nature portfolio

Peer Review File

Purcell-enhanced spin-phonon coupling with a single color-center

Corresponding Author: Dr Graham Joe

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
In this manuscript, the authors present a study based on a specially engineered, microwave-frequency nanomechanical resonator, and they observe the acoustic Purcell effect between SiV spin qubits and the acoustic mode of the nanomechanical resonator. Furthermore, utilizing this system, the phonon spectra of the nanomechanical resonator across a broad frequency range of 9–27 GHz have been measured using SiV center spin. This marks the first observation of the acoustic Purcell effect in single-phonon level spectroscopy, with the highest spin-phonon cooperativity reported to date. Additionally, the nanomechanical resonator structure developed in this manuscript introduces a new regime of control for quantum defects in solids.

Overall, the manuscript presents robust results and offers a detailed explanation of the methodology. However, concerns arise regarding the quality of the data and the treatment of uncertainties, which require further clarification. Additionally, the manuscript does not cite some previous studies on the acoustic Purcell effect, which is essential for the current work. Most importantly, the originality and significance of the article are insufficient to publication in NATURE. Therefore, we conclude that this manuscript is not currently suitable for publication in NATURE. Specific comments are provided below:

The nanomechanical resonator proposed in this manuscript has been previously described in the author's earlier publications (Nano Letters 2024, 24(23), 6831-6837 and Nature Physics 2025, 21, 77-82). Furthermore, the acoustic Purcell effect has also been reported in Phys. Rev. Lett. 120, 114301, and FUNDAMENTAL RESEARCH 4(1), pp. 57-62. These literature indicate that the current manuscript may lack sufficient novelty.
The linear term g_sm (b̂σ̂_++b̂^† σ̂_-) is employed to describe the spin-phonon interaction. However, the nonlinear effects have not been addressed in the manuscript.
Multiple resonance signals are observed in Fig. 4(a) of the manuscript. The authors attribute these to other mechanical modes and provides numerical simulation results in Fig. 19. However, only the first two resonances in this figure align well with the observations, while the last two resonances are not evident in the simulation results.
In Fig.5, the presence of excessively large error bars undermines the validity of the author's conclusion that the transition frequency varies linearly with the magnetic field.
As previously noted, the acoustic Purcell effect has also been reported in Phys. Rev. Lett. 120, 114301 (2018), and FUNDAMENTAL RESEARCH 4(1), pp. 57-62（2024）. These references, however, are not cited within this manuscript.

In conclusion, this work makes valuable contributions to hybrid quantum systems by establishing an acoustic interface with record performance metrics. However, the current evidence for paradigm-shifting advancement over existing approaches remains incomplete. With comprehensive revisions addressing the above concerns, this study would constitute a strong candidate for high-impact publication in the journal such as Nature Communications.

Referee #2

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #3

(Remarks to the Author)
The manuscript “Observation of the acoustic Purcell effect with a color-center and a nanomechanical resonator” by Graham Joe, et al., demonstrates (for the first time, as far as we can tell), the Purcell-enhanced decay of a single atomic defect via the increased acoustic density of states of a phononic microcavity. The defect under consideration here is the SiV in diamond, and the enhanced decay is that of the electron spin in the ground state. This constitutes a seminal demonstration. The results have far-reaching impact and permit development of applications such as transduction for quantum coherent links. Overall, the paper is well-written, the data is convincing, and the supplementary information contains a wealth of background and additional information. We would, however, ask to consider the below comments, which we believe will make the manuscript more approachable

1. Introduction: In the sentence “The physics of an atom in an open continuum also applies to the emission of acoustic waves from an artificial atom in a solid, and is known to limit spin relaxation times in several color-centers [12–15] and rare-earth ion impurities [16, 17] at low temperatures.”, what exactly is the message? That spin relaxation via phonon emission is the dominant decay process at low temperatures?

2. Introduction: In “and control of phonon-induced orbital relaxation processes using phononic crystals [36].”, it might be good to mention that the suppression of the phonon-induced orbital relaxation processes is due the phononic crystal bandgap.

3. Introduction: In “A central challenge in this context has been to realize an experimental system that provides high spin-phonon co-operativity, …”, it might worth replacing “experimental system” with “resonator-defect system” or “resonator-emitter system”.

4. Introduction: “Structuring a solid and crafting an acoustic environment with well-defined modes around an atomic-scale impurity could enable quantum acoustodynamics”, this reads as if quantum acoustodynamics is a field that simply does not exist which is not true (for example [Science 349, 199 (2017)]). What is true is that it opens the field of quantum acoustodynamics to colour centres.

5. Results: “The optical resonator used in this work is single-sided…” There are a few instances in the paper where the reader does not have any knowledge, yet, and is left wondering about how the system is implemented. This is one of them. For a reader, it might be more satisfying to read “The optical resonator used in this work is a 1D photonic crystal resonator…”, of if there was a reference to a figure or a later description.
Another instance is “The resonant frequency of the optical mode can be
frequency-tuned into resonance with the SiV zero-phonon line…”, where the reader is left wondering how the tuning mechanism works. Here, either mentioning gas tuning or referring to the appropriate section would help.

6. Results: During the introduction of the Hamiltonian, it could be mentioned that the photon-phonon interaction is negligible.

7. Results: Some justification of parameters and operating settings are missing. Mainly why a 12 GHz mechanical mode was chosen, or why the authors use a 55° angle for the magnetic field given the strain susceptibility of the SiV spin is magnetic field angle dependent.

