Improving Quantum Clocks via Semidefinite Programming

Michael Mullan[1,2] and Emanuel Knill[1,2]

[1] *National Institute of Standards and Technology*
[2] *University of Colorado at Boulder*

(Dated: July 28, 2011)

The accuracies of modern quantum logic clocks have surpassed those of standard atomic fountain clocks. These clocks also provide a greater degree of control, because before and after clock queries, we are able to apply chosen unitary operations and measurements. Here, we take advantage of these choices and present a numerical technique designed to increase the accuracy of these clocks. We use a greedy approach, minimizing the phase variance of a noisy classical oscillator with respect to a perfect frequency standard after an interrogation step; we do not optimize over successive interrogations or the probe times. We consider arbitrary prior frequency knowledge and compare clocks with varying numbers of ions and queries interlaced with unitary control. Our technique is based on the semidefinite programming formulation of quantum query complexity, a method first developed in the context of deriving algorithmic lower bounds. The application of semidefinite programming to an inherently continuous problem like that considered here requires discretization; we derive bounds on the error introduced and show that it can be made suitably small.

I. QUANTUM CLOCKS

A. The Clock Protocol

Most atomic clocks are designed to lock a noisy classical oscillator to the resonance of an atomic standard. Typically, this is accomplished via the following clock interrogation protocol:

1. **Preparation**: The atomic system is prepared in some initial state.
2. **Query**: The classical oscillator and the atomic system interact. This modifies the atomic state in some way that depends on both the resonant atomic frequency ω_0 and the frequency of the classical oscillator ω.
3. **Measurement**: The atomic system is measured and provides some information about $\omega - \omega_0$.
4. **Correction**: The classical oscillator is adjusted based on this information, ideally reducing $|\omega - \omega_0|$.

This protocol must be repeated indefinitely, as the noisy classical oscillator drifts over time. Furthermore, the information gained in step 3 is always incomplete. Consequently, the frequency of the classical oscillator is never known exactly and must be described by a probability distribution. Figure 1 illustrates how this distribution changes as the clock is run. In this probabilistic perspective, our goal is to maximize our knowledge of the classical oscillator. For the purpose of maintaining an accurate clock, we minimize the phase variance of the classical oscillator with respect to the atomic frequency standard by optimizing over state preparation (step 1) and the post-query measurement (step 3). We also consider interrogations consisting of multiple queries interlaced with unitary control.

A more complete characterization of a clock involves estimating the total time difference between the clock and an ideal clock based on the atomic standard since the clock was started. The time difference is measured in terms of the total phase difference and requires integrating the frequency differences over time. A full Bayesian treatment of this characterization problem requires that we maintain the complete history of the classical oscillator by way of the joint probability distribution, $p(\omega_1, \omega_2 \ldots \omega_t)$, where the marginals $p(\omega_i)$ reflect our knowledge of the classical oscillator at a specific time in the past. However, in this paper we use a greedy approach, namely our optimization procedure does not

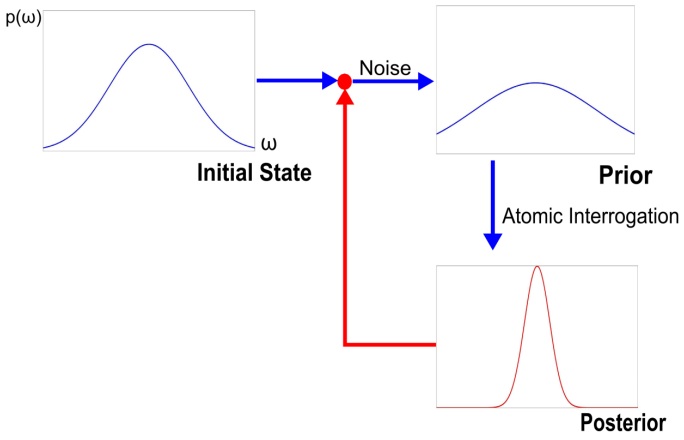


FIG. 1: Evolution of the classical oscillator's probability distribution in a clock protocol. Noise decreases our knowledge of the classical oscillator's frequency, widening $p(\omega)$. The prior probability distribution on the right is used in our optimization procedure, and describes the average frequency of the classical oscillator over the probe time T. A measurement of the atomic standard then yields a measurement outcome a, which can be used to compute the posterior distribution $p(\omega|a)$. Over many iterations of this procedure, we gain knowledge of the history of the frequency differences. This history can be integrated to estimate the total time difference between the classical oscillator and an ideal atomic clock.

take into account this history, and our prior knowledge consists only of the marginal $p(\omega_t)$.

Except for a brief discussion in Sect. II D, we ignore the effects of decoherence on the atomic standard and consider only random fluctuations of the classical oscillator's frequency. We focus on clocks with relatively few atoms and assume full quantum control of the atoms. These clocks are often referred to as quantum clocks. An example is the highly accurate Al^+ quantum clock at NIST [1], which is a candidate for the application of the techniques developed here.

We consider a simplified model of an N-atom clock, whose state is a superposition of the $N+1$ symmetric Dicke states $|0\rangle, \ldots, |N\rangle$ of N identical two-level systems. For example, if $N = 2$, the Dicke states are given by $|0\rangle = |00\rangle_{AB}$, $|1\rangle = (|01\rangle_{AB} + |10\rangle_{AB})/\sqrt{2}$, and $|2\rangle = |11\rangle_{AB}$. Ref. [2] shows that nothing can be gained by considering other states of the two-level systems. The N atoms begin in the state $|0\rangle\langle 0|$ and are then initialized (step 1) by the application of a unitary operator $U(0)$ to the state $\rho(0) = U(0)|0\rangle\langle 0|U(0)^\dagger$. A query (step 2) consists of the application of a second unitary operator that depends on both ω and ω_0; normally, this dependence is only on the difference $\omega - \omega_0$. Then

$$\rho(0) \to \rho(0)' = \Omega(\omega - \omega_0)\rho(0)\Omega(\omega - \omega_0)^\dagger. \qquad (1)$$

Finally, the system is measured (step 3) with a positive operator-valued measure (POVM) $\{P_a\}_a$.

Interrogation is often done via the Ramsey technique [3]. Here, the atoms are subject to two pulses from a classical oscillator of frequency ω, separated by a period of free evolution of length T. These pulses are short enough for their dependence on ω to be neglected, and so for our purposes, their effect can be absorbed into the definitions of $U(0)$ and $\{P_a\}_a$. The period of free evolution is equivalent to a z rotation by an angle of $(\omega-\omega_0)T$. Thus, $\Omega(\omega - \omega_0)$ in Eq. (1) is given by

$$\Omega(\omega - \omega_0) = e^{-iJ_z(\omega_0-\omega)T}, \qquad (2)$$

where J_z is the total z angular momentum operator, $J_z|k\rangle = (k - N/2)|k\rangle$. By making a global phase change, we can write the evolution of the Dicke states as

$$|k\rangle \to e^{-ik(\omega-\omega_0)T}|k\rangle. \qquad (3)$$

Without loss of generality, for the remainder of this paper we assume that $\omega_0 = 0$.

We refer to the action in Eq. (1) as a "clock query". We can generalize such an action by combining multiple queries with interlaced unitary operators. The final clock state can then be written as

$$\begin{aligned}\rho(t_f) &= \Omega(\omega)U(t_f)\ldots\Omega(\omega)U(1) \\ &\quad \times \rho(0)U^\dagger(1)\Omega(\omega)^\dagger\ldots U(t_f)^\dagger\Omega(\omega)^\dagger, \qquad (4)\end{aligned}$$

provided that the relative frequencies do not drift between steps 1 and t_f. We refer to the complete action in Eq. (4) as a "clock interrogation". As written, $U(t)$ acts only on the clock's atoms; below, we consider $U(t)$ that can also act on arbitrary ancilla atoms. Our goal is to minimize the expected cost

$$\begin{aligned}\langle C\rangle &= \int\int C(\omega - f_a)p(a|\omega)p(\omega)d\omega da \\ &= \int\int C(\omega - f_a)\mathrm{tr}(P_a\rho_\omega(t_f))p(\omega)d\omega da. \qquad (5)\end{aligned}$$

Here, f_a is the classical frequency estimate for measurement outcome a with associated POVM operator P_a, $p(\omega)$ is the prior probability distribution of the classical oscillator's frequency, and $p(a|\omega)$ is the probability of measurement outcome a given ω and the interrogation protocol. The subscript ω of ρ indicates that the clock was interrogated by a classical oscillator at frequency ω. If $C(\omega - f_a) = (\omega - f_a)^2$, then a minimization of $\langle C\rangle$ is equivalent to a minimization of the expected posterior variance of the relative phase change ωT. To minimize $\langle C\rangle$ we vary the $U(t)$ in Eq. (4) and the final POVM $\{P_a\}_a$. The operator $\Omega(\omega)$ is considered fixed.

