Peer Review File

Manuscript Title: Realizing spin squeezing with Rydberg interactions in an optical clock

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referees' comments:

Referee #1 (Remarks to the Author):

The manuscript by Eckner et al. reports the first realization of spin squeezing using Rydberg interactions. The spin squeezing is performed in an array of optical-clock atoms 88-Sr, and is achieved by "dressing" the upper clock state with a Rydberg level using a detuned laser beam. This creates interactions between atoms in the upper clock state that are used for spin squeezing following an earlier theoretical proposal. The authors observe ~4 dB of metrological gain for the state, and when they use the spin squeezed state in a Ramsey sequence on the optical transition, they achieve performance ~3.5dB below the standard quantum limit. The interaction as demonstrated does not "scale", i.e. the amount of spin squeezing is essentially the same for 70 atoms as for 4 atoms.

The paper is well written and structured, and I believe that the results are valid. The experimentally realized system, including the sorting of individual atoms into a lattice, Rydberg spin squeezing, and operation of a clock with the squeezed state, are an impressive technical feat. The results are convincing and novel. I recommend that the manuscript, reporting the first operation of an optical clock beyond the standard quantum limit, be published in Nature.
I have just a few remarks/questions that I would like the authors to address:
i) Why are the data in Fig. 2a not described by an ellipse (cosine function), but rater by a "guide to the eye"? On the other hand, in Methods section G, the authors talk about "a numerical cosine fit like the one shown in Fig. 2a". Why is the state not an ellipse, given that the squeezing is rather small? Is it because of the relatively small atom number? Is the cosine a better fit for larger atom number?
ii) Fig. 2c. The anticorrelation at the optimal squeezing angle are about 2% per atom, for ~10 surrounding atoms. This seems to imply a variance reduction of 2x10x2%=40%. Is this consistent with the noise reduction of almost 6dB shown in Fig. 2a?
iii) Fig. 2d, the caption states "Small purple circles and lines show the theoretical prediction based on weak dressing". I thought that the circles in Fig. 2d are the data?
iv) Fig. 3, in the caption the authors state "the ... points correspond to the first 200 points of the raw data from which ...the Allan deviations are calculated". Why is a subset selected rather than using all data? Do the other data show the same performance?
v) It would be helpful to the reader to state the cycle time in the main manuscript, and to specify in methods how much time is spent on what part of the preparation procedure.
vi) Even after repeated reading I did not understand what zero-fill means in the Methods Section B, and which data were removed. Perhaps the authors could reformulate this section?
vii) Can the authors provide a figure in the methods or supplementary material showing the contrast reduction data?

Referee #2 (Remarks to the Author):

The authors report an experimental demonstration of Rydberg-mediated squeezing with a neutral atom clock providing 4 dB of metrological enhancement. The squeezing results from Ising-like interactions, which, thanks to their protocol, approximates the one-axis twisting squeezing interaction in a certain limit. The authors claim their clock has a fractional frequency stability of 10^{-15} at one-second averaging time and around 10^{-17} at half-hour measurement. Besides that, the authors explore the possibility of using the setup for differential phase interferometry via elliptical fitting. In that case, the system exhibits both precision enhancement and diminution due to entanglement.

The writing is excellent, albeit dense, and the analysis is very complete. I have only minor concerns that I'm listing now.