8. Results: There seem to be some discrepancies between the \Deltas that are used throughout the text, or at least they need to be properly introduced.
- In the caption for Fig. 2b, \Delta_m should be defined.
- In the caption for Fig. 2b, \Delta_o is defined as (\kappa/2)/(2\pi). Should \kappa here be \kappa_o? If not, what is \kappa here?
- In Fig. 2c, is \Delta_a the same as \Delta_o?
- In Fig. 3a, how is \Delta_a defined?
- In “When the detuning between the spin transition frequency and the mechanical breathing mode frequency $\Delta m = (\omega s - \omega b)/(2\pi)$ approaches zero,…”, the definition of \Delta_m is different to the one in Fig. 2b, isn’t it? Previously, \Delta_m was the detuning between the analyser and the mechanical mode. Also, for simplicity, \omega_b should be replaced with \omega_m here?

9. Results, Fig. 2 caption: Both DUT and RSA are undefined.

10. Results: “The optically detected noise power spectral density (NPSD) of thermal motion of the mechanical breathing mode at 4K is shown in Fig. 2b.” How is the NSPD obtained from the data?

11. There are many instances where the authors refer to a supplementary section. It would be helpful if instead of just writing “(5.6)”, it would read “(see Sec. 5.6)”, to not confuse with other numbered items.

12. Results: Nowhere in the paper is it mentioned what kind of gas is used for the tuning. Which gas is used? Does the gas form a liquid or a solid at these temperatures – there is both talk about condensation, which would indicate a liquid, but also ice, which would indicate a solid. Would that make a difference to the mechanical Q-factor? Previously, both nitrogen and xenon have been used for tuning photonic crystals. Would the authors expect a difference using a noble gas?

13. Results: It would be more correct to refer to the “ALD tuning” method as “ALD + gas tuning”, both in the text as well as in Fig. 3b – legend.

14. Results: The sentence “While the former is commonly used in cryogenic nanophotonic cavity QED experiments [8], the latter technique allowed us to minimize reduction in the Q-factor of the mechanical resonator upon material deposition.” implies that gas tuning is lowering the mechanical Q-factor, but it might be good explicitly mention this here.

15. Results: “We initialise the SiV spin state by pumping on one of the Zeeman split transitions”. Which one?

16. Results: In “After initializing the spin into the $|\uparrow\rangle$ state, the population decay rate $\gamma_s/(2\pi)$ is proportional to the acoustic density of states…” it would probably be good to add “at these temperatures”.

17. Results, Fig. 3a: Can all 4 peaks be labelled?

18. Results, Fig. 3b: Can the authors explain why there is such a large difference between the right and left sides of the peak in Fig. 3b. Also, since there is so much discussion about the linewidths of these peaks, it would be nice to see how the linewidth was extracted (I am assuming it was a fit).

19. Results, Fig. 3c: Can the extracted decay times be added to the panel?

20. Results, Fig. 3a caption: “to estimate the the spin qubit transition”

21. Results, Fig. 3c caption: It would be good here to refer to the markers in Fig. 3b when describing the 0 GHz and -0.1 GHz datasets.

22. Results: “The difference between this maximum value and the experimentally measured value is well explained by the non-ideal conditions in the studied device, including the presence of static strain (5.2) and the typical expected error in SiV placement.” Was another SiV measured in this device? If yes, what is the quantitative difference between them?

23. Results: “…could attain a maximum value of 9 MHz (5.13).” The 9 MHz maximum spin-phonon coupling seems overly optimistic given (according to Section 5.12) it assumes there is vanishing static strain and a static magnetic field oriented perpendicular to the SiV symmetry axis. However, in these operating conditions, the splitting between the two spin sub-levels also vanishes and is ultimately limited by the Jahn-Teller effect. Therefore, it may not be possible to apply a sufficient magnetic field to tune the SiV splitting to the mechanical resonance. Ideally, a more conservative estimate based on experimentally achievable values would be better.

24. Results, Fig. 4a: What is the expected T1 vs B0 dependency? How does mixing with the orbital states enter here? Could the theoretically expected curve be added in the figure? And where does the jump at 22 GHz come from?

25. Sec. 5.1: “…we get a conversion constant of 2.700±0.186 GHz/kG.” Should this value not be 2.8 GHz/kG? How precisely was the sample mounted in the center of the superconducting magnet?

26. Sec. 5.2: “The data in Fig. 6 was taken after the device was cladding in ALD alumina”, “cladding” should be “clad” here.

27. Sec. 5.3: “Further gas condensation was required ... but a comparatively small amount relative to the bare device.” Is there a way of quantifying this?

28. Sec. 5.4, Figure 8: Please define RSA.

29. Sec. 5.4: The etch chemistry for the silicon nitride is not stated. Additionally, the manuscript states, “OMC pattern is transferred into the diamond by oxygen plasma RIE”. Diamond is very difficult to etch in a standard RIE and typically a ICP-RIE is required. Similarly, it is mentioned “create the mask for the deterministic ion implantation process”. Without saying how, I am assuming EBL but this should be clearly stated.

