Peer Review File

Manuscript Title: Differential clock comparisons with a multiplexed optical lattice clock

Redactions – Mention of other journals

This document only contains reviewer comments, rebuttal and decision letters for versions considered at *Nature*. Mentions of the other journal have been redacted.

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1 - Remarks to the Author:
In the manuscript, Xin Zheng and coauthors demonstrate a perspective approach of forming a few (up to six) ensembles of Sr atoms in one optical lattice and demonstrated a clock transition spectroscopy in these ensembles. Being separated by a few mm distance, atoms in different ensembles experience almost identical laser frequency fluctuations and some other common noise types, so the relative frequency instability measured in different ensembles is extremely low. The accurate value of the relative frequency difference remains unknown. After setting a tight control on systematic shift gradients (e.g. magnetic field, collisional and BBR shifts, etc.), the demonstrated scheme can be used for e.g. gravitational shift measurement on a centimeter and sub-centimeter scales. Moreover, it opens perspectives for highly accurate isotopic shift measurements in different Sr isotopes which is an important tool for search of dark matter and "new physics".
The results are impressive, but I guess that the research is not finalized to the level when it can be published in Nature. After reading the manuscript it was difficult to highlight the main scientific result of the work. The demonstrated level of relative instability of 9x10-20 for two ensembles is indeed impressing, but it is the long way to translate it to e.g. gravitational shift measurements on cm level, since the latter requires accurate knowledge of the absolute frequency difference. It in turn requires exact knowledge of gradient systematic shifts at the requested level of accuracy which is a separate research. The same is valid for the isotope shift measurements. Moreover, the discussion of frequency shifts, which come from spectroscopy of bosonic isotopes, is essential. Authors correctly mention that the presence of magnetic field is necessary for spectroscopy of the bosonic isotopes, but there is no discussion what are the requirements on the magnetic fields (gradients).
Another reason why the manuscript looks not finalized for the Nature publication. Roughly half of the manuscript is basically devoted to fighting against system imperfections:
1. The clock laser. It seems (the authors did not show this data) that the laser coherence is limited by a large cavity frequency drift. For a 12cm-long cavity, the Allan deviation floor (the thermal noise limit) should be at ~3-5x10-16. Of course this is not compatible with the best cryogenic cavities, but should allow longer Rabi pulses. At the same time the idea to measure atom-atom coherence rejecting laser noise is not novel [Young, A.W. et al. Nature 588, 408–413 (2020).]. It this paper the demonstrated coherence time is similar, but the method how to extract the differential frequency instability is given more clearly.
2. The description of "closed-looped" comparisons is too long. This part basically shows that the concept of excitation-correlation comparison works basically demonstrating the proof-of-principle.

The Allan deviation presented (Fig 3c) looks very unusual and confusing. The red point doesn't belong to the Allan deviation and cannot be placed there.

Some minor notes:
1. In abstract: It is strange to see uncertainty of fractional uncertainty.
2. In abstract: "… 26 seconds, a 270-fold improvement over the atom-laser coherence time,

…”. This statement now reads as “compared to previous measurements”, but not “compared to our measurement” of atom-laser coherence.
3. Main text, first paragraph: should be instability and inaccuracy, because we are talking about small values. This mistake repeats in the text. Sometimes term “accuracy” is replaced by “precision”.
4. Page 2, last paragraph: QPN-limited instability of what?
5. Page 4, last paragraph: what does the statement “variance of the Ramsey signal remains high at 100 ms” mean?
6. Later in the same paragraph: “on an ellipse with an opening angle determined by the differential Ramsey phase acquired between the two ensembles” – in my opinion it would be good to already here specify that there is also may be an integer number of \pi/2.
7. Fig2.c and similar figures further: the ellipse fit does not look good. It looks like the longer major axis would do a better fit.
8. In general: not all the plots have specifications of experimental parameters (i.e. dark time).
9. Caption fig1. “Synchronized Ramsey interrogation of the magnetically insensitive” – I would say that “less sensitive” rather than “insensitive”
10. Page 9, top: “By working at shallower lattices and actively controlling atom loading, differential density shifts with uncertainties at the 10-20 level should be feasible.” How much shallower? 20 Er is already quite low, and it is really hard to work with <5Er, so the feasible improvement is a factor of 2-4.
11. Conclusion: “Full characterization of systematic effects such as differential BBR, Stark, and Zeeman shifts will open up possibilities for studying relativistic geodesy at the sub-cm”. In my opinion” studying” is a rather strong statement. For the measurements in the presented configuration (when one cannot move apart atomic ensembles) the anticipated measurement uncertainty even at 100 um level would give only 1% gravitational shift accuracy. For separated optical clock, the systematic shifts would be uncommon.
12. Methods, clock laser beam path section: Could you please comment on why one need to use a dichroic beam splitter to excite \pi-transition in 9/2 – 9/2 transition (when one needs one particular polarization), and a polarizing beam splitter to excite \sigma-transitions, but not vice versa.
13. Extended data figure 4: With a free offset frequency coefficient I do not understand how the U^{5/4} fit can be preferred over simple linear fit.

