coherent states and displacement operators

Coherent states are among the most readily available states as they are produced by a stabilised laser operating well above threshold [62]. To use them in our toolbox, we must construct them in the Fock basis. They can be defined by the action of the displacement operator on the vacuum, $|\alpha\rangle = \hat{D}(\alpha)|0\rangle$, where

$$\hat{D}(\alpha) = \exp(\alpha \hat{a}^\dagger - \alpha^* \hat{a}), \quad (\text{A2})$$

for $\alpha \in \mathbb{C}$, and where \hat{a} (\hat{a}^\dagger) is the annihilation (creation) operator. Alternatively, the following description of a coherent state in the Fock basis can be derived using the normally ordered form of the operator, described in [62]:

$$|\alpha\rangle = \exp(-|\alpha|^2/2) \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \quad (\text{A3})$$

In our simulation we have to truncate the Hilbert space, and therefore in order to simulate accurate states and operators we must limit the magnitude of $|\alpha|$; here we generally use a maximum value of $|\alpha| = 5$.

3. Squeezed states and squeezing operators

Squeezed states have a lower variance (in some quadrature) than coherent states, which makes them useful in a wide variety of applications, such as optical metrology [8]. The single mode squeezed vacuum state is defined by $|\zeta\rangle_i = \hat{S}_i(\zeta)|0\rangle_i$, where $\hat{S}_i(\zeta)$ is given by

$$\hat{S}_i(\zeta) = \exp\left[\frac{1}{2}(\zeta^* \hat{a}_i^2 - \zeta \hat{a}_i^{\dagger 2})\right], \quad (\text{A4})$$

where $\zeta \in \mathbb{C}$ and \hat{a}_i^\dagger and \hat{a}_i are the creation and annihilation operators on mode i [62]. Similarly, the two mode squeezed vacuum state can be defined by $|\zeta\rangle_{ij} = \hat{S}_{ij}(\zeta)|0\rangle_i|0\rangle_j$, where $\hat{S}_{ij}(\zeta)$ is given by

$$\hat{S}_{ij}(\zeta) = \exp\left[\zeta^* \hat{a}_i \hat{a}_j - \zeta \hat{a}_i^\dagger \hat{a}_j^\dagger\right], \quad (\text{A5})$$

where, again, ζ is a complex number which defines the state. As with coherent states, the normally ordered forms of the squeezing operators give rise to a construction of the states in the Fock basis [62].

Both the single mode and the two mode squeezed vacuum states can be created by acting on the vacuum with non-linear optical elements, such as optical parametric oscillation or four-wave mixing. They are further related because the action of a beam splitter on a two mode squeezed state is to produce two single mode squeezed states [62].

Squeezed states are harder to generate than, for example, coherent states, but they are still feasible and are an important resource in many experiments, so we include these in our toolbox and in most of our runs of the algorithm. We also include squeezing operators in the toolbox, although we omit these from most of our runs, as implementing a squeezing operator on an arbitrary state is difficult and rarely available in laboratories, although it is possible [63].

4. Beam splitters and phase shifts

The phase shift operation is given by $\exp(i\hat{n}\phi)$, where $\hat{n} = \hat{a}^\dagger \hat{a}$ is the photon number operator. This is often straightforward to implement as a phase shift can be induced on a mode through changing the length of the optical path it has to travel, when compared to the other modes in the system, for example using a thermo-optic effect [64].

The beam splitter is a vital element of our toolbox as it allows interactions between the modes. They are also readily available, and can be implemented by a partially reflecting mirror (e.g. a half-silvered mirror) or an interface between two glued-together glass prisms, among others [65]. The beam splitter operation is given by

$$\hat{U}_{ij} = \exp\left[\theta\left(\hat{a}_i \hat{a}_j^\dagger - \hat{a}_i^\dagger \hat{a}_j\right)\right], \quad (\text{A6})$$

where the probability of a photon being transmitted through the beam splitter is $T = \cos^2 \theta$ [66]. A ‘50:50’ beam splitter, which mixes the two input modes in equal superposition, refers to the case where $T = 0.5$.