In principle one can consider simultaneously optimizing multiple sequential interrogations with varying probe times T. Because a fixed probe time is unable to distinguish between frequencies differing by multiples of $2\pi/T$, varying the probe time is necessary to avoid undetected frequency hops. Here, we consider only one interrogation at a time and fix the probe time T. In this case, there is a scaling symmetry $\omega \to \alpha\omega$ and $T \to \omega/\alpha$, so we fix $T = 1$ from now on.

B. Background and Summary

A great deal of research has been done on the theory of atomic clock optimization. Of particular interest has been the question of how quantum effects such as entanglement and squeezing can help overcome the atomic shot-noise precision limit of $O(1/\sqrt{N})$ and approach the Heisenberg limit of $O(1/N)$ for N atoms. The possibility that these effects can result in improved precision was raised in [4, 5]. These and related ideas are now at the foundations of the subject of quantum metrology [6]. Our work is based on and extends the analytical studies of Bužek *et al.* [2], who sought to optimize clock interrogations under the assumption of a uniform prior and a family of periodic cost functions. Starting from results of Holevo [7], they obtained a family of initial states that perform well for large numbers of atoms. Recently, Demkovicz-Dobrazanski [8] optimized interrogations with costs determined by the periodic function $C(\omega - f_a) = 4\sin^2\frac{(\omega-f_a)}{2}$ for arbitrary continuous priors. While this approach is largely analytical, its implementation requires the numerical maximization of a trace norm. These works focus on optimizing a single interrogation. Long-term stability has been analyzed for entangled states of a number of atoms in Ref. [9], and

for a family of squeezing protocols of atomic ensembles in Ref. [10]. These studies account for phase noise in the classical oscillator and optimize clock protocols given specific feedback mechanisms and noise models.

The periodic cost functions studied in [2, 8] are convenient for analytic studies of clock optimization but do not penalize phase errors greater than 2π, even though they correspond to frequency estimates far from the true frequency of the oscillator. This issue becomes important when multiple interrogations and long-term clock stability are considered. Given that we do not explicitly consider either, and that in the model described above, phases differing by multiples of 2π cannot be distinguished, this may seem irrelevant. Specifically, when optimizing interrogations consisting of one query or multiple queries with identical probe times as we do here, the prior and the cost function can in principle be folded into the interval $[-\pi, \pi]$. However, this folding results in a cost function that depends on the prior. Because of this and in view of future extensions of this work, we consider non periodic cost functions, particularly the quadratic one.

The goal here is to apply the semidefinite programming strategy originally developed for quantum query algorithms [11, 12] to the problem of optimizing clock interrogations. This enables the greatest flexibility in searching for solutions, as both the prior and the cost function can be arbitrary. The first major obstacle is that the unknown parameter of the queries is continuous rather than discrete and finite as in typical quantum query algorithms. We overcome this by showing how to systematically discretize the parameter spaces while having good control of the discretization errors. The second major obstacle is that the size of the semidefinite program (SDP) grows at a rate of $\Theta\left(t_f|O|^2(|Q|+|F|)\right)$, where $|O|$ is the number of points in our discretization, $|Q|$ is the number of atoms we are simulating, and $|F|$ is the number of POVM elements. This limits the number of atoms and the level of discretization for which general clock interrogations can be optimized, depending on available computational resources. We show that for small but useful numbers of atoms, the SDPs can be implemented and solved given current resources.

The semidefinite programming strategy is formulated in full generality in Sec. II. Our version does not restrict the set of possible queries or the prior over query parameters. We show how to explicitly reconstruct the algorithm and measurements from the solution and discuss how to modify the strategy to account for query noise. This may be of independent interest for interpolating between quantum and classical query complexity. In Secs. III and IV, we specialize the SDPs to the case of quantum clocks. Here we show how to discretize the parameter spaces to obtain finite SDPs while bounding discretization errors. In Sec. V we show the results from applying the discretized SDPs to concrete clock problems. In particular we compare the computational results obtained to prior work, demonstrating both the ability to obtain improvements and to determine bounds on optimal costs.

II. THE SEMIDEFINITE PROGRAMMING FORMULATION OF QUANTUM QUERY COMPLEXITY

A. Constructing the Semidefinite Program

The operation of a quantum clock can be expressed naturally in the query model of quantum computation. Here, we are given an oracle (or black box) chosen from some finite set $\{\Omega(1), \Omega(2)...\Omega(M)\}$. Each oracle $\Omega(x)$ is a unitary operator, selected with probability $p(x)$. The goal is to determine x with queries, that is, with applications of $\Omega(x)$ to quantum states and measurement. Often in this context, one is interested in minimizing the number of queries needed to learn x with near certainty. Here, however, we fix the number of queries and seek to minimize the expected difference between an estimate x' of x and x as quantified by a cost function $C(x, x')$.

We adapt the semidefinite programming formulation of quantum query complexity developed by Barnum *et al.* [11, 12] in the context of proving quantum lower bounds. They cast the problem of determining the number of queries required in terms of a test of feasibility of a semidefinite program. Here we aim to minimize an expected cost, which requires optimizing an objective function with an extended SDP. For a general introduction to semidefinite programming see Ref. [13].

To formulate the SDP, we introduce quantum systems O and Q for the oracle and the querier, respectively. System Q is the one on which the oracle operators $\Omega(x)$ act. An additional system of ancillas A may be used by the querier; for the purposes of the SDP, we normally trace out A. We use the convention that the systems on which an operator acts are denoted by superscripts. For states (density operators) ρ, partially omitted system superscripts imply the partial trace over the omitted systems. Superscripts may be completely omitted if the set of systems being acted on is clear or irrelevant. The dimension of the state space of system S is denoted by $|S|$. Initially, the state of the oracle system is given by $\rho_0 = \sum_x \sum_y \sqrt{p(x)}\sqrt{p(y)}|x\rangle\langle y|$. By representing the oracle probabilities with a pure superposition, we encode the fact that there is no information about x available to the querier (or any system other than O). We define the joint operator $\Omega^{OQ} = \sum_x |x\rangle\langle x|\Omega^Q(x)$, which applies $\Omega(x)$ to system Q conditional on the state $|x\rangle$ of the oracle. In this setting, a multi-query quantum computation is given by the composition

$$\rho^{OQA}(t_f) = \Omega^{OQ} U^{QA}(t_f) \dots \Omega^{OQ} U^{QA}(1) \rho^{OQA}(0) U^{\dagger QA}(1) \Omega^{\dagger OQ} \dots U^{\dagger QA}(t_f) \Omega^{\dagger OQ} \quad (6)$$

followed by a measurement of QA with a POVM. Here $\rho^{OQA}(0)$ is the initial state, with $\rho^O(0) = \rho_0$ pure as defined above, and $\rho^{QA}(0)$ is a state chosen by the querier. The unitary operators $U^{QA}(t)$ are inter-query operators that can be chosen arbitrarily for each step t. The querier's protocol (or algorithm) is determined by the initial state $\rho^{QA}(0)$, the $U^{QA}(t)$, and the POVM. Refs. [11, 12] show that there is a correspondence between these algorithms and solutions to a set of semidefinite constraints. We split the constraints into those that correspond to the initial state and choice of inter-query operators, and those that correspond to the POVM. The first set, S_E, consists of

1. $\rho^O(0) = \rho_0$,

   and the following for $t \in \{0, 1, \dots, t_f\}$:

2. $\rho^{OQ}(t) \geq 0$,
3. $\rho^O(t) = tr_Q(\rho^{OQ}(t))$ and
4. $\rho^O(t) = tr_Q(\Omega^{\dagger OQ} \rho^{OQ}(t-1) \Omega^{OQ})$.

The constraints corresponding to the POVM and the objective to be optimized can be based on a connection to the concept of remote state preparation as described next.

B. Measurement and Remote State Preparation

Remote state preparation [14] involves two systems O and Q. Given a joint state of the two systems, we can prepare states of O conditional on measurements of Q.

Definition 1. *We say that we can* remotely prepare $\{\sigma_a^O\}_a$ *from the state* ρ^{OQ} *on* O *and* Q *if there exists a POVM of* Q *with operators* $\{P_a^Q\}_a$ *such that* $tr_Q(P_a^Q \rho^{QO}) = \sigma_a^O$.

The members σ_a^O of the set that can be remotely prepared are positive operators with trace $0 \leq \mathrm{tr}(\sigma_a^O) \leq 1$. The definition implies that the state $\sigma_a^O / \mathrm{tr}(\sigma_a^O)$ can be conditionally prepared with probability $\mathrm{tr}(\sigma_a^O)$ by means of a fixed POVM on Q. For definiteness, consider a POVM $\{P_a^Q\}$. Any implementation of the POVM has the desired effect. Such an implementation is described by a quantum operation with Kraus operators E_a such that $E_a^\dagger E_a = P_a$. For outcome a, the unnormalized OQ state is $E_a^Q \rho^{OQ} E_a^{\dagger Q}$. Cyclicity of partial trace implies that $\mathrm{tr}_Q E_a^Q \rho^{OQ} E_a^{\dagger Q} = \mathrm{tr}_Q P_a^Q \rho^{OQ} = \sigma_a^O$. The probability of the outcome is $\mathrm{tr}(\sigma_a^O)$.