Summing up, although the results of the manuscript are interesting and important for the specialists, I would recommend publishing in Nature communications or other more specialized journal.

Referee #2 - Remarks to the Author:
“High precision differential clock comparisons with a multiplexed optical lattice clock” reports several noteworthy results and advances:

(1) An interesting and novel twist on the architecture of lattice clock systems: using a moving lattice to create distinct atom ensembles in the same 1D lattice with spatial separations up to a full cm.
(2) Correlated frequency measurements between these ensembles with record relative stability, less than 1e-17 per sqrt of averaging time.
(3) The relative instability reported was achieved without using a heroic ultra-stable laser system. Indeed, the authors report an atom-atom coherence time 270x longer than the laser-atom coherence time. While there are a number of related previous works that exploit correlated measurements to reject limiting noise processes, none can boast the record level of noise/instability rejection demonstrated here.
(4) By observing long atom-atom coherence times (26 seconds), they demonstrate an attractive operational space for future lattice clock efforts that are not completely limited by effects such as tunneling and atomic collisions.
(5) The authors further extend these results by creating six distinct ensembles in their lattice and making measurements between each pair, also with excellent relative stability.
(6) They load different combinations of two strontium isotopes in the 1D lattice, spatially

separated as before.
Beyond these technical achievements and novelties, I find the paper to be extremely well-written. The text is clear, with sufficient detail and depth to support their findings. The analysis presented is thorough and convincing. In short, there is a lot I like about this paper.

Nevertheless, I opine that Nature is not the appropriate journal for these results. My primary concern is that the correlated measurement techniques showcased here relies on extreme cancellation of common mode effects. The number of settings where this cancellation is possible is limited, with the present case involving a scenario where the atom ensembles share the same lattice, vacuum apparatus, spectroscopy laser, readout laser, etc. Consequently, most systematic effects and noise processes are highly common mode. As one relaxes some of these constraints in order to increase the reach of possible scientific application, technical challenges would limit the realized level of cancellation. Indeed, existing work has explored some of these scenarios already.

The authors identify several possible applications of these type of measurements. To be more concrete about the limited range of applicability, I list them here, together with some possible challenges for addressing these applications with the present techniques:

(1) Precision isotope shift measurements: Since each isotope possesses a unique magic wavelength, distinct isotope ensembles in the same lattice would be sensitive to light shifts and/or spectral broadening that are no longer common mode. This could significantly impact measurement precision.
(2) Spatially-resolved characterization of limiting clock systematics: This technique would certainly be useful for exploring systematic gradients in the lattice, which are interesting in some contexts. Furthermore, as the authors demonstrated with the density shift, for those systematic effects where differential conditions in each ensemble can be controlled, systematic effects can be studied. However, most systematics would be common to all ensembles and lacking individualized control, and so could not be evaluated by this method.
(3) Development of clock-based gravitational wave and dark matter detectors: For gravity wave detection, presumably a much larger baseline between the clocks would be required. For most dark matter models, one would also benefit from a larger baseline or vastly different atomic species, both of which would require different techniques.
(4) Measurements of the gravitational redshift at the sub-cm scale: The techniques here are very well-suited for exploring the redshift at sub-cm scales. However, because the shift is so small on these small length scales, the fractional level of measurement precision will certainly be limited.
(5) A platform for exploring spin-squeezing: Because almost all noises except quantum projection noise are cancelled in common mode here, this platform could certainly be well-suited for exploring some spin-squeezing protocols, at least to enhance differential measurement precision.

As described above, this paper has many strengths. I strongly endorse the publication of this paper in a somewhat more specialized journal like Nature Physics. However, I think that the limited scope of these differential techniques make it too specialized for Nature.