30. Sec. 5.10: “at alignments close the SiV axis” should be “at alignments close to the SiV axis”.

31. Sec. 5.12: “…and the components of the static strain tensor ϵ are in the coordinate frame of the SiV center.” Equation 6 does not make sense unless this is the single phonon strain tensor of the resonator. Similarly, the stated resonance condition “\omega_s = \omega_m * g_{sm}” also does not make sense. Was \omega_s = \omega_m intended. Furthermore, the relation “gamma_s B_x = \omega_s / \sqrt{2}” is only an approximation, since tan55° ~ \sqrt{2}. The approximately equals sign should be used.

Referee #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #5

(Remarks to the Author)
The manuscript entitled “Observation of the acoustic Purcell effect with a color-center and a nanomechanical resonator” reports experimental studies on the coupling between highly confined mechanical modes in a diamond optomechanical crystal and the spin states of a SiV color center. The main result is the demonstration of enhanced spin relaxation rate when the spin transition is tuned to a mechanical resonance, which can be viewed as the acoustic or mechanical version of the well-known Purcell effect.

For a SiV center, the direct spin-phonon coupling depends on the B-field induced spin-state mixing and thus depends on the magnitude of the B-field normal to the SiV axis (B_normal or Bx in the supplement). The strong dependence of the spin relaxation rate on the orientation of the magnetic field (for the 21 GHz mechanical mode) shown in 4b (and in Fig. 15 of the supplement) as well as the related resonance behavior provide a convincing demonstration of the acoustic Purcell effect.

Overall, the experimental realization of the acoustic Purcell effect represents a major milestone in the field of spin-mechanics and is a major step toward coherent coupling between spins and phonons at the level of single phonons and single spins, which can open a new frontier for pursuing spin-mechanics based quantum information processing.

I have a few detailed questions and comments in terms of the interpretation and analysis of the experimental results for the authors to consider:

1) For a SiV center, the spin relaxation can take place through the direct spin-phonon coupling (i.e., first order spin-coupling coupling) as well as high order spin-phonon coupling (such as two-phonon processes). The first order coupling is a resonant or nearly resonant process. The corresponding Purcell effect should thus depend on the resonance condition, as demonstrated in the manuscript for a few select mechanical modes). Furthermore, the corresponding enhanced spin relation rate is expected to be roughly proportional to the square of B_normal (due to B-field induced spin-state mixing). Both of these behaviors are nicely shown in Fig. 15 of the supplement.

For the broadband spectrum shown in Fig. 4a, the spin relaxation rates vary with the B-field (i.e., the spin transition frequency). However, there is no overall strong increase in the spin relaxation rate with increasing B-field. This lack of an overall or global B-field dependence suggests possible contributions of higher order spin-phonon interactions, which have no or only a weak dependence on the B-field induced state mixing. Specifically, the spin relaxation can take place via two-phonon spin-flip transitions [see Phys. Rev. B 111, 035421 (2025)]. The authors should give some consideration on how the nanomechanical system affects these two-phonon spin relaxation processes and whether the broad spectrum shown of Fig. 4a can be better understood by including the relevant two-phonon processes.

By the way, why the theoretical calculations in Fig. 19b do not show stronger increase with increasing spin transition frequency (i.e., the B-field)?

2) The manuscript highlights the acoustic Purcell effect due to the 12 GHz mode. Figure 3b shows the expected resonance behavior. I wonder if the author can also show the dependence on B-normal (which they have shown for the 21 GHz mode). I understood that in this case, the spin transition frequency will vary considerably with the orientation of the B-field. Nevertheless, the dependence on B-normal can still be carried out by suitably adjusting the overall magnitude of the B-field (while varying B-normal).

Assuming that the same SiV is used for the results obtained for resonant coupling to the 12 and 21 GHz modes, it would be nice if the authors can show that the Purcell effects including the dependence on B-normal for the two modes can be described by the same set of SiV parameters.

3) For the sample fabrication, can the authors provide the dosage used for the masked implantation, along with the estimated average number of SiV centers per implantation aperture?

4) The relation between spin transition frequency and B-normal used in section 5.13 (see the paragraph above Eq. 6) is only an approximation. The transition frequency might also depend on the Jahn-Teller coupling parameter as well as the spin-orbit coupling rate. Can the authors provide a reference for the limits of the approximation or provide some discussions on the approximation?

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
In this revised manuscript, the authors reply all my comments point to point.

1) For my previous comments on the acoustic Pucell effects being reported previously. I aggree with the authors's reply that the current study is based on a single spin in single phonon level, while the previouse ones only in classical regime. However, I donnot think that they are completely irrelevant. I strongly suggest the authors to cite and discuss them.

2) In present manuscript, the authors move the Fig. 5 to supplementary material as new Fig. 12. I think that this is because the quaulity of the data is not high enough. I donnot think it is proper.

Overall, I donnot think the revised manuscript could be published in Nature.

Referee #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #3

(Remarks to the Author)
The authors have provided very detailed replies to all questions, and we are very happy with their answers.
The only point of contention is maybe the question from Referee #1 on the error bars in Fig. 5 (now Fig. 12). For statistical errors, the error bars are way too large for the observed distribution of data points and are likely overestimated. It seems the authors are using the optical linewidth as the error bar which does not make sense, since from Fig. 3a, the transitions are spectroscopically resolvable, and the fit error would be a much better estimate.
In conclusion, this manuscript constitutes a major advancement, and we support publication in Nature.