Theorem 1 characterizes the set of states that can be remotely prepared according to Def. 1 from a pure state.

Theorem 1. *Let* ρ^{OQ} *be a pure state of systems* O *and* Q. *We can remotely prepare* $\{\sigma_a^O\}_a$ *if and only if* $\sum_a \sigma_a^O = \rho^O$.

The proof of Thm. 1 is given in App. A.

In order to apply Thm. 1 we formulate the expected cost in terms of the POVM-conditional unnormalized states σ_a^O. In the clock problem, the querier associates an estimated frequency f_a with each measurement outcome a. Assume for simplicity that the possible classical oscillator frequencies ω come from a finite set. We can define operators A_a^O in the oracle basis $\{|\omega\rangle\}_\omega$ by $(A_a)_{\omega,\omega'} = \delta_{\omega,\omega'} C(\omega - f_a)$. Eq. (5) can be rewritten as

$$\langle C \rangle = \sum_a \mathrm{tr}(A_a \sigma_a). \quad (7)$$

With this equation as motivation, we consider the general situation where the expected cost of a querier protocol is computed from predetermined cost operators A_a according to the POVM-conditional oracle states as in Eq. (7). This motivates the following SDP $S_M(\rho^O)$ for a given oracle state ρ^O:

$$\text{Minimize } \mathrm{tr}(\sum_a \sigma_a A_a) \text{ subject to: } \begin{cases} \forall a \sigma_a \geq 0, \\ \sum_a \sigma_a = \rho^O. \end{cases} \quad (8)$$

Theorem 2. *Suppose that* ρ^{OQA} *is pure. Then* $S_M(\rho^O)$ *computes the minimum expected cost over measurements on* QA.

Proof. It suffices to observe that according to Thm 1, sets $\{\sigma_a\}_a$ satisfying the constraints of S_M are precisely the sets that can be remotely prepared with access to systems QA. □

The complete SDP for the query optimization problem is obtained by combining $S_Q = S_E \cup S_M(\rho^O(t_f))$. Because query algorithms have access to the non-O systems of a purification of $\rho^O(t_f)$, S_Q computes the optimal average cost of a t_f-query quantum algorithm. The SDP considered in the spectral adversary method [11, 12] is a relaxation of S_Q with a modified objective designed to determine a lower bound on the number of queries needed to obtain some fixed probability of error. There has been significant recent progress in refining this relaxation [15] and demonstrating that it is nearly exact [16]. Note that in our case, we can construct cost operators A_a such that S_Q minimizes the probability of error for fixed t_f.

C. Algorithm and POVM reconstruction

For any solution of S_Q, in particular the optimal one, there is an explicit query algorithm that achieves the as-

sociated cost objective. In particular, given the sequence of density matrices $\rho^{OQ}(t)$ and the conditional operators σ_a, it must be possible to infer the sequence of unitaries in Eq. (6) and the POVM achieving the expected cost. The POVM yielding the conditional operators σ_a can be constructed as in the proof of Thm. 1. To determine the unitaries $U^{QA}(t)$, first extend the querier system with an ancilla A, where A has dimension $|O||Q|$. For each t, construct pure states $\rho'^{OQA}(t)$ and $\rho^{OQA}(t+1)$ by purifying $\rho'^{OQ}(t) = \Omega^{\dagger OQ}\rho^{OQ}(t)\Omega^{OQ}$ and $\rho^{OQ}(t+1)$, respectively. By construction, $\rho'^{O}(t) = \rho^{O}(t+1)$. One can therefore use the Schmidt forms for the pure states $\rho'^{OQA}(t)$ and $\rho^{OQA}(t+1)$ to construct unitaries $U^{QA}(t)$ satisfying

$$\rho^{OQA}(t+1) = U^{QA}(t)\rho'^{OQA}(t)U^{\dagger QA}(t). \tag{9}$$

D. Noise

The SDP S_Q is based on the assumption that there is no noise in the query process. In particular, this excludes decoherence during clock queries. It is possible to adapt S_Q to include the effects of noise. This requires that we extend the states ρ^{OQ} by querier-inaccessible systems E_i modeling the environments causing the noise. The net effect of query and noise can be modeled by an oracle-conditional isometry defined by

$$|x\rangle\langle y|\rho^{Q} \to |x\rangle\langle y|\Omega(x)^{QE}\rho^{Q}|\epsilon\rangle\langle\epsilon|\Omega(y)^{\dagger QE}, \tag{10}$$

where $|x\rangle, |y\rangle$ are the standard oracle basis states, $|\epsilon\rangle$ is a fixed initial state of E and $\Omega(x)^{QE}$ is unitary. Define $\Omega^{OQE} = \sum_x |x\rangle\langle x|\Omega(x)^{QE}$. In the absence of noise, $\Omega(x)^{QE} = \langle x|\Omega^{OQ}|x\rangle$. In many cases one can decompose $\Omega(x)^{QE} = D^{QE}\langle x|\Omega^{OQ} \otimes I^{E}|x\rangle$ for an O-independent unitary D^{QE}. An example is the clock query in the presence of phase decoherence. In a sequence of queries, a new version of E, E_i is introduced at each step by the isometry. Let $E^t = E_1 E_2 \dots E_t$. To account for the noisy query, the SDP S_Q is modified to S_D as follows:

Minimize $\mathrm{tr}(\sum_a \sigma_a^{OE^tf} A_a^O)$ subject to

$$\begin{aligned}
&\sigma_a^{OE^tf} \geq 0\\
&\sum_b \sigma_b^{OE^tf} = \rho^{OE^tf}(t_f)\\
&\rho^{OE^tQ}(t) \geq 0\\
&\rho^{OE^t}(t) = tr_Q(\rho^{OE^tQ}(t))\\
&\rho^{OE^t}(t) = tr_Q\left(\Omega^{OQE_t}\rho^{OE^{t-1}Q}(t-1)|\epsilon\rangle\langle\epsilon|\Omega^{\dagger OQE_t}\right)\\
&\rho^{OE^0}(0) = \rho_0.
\end{aligned}$$

(There is an implicit "for all" over the free variables a and t.)

Phase decoherence is a particularly interesting example of this more general SDP. As mentioned above, for complete phase decoherence, the query isometry factors, in this case giving

$$D^{QE}|i\rangle^{Q}|\epsilon\rangle^{E} = |i\rangle^{Q}|\epsilon_i\rangle^{E}, \tag{11}$$

where the $|\epsilon_i\rangle$ are orthonormal states. The effect is to perfectly correlate the environment at each step with the standard query basis. In the context of standard query algorithms with a cost function that captures the probability of successfully identifying the oracle, the optimal solution to the SDP with complete phase decoherence corresponds to an optimal classical query algorithm for the given number of queries t_f. By modifying D^{QE} to model incomplete phase decoherence, it is possible to interpolate between classical and quantum query algorithms, albeit at the large cost of adding the systems E_i.

Note that it is not possible to simply trace out the E_i in the SDP: As can be seen from the method of reconstructing the algorithm from a solution of the SDP, this would be equivalent to giving the querier access to the E_i. That is, the querier has implicit access to anything that gets traced out, since traced out systems are not constrained by the SDP. When the noisy query factors, this is equivalent to not having had any noise at all, because the querier can just undo the noise isometries.

III. THE CLOCK SDP

One can apply the SDP S_Q to the clock problem, but the result is not finitely implementable because the oracle system O is continuous and the cost operators are continuously indexed. To make the SDP finite, we discretize both the possible oracle frequencies and the frequency estimates that determine the cost operators. In particular, we constrain $\omega \in \{\omega_1, \dots, \omega_d\}$, and restrict the measurement outcomes to a finite set $a \in \{1, \dots, m\}$ associated with the set of frequency estimates $F = \{f_a\}_{a=1,\dots,m}$. The frequency estimates need not be among the ω_i. The query system Q's Hilbert space is spanned by the Dicke states $|k\rangle$. The oracle initial state is given by

$$\rho^{O}(0) = \sum_{x,y} \sqrt{p(\omega_x)p(\omega_y)}|x\rangle\langle y|, \tag{12}$$

where $p(\omega_x)$ is the prior probability of ω_x, a discretization of the continuous prior, and $|x\rangle$ is the oracle basis state corresponding to classical oscillator frequency ω_x. The operator Ω^{OQ} is now given by

$$\Omega^{OQ} = \sum_x |x\rangle\langle x| \exp(-iJ_z^Q\omega_x), \tag{13}$$

up to irrelevant x-dependent phases. As noted before Eq. (7), the cost operators are given by

$$(A_a)_{x,y} = C(\omega_x - f_a)\delta(x,y). \tag{14}$$

Since the SDP depends on p and F, we denote it by $S_C(p,F)$. We also use $S_C(p,F)$ to denote the optimum

achievable cost given p and F. For discretized prior and frequency estimates, this is the cost computed by the SDP.

