
Supplementary information

Optimal metrology with programmable quantum sensors

In the format provided by the
authors and unedited

Supplementary Material: Optimal metrology with programmable quantum sensors

Christian D. Marciniak^{1,*}, Thomas Feldker^{1,*}, Ivan Pogorelov¹, Raphael Kaubrügger^{2,3}, Denis V. Vasilyev^{2,3}, Rick van Bijnen^{2,3}, Philipp Schindler¹, Peter Zoller^{2,3}, Rainer Blatt^{1,2}, and Thomas Monz^{1,4†}

¹ *Institut für Experimentalphysik, 6020 Innsbruck, Austria*

² *Institute for Quantum Optics and Quantum Information, 6020 Innsbruck, Austria*

³ *Center for Quantum Physics, 6020 Innsbruck, Austria and*

⁴ *Alpine Quantum Technologies (AQT), 6020 Innsbruck, Austria*

SUPPLEMENTARY METHODS

S1. State preparation and readout

In the AQTION platform $^{40}\text{Ca}^+$ ions are hosted in a Paul trap forming a linear ion crystal under appropriate cooling. Optical qubits are encoded in the ground state $|4S_{1/2}, m_J = -1/2\rangle$ and excited state $|3D_{1/2}, m_J = -1/2\rangle$, which are connected via an electric quadrupole transition near 729 nm.

Further lasers at 397 nm, 854 nm, and 866 nm are required for state preparation and cooling. All ions are prepared in the qubit ground state via optical pumping before the start of each Ramsey sequence. The chain is cooled close to the motional ground state using resolved sideband cooling. Readout of the qubit system is performed optically via state-selective fluorescence near 397 nm, rendering the ground state bright and excited state dark. We perform site-selective readout of each ion using spatially-resolved detection on a camera through a high-numerical-aperture objective solely to increase detection fidelity; the scheme itself requires only collective measurements. All qubit operations as part of the Ramsey sequence are performed via a single laser beam oriented along the ion chain. The beam is aligned to interact identically with all ions.

S2. Implementation and calibration of unitaries

Qubit rotations $\mathcal{R}_{x,y}$ are implemented via resonant excitation on the qubit transition. The relative phase of pulses used to implement driven excitations sets the basis in which the rotation takes place, where a phase shift of $\pi/2$ changes between x and y . \mathcal{R}_z rotations may be implemented using AC Stark shifts induced by off-resonant excitation, but are not required here.

The infinite-range one axis twistings are implemented using the Mølmer-Sørensen interaction, whose unitary propagator takes the form $\exp(-i\chi\hat{J}_x^2)$ once appropriate conditions are met [1]. The interaction is generated by a bichromatic light field detuned from the ion chain's first

axial center-of-mass vibrational sidebands by Δ_{MS} . An optical phase shift is again used to change between the x and y basis. Negative rotation angles may be implemented by inverting both tones' detunings with respect to the chosen mode's sidebands, but we restrict ourselves to positive twisting angles for ease of implementation.

The twisting angles implemented by the Mølmer-Sørensen interactions are given by the geometric phase χ enclosed during the operation. It is challenging to optimize this phase to an arbitrary value given that it has no easily accessible observable associated with it. Additionally, we require that the Mølmer-Sørensen operation leave no residual spin-motional entanglement which puts further conditions on experimental parameters that achieve any given value of χ . Consequently, our strategy has two steps. First, we optimize a fully entangling Mølmer-Sørensen gate which implements a geometric phase of $\chi = \pi/2$ and creates a GHZ state. The state fidelity of the GHZ state allows for careful calibration of χ at this value. After this, we scale all parameters relative to the obtained values as required to maintain loop closure to the desired geometric phases.

In particular, we close eight loops in phase space after the $\tau_{\text{MS}} = 1600\mu\text{s}$ gate duration. A single loop in $200\mu\text{s}$ at the same detuning thus implements a twisting angle of $\pi/16$. All required twisting angles are then obtained from this calibration by scaling the gate duration and detuning together fixing $\tau_{\text{MS}} = 2\pi/\Delta_{\text{MS}}$, noting that then $\chi \propto \tau_{\text{ms}}^2$. The laser-induced AC Stark shifts during the Mølmer-Sørensen interaction are mitigated using power imbalancing [2] rather than centerline detuning. This dramatically reduces the complexity of the sequences since these phase shifts depend on the twisting angles, which are varied across the sequences, thus requiring individual measurement and compensation.