1. According to the paper, the squeezing parameter (metrological enhancement) is determined from Eq. (3) by measuring σ^2 and fitting the contrast C. One alternative way of doing this would be to calculate the enhancement directly by performing Ramsey measurements for several values of the phase, plot dz as a function of the phase, and then estimate the sensitivity via $\delta \theta^2 = \Delta d_z^2 / |dz/d\theta|^2$, i.e., calculating the slope at the working point by a fit. In that way, it is not necessary to determine C. Indeed, that is how the metrological gain is demonstrated in Ref. 33. That may be a more realistic way of estimating the enhancement, as it simulates the actual working of the device. I think both are only equal in perfect conditions. That also may explain why the enhancement in the Allan deviation plot is different from the one predicted by the squeezing parameter, as in the Allan deviation case, they perform an actual Ramsey interferometry sequence.
2. The authors state in the introduction that all-to-all cavity-mediated interactions are a route to scalable spin squeezing, yet they fail to provide access to high-fidelity rotations necessary to boost precision even more. The current approach may give great control over the system. However, the results show that it scales poorly with N. Instead, the metrological gain saturates to a certain metrological enhancement. The authors comment on this.
3. The theoretical model fails to describe the metrological gain for large values of N, as shown in Fig. 2 and the supplementary material. In the same line as the previous point, it is necessary to develop a good theoretical description before more sophisticated protocols can be developed to exploit the controllability of the device to increase sensitivity further.
4. I think it may be worth pointing out the scaling with N of the squeezing parameter predicted by the theoretical model employed.
5. In the context of the elliptical fitting, it is unclear whether squeezing should be beneficial or detrimental to the phase estimation, as, at times, squeezing and anti-squeezing will contribute to the measurement. On the other hand, Fig. 4b shows a contrast loss. That could explain Fig. 4c. Furthermore, there is literature suggesting that enhancement is possible under certain conditions (see Quantum-enhanced differential atom interferometers and clocks with spin-squeezing swapping. Quantum, 7, 965 (2023). That kind of analysis could explain the experimental results.
5. Physical Review Letters 126, 113401 (2021) deals with a similar family of ideas (not exactly the same system), demonstrating 4.9 dB of enhancement via Ising-like interactions. I think the authors should contrast their findings with theirs.

In conclusion, the data demonstrate metrologically useful Rydberg-mediated squeezing, and the reported frequency stability is a significant step towards a competitive neutral atom clock. The analysis is comprehensive, and it will attract a broad audience. I think the work merits publication, subject to the minor revisions and clarifications I have suggested.

Author Rebuttals to Initial Comments:

Response to report from the first referee (A)

> The manuscript by Eckner et al. reports the first realization of spin squeezing using Rydberg interactions. The spin squeezing is performed in an array of optical-clock atoms 88-Sr, and is achieved by "dressing" the upper clock state with a Rydberg level using a detuned laser beam. This creates interactions between atoms in the upper clock state that are used for spin squeezing following an earlier theoretical proposal. The authors observe 4 dB of metrological gain for the state, and when they use the spin squeezed state in a Ramsey sequence on the optical transition, they achieve performance 3.5dB below the standard quantum limit. The interaction as demonstrated does not "scale", i.e. the amount of spin squeezing is essentially the same for 70 atoms as for 4 atoms.
> The paper is well written and structured, and I believe that the results are valid. The experimentally realized system, including the sorting of individual atoms into a lattice, Rydberg spin squeezing, and operation of a clock with the squeezed state, are an impressive technical feat. The results are convincing and novel. I recommend that the manuscript, reporting the first operation of an optical clock beyond the standard quantum limit, be published in Nature. I have just a few remarks/questions that I would like the authors to address:
> i) Why are the data in Fig. 2a not described by an ellipse (cosine function), but rater by a "guide to the eye"? On the other hand, in Methods section G, the authors talk about "a numerical cosine fit like the one shown in Fig. 2a". Why is the state not an ellipse, given that the squeezing is rather small? Is it because of the relatively small atom number? Is the cosine a better fit for larger atom number?

(A1) It was not our intention to suggest that the state is not an ellipse. We referred to the cosine fit in Fig. 2a as a guide to the eye since we do not use this particular curve to extract any information about quoted variance reductions. We have adjusted the corresponding caption, and now state that the guide to the eye is a cosine fit to the data.

> ii) Fig. 2c. The anticorrelation at the optimal squeezing angle are about 2% per atom, for 10 surrounding atoms. This seems to imply a variance reduction of 2x10x2%=40%. Is this consistent with the noise reduction of almost 6dB shown in Fig. 2a?