As written, the size of the SDP S_Q is $\Theta\left(t_f|O|^2(|Q|^2+|F|)\right)$, where $F = \{f_a\}$. Our implementation of the SDP takes advantage of the fact that for the clock problem, Ω is diagonal in the Dicke basis, thus reducing the size and therefore the memory and time resources required to solve the SDP. In particular, the set of solutions is invariant under the transformation $\rho^{OQ}(t) \to U(t)^Q \rho^{OQ}(t) U(t)^{\dagger Q}$ for operators $U(t)$ diagonal in the Dicke basis. It follows that we can restrict the SDP by assuming $\rho^Q(t)$ is diagonal. The restricted SDP's size is $\Theta\left(t_f|O|^2(|Q|+|F|)\right)$.

IV. DISCRETIZATION ERROR

When solving the SDP, it is necessary to make an estimate of the difference between the discretized SDP's optimal cost and that of the infimum of the costs of solutions to the continuous problem. We separately bound the error due to discretizing the prior (*oracle discretization*) and that due to discretizing the measurement outcomes (*querier discretization*). To deal with the oracle discretization error, we solve clock SDPs with random oracle discretizations so that the average cost for different discretizations is a lower bound on the optimum cost of the continuous problem. An upper bound is obtained by a cost integral applied to any of the SDPs solved. For the quadratic cost function, the querier discretization error can be bounded in a way that depends only on the set $F = \{f_a\}_a$ and can be made to go to zero by increasing the size and resolution of F. We start with querier discretization.

A. Querier Discretization Error

To guarantee that we determine the optimum cost for a frequency prior p, the set of allowed querier frequency estimates F should be all of $\mathbb{R}$. To solve the clock SDP numerically, we restrict F to a finite subset. By definition, $S_C(p,F) \geq S_C(p,\mathbb{R})$. The goal of this section is to obtain a lower bound on $S_C(p,\mathbb{R})$ depending on F, and to explain an iterative strategy for locally optimizing F. Let $F = \{f_1, \ldots, f_N\}$ with $f_j < f_{j+1}$.

Theorem 3. *Let C be a second differentiable, non-negative function satisfying $C''(x) \leq b$, $C(0) = 0$ and C is monotone on $[0,\infty)$ and $(-\infty, 0]$. Define $M(\omega)$ by*

$$M(\omega) = \begin{cases} C(\omega - f_1) & \text{for } \omega \leq f_1 \\ C(\omega - f_N) & \text{for } \omega \geq f_N \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

We have the following inequality:

$$S_C(p,F) - S_C(p,\mathbb{R}) \;\leq\; \max_j \frac{b}{8}(f_{j+1} - f_j)^2 + \int M(\omega)p(\omega)d\omega. \tag{16}$$

The proof is in App. B. A strategy for obtaining an initial choice of an N-point F is to optimize the right-hand-side of Eq. (16).

An important question is how many frequency estimates f_a are needed so that $S_C(p,F) = S_C(p,\mathbb{R})$. One can obtain an upper bound by observing that the rank r of the final oracle density matrix $\rho^O(t_f)$ is bounded by $|Q|^{t_f}$ and $|O|$, where $|Q|$ and $|O|$ are the dimensions of the query and oracle systems. From this one finds that the optimal POVM can always be reduced to at most r^2 elements, implying that a set F with $|F| \leq r^2$ suffices. There is evidence that a smaller set is optimal for the clock problem. For $t_f = 1$ and for some costs and priors, we need only $|F| = r$ elements. [8, 17].

At present, we have no provably correct way of choosing the finite set of frequency estimates optimally. However, for the quadratic cost function $C(x) = x^2$, the following heuristic is often effective at optimizing the choices given an oracle discretization. We begin with any set of estimates. We then run the SDP and use the solution to compute the posterior distribution for each measurement outcome:

$$p(\omega|f_a) = \frac{p(f_a,\omega)}{p(f_a)} = \frac{(\sigma_a)_{\omega,\omega}}{\mathrm{tr}(\sigma_a)}, \tag{17}$$

where the σ_a are the measurement-conditional unnormalized oracle states at the end of the algorithm. Next, we compute the mean of each of these distributions:

$$\langle\omega|a\rangle = \sum_x \frac{(\sigma_r)_{x,x}}{\mathrm{tr}(\sigma_a)}\,\omega_x, \tag{18}$$

where x indexes the finite set of frequencies and the expression in angle brackets is the average of ω given that measurement outcome a was obtained. Replacing the original estimates f_a by their posterior means is guaranteed to improve the cost without having to change the algorithm. We then run the SDP again, replacing the frequency estimate f_a with $\langle\omega|a\rangle$. This procedure is repeated until each estimate is numerically close to its posterior mean. The procedure can be adapted to other costs, but the mean must be replaced by a statistic appropriate for the cost. For example, for $C(x) = |x|$ we compute the median instead of the mean.

B. Oracle Discretization Error

Let $p(\omega)$ be the probability density of the prior distribution of clock frequencies. For simplicity, we assume that the prior distribution is absolutely continuous with

respect to Lebesgue measure. Let $P(\omega)$ be the cumulative distribution function of $p(\omega)$, and P^{-1} its inverse on $(0,1)$. Given an offset $o \in (0, 1/d)$, we can define the probability distribution

$$p_o(\omega) = \sum_{k=0}^{d-1} \frac{1}{d} \delta_{P^{-1}(o+k/d)}(\omega). \tag{19}$$

This is an instance of a discretized prior. Define $\omega(k,o) = P^{-1}(o+k/d)$. The distribution p_o is designed to approximate p as d goes to infinity. If we choose o uniformly at random from $(0,1/d)$, it is identical to p. Specifically,

$$\begin{aligned} &\int_{o=0}^{1/d} p_o(\omega) p(o) do \\ &= \int_{o=0}^{1/d} d \sum_{k=0}^{d-1} \frac{1}{d} \delta_{\omega(k,o)}(\omega) do = p(\omega). \end{aligned} \tag{20}$$

This follows from inverse transform sampling [18] (pg. 28).

To estimate the optimum cost $S_C(p,F)$, we estimate the average $\langle S_C(p_o,F)\rangle_{o\in(0,1/d)}$ by solving $S_C(p_o,F)$ for a number of offsets o chosen uniformly at random from $(0,1/d)$. According to the next theorem, this gives a lower bound on $S_C(p,F)$. We assume that the frequency estimates have already been discretized, $F = \{f_a\}_a$.

Theorem 4. *The following inequality holds:* $\langle S_C(p_o,F)\rangle_o \leq S_C(p,F)$.

Proof. Consider an arbitrary query algorithm $\mathcal{Q}$. Given a prior $r(\omega)$, $\mathcal{Q}$ results in the measurement-conditional, unnormalized oracle states $\sigma_a(r,\mathcal{Q})$. Because the cost operators are diagonal, we can consider just the diagonals of the $\sigma_a(r,\mathcal{Q})$, which define the joint probability distributions $r(a,\omega|\mathcal{Q})$. Because the oracle operators are conditional on the standard oracle basis, $r(a,\omega|\mathcal{Q})$ factors as

$$r(a,\omega|\mathcal{Q}) = q(a|\omega,\mathcal{Q}) r(\omega), \tag{21}$$

where, as indicated, the distribution q does not depend on r.