The proposal's original unitaries contain twisting operations partly along the z axis. Unfortunately, \mathcal{T}_z cannot be natively implemented using the Mølmer-Sørensen interaction. However, we may use a simple basis change on the entire sequence to transform into solely using \mathcal{T}_x and \mathcal{T}_y . The Ramsey sequence unitary then takes the form

$$\mathcal{U}'_{\text{Ramsey}} = \mathcal{U}_{\text{De}}^{x,y} \mathcal{R}_y(\phi) \mathcal{U}_{\text{En}}^{x,y} \mathcal{R}_y(\pi/2), \quad (1)$$

where the superscripts indicate the two bases used for twisting and rotation. In atomic clock settings the phase ϕ accumulated due to a detuning from the atomic reference frequency leads to a rotation $\mathcal{R}_z(\phi)$,

* These authors contributed equally to this work

† thomas.monz@uibk.ac.at

which may be implemented using the identity $\mathcal{R}_y(\phi) = \mathcal{R}_x(-\pi/2)\mathcal{R}_z(\phi)\mathcal{R}_x(\pi/2)$.

S3. Restrictions on sequence parameters

It is difficult to carefully calibrate twisting angles that differ strongly in magnitude. Consequently, we restrict the optimizer search to twisting angles $\varphi_{\min} \leq \varphi \leq \pi/8$. For small particle numbers, this comes at little cost to the predicted BMSE. φ_{\min} in turn is a lower bound on the twisting angles that is set by the minimal duration a Mølmer-Sørensen pulse can have in the experiment. This limitation is not set by experimental hardware, but instead by the pulse shaping necessary to minimize off-resonant carrier excitations [2]. φ_{\min} depends on the axial trap frequency, and we set it conservatively to $\varphi_{\min} = \pi/160$. The absolute limit is near $\varphi_{\min} = \pi/576$ at which point the pulse will be entirely made up of shaping slopes. As an example, this $\pi/160$ limit means that 2 of the 6 twistings in the (1, 2) sequence are dropped for up to 26 ions, with changes in the theoretical BMSE well below the experimental uncertainty. For the (1, 0) with up to 26 ions 1 of the 2 twistings are dropped, making this identical to a spin-squeezed state.

Rotations are not in principle limited in this fashion, but due to technical constraints in the hardware must also have a minimum duration. We mitigate this by calibrating multiple rotation objects with different Rabi rates for large and small rotations. Rotation angles below a certain threshold will nonetheless be skipped, adding experimental error for small rotation angles.

Adding additional entangling and decoding layers to a sequence increases the prospective gain relative to the CSS baseline Ramsey sequence. However, these gains diminish for ever larger layer depths at a fixed particle number as the sequence converges towards the optimal quantum interferometer, while both sequence overhead and gate error accumulation do not saturate in this manner. Consequently, there is a break-even point after which adding additional layers will decrease performance. For the particle numbers considered here this is at the (1, 2) sequence, as seen in Fig. 2 c and its inset, where the gain in adding an additional layer is already below the deviation of the data from the theory.

S4. Experimental calculation of cost

The measurement record from the experiment is given by binary bit strings representing the z projections of the ion chain state during the measurement. An example would be a string 0010 = DDSD representing a four ion register where ions 1, 2, and 4 are in the excited D state and thus do not fluoresce, and ion 3 is in the bright ground state S.

For the measurement protocol we only require knowledge of collective measurement outcomes. Consequently,

we calculate the probability $p(k|\phi_{\text{exp}})$ of finding k excitations among all bit strings for a given phase ϕ_{exp} from the histogram of outcomes. This is identical to finding the probability $p(m|\phi_{\text{exp}})$ to measure the excitation imbalance $m = (2k - N)/2$. Here ϕ_{exp} is the phase ϕ we want to implement up to a phase shift $\tilde{\phi}$ that arises in the implementation due to experimental imperfections such as timing jitter or residual AC Stark shifts.