As the authors point out, another work using strontium in a lattice recently reported differential measurements between two regions within a single atom ensemble. It seems that these two papers even appeared on the arXiv the same day! And while both papers have a lot of overlap, I found the two complementary and both are interesting in their own right. That other paper carries out a more detailed analysis of systematic effects, and combined with even better relative stability, reports the observation of the redshift on the mm scale of their atomic ensemble. This paper achieves larger spatial separations (cm scale) between two or more ensembles, and deeper cancelation of clock laser noise in the differential measurements.

Below, please find a few miscellaneous comments for consideration by the authors:

(A) Speaking generally, the authors do a very nice job of clearly identifying that the precision reported is relative (between one atom sample to another), rather than something in absolute

terms. This is to the authors' credit, as many of the other referenced papers using differential techniques can be confusing on this point. And while it can be tiresome to use technical or cumbersome terminology throughout the paper, there are a few places that clarity might be improved with the addition of a word or two. This distinction can
Bottom of Page 2: "These results illustrate that simultaneous differential clock comparisons enable record-setting stability and precision without requiring state-of-the-art mHz linewidth clock lasers …" Add 'relative' before 'stability'.
Page 6: "To characterize the stability of the multiplexed OLC, …" Add 'relative' before 'stability'.
Page 7: "This demonstration of precision below the 10-19 level with a rack-mounted, …" Add 'relative' before 'precision'.

(B) The authors elect to use the term 'uncertainty' when referring to relative stability at some long averaging time. This is perhaps unconventional, and clarity could be improved by referring to it as 'relative statistical uncertainty' or simply 'relative long-term stability'.

(C) The authors note that the retro-reflected lattice beam has 50% the power of the incident beam, mostly due to the efficiency of the double pass AOMs employed. Therefore, in addition to the lattice standing wave, the atoms experience an additional running wave. Is the observed 3P0 quenching rate consistent with past measurements, when accounting for both the standing and running wave? Does the running wave have any impact on the moving lattice?

Referee #3 - Remarks to the Author:
The authors describe the results from their new optical lattice clock machine where more than one atomic ensembles can be trapped and interrogated by a single clock laser to cancel out the effect of the decoherence between the atoms and the laser. Although the whole description is somewhat on the technical side, the ultimate Allan deviation of 8.9x10^-20 they reached has an unprecedentedly high precision, and their system has unique features of trapping and comparing atomic ensembles millimeters away that can be utilized for various purposes in the future. Comparing six atomic ensembles simultaneously is another achievement for the first time. The methodology is well described in details, and their logical flow towards the conclusions is clearly described.

The manuscript sometimes includes contents that themselves are impressive but are irrelevant to the main context of the high precision differential clock comparison. It would be better for these contents to be published as a more technical instrumentation paper and such paper should be cited in this manuscript, if necessary.

1. From the end of page 3 to the beginning of page 4, the feature of the moving lattice is descibed in details, including the maximum acceleration achievable, but except for the fact that the lattice is moving lattice and this is the key point for trapping multiple atomic ensembles, all other details is irrelevant to the context of this paper.

2. Trapping multiple isotope simultaneously is out of the context of this paper. It is recommended to publish another technical paper about this topic and part of the abstract, Fig. 4d, 2nd paragraph of page 9, and supplementary information F should be moved to this paper, as the technical details of this part themselves are interesting to a part of readers. In case authors like to keep this topic of trapping multiple isotopes in the manuscript, a certain kind of differential clock comparison (e.g., Allan deviation plot for the frequency ratio measurement of the same transition in two different isotopes) is desired to match the context.

Also, following points need to be clarified for readers' better understanding.
3. What limited the length of integration time to 3.3 hours? Typically for the state-of-the-art atomic clocks with uncertainty budgets, they show an Allan deviation plot down to their systematic uncertainties. However, in this manuscript such limitation does not apply, as the systematic shifts are currently under investigation.

Overall, the manuscript reaches the level that can be accepted for publication, but the points mentioned above need to be elaborated further before the final decision.

There are some minor comments on typos and so on.

4. In the 5th line of the second paragraph of page 6, the number with long digits should be separated with spaces, not commas.

5. In page 17, section Units and errors, the acronym s.d. is not defined (though it is quite obvious that it stands for standard deviation).

6. In the 3rd paragraph of page 29, rec of Erec should be subscription.

7. Regarding "experimental as input, that we can bound the bias error" in the end of page 31, either the comma should be removed or that should be changed to which.