The average cost for $\mathcal{Q}$ and r is given by

$$C(r,\mathcal{Q}) = \sum_a \int C(\omega - f_a) r(a,\omega|\mathcal{Q}) d\omega. \tag{22}$$

If $\mathcal{Q}_{\min}$ is an optimal algorithm for $S_C(r,F)$, then for any algorithm $\mathcal{Q}$, it follows that $C(r,\mathcal{Q}) \geq C(r,\mathcal{Q}_{\min}) = S_C(r,F)$. Let $\mathcal{Q}_o$ and $\mathcal{Q}_{\text{opt}}$ be optimal algorithms for $S_C(p_o,F)$ and $S_C(p,F)$, respectively. Then

$$\begin{aligned} \langle S_C(p_o,F)\rangle_o &= \langle C(p_o,\mathcal{Q}_o)\rangle_o \\ &\leq \langle C(p_o,\mathcal{Q}_{\text{opt}})\rangle_o, \end{aligned} \tag{23}$$

where the subscript on the expectations indicates that they are taken with respect to the distribution over offsets o. The intuition here is that we can obtain a lower cost if we are able to choose a different algorithm $\mathcal{Q}_o$ for different choices of o, than if we are forced to use the same algorithm $\mathcal{Q}_{\text{opt}}$ in all cases. We can continue from the last line of Eq. (23) as follows:

$$\begin{aligned} &\langle C(p_o,\mathcal{Q}_{\text{opt}})\rangle_o \\ &\quad= \left\langle \sum_a \int C(\omega - f_a) p_o(a,\omega|\mathcal{Q}_{\text{opt}}) d\omega \right\rangle_o \\ &\quad= \left\langle \sum_a \int C(\omega - f_a) q(a|\omega,\mathcal{Q}_{\text{opt}}) p_o(\omega) d\omega \right\rangle_o \\ &\quad= \sum_a \int C(\omega - f_a) q(a|\omega,\mathcal{Q}_{\text{opt}}) \langle p_o(\omega)\rangle_o \, d\omega. \end{aligned} \tag{24}$$

Combining Eqs. (20),(23),(24) then gives

$$\begin{aligned} \langle S_C(p_o,F)\rangle_o &\leq \sum_a \int C(\omega - f_a) q(a|\omega,\mathcal{Q}_{\text{opt}}) p(\omega) d\omega \\ &= C(p,\mathcal{Q}_{\text{opt}}) = S_C(p,F), \end{aligned} \tag{25}$$

proving the claim of the theorem. □

We note that Thm. 4 generalizes to arbitrary oracle problems where the cost operators A_a are diagonal. Furthermore, the proof works for any family of probability distributions p_o such that p is a mixture of the p_o.

Upper bounds on the optimal expected costs can be obtained by applying the inequality in the next Theorem.

Theorem 5. *Let $\mathcal{Q}$ be an algorithm for $S_C(r,F)$. Then*

$$S_C(p,F) \leq \sum_a \int C(\omega - f_a) q(a|\omega,Q) p(\omega) d\omega. \tag{26}$$

Proof. The right-hand-side is the expected cost for algorithm $\mathcal{Q}$ given oracle prior p. Since $\mathcal{Q}$ is optimal for prior r but not necessarily for p, this must be greater than $S_C(p,F)$, which is the optimum expected cost for prior p. □

In view of the results of this section, we adopt the following procedure $\mathcal{A}(p,F)$ for estimating a lower bound and calculating an upper bound of $S_C(p,F)$.

1. Choose discretization parameter d and the number of random samples k. Large d should tighten the bounds. Large k improves the statistical estimate of the lower bound.

2. Independently choose $o_j \in (0,1/d)$, $j \in \{1,\ldots,k\}$ uniformly at random.

3. Do the following for each $j \in \{1,\ldots,k\}$:

   (a) Compute the optimum cost $C_j = S_C(p_{o_j},F)$, and from the SDP solution, reconstruct an optimal algorithm $\mathcal{Q}_j$ achieving this cost.

(b) From $\mathcal{Q}_j$ derive an algorithm for evaluating $q(a|\omega, \mathcal{Q}_j)$.

(c) Using this algorithm, evaluate

$$\overline{C_j} = \sum_a \int C(\omega - f_a) q(a|\omega, \mathcal{Q}) p(\omega) d\omega$$

by numerical integration.

4. Return $\frac{1}{k}\sum_j C_j$ as a statistical estimate of a lower bound (together with its estimated error) and $\min_j(\overline{C_j})$ as a numerical upper bound.

C. Procedure for Solving the Clock SDP

Combining the ideas of this section, we use the following procedure for approximately solving the clock SDP $S_C(p, \mathbb{R})$:

1. Choose a discretization F, $|F| = m$, of the frequency estimates. If the cost function is suitable, we can optimize the right-hand-side of Eq. 16 and let ϵ_q be the corresponding querier discretization error bound.

2. Apply procedure $\mathcal{A}(p, F)$ and let c_l be the statistically estimated lower bound with estimated standard error s_l, and c_u the numerical upper bound obtained.

3. Give the estimated cost of $S_C(p, \mathbb{R})$ in the form $((c_l - \epsilon_q) \pm s_l, c_u)$.

If the cost function is not suitable, we set $\epsilon_q = 0$ and describe the discretization F for which the bounds apply.

V. RESULTS

We begin by considering one-query clock protocols ($t_f = 1$). In this case the protocol consists of an initial query state to be prepared and a final measurement. Figures 2 and 3 illustrate the importance of taking into account prior knowledge when deriving optimal clock protocols. Here our technique is compared to that of Ref. [2], which derives protocols under the assumption that ω is uniformly distributed on $[-\pi, \pi]$. We consider Gaussian priors of various widths and see that solving the clock SDP can substantially reduce the expected cost. If we use the periodic cost function $4\sin^2(\frac{\omega-\omega_a}{2})$ considered in [2], in the limit of wide prior, we obtain identical protocols. This is illustrated by Table I, which lists initial 2-atom states that optimize this cost function; the final row corresponds to the state computed in Ref. [2].

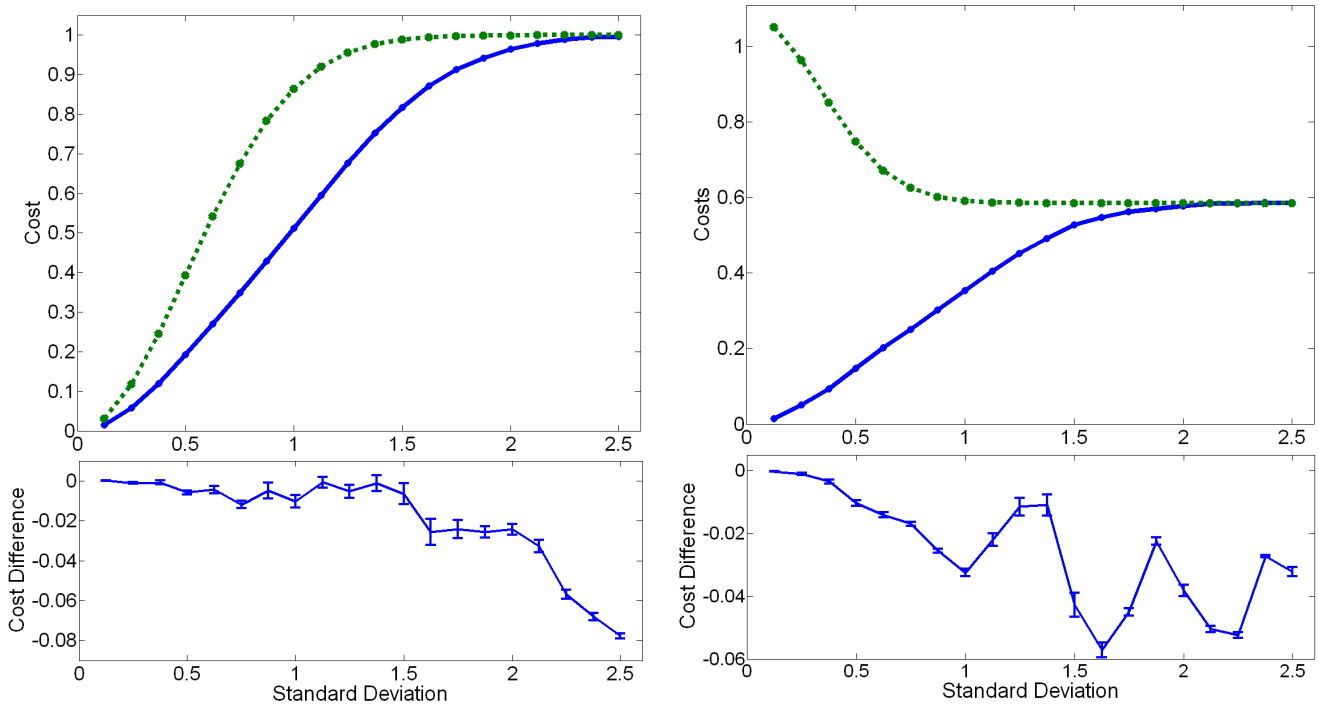
For computing the graphs of Figs. 2 and 3, and the initial states in Table I, we did not optimize the frequency estimates according to the iterative technique described at the end of Sect. IV A. In order to achieve sufficiently small discretization error, we discretized the prior frequency distribution with 15 points and used 20 frequency estimates chosen by minimizing Eq. (16). The bounds are based on 100 random discretizations to obtain sufficiently good statistics on the lower bound.

Standard Deviation	$\|0\rangle$	$\|1\rangle$	$\|2\rangle$
.25	.7071	0	.7071
.75	.5626	.6058	.5626
1.25	.5170	.6823	.5170
1.75	.5025	.7035	.5025
2.25	.5000	.7071	.5000

TABLE I: Initial states that minimize the cost function $C(\omega - f_a) = 4\sin^2(\frac{\omega-f_a}{2})$ for one-query protocols assuming different width Gaussian priors. SDP parameters are identical to those of Fig. 2

To verify the technique for obtaining error bounds, we compare our upper and lower bounds to the optimal solution obtained according to the formulas in Ref. [8]; this is illustrated in Fig. 4. We use the optimal set of classical frequency estimates derived in Ref. [8]; therefore, in the continuous limit, the clock SDP and that of Ref. [8] should yield identical costs. Consequently, the deviation depicted in Fig. 4 is due entirely to discretization error and the limitations of our bounds.