To calculate the MSE and BMSE we require the phase estimator $\phi_{\text{est}} = am$, and knowledge of $\phi = \phi_{\text{exp}} - \tilde{\phi}$ to obtain the conditional probabilities at the true phase ϕ . To do this, we use the raw measurement data to calculate the conditional excitation imbalance probabilities and the MSE with both slope a and $\tilde{\phi}$ as free parameters, at a fixed grid of phase values. The parameter-dependent MSE errors are integrated using the Simpson rule and the resulting BMSE is minimized in terms of the free parameters. Typically we find $\tilde{\phi} \lesssim 0.01$, and the slope a is close to the theoretically predicted value. As an example, SI Tab. I shows $1/a$ for the $\delta\phi = 1$ data point in Fig. 2 c.

$(n_{\text{En}}, n_{\text{De}})$	Exp. $1/a$	Theory $1/a$
(0, 0)	4.731(12)	4.745
(1, 0)	4.363(9)	4.378
(0, 2)	3.042(4)	3.001
(1, 2)	3.023(4)	2.963

SI Table I. Comparison of linear phase estimator slopes a determined experimentally via numerical optimization of raw data and from theory for $\delta\phi = 1$ data in Fig. 2 c.

The cost function evaluation during the optimization is accelerated by measuring at ϕ values chosen to coincide with the nodes of the Hermite-Gaussian polynomials up to twelfth order. We further leverage the symmetry of the MSE around $\phi = 0$ to perform Hermite-Gauss quadrature (HGQ) integration for $\phi \geq 0$ only. The HGQ proves particularly useful if only a small number of ϕ points is used for the numerical integration, where equidistantly spaced integration schemes can lead to a biased estimate of the cost function.

S5. Optimizer and meta-model design and tuning

On-device optimization is essentially a classical optimization problem, with data generated on a quantum machine. This is a well-studied problem for which several solutions have been demonstrated [3]. For the optimization we employ a modified version of the Dividing Rectangles algorithm (DIRECT) [4–6] which is a gradient-free algorithm that guarantees global convergence for sufficiently many iterations. DIRECT divides the search space into hyperrectangles, which we will refer to as ‘cells’. Each cell is represented by a single cost function evaluation taken in its interior. Promising cells are sampled at a finer scale by subdividing further, prioritizing cells with low cost function values, as well as cells with large size.

During the optimization the algorithm maintains an internal representation, denoted meta-model, of the cost function landscape in the form of a Gaussian process [7], modelling the data as jointly-distributed Gaussian variables with a trigonometric covariance kernel. For this meta-model we may expand the cost for each gate in a series of trigonometric functions of the interaction parameter since all interactions in the sequence are global. The cost function is then a product of $3(n_{\text{En}} + n_{\text{De}})$ trigonometric series which generally have non-vanishing expansion coefficients and argument scalings for both even and odd terms. The number of terms in each series as well as the scalings depend on the differences of eigenvalues of the generators of the gate, that is the differences of eigenvalues of $J_{x,y,z}$, and $J_{x,y,z}^2$. For rotations, these differences grow linearly in number, while for twistings they grow quadratically. The number of terms and scalings are known a priori for a given number of particles, so the cost has $6(n_{\text{En}} + n_{\text{De}})$ free parameters whose optimal match to the experimental data is continuously adapted with incoming measurement data.

The algorithm is provided with a finite measurement budget that it can spend during the optimization process. We invest only a relatively low number of measurements at each point to use the budget economically. If the variance of the cost function measurements is above a certain threshold, the algorithm can request refinement steps at points already sampled in order to increase the probability of correctly deciding which cells to subdivide. The algorithm selects when to perform the refinement steps, and how many measurements to spend in this stage. These decisions are based on methods from decision theory and optimal computing budget allocation [5, 8].

To start the optimization we initialize search boxes ranging from 50% to 150% of the theoretically optimal parameter values. This situation is representative if there is some degree of trust in the operations implemented on the quantum machine. We further expect that for classically not simulatable resources, like finite-range interactions, the optimal parameters will smoothly change with the particle number, allowing extrapolation of similar constraints for the search space. We initialize the search by adding a random displacement to the search box relative to the theoretical optimum to prevent the optimization algorithm from trivially finding the theoretically optimal parameters by initially sampling the parameter space in the center of the search box. An example set of optimal parameters after a complete search together with theory predictions is given in SI Tab. II.