5. Homodyne detection

To define homodyne detection we first require the generalised quadrature operator [62]:

$$\hat{x}_\lambda = \frac{1}{\sqrt{2}} \left[\hat{a} \exp(-i\lambda) + \hat{a}^\dagger \exp(i\lambda) \right], \quad (\text{A7})$$

where λ is the quadrature angle. Setting $\lambda = 0$ and $\lambda = \pi/2$ in the above expression gives the usual ‘position’ and ‘momentum’ quadrature operators, respectively. Homodyne detection is a projection onto the eigenstates of these quadrature operators. However, it is not possible to construct normalised eigenstates of the quadrature operators. Instead we construct states, $|x_\lambda\rangle$, defined by a complex number $x_\lambda = |x_\lambda| \exp i\lambda$, which obey the eigenvalue equation

$$\hat{x}_\lambda |x_\lambda\rangle = |x_\lambda| |x_\lambda\rangle. \quad (\text{A8})$$

These are not orthogonal but instead their overlap is given by the Dirac delta function

$$\langle x_\lambda | x'_\lambda \rangle = \delta(x_\lambda - x'_\lambda). \quad (\text{A9})$$

These eigenstates may also be constructed by operating on the vacuum state as follows [62]

$$|x_\lambda\rangle = \pi^{1/4} \exp\left[-\frac{1}{2}|x_\lambda|^2 + \sqrt{2}x_\lambda \hat{a}^\dagger - \frac{1}{2}\exp(2i\lambda)\hat{a}^{\dagger 2}\right] |0\rangle.$$

Perfect homodyne detection is the projection onto a line in phase space, characterised by the projector $|x_\lambda\rangle\langle x_\lambda|$.

6. Simulating noisy experiments

Photon loss can be modelled by mixing the noisy mode with the vacuum using a beam splitter of transmissivity $T = 1 - \gamma$ (and hence loss rate γ) [66]. However, this is computationally expensive as it requires simulating an additional mode, so here we simulate loss using Kraus operators as follows [67]

$$\rho_{out} = \sum_{k=0}^{\infty} K_k |\psi_{in}\rangle \langle \psi_{in}| K_k^\dagger, \quad (\text{A10})$$

where ρ_{out} is the output state, $|\psi_{in}\rangle$ is the input state, and

$$K_k = \sum_{n=0}^{\infty} \sqrt{\binom{n}{k}} \sqrt{(1-\gamma)^{n-k} \gamma^k} |n-k\rangle \langle n|. \quad (\text{A11a})$$

The Kraus operator K_k represents the loss of k photons to the environment with loss rate of γ .

We can include loss in the measurements by modifying the POVM elements to [68, 69]

$$E_n = \sum_{k=n}^{\infty} \binom{k}{n} (1-\gamma)^n \gamma^{k-n} |k\rangle \langle k|. \quad (\text{A12})$$

For the bucket measurement, we use [69, 70]

$$E_0 = \sum_{n=0}^{\infty} \gamma^n |n\rangle \langle n|, \quad (\text{A13a})$$

$$E_1 = \mathbb{I} - E_0, \quad (\text{A13b})$$

where E_0 corresponds to a measurement of no photons and E_1 is the measurement of one or more photons. Finally, for the multiplex detector, we construct a set of POVMs by using Eq. (A1) but with $|c\rangle\langle c|$ replaced with $E_{n=c}$ from Eq. (A12).

Appendix B: Genetic algorithms

Genetic algorithms take inspiration from biological evolution [23, 71, 72]. The aim of the algorithm is to maximise (or minimise) the fitness function – a summary of how this is achieved is presented in Fig. 3. Before starting the algorithm, the fitness function must be defined. This is the function to be optimised, known as the objective function in standard optimisation algorithms, which must produce a real number (the *fitness* or *objective value*) from a collection of real variables. Any point that the fitness function can be applied to, i.e. a list of values for each of the variables, is known as a *genome*, and the individual values of variables in this genome are known as the genes.

The algorithm works by varying a group of points, known as the population. Hence, the first step is to generate the initial population. For each genome in the population, each variable is generated at random within the provided bounds. Then the algorithm must iterate the population, hoping to optimise the fitness function by finding genomes that improve on the best value of the fitness function of previous generations. This iteration has the following steps, which repeat until the stopping conditions are reached:

Evaluation The value of the fitness function is found for each genome in the current population, these are known as the raw fitness scores. At this point the stopping conditions are also evaluated.