Referee #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #5

(Remarks to the Author)
I thank the authors for their detailed response to my questions and comments. Most of these have been adequately addressed. The remaining one is with regard to the B_normal dependence of the acoustic Purcell effect for the 12 GHz mode.

The main reason I have asked for this additional data is that Fig. 19 shows good agreements between the experiment and simulation for the 12 GHz mode but does not show the strong Purcell effect observed for the 21 GHz mode. On the other hand, the B_normal dependence (see Fig. 13), which I consider as the most conclusive data for the acoustic Purcell effect, is demonstrated for the 21 GHz mode, but not for the 12 GHz mode.

I understood that the B_normal dependence is much harder to obtain for the 12 GHz mode than that for the 21 GHz mode and that the leading author has since left the group. In this context, I wonder if the authors can do the following:

1) Revise/expand Fig. 19 to reproduce the strong Purcell effect observed for the 21 GHz mode in the simulation.
2) Include experimental data on the characterization of the 21 GHz mechanical mode as well as the 12 GHz mechanical mode in the supplement.

I would be glad to recommend the publication of the manuscript in Nature after the suggested revisions.

Version 2:

Reviewer comments:

Referee #3

(Remarks to the Author)
The thank the authors for considering all of the reviewers' questions and comments and providing detailed replies. We are satisfied on all points and recommend publication.

Referee #4

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #5

(Remarks to the Author)

The authors present one additional data point related to the 12 GHz mode. Figure 14 now compares the data obtained when the B field is 90 and 55 degrees away from the SiV axis. The new data point provides additional supporting evidence, though it does not have the systematic and conclusive nature of Fig. 13, as I pointed out earlier.

The authors noted in their response the absence of direct characterization data for the 21 GHz mode and the lack of clear confirmation of the observed enhancement for the 21 GHz mode in the numerical simulation due to various complications.

Overall, I would like to give the authors the benefit of the doubt and believe that their interpretation of the experimental data is likely to be correct.

Response Letter to Referee Comments

Text Format: Original comments made by referees are shown in black text, our response is given in blue text, and any quotations and/or changes made to our manuscript are shown in red text.

Any citations made in blue refer to the bibliography of this document, any citations made in red refer to the bibliography of the manuscript. Please also note that we have changed the ordering of sections in the Supplementary Information (SI) to reflect a more logical order. Any changes to section or figure numbering in the referee comments have been noted in blue, all section and figure numberings in our responses and quotations use the ordering of the current manuscript version.

Referees' comments:

Referee #1 (Remarks to the Author):

In this manuscript, the authors present a study based on a specially engineered, microwave-frequency nanomechanical resonator, and they observe the acoustic Purcell effect between SiV spin qubits and the acoustic mode of the nanomechanical resonator. Furthermore, utilizing this system, the phonon spectra of the nanomechanical resonator across a broad frequency range of 9–27 GHz have been measured using SiV center spin. This marks the first observation of the acoustic Purcell effect in single-phonon level spectroscopy, with the highest spin-phonon cooperativity reported to date. Additionally, the nanomechanical resonator structure developed in this manuscript introduces a new regime of control for quantum defects in solids.

Overall, the manuscript presents robust results and offers a detailed explanation of the methodology. However, concerns arise regarding the quality of the data and the treatment of uncertainties, which require further clarification. Additionally, the manuscript does not cite some previous studies on the acoustic Purcell effect, which is essential for the current work. Most importantly, the originality and significance of the article are insufficient to publication in NATURE. Therefore, we conclude that this manuscript is not currently suitable for publication in NATURE. Specific comments are provided below:

We thank the reviewer for their careful evaluation and appreciate their acknowledgment of the robustness of our results and methodology. We regret to hear that they do not find the original manuscript suitable for publication in Nature. While we respectfully disagree with the assessment that the original manuscript lacked sufficient novelty for publication in *Nature*, we believe that the revised version - substantially strengthened by the constructive feedback from all reviewers - rises well above this threshold for publication. We address each concern in detail below.

The nanomechanical resonator proposed in this manuscript has been previously described in the author's earlier publications (Nano Letters 2024, 24(23), 6831-6837 and Nature Physics 2025, 21, 77-82).

We thank the reviewer for this comment, but we respectfully disagree that the current manuscript lacks sufficient novelty. We would like to clarify that the novelty of this manuscript lies in the *observation of the acoustic Purcell effect with a **spin qubit** using a nanomechanical resonator*—a regime not previously explored.

Our prior work in Nature Physics 2025, 21, 77-82 focused on SiV centers embedded in 1D phononic crystals with a bandgap commensurate with the ***orbital*** ground state splitting of the SiV center, as is mentioned in our introduction.

"... and control of phonon-induced orbital relaxation processes using the bandgap of phononic crystals [38]."

That study demonstrated ***orbital*** lifetime enhancement due an engineered phononic bandgap; however, ***it did not involve spin-related*** physics or acoustic cavity resonances. In contrast to the orbital degrees of freedom, **spins in diamond are highly technologically and scientifically relevant** as leading solid state quantum memory platforms. In fact, for this reason, in our Nature Physics work we tried very hard to observe spin-phonon physics but were ultimately unsuccessful due to the absence of nanophotonic cavities coupled to the SiV centers. Our current work shifts focus from the orbital degree of freedom to the much more useful spin degree of freedom, and demonstrates for the first time its coupling to a discrete mechanical mode. This is a fundamentally distinct, more experimentally demanding, and significantly more scientifically and technologically relevant study as solid-state spins have emerged as leading quantum memory platforms.