Author Rebuttals to Initial Comments:

We thank the three referees for thoroughly reviewing our manuscript, and for their insightful comments and helpful feedback. We have revised our manuscript to reflect and address the input of all three referees. We believe that the manuscript has been significantly improved as a result. Below are point-by-point responses to each of the reviewers' comments (original comments in italics):

Referee #1:

The results are impressive, but I guess that the research is not finalized to the level when it can be published in Nature. After reading the manuscript it was difficult to highlight the main scientific result of the work. The demonstrated level of relative instability of 9x10^{-20} for two ensembles is indeed impressing, but it is the long way to translate it to e.g. gravitational shift measurements on cm level, since the latter requires accurate knowledge of the absolute frequency difference. It in turn requires exact knowledge of gradient systematic shifts at the requested level of accuracy which is a separate research. The same is valid for the isotope shift measurements. Moreover, the discussion of frequency shifts, which come from spectroscopy of bosonic isotopes, is essential. Authors correctly mention that the presence of magneticfield is necessary for spectroscopy of the bosonic isotopes, but there is no discussion what are the requirements on the magnetic fields (gradients.)

We are gratified that the referee finds our results impressive. However, we respectfully disagree with the spirit of the referee's comment that a full systematic evaluation of the sources of the absolute frequency difference is required in order for our results to be highly impactful. With the help of the comments of all three referees, we have revised the manuscript to better communicate and reflect our main scientific results, which are achieving atom-atom coherence time of 26 seconds using a vertical, shallow, one-dimensional optical lattice, and the demonstration of extremely low levels of instability and statistical uncertainty for a frequency difference between two clocks using synchronized differential comparisons with a Hz linewidth clock laser in a novel measurement platform with unique capabilities. A thorough analysis and discussion of the systematics for a gravitational redshift measurement or isotope shift comparison are outside the scope of this work, and will be included in future manuscripts that focus on each of these studies.

The clock laser. It seems (the authors did not show this data) that the laser coherence is limited by a large cavity frequency drift. For a 12cm-long cavity, the Allan deviation floor (the thermal noise limit) should be at ~3-5x10-16. Of course this is not compatible with the best cryogenic cavities, but should allow longer Rabi pulses.

We thank the referee for raising this important point. The Allan deviation of our clock laser is ~8x10^-16 at 1 second as measured and specified by the manufacturer (Menlo Systems), corresponding to a measured instantaneous linewidth of ~1 Hz. We agree with the referee that this is not quite at the thermal noise limit for a 12 cm ULE cavity,

and we believe this is likely the result of residual sensitivity to both acceleration noise and thermal drifts (the cavity is housed in a standard rack rather than on an optical table.) Due to both the dead time between experiments required to recool, trap, prepare and reading-out the strontium atoms and the need to perform multiple measurements to determine a linewidth or coherence time, we do observe an atom-laser coherence time of 100 ms and a measured Rabi linewidth of 10 Hz, roughly a factor of 10 worse than the instantaneous laser linewidth. This is consistent with other prior optical clock experiments, such as *Young, A.W. et al. Nature 588, 408–413 (2020)*, where a laser-atom coherence time of 3 s was observed despite the use of an ~8 mHz linewidthlaser with an instability of 4x10^-17 at 1 s.

(1) At the same time the idea to measure atom-atom coherence rejecting laser noise isnot novel [Young, A.W. et al. Nature 588, 408–413 (2020).]. It this paper the demonstrated coherence time is similar, but the method how to extract the differential frequency instability is given more clearly.

We agree with the referee that our work is not the first work to suggest or demonstrate the use of simultaneous correlated measurements to reject clock laser noise and probe beyond the measured atom-laser coherence time. However, while the atom-atom coherence times in the *Young et al.,* are similar to (and even exceed) the atom-atom coherence times we measure, we emphasize that the measurements in *Young et al.,* were performed with an 8 mHz linewidth laser [the same laser used in *Campbell, S. L.et al., Science **358**, 90-94 (2017)* and *Marti, G. E. et al., Phys. Rev. Lett. **120**, 103201 (2018)*], so the atom-atom coherence times they measure likely do not exceed the instantaneous coherence time of the local oscillator. In addition, the trapping geometryin our work is completely different. In optical tweezers, single-atom-per-site occupation rejects the atomic interactions, while the tight confinement limits the impactof the finite atom temperatures. Coupling between axial and radial motional modes andp-wave collisions and the associated density shift have been considered major limitingfactors in one-dimensional (1D) lattices from the previous studies *[Campbell, G. K. etal., Science **324**, 360-363 (2009), Lemke, N. D. et al., Phys. Rev. Lett. **107**, 103902 (2011), Martin, M. J. et al., Science **341**, 632-636 (2013), Campbell, S. L. et al., Science**358**, 90-94 (2017)]*. We therefore feel our demonstration of 26 seconds atom-atom coherence time using a vertical, shallow 1D lattice is a highly non-trivial result. Our results show that 1D optical lattice clocks need not be limited by atomic interactions, and that high precision differential clock comparisons do not require sub-Hz linewidthclock lasers. We have revised our manuscript to better communicate these points.