Selection Using the results of the evaluation, a subset of the population known as the parents are selected.

Reproduction The next generation is generated using the selected parents. This is done using three methods: elite, crossover and mutation.

The selection step works by first scaling the raw fitness scores to convert them into a more usable range of values. These scaled scores are then used to select the parents of the next generation. The selection function assigns a higher probability of selection to genomes with higher scaled fitness scores. This mimics natural selection, or ‘survival of the fittest’, as the ‘fitter’ genomes are more likely to reproduce.

The reproduction step creates genomes of the next generation (‘children’) using the parents of the current generation through three methods. The first are elite children, which form only a small percentage of the next generation. These are the parents with the highest fitness scores, which survive to the next generation unchanged. The second are crossover children. Here, two parents are chosen at random and each gene of the child is produced by copying the gene from one of the two parents. The final method to create children is mutation. Here, a single parent is chosen and random changes are made to its genome to produce the child. In Section C2 we introduce the specific mutation and crossover functions we use.

These children together form the next generation and the process repeats until the stopping conditions are met (see [73] for a visual introduction to genetic algorithms). Possible stopping conditions include meeting a maximum number of generations over which the best fitness score does not change, within tolerance. There is no proof that genetic algorithms must converge [23], but they have been known to perform well in situations where standard, gradient-based, optimisation algorithms have failed. In quantum circuit design, genetic algorithms have been successfully used to design quantum logic gates [74]. In other fields, examples of problems where genetic algorithms have been used include a NASA design of a radio antenna to pick up signals in space [75], computer models for walking for bipedal creatures [76] and optimising the aerodynamics of hypersonic space vehicles [77].

Appendix C: Running AdaQuantum to obtain our results

1. Parameters that affect the running speed

In general it is not easy to say *quantitatively* how the speed and performance of the algorithm changes when changing different parameters, such as the number of modes or the size of the selected toolbox (which in turn determine the genome length). The reason for this is that the dominant contribution to the run-time is the truncation. But, as discussed in Section IC, in Stage 3 of our algorithm the truncation changes with each simulation.

With this in mind, increasing the number of allowed operators (labeled m in Fig. 2) in itself would (approximately) linearly increase the run-time. But in addition, having more operators allows for the creation of larger states (e.g. by applying multiple displacement operators), so the point where we truncate in general needs to increase to accommodate the higher-energy states that will be produced. For the most of the results in this paper, we allowed for $m = 3$ operators. We also ran the algorithm for $m = 5$ for some of the no-loss runs, but didn’t see any improvement. One important point is that we included the *identity* in all runs of the algorithm, which explains why many of the results in this paper only have 2 operators. Another dominant factor in the simulation speed is the toolbox selection, particularly in the choices of parameters. Allowing for large states and operators that increase the energy slows down the simulation significantly. However, in our runs we always limited the size of the output state to either 1 or 2 photons on average (this can be done in the user interface for selecting the fitness function), which means that states that are too large *after* the measurement are discarded anyway.

All the runs of AdaQuantum presented here are for two-mode experiments. We attempted three-mode runs but were unable to outperform the two-mode experiments. The reason for this is that the simulation becomes exponentially slower when we increase the number of modes, and furthermore the search space becomes much larger. As elaborated on in Section III, improving the genetic algorithm so that it can effectively search for three-mode experiments will be a central focus of our future work.

2. Hyperparameters and settings for the genetic algorithm

Mutation and crossover are central to genetic algorithms, and the choices of mutation and crossover functions, as well as the hyperparameters of these functions, plays a large part in determining the effectiveness of the search. In our runs we experimented with three different crossover functions that are all inbuilt into Matlab [23]: `scattered`, `single point` and `two point` crossover. None of the crossover functions exploit the division into states, operators and measurements, for example by mixing the input states for one experiment with the operators and measurements of another. It would be interesting to see whether exploiting such a division improves the results. In general, it isn’t intuitively clear why crossover works, given the nature of quantum optics experiments, but our tests demonstrated that it was an important element of the search process, increasing diversity and enabling a larger space to be explored. Despite this, we suspect that more in-depth tests (that we reserve for future work) might reveal that the advantage of crossover comes more from the occasions when it mixes similar experiments with each other. In the next subsection we give the specific values of the genetic algorithm settings and hyperparameters for the results presented in this paper.