Although a similar experimental platform is described in our work in Nano Letters 2024, 24(23), 6831-6837, that study only achieved independent characterization of mechanical modes and intra-cavity SiV centers. Crucially, **no spin-mechanical coupling was observed**, primarily due to experimental limitations stemming from optical heating in the earlier device and the experiment's architecture. Specifically, in that work, mechanical modes could be brought into resonance with spin, but the optical mode was highly off-resonant from the optical transition (chosen for experimental simplicity), which led to a highly inefficient spin-photon interface with experimentally debilitating excess heating. In fact, to be blunt, our NanoLetters work is in a way a "negative result" which informs the community about various issues one needs to overcome before observing spin-phonon coupling, and thus paves the way to our current work. To clarify the novelty of this work relative to our previous work in Nano Letters 2024, 24(23), 6831-6837 we have modified the following sentence to read:

"This design provides a minimally invasive window to probe the spin using single-photon-level laser pulses, thereby mitigating the laser-induced heating of the acoustic environment at cryogenic temperatures [22] **which limited our previous work** [40]."

The present manuscript introduces several key innovations that enabled direct observation of the acoustic Purcell effect in a spin-defect system - a regime not previously explored. By engineering the optical resonator frequency to closely match the SiV center's optical transition and by employing gas and ALD-based frequency tuning techniques, we established a highly efficient spin-photon interface. This allowed us to probe spin relaxation at ultra-low input powers in the picowatt range, allowing comprehensive spin-relaxation measurements at millikelvin device temperatures - previously unattainable in our system.

Furthermore, the acoustic Purcell effect has also been reported in Phys. Rev. Lett. 120, 114301, and FUNDAMENTAL RESEARCH 4(1), pp. 57-62. These literature indicate that the current manuscript may lack sufficient novelty.

We thank the referee for pointing these out to us, but we find these **completely irrelevant for our work.** We note that those studies report acoustic Purcell-like effects in ***classical***, ***macroscopic resonator systems*** (in one of them they use a speaker to excite acoustic modes) as opposed to ***the quantum nanomechanical system presented here.*** Given these **fundamental differences**, we believe these works are not directly relevant to the quantum spin-phonon interface we demonstrate.

To emphasize again, our manuscript marks the ***first demonstration of direct spin-phonon coupling on the single phonon level, a crucial regime for the realization of emergent quantum phononic networks.***

Furthermore, the presence of a high-efficiency optical readout channel introduces a novel mechanism for probing cavity quantum acoustodynamics in solid-state quantum systems—one that may enable a new generation of hybrid quantum technologies.

We appreciate the reviewer's thoughtful comments and have revised the manuscript to more clearly articulate the unique contributions and scope of our work. We believe this study represents a meaningful advance in the field of engineered spin-phonon quantum systems.

The linear term g_sm (bˆσˆ_++bˆ^† σˆ_-) is employed to describe the spin-phonon interaction. However, the nonlinear effects have not been addressed in the manuscript.

We thank the reviewer for raising this point. Our system operates far away from the strong coupling regime as the spin-phonon coupling strength $g_{sm} \ll \omega_s$. In this limit, the rotating-wave approximation (RWA) [1] is well justified, and the nonlinear counter-rotating terms are negligible. As this is a standard and widely accepted approximation in the field, we do not believe it is necessary to explicitly elaborate on it in the manuscript. Additionally, in this work we investigate spin-phonon coupling near the single-phonon level; all nonlinear effects of the mechanical resonator are negligible at such low acoustic powers.

Multiple resonance signals are observed in Fig. 4(a) of the manuscript. The authors attribute these to other mechanical modes and provides numerical simulation results in Fig. 19. However,

only the first two resonances in this figure align well with the observations, while the last two resonances are not evident in the simulation results.

We agree that the simulations in Fig. 19 do not fully reproduce all resonance features observed in Fig. 4(a), particularly at higher frequencies. As shown in Fig. 19, the simulated spectrum is highly sensitive to the precise spatial position of the SiV center. While this sensitivity could, in principle, be used to determine the precise SiV location, in practice, the complexity of the spectrum and the difficulty of modeling real-world imperfections—such as Q-factor degradation from ALD alumina and deposited gas (as shown in Fig. 3(b) and Fig. 7)—limit the accuracy of such simulations.

Despite these challenges, we emphasize the strong and consistent agreement of the simulated 12 GHz mode with both optomechanical spectroscopy and spin-decay measurement. This consistency supports the integrity of our interpretation. Furthermore, many qualitative features of the spin decay spectrum are reproduced in the simulations.

To address Reviewer's comments we have modified the following sentences in the Simulation of Spin-Phonon Coupling to the Broadband Acoustic Density of states section (SI section 5.14) of our Supplementary Material to further emphasize the challenges associated with modeling our system, as well as similarities and differences between theory and experiments:

The highly local nature of spin-phonon coupling poses a modelling challenge due to the uncertain position of the SiV center within the implantation aperture, which is 50 nm wide in both the x and y directions. As seen in Fig. 19, even a few nanometer change of the in-plane location of the SiV center results in qualitatively distinct spin decay spectra. **While this sensitivity could, in principle, be used to determine the precise SiV location, in practice, the complexity of the spectrum and the difficulty of modeling real-world imperfections—such as the aforementioned Q-factor degradation** from ALD alumina and deposited gas (as shown in Fig. 3(b) and Fig. 11)—limit the accuracy of such simulations. Nonetheless, many of the experimentally observed features are seen in our simulations. In particular, a resonance in the spin decay rate due to the 12 GHz breathing mode of the device can be observed consistently. **The agreement of these simulations with both the spin-decay spectroscopy and optomechanical spectroscopy measurements on the 12 GHz breathing mode support the validity of this simulation approach.**"

In Fig.5 (now Fig. 12), the presence of excessively large error bars undermines the validity of the author's conclusion that the transition frequency varies linearly with the magnetic field.