As discussed, our procedure can also optimize sequences of two or more clock queries ($t_f \geq 2$). If these queries are fully coherent, the algorithm is of the form given in Eq. (3), and the SDP implicitly optimizes the initial state, the $U(t_i)$ and the measurement. Alternatively, we can combine the two queries classically. In this case we update our knowledge of the clock's phase using Bayes' rule between the queries. That is, after the first query, we compute posterior distributions for each measurement outcome, as in Eq. (17). We then run the SDP again, once for each outcome, using the corresponding posterior distribution as the new prior. We compute a new cost by averaging each of the costs obtained, weighted by the probability of obtaining the corresponding measurement outcome, $\mathrm{tr}(\sigma_a)$. Here we assume that there is no noise between sequential queries. Any noise would affect the intermediate prior distributions. Fig. 5 compares these two methods for a sequence of two queries with two atoms. Here we used $|F| = 25$ possible frequency estimates and the iterative technique described in Sect. IV A for optimizing them. We found that the number of frequency estimates needed was substantially reduced after optimization. For one query ($t_f = 1$), and when such queries are combined classically, three estimates per query suffice after applying estimate optimization. When two queries are combined quantumly, five estimates suffice. It appears that, as expected, while we gain information by combining queries classically, fully coherent queries provide the greatest advantage. However, the technique developed for computing lower bounds cannot be applied to classical combinations of queries, so this advantage remains to be proven.



(a) One atom (b) Two atoms

FIG. 2: Comparison of the protocol derived in Ref. [2] and those derived by our method for one and two atoms and the cost function $C(\omega - f_a) = 4\sin^2(\frac{\omega - f_a}{2})$. We use a 15-point oracle discretization, 20 frequency estimates, and simulate Gaussian priors of various widths. The figures on the top plot the cost computed in Ref. [2] (dashed line) and our numerical upper bound (solid line), which as discussed, is equivalent to the minimum continuous cost obtained by one of our extracted algorithms. The figures on the bottom plot the difference between our lower bound and upper bound, $c_l - c_u$, illustrating both the strength of our bounds and how much lower the continuous cost could potentially be. The lower bound was computed by averaging 100 discretizations; error bars show the estimated standard error of the average thus obtained. Our querier discretization bounds cannot be applied to the periodic cost function used here, so the lower bounds are for the discretizations chosen. The lines connecting the data points in our figures are to guide the eyes.

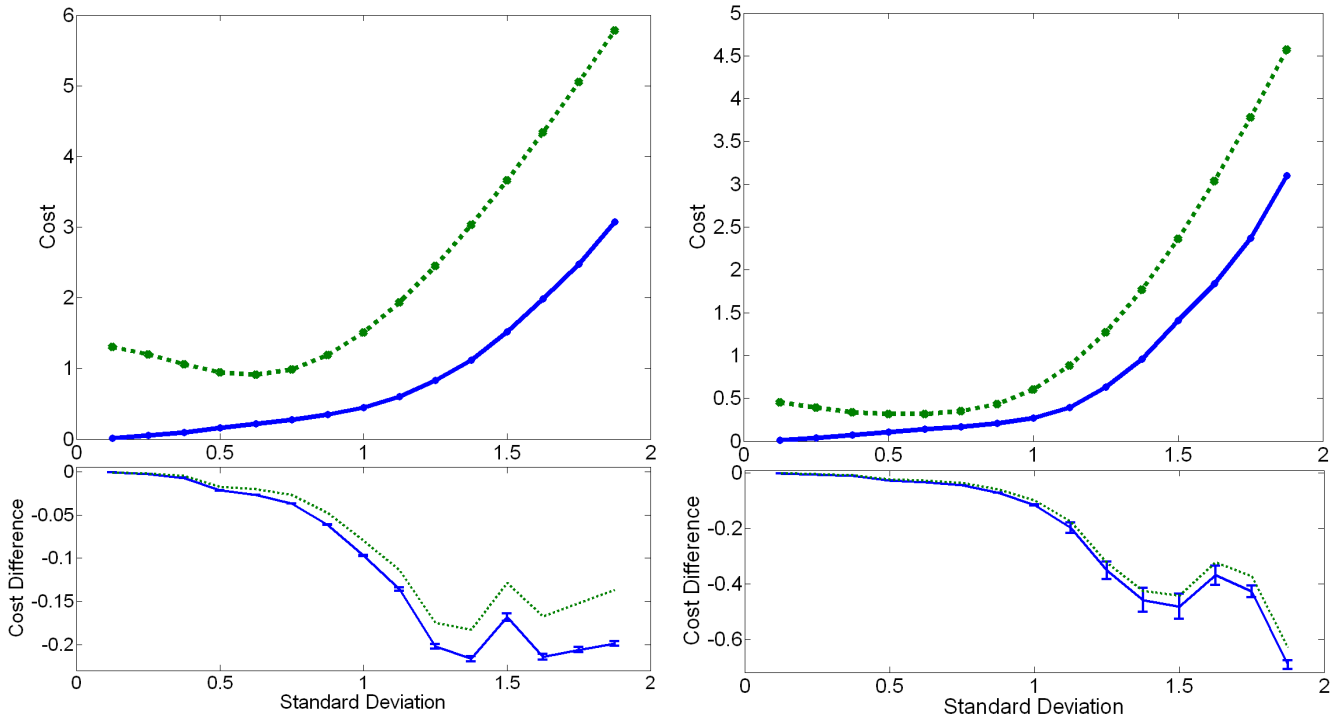
Table II examines the effect of adding more atoms to the clock. Additional atoms and additional queries always provide an advantage, as can be seen by reading down the table. The far right column gives the computational time required to run the SDP for a single discretization on a quad core 2.8 GHz machine with 12 GiB of RAM. Note that we are able to simulate more atoms than Table II may imply. For example, a simulation of 10 atoms using a uniform 15-point oracle discretization and $t_f = 1$ was computed in 6 minutes, 3 seconds, and yields a cost of 0.0785. Computing bounds for a system of this size, however, would require a great deal of computational time.

Notice that on the far right of Fig. 3(b), and in the last two rows of Table II, the gap between the lower bound and the upper bound becomes quite large. We need to increase the number of points in our discretizations if we wish to compute better approximations to the optimal solution for the continuous problem. This is as expected, since having more queries or atoms enables finer resolution of oracle frequencies. Fig. 6 illustrates the effect on our bounds of increasing the number of points in the oracle discretization. Here we are reanalyzing the last point in Fig. 3(b).

In summary, the method presented here is a general way of deriving better quantum clock protocols. Discretization is necessary, but the error introduced can

Number of Atoms	Number of Queries	c_l	c_u	s_l	ϵ_q	time (min:sec)
1	1	.6010	.6321	.0127	.0152	2:24
2	1	.4083	.4379	.0109	.0164	2:33
3	1	.2885	.3263	.0105	.0177	3:07
4	1	.1974	.2563	.0045	.0192	2:46
1	2	.4144	.4379	.0132	.0164	2:14
2	2	.1957	.2565	.0047	.0192	3:35
3	2	.1071	.2119	.0020	.0229	4:07
4	2	.0902	.2657	.0022	.0229	4:58

TABLE II: Costs for various combinations of atoms and queries for a Gaussian prior with a standard deviation of 1. SDP parameters are identical to those in Fig. 5. The times are given for solving one instance of the SDP. 100 instances were solved for the estimates of c_l and c_u.



(a) Two atoms (b) Four atoms

FIG. 3: Comparison of the protocol derived in Ref. [2] and those derived by our method for two and four atoms and the cost function $C(\omega - f_a) = (\omega - f_a)^2$. SDP parameters are identical to those in Fig. 2. The plots are also as in Fig. 2, except that for this cost function, we can compute querier discretization errors ϵ_q. Thus, the lower plots show $c_l - c_u$ as dotted lines above the line for $c_l - c_u - \epsilon_q$, which gives the lower bounds for the continuous problem.

be controlled. While the complexity of the SDPs to be solved limits the application of the method to small quantum systems, there are very promising atomic clocks, such as the ion-based ones, that use only a small number of atoms.

The optimization strategy pursued here is greedy, taking into account only the most recent prior. Therefore, it does not necessarily minimize long-term variance. Further research is required to develop more realistic strategies, taking advantage of knowledge of the clock's history.