S6. Coarse and fine optimizer scans

Any optimizer intrinsically requires many evaluations of the same parameter landscape over time. In real experiments the quantum operations are subject to drift. Consequently, we perform ‘coarse’ evaluations to speed up the procedure with fewer repetitions (typically 100)

per point and fewer sampling points in ϕ per evaluation (typically 5 - 11 nodes). Once a good parameter candidate has been identified, the optimizer requests a ‘fine’ scan of these parameters with finer sampling (typically 20 or 21 nodes) on both sides of $\phi = 0$, and more repetitions (typically 250).

Longer ion chains are subject to more frequent collisions with background gas in the trap than smaller chains, leading to frequent melting of the crystal. A molten crystal has to be recrystallized before the next state initialization can begin - a process typically referred to as refreezing. Refreezing the crystal takes substantially longer than state preparation and sequence execution. Refreeze events requiring remeasuring of a given point are more frequent in long chains, and fine scans with many repetitions per measurement point on long chains can thus become time consuming. Additionally, a detected refreeze event will discard all measured repetitions and start the measurement process anew. We alleviate this problem by batching measurement requests from the optimizer into multiples of 50-repetition measurements, so the maximum number of discarded measurements is 50.

S7. Noise frequency reconstruction

We try to reconstruct the known value of injected noise (detuning) by performing phase estimation with two different Ramsey sequences, while also accounting for the drift in the laser’s reference cavity.

A frequency detuning $\Delta\omega$ is drawn at random from a normal distribution with width $\Delta\omega_N = 2\pi \cdot 40$ Hz and added to the laser frequency using an acousto-optic modulator. The distribution is truncated at $2\Delta\omega_N$ to reduce the influence of rare large detunings on the measurement outcome from the limited statistics of 200 different samples per Ramsey time.

We perform three Ramsey sequences successively where we impart the phase ϕ via $\mathcal{R}_x(-\pi/2)\mathcal{R}_z(\phi = \Delta\omega T_R)\mathcal{R}_x(\pi/2)$, with a detuning $\Delta\omega$: First, a standard Ramsey sequence without detuning ($\Delta\omega = 0$), with $T_R = 15$ ms and averaging over 50 repetitions. This is used to correct for the slow drift of our qubit laser’s cavity. Second, 50 measurements with detuning $\Delta\omega$ using a standard CSS Ramsey sequence where the Ramsey time is varied and the laser frequency is estimated from evaluation of single measurements. Lastly, a sequence optimized for a prior width $\delta\phi = 0.6893$ with depths (1, 2) is used analogously to the second CSS sequence. This value is chosen as the equivalent Ramsey time at which both the CSS and the optimized sequence approximately have their respective minima in the BMSE. We repeated this measurement 200 times for each Ramsey time. The order of CSS and (1, 2) sequence was reversed after 100 repetitions to exclude a systematic bias, but no difference in outcome was found.

$(n_{\text{En}}, n_{\text{De}})$	Approach	Entangling			Decoding 1			Decoding 2		
		\mathcal{T}_y	\mathcal{T}_x	\mathcal{R}_x	\mathcal{R}_x	\mathcal{T}_x	\mathcal{T}_y	\mathcal{R}_x	\mathcal{T}_x	\mathcal{T}_y
(1, 0)	Theory	0.0551	0	-1.0699	0	0	0	0	0	0
	Optimizer	0.1080	0	-1.3090	0	0	0	0	0	0
(1, 2)	Theory	0.0626	0	-1.2725	0.6938	0.0196	0.0631	-0.6518	0.0196	0
	Optimizer	0.0896	0	-1.3818	0.5841	0.0213	0.0813	-0.6357	0.0196	0

SI Table II. Twisting and rotational angles (in rad) of variational unitaries for two variational sequence types with 26 particles. Optimizer parameters are for the data displayed in Fig. 3 in the main text. Twisting angles of 0.0196 correspond to the lower boundary of $\pi/160$ as described in section S3.

To explain the difference between the ideal measurement and our experiment we identify three main sources of error: First, frequency flicker noise of our laser. To determine the frequency noise we perform the measurement described above, without detuning $\Delta\omega$. For long Ramsey times the standard deviation of the measured frequencies converges to $b_\alpha \approx 2\pi \cdot 6\text{ Hz}$. Second, the non-ideal BMSE measured in our system. Third, the BMSE increases with detuning $\Delta\omega$, so that the distribution width $\Delta\omega_N$ is a compromise between reducing the influence of laser frequency noise on the measurement ($\Delta\omega \gg b_\alpha$) and avoiding gate errors to achieve the best BMSE ($\Delta\omega \rightarrow 0$). Consequently the achievable gain is directly correlated to the laser frequency noise.