We appreciate the reviewer's concern. The linear dependence of the SiV spin transition frequency on magnetic field strength is well established in the literature [2], [3], [4]. In our work, we use the vector magnetic field magnitude as a precise but limited accuracy measurement of the SiV center's spin transition frequency. The reproducibility of sharp, distinct spin-resonance features as a function of vector magnetic field magnitude provides strong evidence of this precise control.

The dominant uncertainty in our experiment is in the conversion from field magnitude to spin transition frequency due to the reliance on the somewhat imprecise optical detuning calibration measurements of Purcell-enhanced optical transitions. The error bars in the left panel of Fig. 5 of the original manuscript (now Fig. 12) represented the FWHM linewidths of these transitions. This is not the conventional definition of uncertainty and is perhaps misleading. We have chosen to remove the error bars from the left panel because of this. However, the FWHM optical linewidth does accurately estimate at least an upper bound on the uncertainty in the conversion from optical detuning to spin transition frequency; we have chosen to leave the right panel of Fig. 12 as it was in the original version of the manuscript. The referee states that this undermines the claim that the spin transition frequency varies linearly with magnetic field magnitude. We do not believe that this is the case due to the numerous thorough studies of this using much more accurate transition frequency measurement techniques like coherent population trapping [5] (possible but technically challenging in our experimental system) and optically detected magnetic resonance (ODMR) measurements [4] (possible in our system if microwave striplines are integrated with the device). The large uncertainty in our measurement is simply a limitation of the measurement technique that we used.

This uncertainty causes difficulties in comparing measurements under different experimental conditions (magnetic field orientation and static strain conditions) and in accurately reporting the spin transition frequency. In this work, we use optical detuning calibrations as a coarse spin-transition frequency measurement, and then use sharp identifiable features in the observed spectrum to line up and compare measurements under different conditions, as shown clearly in Fig. 11 for the broad spectrum and Fig. 13 for a narrow feature and multiple magnetic field orientations. Additionally, given the alignment in frequency between the ~12 GHz peak in the spin decay spectrum and the optomechanical spectrum, and the high degree of accuracy and precision in the optomechanical frequency measurement, which is performed on a real-time spectrum analyzer, we have a high degree of confidence in the reported spin transition frequencies despite the perhaps surprisingly large error bars in Fig. 12.

To address Reviewer concerns we have included following text to the Calibration of Spin Transition Frequency section of our supplementary material:

“Due to the wide optical linewidths, the uncertainty in this conversion constant is somewhat large, reflecting the high precision but limited accuracy of the magnetic field magnitude as an estimator of the spin transition frequency. Here, we use these calibrations as a coarse spin-transition frequency measurement and then use sharp identifiable features in the observed spectra to line up and compare measurements under different conditions. Observation of the ~12 GHz mode (whose frequency is known to high precision from optomechanical measurements) then offers a useful reference to identify the frequency of other features with high confidence.”

As previously noted, the acoustic Purcell effect has also been reported in Phys. Rev. Lett. 120, 114301 (2018), and FUNDAMENTAL RESEARCH 4(1), pp. 57-62（2024）. These references, however, are not cited within this manuscript.

As we explained in response to the reviewer's first comments, ***these papers are not relevant to our work at all***, and we have therefore chosen not to cite them in our revised manuscript.

In conclusion, this work makes valuable contributions to hybrid quantum systems by establishing an acoustic interface with record performance metrics. However, the current evidence for paradigm-shifting advancement over existing approaches remains incomplete. With comprehensive revisions addressing the above concerns, this study would constitute a strong candidate for high-impact publication in the journal such as Nature Communications.

We thank the reviewer for the words of praise, but we respectfully disagree with the conclusion. We reiterate that this work demonstrates, for the first time, an **acoustic interface between a spin qubit and a nanomechanical resonator at the single phonon level** in a fully nanoscopic system, with record spin-phonon cooperativity. These results represent a meaningful advancement in spin-mechanical quantum systems and open new avenues for hybrid quantum technologies. We believe that our revised manuscript warrants consideration for publication in *Nature*.

Referee #2 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

We thank the reviewer for all of the questions and comments. Anything raised by this reviewer in the listed reports is addressed in that report.