Finally, we note that the general form of the SDP and the discretization analysis given here can be used to extend the adversary method to continuous, noisy, and classical problems. It may also find applications to quantum parameter estimation problems other than phase or frequency estimation, where we wish to estimate the value of a parameter x of an arbitrary but known family of unitary operators $U(x)$ that can be applied to quantum states of our choice.

Acknowledgments

This paper is a contribution of the National Institute of Standards and Technology and not subject to U.S. copyright. We thank Till Rosenband and Didi Leibfried for pointing us to the clock problem.

[1] P. O. Schmidt, T. Rosenband, C. Langer, W. M. Itano, J. C. Bergquist, and D. J. Wineland, Science **309**, 749 (2005).
[2] V. Bužek, R. Derka, and S. Massar, Physical Review Letters **82**, 2207 (1999), arXiv:quant-ph/9808042.
[3] N. Ramsey, *Molecular Beams* (Oxford University Press, 1956).
[4] D. J. Wineland, J. J. Bollinger, W. M. Itano, and D. J. Heinzen, Phys. Rev. A **50**, 67 (1994).
[5] J. J. Bollinger, W. M. Itano, D. J. Wineland, and D. J. Heinzen, Phys. Rev. A **54**, R4649 (1996).
[6] V. Giovannetti, S. Lloyd, and L. Maccone, Nat. Phot. **5**, 222 (2011).
[7] A. S. Holevo, *Probablistic and Statistical Aspects of Quantum Theory* (North-Holland, Amsterdam, 1982).
[8] R. Demkowicz-Dobrzanski (2011), arXiv:1102.0786.
[9] D. J. Wineland, C. Monroe, W. M. Itano, D. Leibfried, B. E. King, and D. M. Meekhof, J. Res. NIST **103**, 259

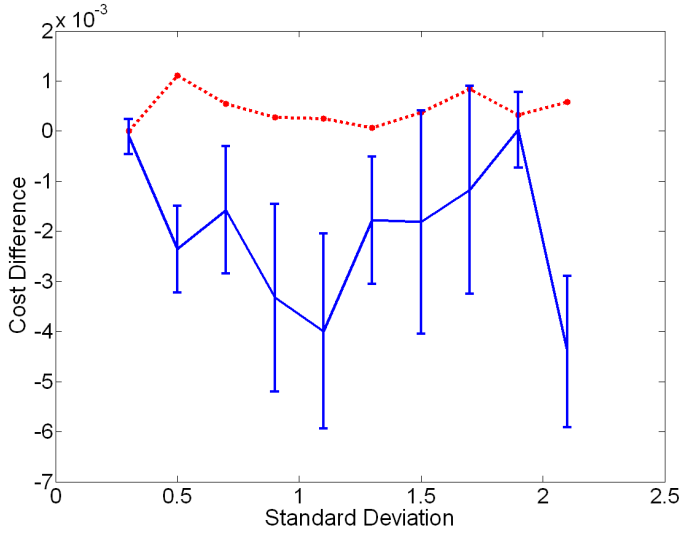


FIG. 4: Comparison of the optimal protocol derived in Ref. [8] and the one obtained by our SDP for two atoms. The dashed line depicts the difference between our upper bound and the cost computed in Ref. [8], and the solid line depicts the difference between our lower bound and the cost computed in Ref. [8]. Error bars again correspond to the estimated standard error in the lower bound. We used a 15-point prior discretization and set F to the optimal set of frequency estimates derived in [8], so there is no querier discretization error. The lower bound was computed by averaging 500 random oracle discretizations.

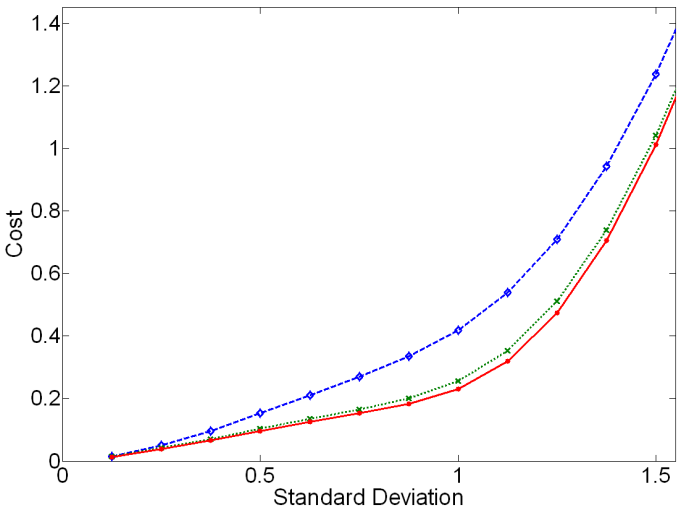


FIG. 5: Average cost after one query (top, dashed line), two queries combined classically (middle, dotted line), and two coherent quantum queries (bottom, solid line). We used a 15-point oracle discretization, $|F| = 25$, the cost function $C(\omega - f_a) = (\omega - f_a)^2$, and simulated two atoms.

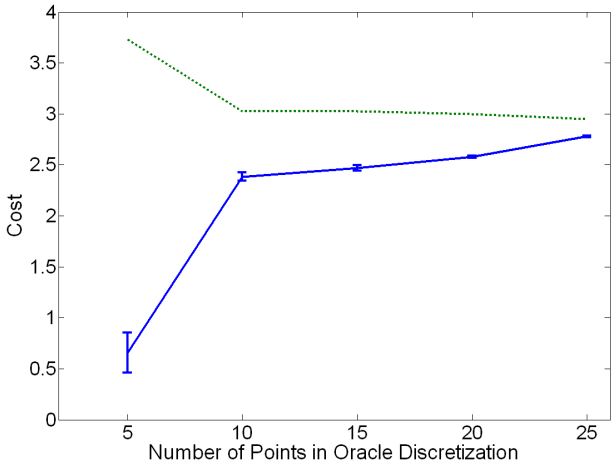


FIG. 6: Upper (c_u, dotted line) and lower (c_l, solid line) bounds on cost for oracle discretizations with varying numbers of points. We are simulating the last point of Fig. 3(b) - that is, 4 atoms, a standard deviation of 1.875, and the quadratic cost function. The lower bound is computed by averaging 32 random oracle discretizations. Note that the values in Fig. 3 are for a 15 point oracle discretization.

(1998).
[10] A. André, A. S. Sørensen, and M. D. Lukin, Phys. Rev. Lett. **92**, 230801 (2004).
[11] H. Barnum, M. E. Saks, and M. Szegedy, in *IEEE Conference on Computational Complexity'03* (2003), pp. 179–193.
[12] H. Barnum (2007), arXiv:quant-ph/0703141.
[13] L. Vandenberghe and S. Boyd, SIAM Review **38**, pp. 49 (1996), URL `http://www.jstor.org/stable/2132974`.
[14] L. Hughston, R. Jozsa, and W. Wootters, Phys Lett A **183**, 14 (1993).
[15] P. Hoyer, T. Lee, and R. Spalek, in *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing* (ACM, New York, NY, USA, 2007), STOC '07, pp. 526–535, URL `http://doi.acm.org/10.1145/1250790.1250867`.
[16] T. Lee, R. Mittal, B. W. Reichardt, and R. Spalek (2010), arXiv:1011.3020v1.
[17] R. Derka, V. Buz˘ek, and A. K. Ekert, Phys. Rev. Lett. **80**, 1571 (1998).
[18] L. Devroye, *Non-Uniform Random Variate Generation* (Springer-Verlag, 1986).
[19] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2001).

Appendix A: Remote State Preparation

Here we prove Thm. 1.

Proof. Assume that we can remotely prepare $\{\sigma_a^O\}$ from ρ^{OQ}, and let $\{P_a^Q\}_a$ be the required POVM. Then $\sum_a \sigma_a^O = \sum_a \mathrm{tr}_Q P_a^Q \rho^{OQ} = \mathrm{tr}_Q(\sum_a P_a)\rho^{OQ} = \rho^O$, by the definition of a POVM.