2) The description of "closed-looped" comparisons is too long. This part basically shows that the concept of excitation-correlation comparison works basically demonstrating the proof-of-principle.

We agree with the referee. We have simplified and condensed this discussion in the revised manuscript.

3) The Allan deviation presented (Fig 3c) looks very unusual and confusing. The red point doesn"t belong to the Allan deviation and cannot be placed there.

We thank the referee for pointing this out. We agree with the referee. This point is removed in the revised manuscript.

4) In abstract: It is strange to see uncertainty of fractional uncertainty.

We thank the referee for pointing this out. We have removed the uncertainty for fractional uncertainty in the revised manuscript.

5) In abstract: "… 26 seconds, a 270-fold improvement over the atom-laser coherence time, …". This statement now reads as "compared to previous measurements", but not "compared to our measurement" of atom-laser coherence.

We thank the referee for pointing this out. This is corrected as "compared to our measured atom-laser coherence time" in the revised manuscript.

6) Main text, first paragraph: should be instability and inaccuracy, because we are talking about small values. This mistake repeats in the text. Sometimes term "accuracy" is replaced by "precision".

We thank the referee for raising this issue. This is corrected as "instability and inaccuracy" in the revised manuscript.

7) Page 2, last paragraph: QPN-limited instability of what?

We thank the referee for pointing this out. This has been clarified as "instability of 8.9x10^{-18} consistent with the QPN limit" in the revised manuscript.

8) Page 4, last paragraph: what does the statement "variance of the Ramsey signal remains high at 100 ms" mean?

We thank the referee for raising this issue. This is now clarified as "the variance of excitation fraction remains large at 100 ms (Fig.2b, inset2), implying that for that interrogation time the atoms within the ensemble remain phase coherent with each other [27]" in the revised manuscript.

9) Later in the same paragraph: "on an ellipse with an opening angle determined by the differential Ramsey phase acquired between the two ensembles" – in my opinion it would be good to already here specify that there is also may be an integer number of \pi/2.

We thank the referee for pointing this out. This is now included in the revised manuscript.

10) Fig2.c and similar figures further: the ellipse fit does not look good. It looks like the longer major axis would do a better fit.

We thank the referee for pointing out this. This is due to the bias in the least-squares ellipse fitting in the presence of QPN, which results in relatively poor fitting performance at 0 or \pi phase. This is clarified in the revised manuscript.

11) In general: not all the plots have specifications of experimental parameters (i.e. dark time).

We thank the referee for pointing this out. The specifications of experimental parameters, such as dark time and lattice trap depth, are now included into the revised manuscript.

12) Caption fig1. "Synchronized Ramsey interrogation of the magnetically insensitive" – I would say that "less sensitive" rather than "insensitive"

We agree with the referee. This is corrected as "less sensitive" in the revised manuscript.

13) Page 9, top: "By working at shallower lattices and actively controlling atom loading, differential density shifts with uncertainties at the 10-20 level should be feasible." How much shallower? 20 Er is already quite low, and it is really hard to work with <5Er, so the feasible improvement is a factor of 2-4.

We agree with the referee. This statement is now removed in the revised manuscript.