We note that the deterioration of the BMSE with detuning is an artifact of our measurement. In an optical clock the detuning $\Delta\omega$ is small and T_R is consequently much larger to reach the same $\delta\phi$. The large values of $\Delta\omega$ here cause deterioration of the gate-based implementation of the twistings, which is a consequence of the technical limitations of the implementation, not the scheme itself.

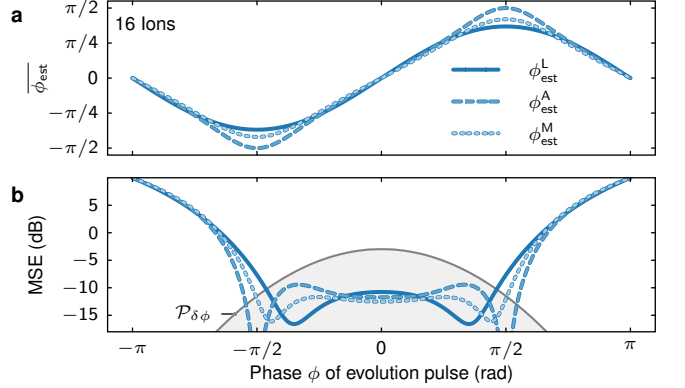
S8. Sequence overhead

The variational sequences contain more operations than the archetypical Ramsey sequence, which produces time overhead in state preparation and measurement relative to the coherent spin state interferometer. These effects are typically negligible for high-precision trapped-ion or atomic systems, in particular for atomic clocks, where Ramsey times of order 1 s are reached. Even in our case the increase in sequence length accounts for less than 10 % of the state preparation time, below the gain in T_R . Specifically, for the sequences (1, 0) (SSS) our overhead is 165 μs , the (1, 2) takes an additional 250 μs , and (2, 1) (GHZ) takes 300 μs extra compared to the CSS.

SUPPLEMENTARY DISCUSSION

S9. Comparison of phase estimator functions

Estimator functions are used to produce an estimate of the desired parameter based on measurement out-



SI Figure 1. Theoretical performance of a conventional Ramsey sequence $(n_{\text{En}}, n_{\text{De}}) = (0, 0)$ for different estimator functions. **a** Expectation value ϕ_{est} of phase estimators as a function of evaluation pulse phase ϕ . Any deviation from a line with unit slope is a manifestation of bias. **b** MSE calculated for the different estimator functions. Overlaid in grey is the prior distribution $\mathcal{P}_{\delta\phi}(\phi)$ with $\delta\phi \approx 0.79$. $\phi_{\text{est}}^{\text{L}}$ and $\phi_{\text{est}}^{\text{M}}$ are optimized for this particular prior distribution, while $\phi_{\text{est}}^{\text{A}}$ is independent of it. $\Delta\phi/\delta\phi$ values for linear, arcsine, and MBMSE corresponding the MSE and $\mathcal{P}_{\delta\phi}$ in **b** are: -4.01 dB, -3.94 dB, -4.20 dB.

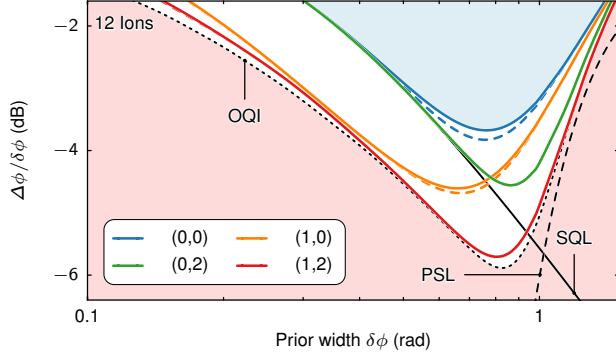
comes. Choosing an appropriate estimator function is paramount in metrology. This choice is informed not only by suitability in the ideal case, but also implementation considerations such as resilience against noise, or practicability of computation in systems with time constraints.