(A2) The referee raises an interesting question, but the data in Fig. 2a and 2c cannot be compared directly for several reasons:

1. The points in Fig. 2a and Fig. 2c do not come from the same data set and correspond to different system sizes ($N = 4 \times 4 = 16$ and $5 \times 14 = 70$ atoms). Note that the maximum achievable squeezing observed in the experiment differs for the two cases.
2. As the caption of Fig. 2a state, the $N = 16$ data is obtained for $t_{\text{int}} = 2.4\,\mu\text{s}$. This interaction time is not optimal in terms of the smallest Wineland parameter achievable in the experiment. This can be directly seen by a comparison to Fig. 2b which shows the time-dependece of the Wineland squeezing parameter for $N = 4 \times 4 = 16$ in the experiment. In contrast, the data in Fig. 2c has been taken close to the optimal interaction time.

However, we can use the correlators shown in Fig. 2c to calculate the variance of a single subarray according to the equation

$$\text{Var}\left[\hat{S}_z^{(A)}\right] = \sum_{i \neq j \in A} g_{i,j}^{(2)} + \frac{1}{4N_A^2} \sum_{i \in A} \left(1 - \langle \hat{\sigma}_z^{(i)} \rangle^2\right). \quad \text{(R1)}$$

We find that the results of this calculation agree with a calculation of the variance from the ensemble measurement, as shown in Fig. R1.

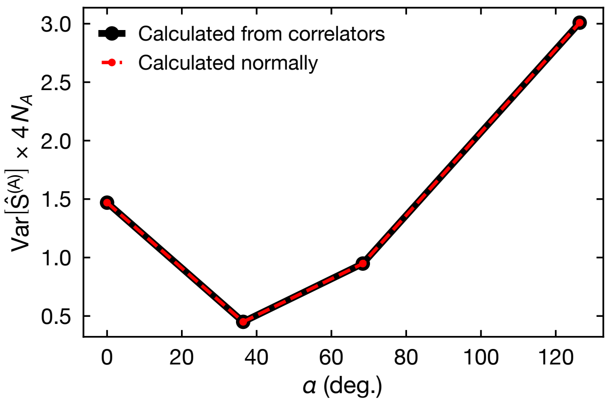

FIG. R1. **Single subarray variance from data in Fig. 2c.** In red, we plot the calculated values for $\mathrm{Var}\left[\hat{S}_z^{(A)}\right]$, as defined in the main text. The QPN-limit for this quantity is $1/(4N_A)$, to which we normalize calculated values. Using the same data, we can also calculate this variance by first computing the correlators $g_{i,j}^{(2)}$, as shown in Fig. 2c, and then calculating the variance according to Eq. (R1). Exact agreement between the calculated variance from these two approaches serves as a good consistency check on the calculation for the correlator presented in the main text.

> iii) Fig. 2d, the caption states "Small purple circles and lines show the theoretical prediction based on weak dressing". I thought that the circles in Fig. 2d are the data?

(A3) In Fig. 2d, the two-toned, larger purple circles are the data. "Small purple circles" was intended to refer to the smaller points connected by the solid purple line. Since both sets of points are circles, the caption left some room for ambiguity. We thank the reviewer for pointing this out, and we have reworded the caption to be more precise.

> iv) Fig. 3, in the caption the authors state "the ... points correspond to the first 200 points of the raw data from which ...the Allan deviations are calculated". Why is a subset selected rather than using all data? Do the other data show the same performance?

(A4) To begin, we would like to emphasize that the data for the Allan deviation plots in Fig 3. a,b use all collected data, and are not post-selected (with the exception of one trial in which zero atoms were loaded, as mentioned explicitly in the Methods section titled *Post-selection*). It is for purely visual purposes that we only display the first 200 points in the smaller plots on the right hand side of each subplot.

Related to this point, the referee has helpfully identified a sentence in the caption to Fig. 3 that could easily be misinterpreted. We have rewritten this sentence in the caption and tried to make it clearer.