14) Conclusion: "Full characterization of systematic effects such as differential BBR, Stark, and Zeeman shifts will open up possibilities for studying relativistic geodesy at the sub-cm". In my opinion" studying" is a rather strong statement. For the measurements in the presented configuration (when one cannot move apart atomic ensembles) the anticipated measurement uncertainty even at 100 um level would give *only 1% gravitational shift accuracy. For separated optical clock, the systematic shifts would be uncommon.*

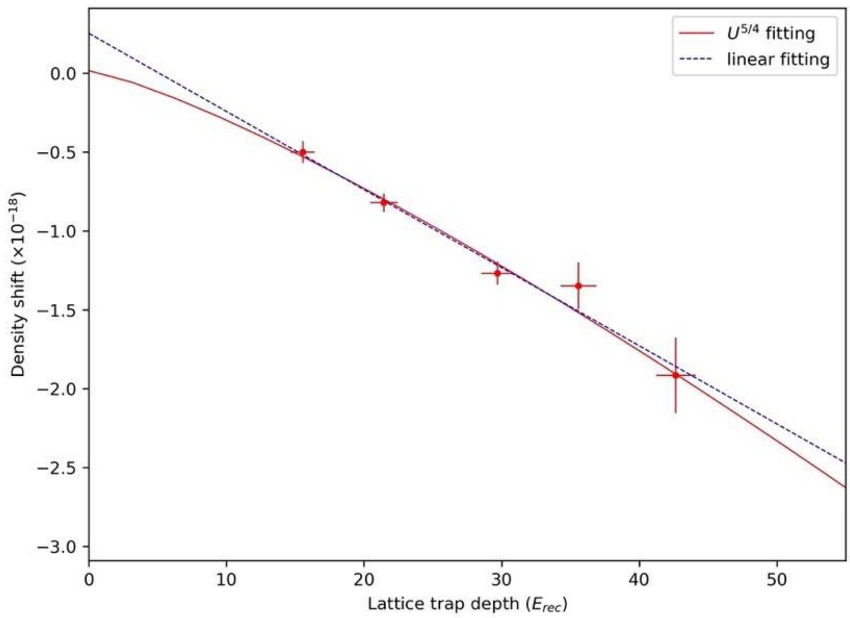
We agree with the referee. “Measuring” is now used in the revised manuscript.

15) Methods, clock laser beam path section: Could you please comment on why one need to use a dichroic beam splitter to excite \pi-transition in 9/2 – 9/2 transition (when one needs one particular polarization), and a polarizing beam splitter to excite \sigma-transitions, but not vice versa.

We thank the referee for pointing this out. Linearly polarized clock beam with polarization parallel (perpendicular) to the bias magnetic field is used to excite the \pi (\sigma) transition. This is elaborated more explicitly in the both the inset and caption of Extended Data Fig.1 in the revised manuscript.

16) Extended data figure 4: With a free offset frequency coefficient I do not understand how the U^{5/4} fit can be preferred over simple linear fit.

We thank the referee for pointing this out. Below please find a comparison between the (a U^{5/4} + b) fitting and the simply linear (a U + b) fitting to the same data, in which a greater offset from zero is observed in linear fitting.



Referee #2:

1) Nevertheless, I opine that Nature is not the appropriate journal for these results. My primary concern is that the correlated measurement techniques showcased here relies on extreme cancellation of common mode effects. The number of settings where this cancellation is possible is limited, with the present case involving a scenario where the atom ensembles share the same lattice, vacuum apparatus, spectroscopy laser, readout laser, etc. Consequently, most systematic effects and noise processes are highly common mode. As one relaxes some of these constraints in order to increase the reach of possible scientific application, technical challenges would limit the realized level of cancellation. Indeed, existing work has explored some of these scenarios already.

We thank the referee for raising this issue. We agree that our results rely on extreme cancellation of common mode effects, and in particular on cancellation of the noise on the clock laser used to interrogate the atom ensembles. However, we respectfully disagree with the spirit of the referee's comment that this reduces the impact of our work. On the contrary, we believe that our work highlights the advantages of engineering situations in which as many effects are made common mode as possible. The referee correctly points out that other works have already explored some of these scenarios and demonstrated the feasibility of this approach in those contexts, including the use of the same vacuum chamber for shared environmental perturbations in interleaved isotope shift measurements *[Takano, T. et al., Appl. Phys. Express **10** 072801 (2017)]* and the use of the same clock laser to probe two ions in separate vacuum chambers beyond the coherence time of the laser *[Clements, E. R. et al., Phys. Rev. Lett. **125**, 243602 (2020)]*. Furthermore, as the referee notes below, there are a number of (in our view) significant research directions that can be explored using our approach, which we discuss further below. We have revised the manuscript to clarify and highlight these points.

2) Precision isotope shift measurements: Since each isotope possesses a unique magic wavelength, distinct isotope ensembles in the same lattice would be sensitive to light shifts and/or spectral broadening that are no longer common mode. This could significantly impact measurement precision.