Toolbox Selection

States

Fock states

Number
Min. 0 Max. 2

Coherent states

|alpha|
Min. 0 Max. 4

arg(alpha)
Min. 0 Max. 6.2831

Squeezed states

|z|
Min. 0 Max. 1.3

arg(z)
Min. 0 Max. 6.2831

Two mode squeezed states

|z|
Min. 0 Max. 1.3

arg(z)
Min. 0 Max. 6.2831

Operations

Beam splitter

T
Min. 0 Max. 1

Displacement

|alpha|
Min. 0 Max. 4

arg(alpha)
Min. 0 Max. 6.2831

Phase shift

theta
Min. 0 Max. 6.2831

Single mode squeezing

|z|
Min. 0 Max. 1.3

arg(z)
Min. 0 Max. 6.2831

Two mode squeezing

|z|
Min. 0 Max. 1.3

arg(z)
Min. 0 Max. 6.2831

Measurements

On/off detector

on/off
Min. 0 Max. 1

Detector loss 0

Number resolving detector

n
Min. 0 Max. 2

Detector loss 0

Homodyne measurement

|x_lambda|
Min. 0 Max. 5

lambda
Min. 0 Max. 6.2831

Multiplex detector

n
Min. 0 Max. 2

number of detectors 5

Detector loss 0

Other options

Number of modes 2

Max. number of operations 5

Max. Truncation 140

Loss on output state 0

Load settings Save settings Next

FIG. 9: The user interface that opens when first running our program, allowing the user to choose the elements of the toolbox they have available, and the parameter limit for these elements. After pressing “Next”, the user is taken to a similar user interface that allows the fitness function to be specified. After this has been chosen the genetic algorithm runs and generates optimised experiment designs.

The mutation functions we used mutated both the experiment arrangements and the parameters, and therefore our search is performed simultaneously over these quite diverse characteristics of an experiment. Matlab did not have a mutation function with enough flexibility for our purposes, so we introduced two mutation functions, `power mutation` and `power selection`. Both are based on [78]. In short, `power mutation` mutates every gene in the genome by a random distance, whose maximum magnitude is determined by the value of a hyperparameter named `power`, for which `power=1` will mutate each gene to a completely random new value, whereas `power=∞` doesn’t mutate at all. `power selection` only mutates a fraction (given by `rate`) of the genes. See [27] and [78] for further details.

Throughout we used tournament selection to select the parents for the next generation.

3. Obtaining the results in this paper

We mainly optimised our search algorithm using the toolbox labelled ‘Tool 1’ in Fig. 4 and Table. II. The main reason for this being that this allowed us to compare directly with the results in [1]. For our initial runs, using Matlab’s inbuilt genetic algorithm with a population size of 200, the search often converged to the same results in [1]. But after increasing the population size, using our 3-stage algorithm, and optimising the hyperparameters, we soon found the greatly-improved results presented in this paper. Having found a variety of genetic algorithm settings and hyperparameters, we then ran AdaQuantum for all the toolboxes presented in this paper. To obtain each result, we ran the AdaQuantum for 96 hours on 16 cores of the University of Nottingham’s High Performance Computing facility. For most of the toolboxes and fitness functions, we ran the algorithm twice with different settings/hyperparameters, then selected the better of the two results. The settings and hyperparameters are shown in Table. V. The population sizes in stages 1, 2 and 3 were 10^7 , 10^5 and 2×10^4 , respectively.