In this study we judge the performance of different types of variational sequences relative to one another based on our cost function which in turn depends on the choice of estimator function. In the main text we restrict ourselves to considering a linear estimator of the phase ϕ of the form

$$\phi_{\text{est}}^{\text{L}}(m) = a m. \quad (2)$$

However, other choices of estimator function are possible, and may at first glance seem superior. For sinusoidally varying measurement outcomes one may consider using the analytic inverting function, that is using an arcsine estimator

$$\phi_{\text{est}}^{\text{A}}(m) = \arcsin(m \frac{2}{N}), \quad (3)$$



SI Figure 2. Theoretical $\Delta\phi/\delta\phi$ as a function of prior width $\delta\phi$ for 12 ions in the four variational sequences. The blue shaded region corresponds roughly to classical Ramsey sequences. The red shaded region is inaccessible using this scheme, where the boundary corresponds to the optimal quantum interferometer (OQI). The OQI, the standard quantum limit (SQL) and the phase slip limit (PSL) are indicated by black lines. Solid lines are obtained for a linear phase estimator while the dashed lines are obtained for the estimator that minimizes the $\Delta\phi$ over all possible estimators at the respective prior width.

which ideally is an unbiased estimator for all ϕ for sufficiently large particle number N . It is clear that any loss of amplitude through e.g. decoherence effects will cause bias everywhere again, i.e. this estimator feature is not stable in the presence of noise.

Another choice is the minimum Bayesian mean squared error (MBMSE) estimator given by

$$\phi_{\text{est}}^{\text{M}}(m) = \int_{-\infty}^{\infty} d\phi \phi p(\phi|m), \quad (4)$$

where $p(\phi|m) = p(m|\phi)\mathcal{P}_{\delta\phi}(\phi)/p(m)$ may be obtained from measured conditional probabilities $p(m|\phi)$ using Bayes' theorem, and with $p(m) = \int_{-\infty}^{\infty} d\phi p(m|\phi)$. This estimator can be shown to minimize the cost function, i.e. the BMSE, over all possible estimators [9], but is computationally expensive.

Both the arcsine and MBMSE estimators extend the dynamic range of an ideal, conventional (0,0) Ramsey interferometer relative to the linear estimator, that is they remain approximately unbiased for a larger range of parameter values. At 16 ions the MBMSE extends only slightly further than the linear estimator, while the arcsine extends the dynamic range to almost the interval $(-\pi/2, \pi/2)$ (SI Fig. 1 a). This ordering of dynamic ranges is preserved as the particle number is increased. However, the MSE as a metric to judge estimator performance includes not only the bias but also the sensitivity of the interferometer. Consequently, when $\phi = \pm\pi/2$ and the interferometer is in an z eigenstate, projection noise vanishes and at zero bias the arcsine estimator MSE vanishes as well, while the linear and MBMSE estimators remain finite (SI Fig. 1 b). While in the immediate vicinity of $\phi = 0$ the arcsine estimator performs better than

the linear, its MSE remains at a similar level further out. The sensitivity of the linear estimator on the other hand is generally higher in the region where the prior distribution is appreciable. As a result, the arcsine estimator performs overall worse than both linear and MBMSE estimators in our operational cost function \mathcal{C} , which averages over the MSE weighted by the prior distribution $\mathcal{P}_{\delta\phi}$.

Returning to the normalized cost $\Delta\phi/\delta\phi$ as a metric we see that the MBMSE estimator retains a small advantage over the simple linear estimator at prior widths close to the sequences' optimal widths (SI Fig. 2). This gain decays quickly for prior widths away from the optima, and is completely lost for sequences that feature decoding layers, i.e. $n_{\text{De}} > 0$. We can thus conclude that concentrating on the simple and stable linear estimators does not qualitatively change the analysis, and is near-optimal for the more complicated sequences.

S10. Application to atomic clocks

A primary example where programmable quantum sensors may directly benefit a crucial application in the near-term is the improvement of atomic or ionic clocks. The resources and interactions available in for example optical lattice clocks have already been studied theoretically [10, 11], and their application demonstrated in the context of squeezing for enhanced clock operation [12]. Programmable quantum sensors offer additional benefits to atomic clocks in realistic metrology settings. The tailored entanglement extends the interferometer dynamic range, for example, which increases the optimal Ramsey time T_R at a fixed noise level. Longer Ramsey times can further increase the duty cycle of atomic clocks which reduces the optical Dick effect [13, 14]. In SI Tab. III we show the potential gain of the scheme for 12 and 26 ions as demonstrated here compared to 362 atoms as in Ref. [12] in an optical lattice.