We thank the referee for raising this point. We agree with the referee that each isotope will have a unique magic wavelength that will be shifted by hundreds of MHz from that of the other isotopes. However, we anticipate that isotope shift measurements will still benefit from simultaneous interrogation of both isotopes in a shared optical lattice. As we demonstrate in our manuscript, the atomic coherence is still reasonably good (>55% contrast) at +/- 250 MHz lattice detunings from the magic wavelength at 20 E_{rec} trap depth for a 10 s Ramsey interrogation (Fig.2**e** in the main text). In addition, the magic wavelength for 87Sr can be tuned by selecting specific hyperfine transitions to adjust the tensor lattice shift and by applying circularly polarized lattice light to introduce and vary the vector lattice shift. In addition, in *Takano, T. et al., Appl. Phys. Express **10** 072801 (2017)* they interleaved measurements between ^{88}Sr and ^{87}Sr in the same lattice and vacuum chamber, and showed that higher order lattice light shifts remain common mode, leaving linear differential shifts that can be relatively easily characterized and extrapolated to zero lattice intensity.

3) Spatially-resolved characterization of limiting clock systematics: This technique would certainly be useful for exploring systematic gradients in the lattice, which are interesting in some contexts. Furthermore, as the authors demonstrated with thedensity shift, for those systematic effects where differential conditions in each ensemblecan be controlled, systematic effects can be studied. However, most systematics wouldbe common to all ensembles and lacking individualized control, and so could not be evaluated by this method.

We thank the referee for pointing this out. We agree that the study of spatial gradients is becoming increasingly important in evaluating and improving the accuracy of optical

clocks. In addition, as the referee points out, as long as a specific systematic can be introduced in a differential manner between the ensembles it can be studied while still benefiting from common mode cancellation of laser noise and all other common systematics. Magnetic field gradients, electric field gradients and thermal gradients can be controllably applied in order to study absolute Zeeman, dc Stark, and BBR shifts. To study ac Stark shifts from lattice light, clock light, and other wavelengths of interest, laser light can be selectively applied to a single ensemble using a focused beam applied perpendicular to the lattice axis. Density shifts and the effects of finite atom temperature can be studied by altering the loading sequence. We therefore argue that the multiplexed OLC technique can in principle be used to study most (but not all) limiting clock systematics.

4) Development of clock-based gravitational wave and dark matter detectors: For gravity wave detection, presumably a much larger baseline between the clocks would be required. For most dark matter models, one would also benefit from a larger baseline or vastly different atomic species, both of which would require different techniques.

We thank the referee for pointing this out. Gravitational wave detection would require a spaced-based clock laser which is likely to be orders of magnitude worse than the state-of-the-art cryogenic silicon cavity stabilized laser. Demonstration of long atomic coherence and low relative instability using a Hz-linewidth, rack-mounted local oscillator is an important proof-of-principle step and will enable us and others to test protocols for these applications. Differential spectroscopy and improved clock comparison between differential atomic species have also recently been demonstrated in *Kim, M. et al., arXiv: 2109.09540 (2021)*, therefore we see no fundamental limit to achieving comparable degree of laser noise cancellation across short baselines. The referee is correct that other systematics will no longer necessarily be common mode in these experiments, and that long optical baselines will add photon shot-noise.

5) Measurements of the gravitational redshift at the sub-cm scale: The techniques here are very well-suited for exploring the redshift at sub-cm scales. However, because the shift is so small on these small length scales, the fractional level of measurement precision will certainly be limited.

We thank the referee for pointing this out. We agree the fractional level of measurement precision at sub-cm scale will be limited, however, we believe that a testof relativity at a fractional precision of ~1% at the cm length scale is still of interest. Inaddition, measuring the gravitational redshift at these scales is also of interest for exploring the achievable limits of relativistic geodesy with optical clocks. We have revised the manuscript to better reflect this point.

6) A platform for exploring spin-squeezing: Because almost all noises except quantum projection noise are cancelled in common mode here, this platform could certainly be *well-suited for exploring some spin-squeezing protocols, at least to enhance differential measurement precision.*

We agree with the referee. We have revised the manuscript to emphasize this point.

7) Bottom of Page 2: "These results illustrate that simultaneous differential clock comparisons enable record-setting stability and precision without requiring state-of-the-art mHz linewidth clock lasers …" Add „relative" before „stability".

We thank the referee for clarifying this. This has been added in the revised manuscript.

8) Page 6: "To characterize the stability of the multiplexed OLC, …" Add „relative" before „stability".