> v) It would be helpful to the reader to state the cycle time in the main manuscript, and to specify in methods how much time is spent on what part of the preparation procedure.

(A5) We thank the referee for this suggestion, and have added a statement of the cycle time for data in Fig. 3 in the corresponding caption. Additionally, in the Methods, we have stated approximate durations for initial atom loading, rearrangement, and imaging.

> vi) Even after repeated reading I did not understand what zero-fill means in the Methods Section B, and which data were removed. Perhaps the authors could reformulate this section?

(A6) We thank the reviewer for identifying this confusing terminology. By "zero-fill," we mean that a given experimental run did not prepare any atoms. We have adjusted relevant Methods section, and no longer use this term.

> vii) Can the authors provide a figure in the methods or supplementary material showing the contrast reduction data?

(A7) In the first column of Fig. S4 in the supplementary information, we plot the contrast versus interaction time.

Response to report from the second referee (B)

> The authors report an experimental demonstration of Rydberg-mediated squeezing with a neutral atom clock providing 4 dB of metrological enhancement. The squeezing results from Ising-like interactions, which, thanks to their protocol, approximates the one-axis twisting squeezing interaction in a certain limit. The authors claim their clock has a fractional frequency stability of 10^{-15} at one-second averaging time and around 10^{-17} at half-hour measurement. Besides that, the authors explore the possibility of using the setup for differential phase interferometry via elliptical fitting. In that case, the system exhibits both precision enhancement and diminution due to entanglement.
> The writing is excellent, albeit dense, and the analysis is very complete. I have only minor concerns that I'm listing now.
> 1. According to the paper, the squeezing parameter (metrological enhancement) is determined from Eq. (3) by measuring σ^2 and fitting the contrast C. One alternative way of doing this would be to calculate the enhancement directly by performing Ramsey measurements for several values of the phase, plot dz as a function of the phase, and then estimate the sensitivity via $\delta\theta^2 = \Delta d_z^2/|dz/d\theta|^2$, i.e., calculating the slope at the working point by a fit. In that way, it is not necessary to determine C. Indeed, that is how the metrological gain is demonstrated in Ref. 33. That may be a more realistic way of estimating the enhancement, as it simulates the actual working of the device. I think both are only equal in perfect conditions. That also may explain why the enhancement in the Allan deviation plot is different from the one predicted by the squeezing parameter, as in the Allan deviation case, they perform an actual Ramsey interferometry sequence.

(B1) We thank the referee for this discussion. It is indeed important to note that the squeezing parameter, as quantified by Eq. 3, can differ from the provided metrological enhancement in an interferometry sequence, which might depend on a variety of decoherence mechanisms. Therefore, measurements of the squeezing parameter, as presented in Fig. 2, provide information on the quality of our Rydberg-dressing dynamics and state preparation. Measurements of improved sensitivity as quantified by the Allan deviation, presented in Fig. 3, then demonstrate that this metrologically useful entanglement can improve the performance of an actual Ramsey sequence.

> 2. The authors state in the introduction that all-to-all cavity-mediated interactions are a route to scalable spin squeezing, yet they fail to provide access to high-fidelity rotations necessary to boost precision even more. The current approach may give great control over the system. However, the results show that it scales poorly with N. Instead, the metrological gain saturates to a certain metrological enhancement. The authors comment on this.

(B2) We thank the referee for this comment, which addresses an important point. The demonstrated squeezing is roughly similar for the atom numbers explored in this work.

> 3. The theoretical model fails to describe the metrological gain for large values of N, as shown in Fig. 2 and the supplementary material. In the same line as the previous point, it is necessary to develop a good theoretical description before more sophisticated protocols can be developed to exploit the controllability of the device to increase sensitivity further.