Referee #3 (Remarks to the Author):

The manuscript “Observation of the acoustic Purcell effect with a color-center and a nanomechanical resonator” by Graham Joe, et al., demonstrates (for the first time, as far as we can tell), the Purcell-enhanced decay of a single atomic defect via the increased acoustic density of states of a phononic microcavity. The defect under consideration here is the SiV in diamond, and the enhanced decay is that of the electron spin in the ground state. This constitutes a seminal demonstration. The results have far-reaching impact and permit development of applications such as transduction for quantum coherent links. Overall, the paper is well-written, the data is convincing, and the supplementary information contains a wealth of background and additional information. We would, however, ask to consider the below comments, which we believe will make the manuscript more approachable

We sincerely thank the reviewer for their thoughtful and thorough evaluation of our manuscript, as well as for the constructive suggestions aimed at improving its clarity and accessibility. ***We are truly encouraged and honored by the reviewer’s assessment of our work as a seminal demonstration in the field.*** To the best of our knowledge, this indeed represents the first observation of Purcell-enhanced decay of a single atomic defect mediated by a nanomechanical resonator. In response to the reviewer’s comments, we have carefully revised the manuscript to enhance its clarity and approachability, and we believe these revisions have further strengthened the presentation of our results.

1. Introduction: In the sentence “The physics of an atom in an open continuum also applies to the emission of acoustic waves from an artificial atom in a solid, and is known to limit spin relaxation times in several color-centers [12–15] and rare-earth ion impurities [16, 17] at low temperatures.”, what exactly is the message? That spin relaxation via phonon emission is the dominant decay process at low temperatures?

We thank the reviewer for this comment, this sentence was intended to convey two distinct points. First, building on the preceding discussion of atoms in an open optical continuum, we are drawing an analogy to acoustic systems - specifically suggesting that an acoustic analogue of the Purcell effect may be observable in solid-state spin defects. The second point is indeed that spin relaxation via phonon emission has been shown to be the dominant decay mechanism at low temperatures for many systems including color centers and rare-earth ion impurities.

To clarify the message in the manuscript, we have revised the sentence to read:

“The physics of an atom in an open continuum also applies to the emission of acoustic waves from an artificial atom in a solid**. Such a process of phonon emission is particularly prominent in** several color-centers [12–15] and rare-earth ion impurities [16, 17] **where it limits spin relaxation times** at low temperatures.”

2. Introduction: In “and control of phonon-induced orbital relaxation processes using phononic crystals [36].”, it might be good to mention that the suppression of the phonon-induced orbital relaxation processes is due the phononic crystal bandgap.

We thank the reviewer for this suggestion, we have revised this sentence to read:

“... and control of phonon-induced orbital relaxation processes using **the bandgap of** phononic crystals [38].”

3. Introduction: In “A central challenge in this context has been to realize an experimental system that provides high spin-phonon co-operativity, …”, it might worth replacing “experimental system” with “resonator-defect system” or “resonator-emitter system”.

We thank the reviewer for this suggestion, we have revised this sentence to read:

“A central challenge in this context has been to realize ~~an experimental~~ **a resonator-defect** system that provides high spin-phonon co-operativity, with the threshold for quantum-coherent interactions being unity.”

4. Introduction: “Structuring a solid and crafting an acoustic environment with well-defined modes around an atomic-scale impurity could enable quantum acoustodynamics”, this reads as if quantum acoustodynamics is a field that simply does not exist which is not true (for example [Science 349, 199 (2017)]). What is true is that it opens the field of quantum acoustodynamics to colour centres.

We thank the reviewer for this comment. We certainly did not mean to imply that the field of quantum acoustodynamics does not exist. To address this, we have modified this sentence to read:

“Structuring a solid and crafting an acoustic environment with well-defined modes around **a color center qubit, could endow the field of** quantum acoustodynamics **[18, 19] with atomic-scale, long-lived memories. This situation is** analogous to cavity QED with light and microwaves.”

5. Results: “The optical resonator used in this work is single-sided…” There are a few instances in the paper where the reader does not have any knowledge, yet, and is left wondering about how the system is implemented. This is one of them. For a reader, it might be more satisfying to read “The optical resonator used in this work is a 1D photonic crystal resonator…”, of if there was a reference to a figure or a later description.
Another instance is “The resonant frequency of the optical mode can be frequency-tuned into resonance with the SiV zero-phonon line…”, where the reader is left wondering how the tuning mechanism works. Here, either mentioning gas tuning or referring to the appropriate section would help.

We thank the reviewer for these comments. We have modified the first sentence in accordance with the reviewer's suggestion to read:

"The optical resonator used in this work is **a** single-sided **1D photonic crystal cavity**: the mirror on the left of Fig. 1a is less reflective than the mirror on the right such that light can be coupled in and out of the cavity via an optical waveguide mode to maximize photon collection efficiency."

We note that there is already a reference to Fig. 1a in this sentence.

For the second sentence we have chosen to add a reference to the appropriate "Frequency Tuning" section of the SI so as not to interrupt the flow of the sentence, which is speaking in generality about frequency tuning. The sentence now reads as follows:

"The resonant frequency of the optical mode is tuned into resonance with the SiV zero-phonon line (ZPL) at 737 nm (406 THz) using a gas tuning technique **(see SI section 5.7),** giving rise to an efficient spin-photon interface."

Elsewhere throughout the manuscript, we have attempted to implement the reviewer's suggestion by adding references to the appropriate figure or SI section, wherever relevant.

6. Results: During the introduction of the Hamiltonian, it could be mentioned that the photon-phonon interaction is negligible.

We thank the reviewer for this comment, we agree this is valuable to mention. We have added the following sentence to the introduction of the optical Hamiltonian:

"In the very low intracavity photon number regime ($n_c \ll 1$) used in this work for all experiments involving the SiV center, the optomechanical photon-phonon interaction is negligible."