N	Approach	(1,0)	(1,2)	Gap to OQC
12	Theory	1.49(0) dB	2.13(0) dB	0.04(0) dB
	Direct	1.38(1) dB	1.75(2) dB	0.50(5) dB
26	Theory	2.12(0) dB	2.70(0) dB	0.80(0) dB
	Direct	1.47(8) dB	2.02(8) dB	1.66(8) dB
	Optimizer	1.54(9) dB	1.77(8) dB	1.90(8) dB
362	Theory	4.53(0) dB	7.50(0) dB	1.47(0) dB

SI Table III. Gain over CSS in Allan deviation for 12 and 26 ions (theoretical and closest-measured value of $\delta\phi$) and 362 atoms (theoretical) for two different interferometer states at their respective optimal prior widths (Ramsey times). Gap of (1,2) sequence to optimal quantum clock (OQC) shown alongside. Deeper circuits with more layers approach the OQC more closely.

The potential metrological gain in the clock setup can be increased by 3 dB by using the slight additional resource overhead of the decoding layers, but notably using

the same operations as before. Similarly, Ref. [12] points out the first squeezing demonstrated on an optical transition with a Wineland squeezing parameter of 4.4(6) dB with ≈ 360 atoms. We directly measure a Wineland squeezing parameter [15] of 5.29(5) dB below the stan-

dard quantum limit using only 26 particles with the (1, 2) sequence. Second, the robustness of the scheme with respect to variations of the prior width (Fig. 4) enables reliable operation when shot-per-shot measurements exhibit fluctuations in the number of particles as commonly encountered in cold atom experiments.

-
- [1] Roos, C. F. Ion trap quantum gates with amplitude-modulated laser beams. *New J. Phys.* **10**, 013002 (2008).
 - [2] Kirchmair, G. *et al.* Deterministic entanglement of ions in thermal states of motion. *New J. Phys.* **11**, 023002 (2009).
 - [3] Cerezo, M. *et al.* Variational quantum algorithms. *Nat. Rev. Phys.* **3**, 625–644 (2021).
 - [4] Jones, D. R., Perttunen, C. D. & Stuckman, B. E. Lipschitzian optimization without the Lipschitz constant. *J. Optim. Theory Appl.* **79**, 157–181 (1993).
 - [5] Nicholas, P. A dividing rectangles algorithm for stochastic simulation optimization. In *Proc. Inform. Comput. Soc. Conf.*, vol. 14, 47–61 (2014).
 - [6] Kokail, C. *et al.* Self-verifying variational quantum simulation of lattice models. *Nature* **569**, 355–360 (2019).
 - [7] Rasmussen, C. E. *Gaussian Processes in Machine Learning* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004).
 - [8] Fu, M. C., Chen, C.-H. & Shi, L. Some topics for simulation optimization. In *Proceedings of the 40th Conference on Winter Simulation*, 27–38 (Winter Simulation Conference, 2008).
 - [9] Demkowicz-Dobrzański, R., Jarzyna, M. & Kolodyński, J. *Quantum Limits in Optical Interferometry*, vol. 60 of *Progress in Optics* (Elsevier, 2015).
 - [10] Kaubruegger, R. *et al.* Variational spin-squeezing algorithms on programmable quantum sensors. *Phys. Rev. Lett.* **123**, 260505 (2019).
 - [11] Kaubruegger, R., Vasilyev, D. V., Schulte, M., Hammerer, K. & Zoller, P. Quantum Variational Optimization of Ramsey Interferometry and Atomic Clocks. *Phys. Rev. X* **11**, 041045 (2021).
 - [12] Pedrozo-Peñafiel, E. *et al.* Entanglement on an optical atomic-clock transition. *Nature* **588**, 414–418 (2020).
 - [13] Takamoto, M., Takano, T. & Katori, H. Frequency comparison of optical lattice clocks beyond the Dick limit. *Nat. Photonics* **5**, 288–292 (2011).
 - [14] Schioppo, M. *et al.* Ultrastable optical clock with two cold-atom ensembles. *Nat. Photonics* **11**, 48–52 (2017).
 - [15] Wineland, D. J., Bollinger, J. J., Itano, W. M. & Heinzen, D. Squeezed atomic states and projection noise in spectroscopy. *Phys. Rev. A* **50**, 67 (1994).