(B3) As the referee points out, it will be important for future work to have a better understanding of potential collective loss mechanisms and multi-body interactions. Broadly, we expect that these investigations could provide exciting directions for future work, and reveal interesting effective interactions that are not entirely captured by two-body terms. While these effects are likely not beneficial for squeezing dynamics, a better understanding could point the way toward a more optimal principal quantum number for dressing experiments. Furthermore, protocols for generating other types of metrologically useful states, such as GHZ states, could be realized with resonant-Rydberg dynamics, which might allow for protocols that can be understood with exact numerics up to larger atom numbers than is possible with strong-dressing experiments.

> 4. I think it may be worth pointing out the scaling with N of the squeezing parameter predicted by the theoretical model employed.

(B4) We thank the referee for this suggestion. In the Methods, we now specify the limiting value for squeezing with the theoretical model for weak dressing in the limit $N \to \infty$.

> 5. In the context of the elliptical fitting, it is unclear whether squeezing should be beneficial or detrimental to the phase estimation, as, at times, squeezing and anti-squeezing will contribute to the measurement. On the other hand, Fig. 4b shows a contrast loss. That could explain Fig. 4c. Furthermore, there is literature suggesting that enhancement is possible under certain conditions (see Quantum-enhanced differential atom interferometers and clocks with spin-squeezing swapping. Quantum, 7, 965 (2023). That kind of analysis could explain the experimental results.

(B5) We thank the referee for bringing our attention to this paper, which points the way toward interesting directions for differential frequency comparisons, similar to the ones performed in our work.

> 6. Physical Review Letters 126, 113401 (2021) deals with a similar family of ideas (not exactly the same system), demonstrating 4.9 dB of enhancement via Ising-like interactions. I think the authors should contrast their findings with theirs.

(B6) The referee has pointed out a very interesting proposal for generating spin-squeezed states in a dipolar molecular gas. We can compare the interactions in this work to both 1. weak and 2. strong dressing interactions.

1. The spin Hamiltonian in the cited work takes on a different form than the theoretical model for weak dressing that we present. While weak dressing provides short-range Ising interactions, the molecular interactions in the cited work simulate an XXZ model. Ref. 63 studies this XXZ Hamiltonian intensively, and demonstrates that XXZ-type models do provide scalable squeezing. This is in contrast to pure Ising dynamics, which have also been the subject of extensive theoretical study, and which do not provide scalable squeezing.
2. A comparison to strong dressing dynamics is more challenging, as there are fewer theoretical works that can provide deep insight into the problem. This highlights one practical difference, which is that a non-perturbative numerical study of strong Rydberg dressing dynamics likely requires studying the three-level dynamics of the states $\{|g\rangle, |e\rangle, |r\rangle\}$ (as defined in the main text). Therefore, strong dressing cannot, to the best of our knowledge, be studied with the truncated-Wigner approximation, a technique exployed in the paper the referee points out, and which can be used to numerically study spin squeezing in two-level spin models up to the system sizes we achieve experimentally.

> In conclusion, the data demonstrate metrologically useful Rydberg-mediated squeezing, and the reported frequency stability is a significant step towards a competitive neutral atom clock. The analysis is comprehensive, and it will attract a broad audience. I think the work merits publication, subject to the minor revisions and clarifications I have suggested.

Changes related to questions, comments, and suggestions from referees

- (A1) We have adjusted the caption to Fig. 2a, and now state that the guide to the eye is a cosine fit to the data.
- (A3) We have reworded the caption to Fig. 2d to be more precise.
- (A4) We have rewritten the caption to Fig. 3 to make it clearer.

- (A5) have added a statement of the cycle time for data in Fig. 3 in the corresponding caption. Additionally, in the Methods, we have stated approximate durations for initial atom loading, rearrangement, and imaging.
- (A6) We have adjusted the relevant Methods section, and no longer use the term "zero-fill." The section of *Post-selection* has largely been rewritten to accommodate this change.
- (B4) In the Methods, we now specify the limiting value for squeezing with the theoretical model for weak dressing in the limit $N \to \infty$.