It is important to stress here that the focus of this paper is not on optimising the genetic algorithm. Rather, our goal was to create a search algorithm that worked reliably for a number of different fitness functions, and to find a range of hyperparameters and settings such that the algorithm works consistently. Researchers who use AdaQuantum with

FF & Toolbox	crossover function	crossover fraction	tournament size	mutation function	power	rate
Pure QFI & Tool 1	scattered	0.3	8	power mutation	10	-
Pure QFI & Tool 2	single point	0.3	8	power selection	4	0.2
Pure QFI & Full	two point	0.3	8	power mutation	10	-
Pure QFI & No PNRD	two point	0.3	8	power mutation	10	-
QFI, loss = 0.05	single point	0.3	8	power selection	10	0.5
QFI, loss = 0.1	two point	0.3	8	power selection	20	0.1
QFI, loss = 0.2	scattered	0.3	8	power mutation	5	-
QFI, loss = 0.3	scattered	0.3	8	power mutation	5	-
BMSE, $\mu = 1$	scattered	0.3	8	power mutation	10	-

TABLE V: The settings and hyperparameters for the results in this paper. FF = fitness function. For all runs the number of elite children was 10. Despite the crossover rate being 0.3 for all results here, we found that values between 0.2 and 0.5 often gave comparable results, suggesting the crossover does provide some benefit. Unfortunately we do not have available the hyperparameters for the BMSE runs for the remaining values of μ . However, we found that for the BMSE the search converged quickly for all the hyperparameters.

a different fitness function will likely have to experiment with different settings/hyperparameters, but our choices in Table. V should serve as a useful guide. In ongoing work we are performing a thorough optimisation of the settings/hyperparameters using a range of toolboxes and fitness functions, and comparing this to a range of different search algorithms.

We made a number of observations when running the algorithm. Some of the runs for pure states (in particular for the larger toolboxes) did not fully converge after the 96 hours time, suggesting that improvements over the results here might be possible. Furthermore, for most of the pure-QFI fitness functions, different runs of the algorithm often produced different experiments with similar fitness values, suggesting there are whole classes of states with large QFI values. In contrast, for the mixed state runs and the BMSE the algorithm often converged quickly (long before the 96 hour running time) to the same or similar experiments, even for a range of hyperparameters/settings.

Appendix D: Comparison to previous work

Ref. [1] is most similar to our own, as it too uses a search algorithm to optimise the arrangement and parameters in a quantum optics setup to find quantum states for metrology. However, our current algorithm has numerous significant advantages over [1]: i) The search algorithm in [1] is less refined, and is in effect a stochastic hill-climbing algorithm with random restart. ii) The simulation in our current algorithm has overcome various numerical challenges present in [1], which meant that the search algorithm in [1] had to truncate the number basis at 30 photons, in comparison to our current work that allows the truncation to reach 170. The combined effect of the developments i) and ii) is that the results we present here greatly exceed those in [1], with up to a 5-fold QFI improvement (and note that truncating at 30 photons does not even allow any of the states we found here to be properly simulated).

Further advantages over [1] include: iii) In our current work we model the dominant experimental noise in typical quantum optics experiments (photon loss on the output state and imperfect detectors). iv) The algorithm we present here is available on GitHub [27], and as we detail in Section ID it has been designed specifically for flexibility (e.g. AdaQuantum can optimise with different numbers of modes and toolboxes), as illustrated by the three different fitness functions we study here (compared to just the QFI in [1]). AdaQuantum can also be easily modified, so that researchers can add/change both the quantum optics elements to be searched over, and the fitness function to be optimised. A long term goal of our research is to construct an algorithm that can fully model, and then optimise and design, realistic and general quantum optics experiments. Together iii) and iv) mean that AdaQuantum has taken a major step towards this goal.

Refs. [2] and [3] differ from ours in both the optics experiments of interest (they search for higher-dimensional states with angular momentum, and look for certain entangled states) and the algorithms used ([2] uses random search with learning, and [3] uses reinforcement learning). At present the optics settings are different enough so that our algorithm

cannot be directly applied to their research, and vice versa, though there is no fundamental reason why future work cannot apply global search algorithms to their setting, and reinforcement learning to ours.

Finally, in Refs. [4] and [5] a machine learning algorithm can be used to optimise the parameters in an optics experiment. [4, 5] have the advantage that they use gradient descent to learn the parameters, rather than our stochastic algorithm, but this seems to come at the expense of the speed of their simulation (we have not done a full comparison, but the online version of the algorithm they use, **Strawberry Fields** [26], has to truncate the Fock basis at 20 photons). Furthermore, we optimise over the arrangement of elements in addition to the parameters; our algorithm is specifically designed so that the fitness function can be easily modified; and we have realistic experimental setups in mind (though we believe these latter two features could be incorporated into [26]).