The converse is a generalization of the GHJW theorem [14] to mixed density operators. Suppose that $\sum_a \sigma_a^O = \rho^O$. To construct the required POVM, first write each σ_a^O as an explicit mixture of pure states

$$\sigma_a^O = \sum_m p_{a,m} |\psi_{am}\rangle\langle\psi_{am}|.$$

By assumption, $\rho^O = \sum_{a,m} p_{a,m} |\psi_{a,m}\rangle\langle\psi_{a,m}|$. By filling in mixture terms with $p_{a,m} = 0$ if necessary, we can assume that the range of the index m is independent of a. Since $\rho^{OQ} = |\Phi\rangle\langle\Phi|$ is pure, we can write the OQ state in Schmidt form

$$|\Phi\rangle = \sum_j \sqrt{q_j} |\phi_j\rangle|\varphi_j\rangle, \tag{A1}$$

where the $|\phi_j\rangle$ and $|\varphi_j\rangle$ are orthonormal in the Hilbert spaces of O and Q, respectively. Thus ρ^O can also be written as the mixture $\rho^O = \sum_j q_j |\phi_j\rangle\langle\phi_j|$. By unitary freedom (for example, see [19], pg. 103),

$$\sqrt{p_{am}}|\psi_{am}\rangle = \sum_j u_{am,j}\sqrt{q_j}|\phi_j\rangle, \tag{A2}$$

where the $u_{am,j}$ are the entries of a unitary matrix. We can now define

$$P_a^Q = \sum_m \sum_{j,j'} u^*_{am,j} u_{am,j'} |\varphi_j\rangle\langle\varphi_{j'}|. \tag{A3}$$

That $\sum_a P_a = \mathbb{1}$ follows from the unitarity condition for $u_{am,j}$. To verify the partial trace condition, compute

$$\begin{aligned}
\mathrm{tr}_Q P_a^Q \rho^{OQ} &= \sum_m \sum_{j,j'} u^*_{am,j} u_{am,j'} \\
&\quad \mathrm{tr}_Q \left(|\varphi_j\rangle\langle\varphi_{j'}| \sum_{l,l'} \sqrt{q_l q_{l'}} |\phi_l\rangle\langle\phi_{l'}||\varphi_l\rangle\langle\varphi_{l'}| \right) \\
&= \sum_m \sum_{j,j'} u^*_{am,j} u_{am,j'} \sqrt{q_j q_{j'}} |\phi_{j'}\rangle\langle\phi_j| \\
&= \sum_m p_{am} |\psi_{am}\rangle\langle\psi_{am}| \\
&= \sigma_a^O,
\end{aligned} \tag{A4}$$

as desired. □

Appendix B: Querier Discretization Bound

Here is the proof of Thm. 3.

Proof. Define $C(p, \mathcal{Q}) = \sum_a \int C(\omega - g_a) p(g_a, \omega|\mathcal{Q}) d\omega$, where the g_a are $\mathcal{Q}$'s frequency estimates. Then $C(p, \mathcal{Q})$ is the expected cost of $\mathcal{Q}$ given prior p. Let $g(a)$ be defined by

$$g(a) = \mathrm{argmin}_g \left(\int C(\omega - g) p(a|\omega, \mathcal{Q}) p(\omega) d\omega \right). \tag{B1}$$

Then $g(a)$ is the optimum frequency estimate $\mathcal{Q}$ could make given measurement outcome a. Let $\mathcal{Q}_g$ be $\mathcal{Q}$ modified to make the frequency estimates $g(a)$.

Let B be the expression on the right-hand-side of Eq. (16). We show that $C(p, \mathcal{Q}_g) \geq S_C(p, F) - B$ for any algorithm $\mathcal{Q}$. Since $S_C(p, \mathbb{R}) = \inf_{\mathcal{Q}} C(p, \mathcal{Q}) = \inf_{\mathcal{Q}} C(p, \mathcal{Q}_g)$, the result follows. We prove the bound in two steps. In the first step we force the frequency estimates to lie in $[f_1, f_N]$ and in the second we change them to lie in F.

For the first step, let $\tilde{g}(a)$ be the value in $[f_1, f_N]$ nearest to $g(a)$. If $\tilde{g}(a) = g(a)$, then $C(\omega - g(a)) \geq C(\omega - \tilde{g}(a)) - M(\omega)$ since $M(\omega) \geq 0$. If $\tilde{g}(a) = f_1$, then one of the following holds: 1. $\omega \geq f_1$, in which case $\tilde{g}(a)$ is nearer ω and on the same side, so that $C(\omega - g(a)) \geq C(\omega - \tilde{g}(a)) \geq C(\omega - \tilde{g}(a)) - M(\omega)$. 2. $\omega < f_1$, in which case $C(\omega - g(a)) \geq 0 = C(\omega - f_1) - C(\omega - f_1) = C(\omega - \tilde{g}(a)) - M(\omega)$. A similar argument works for $\tilde{g}(a) = f_N$. Substituting the inequalities in the integral for $C(p, \mathcal{Q})$ we get

$$\begin{aligned}
C(p, \mathcal{Q}_g) &\geq \sum_a \int C(\omega - \tilde{g}(a)) p(a|\omega, \mathcal{Q}) p(\omega) d\omega \\
&\quad - \sum_a \int M(\omega) p(a|\omega, \mathcal{Q}) p(\omega) d\omega \\
&= C(p, \mathcal{Q}_{\tilde{g}}) - \int M(\omega) d\omega.
\end{aligned} \tag{B2}$$

For the second step, we modify $Q_{\tilde{g}}$ to $Q_{\tilde{f}}$, where $\tilde{f}(a)$ is one of the elements of F on either side of $\tilde{g}(a)$. That is, because $f_1 \leq \tilde{g}(a) \leq f_N$, there exists a unique j such that $f_j \leq \tilde{g}(a) \leq f_{j+1}$, and we set $\tilde{f}(a)$ to either f_j or f_{j+1}. Define $\lambda \in [0, 1]$ by $\tilde{g}(a) = \lambda f_j + (1 - \lambda) f_{j+1}$. It is convenient to let $Q_{\tilde{f}}$ be a "mixed" (randomized) algorithm, where $\tilde{f}(a) = f_j$ with probability λ and f_{j+1} with probability $1 - \lambda$. Note that a mixed algorithm of this sort cannot be better than the optimal one, that is $C(p, Q_{\tilde{f}}) \geq S_C(p, F)$. To bound the cost, we consider a given a and ω and estimate the quantity

$$\begin{aligned}
c(\omega, a) &= \lambda C(w - f_j) + (1 - \lambda) C(w - f_{j+1}) \\
&\quad - C(\omega - \tilde{g}(a)) \\
&= \lambda \left(C(w - f_j) - C(\omega - \tilde{g}(a)) \right) \\
&\quad + (1 - \lambda) \left(C(w - f_{j+1}) - C(\omega - \tilde{g}(a)) \right).
\end{aligned} \tag{B3}$$

Define $\omega_0 = \omega - \tilde{g}(a)$, $\omega_l = \omega - f_{j+1}$ and $\omega_u = \omega - f_j$. We can estimate

$$\begin{aligned}
C(\omega) - C(\omega_0) &= (\omega - \omega_0) C'(\omega_0) \\
&\quad + \int_0^{\omega - \omega_0} \int_0^x C''(\omega_0 + y) dy dx \\
&\leq (\omega - \omega_0) C'(\omega_0) \\
&\quad + \frac{1}{2} (\omega - \omega_0)^2 \max_y C''(y) \\
&\leq (\omega - \omega_0) C'(\omega_0) + \frac{b}{2} (\omega - \omega_0)^2.
\end{aligned} \tag{B4}$$

Substituting this bound for each summand of Eq. (B3)

gives

$$\begin{aligned} c(\omega,a) &\le \lambda((\omega_u-\omega_0)C'(\omega_0)+\frac{b}{2}(\omega_u-\omega_0)^2) \\ &\quad +(1-\lambda)((\omega_l-\omega_0)C'(\omega_0)+\frac{b}{2}(\omega_l-\omega_0)^2) \\ &= \frac{b}{2}\left(\lambda(\omega_u-\omega_0)^2+(1-\lambda)(\omega_l-\omega_0)^2)\right) \\ &= \frac{b}{2}\left(\lambda(\omega_u-\omega_0)(\omega_u-\omega_l)\right) \\ &\le \frac{b}{2}\frac{(f_{j+1}-f_j)^2}{4}, \end{aligned} \tag{B5}$$

where we first applied $\lambda\omega_u+(1-\lambda)\omega_l=\omega_0$. The next identity requires applying $(1-\lambda)(\omega_l-\omega_0)=-\lambda(\omega_u-\omega_0)$ to the second summand, and the final inequality is obtained by noting that $\lambda(\omega_u-\omega_0)/(f_{j+1}-f_j)$ is maximized at $\lambda=1/2$. We can apply the above inequalities to bound $C(p,\mathcal{Q}_{\tilde{g}})$ as follows:

$$\begin{aligned} C(p,\mathcal{Q}_{\tilde{g}}) &= \sum_a \int C(\omega-\tilde{g}(a))p(a|\omega,\mathcal{Q})p(\omega)d\omega \\ &\ge \sum_a \int \big(\lambda C(\omega-f_j)+(1-\lambda)C(\omega-f_{j+1}) \\ &\quad -\max_j \frac{b}{8}(f_{j+1}-f_j)^2\big)p(a|\omega,\mathcal{Q})p(\omega)d\omega \\ &= C(p,\mathcal{Q}_{\tilde{f}})-\max_j \frac{b}{8}(f_{j+1}-f_j)^2 \\ &\ge S_C(p,F)-\max_j \frac{b}{8}(f_{j+1}-f_j)^2. \end{aligned} \tag{B6}$$

To finish the proof, we combine Eqs. (B2) and (B6). □