We thank the referee for clarifying this. This is added in the revised manuscript.

8) Page 7: "This demonstration of precision below the 10-19 level with a rack-mounted, …" Add „relative" before „precision".

We thank the referee for clarifying this. This is added in the revised manuscript.

9) The authors elect to use the term „uncertainty" when referring to relative stability at some long averaging time. This is perhaps unconventional, and clarity could be improved by referring to it as „relative statistical uncertainty" or simply „relative long-term stability".

We agree with the referee. This is now included in the revised manuscript.

10) The authors note that the retro-reflected lattice beam has 50% the power of the incident beam, mostly due to the efficiency of the double pass AOMs employed. Therefore, in addition to the lattice standing wave, the atoms experience an additional running wave. Is the observed 3P0 quenching rate consistent with past measurements, when accounting for both the standing and running wave? Does the running wave have any impact on the moving lattice?

We thank the referee for raising this point. From preliminary measurements, the 3P0 Raman scattering rate is consistent with the rate reported in *Dörscher, S. et al., Phys. Rev. A* ***97****, 063419 (2018)*, after accounting for both the running and standing wave patterns. We do not observe any impact on the moving lattice from the running wave, which acts as a very weak dipole trap on top of the lattice.

Referee #3:

1) From the end of page 3 to the beginning of page 4, the feature of the moving lattice is descibed in details, including the maximum acceleration achievable, but except for the fact that the lattice is moving lattice and this is the key point for trapping multiple atomic ensembles, all other details is irrelevant to the context of this paper.

We thank the referee for pointing this out. We agree with the referee. The technical details regarding the moving lattice have been moved to the Methods section in the revised manuscript.

2) Trapping multiple isotope simultaneously is out of the context of this paper. It is recommended to publish another technical paper about this topic and part of the abstract, Fig. 4d, 2nd paragraph of page 9, and supplementary information F should be moved to this paper, as the technical details of this part themselves are interesting to a part of readers. In case authors like to keep this topic of trapping multiple isotopes in the manuscript, a certain kind of differential clock comparison (e.g., Allan deviation plot for the frequency ratio measurement of the same transition in two different isotopes) is desired to match the context.

We thank the referee for raising this point. As Referee #2 pointed out, isotope shift comparisons are one example of an interesting differential frequency shift between two otherwise identical ensembles in the same environment, and in our opinion therefore represent an important use case for the techniques we demonstrate in this work. To our knowledge the loading of multiple isotopes in the same lattice in this manner has not previously been realized (although mixtures of strontium isotopes have previously been loaded in dipole traps in e.g. *Tey, M.K. et al., PRA* ***82*** *011608 (2010)*.) We therefore feel this capability is important to highlight, and opt to leave it in the manuscript. As noted in the manuscript, a differential comparison between two isotopes will require larger magnetic fields and two simultaneous clock laser frequencies, and is therefore left for a future work.

3) What limited the length of integration time to 3.3 hours? Typically for the state-of-the-art atomic clocks with uncertainty budgets, they show an Allan deviation plot down to their systematic uncertainties. However, in this manuscript such limitationdoes not apply, as the systematic shifts are currently under investigation.

We thank the referee for raising this point. The 3.3-hour operation was chosen such that a set of data averaging time is greater than 10000 seconds. We anticipate differential Zeeman shifts associated with the fluctuations in ambient magnetic field gradient and differential density shift associated with ensemble loadings due to temperature drifts would limit us at the mid-10^{-20} level, but have not yet averaged beyond 3.3 hours to reach this point.

4) In the 5th line of the second paragraph of page 6, the number with long digits shouldbe separated with spaces, not commas.

We thank the referee for pointing this out. This is corrected in the revised manuscript.

5) In page 17, section Units and errors, the acronym s.d. is not defined (though it isquite obvious that it stands for standard deviation).

We thank the referee for pointing this out. This is corrected in the revised manuscript.

6) In the 3rd paragraph of page 29, rec of Erec should be subscription.

We thank the referee for pointing this out. This is corrected in the revised manuscript.

7) Regarding "experimental as input, that we can bound the bias error" in the end ofpage 31, either the comma should be removed or that should be changed to which.

We thank the referee for clarifying this. We remove "that" in the revised manuscript.

We trust that these changes satisfactorily address the above comments, and thank the referees once again for their time, effort, insightful comments, and thoughtful consideration of this manuscript.

Xin Zheng and Shimon Kolkowitz, on behalf of the authors