7. Results: Some justification of parameters and operating settings are missing. Mainly why a 12 GHz mechanical mode was chosen, or why the authors use a 55° angle for the magnetic field given the strain susceptibility of the SiV spin is magnetic field angle dependent.

We thank the reviewer for highlighting the need for justification of the study of the 12 GHz mechanical breathing mode and the choice of a 55 degree angle for the magnetic field. The 12 GHz mechanical breathing mode was selected because it exhibits the strongest optomechanical coupling among the available mechanical modes (see SI section 5.2). This enables its characterization independent of its coupling to the SiV center, as noted in the Introduction.

Regarding the 55° magnetic field angle, this orientation corresponds to the laboratory z-axis, allowing measurements to be performed using a single coil of the superconducting vector magnet for experimental simplicity. While a magnetic field aligned at 90° would maximize the SiV spin decay rate due to strain susceptibility (see the results of Fig. 4b in the main text), the

55° configuration still provides a substantial perpendicular field component (B⊥), making it sufficient for most measurements presented in this manuscript. We note that for different magnetic field angles, the qualitative spectral features remain similar as shown in Fig. 13. To clarify the reason for our choice of magnetic field angle, we have modified the first sentence in our discussion of Fig. 3 to read:

"To define a spin qubit with the SiV center, we apply a magnetic field of 2.5 kG along the z- axis in the lab frame (θ ≈ 55° with respect to the SiV center's high symmetry axis**, chosen for measurement simplicity, see SI section 5.9**) and lift the spin degeneracy."

At the front of the SI section 5.9 we've added the following paragraph for justification:

"Most of the measurements in this work used a magnetic field angle of 55° relative to the high symmetry axis of the SiV center. This orientation corresponds to the laboratory z-axis, allowing measurements to be performed using a single coil of the superconducting vector magnet for experimental simplicity. Although a magnetic field aligned at 90° would maximize the SiV spin decay rate due to strain susceptibility (see the results of Fig. 4b in the main text), the 55° configuration still provides a substantial perpendicular field component (B⊥), making it sufficient for most measurements presented in this manuscript."

8. Results: There seem to be some discrepancies between the \Deltas that are used throughout the text, or at least they need to be properly introduced.
- In the caption for Fig. 2b, \Delta_m should be defined.
- In the caption for Fig. 2b, \Delta_o is defined as (\kappa/2)/(2\pi). Should \kappa here be \kappa_o? If not, what is \kappa here?
- In Fig. 2c, is \Delta_a the same as \Delta_o?
- In Fig. 3a, how is \Delta_a defined?
- In "When the detuning between the spin transition frequency and the mechanical breathing mode frequency Δm = (ωs−ωb)/(2π) approaches zero,…", the definition of \Delta_m is different to the one in Fig. 2b, isn't it? Previously, \Delta_m was the detuning between the analyser and the mechanical mode. Also, for simplicity, \omega_b should be replaced with \omega_m here?

We thank the reviewer for these astute observations. We have modified the caption of Fig. 2b to define Δ_m as $\Delta_m=(\omega_a-\omega_m)/(2\pi)$.

Also in the caption of Fig. 2b, κ should in fact be κ_o, we have modified the caption to read as such.

In the caption of Fig. 2c we have defined Δ_a as $\Delta_a=(\omega_l-\omega_a)/(2\pi)$ and defined $\omega_a/(2\pi)$ as the zero magnetic field C-line transition frequency.

In the case where the optical mode is tuned into resonance with the C-line transition frequency, Δ_a and Δ_o would have the same meaning. However, since the mechanical mode can only be measured in the untuned case, between Fig. 2b and Fig. 2c, Δ_o and Δ_a would technically have

different meanings. For precision, we believe that it is important to separately use Δ_o and Δ_a where appropriate.

Indeed Δ_m in the sentence:

"When the detuning between the spin transition frequency and the mechanical breathing mode frequency $\Delta_m = (\omega_s - \omega_b)/(2\pi)$ approaches zero,…"

In the original manuscript is a slightly different quantity than the Δ_m previously defined in the caption of Fig. 2b, to distinguish them we have replaced the symbol in the above sentence with $\Delta_{m,s}$. The Δ_m in the caption of Fig. 3b and in the x-axis of Fig. 3b have also been changed to $\Delta_{m,s}$ in accordance with this.

We agree with the reviewer that ω_b should be replaced with ω_m for simplicity. We have replaced ω_b with ω_m in the main text.

9. Results, Fig. 2 caption: Both DUT and RSA are undefined.

We thank the reviewer again for the astute observations, we have defined DUT as the device under test and RSA as real-time spectrum analyzer in the caption of Fig. 2a.

10. Results: "The optically detected noise power spectral density (NPSD) of thermal motion of the mechanical breathing mode at 4K is shown in Fig. 2b." How is the NSPD obtained from the data?

We thank the reviewer for requesting clarification on how the noise power spectral density (NPSD) is extracted.

In our setup, a high-bandwidth photodetector converts the optical fluctuations induced by the mechanical breathing mode into a voltage signal, which is sent directly to a real-time spectrum analyzer (RSA). The NPSD is directly proportional to the direct output of the real time spectrum analyzer connected.

Technically the direct output is the power spectrum which must then be divided by the resolution bandwidth to obtain the power spectral density. In this case, it is a noise power spectral density since we are analyzing a noise signal. This is a widely used technique in the field of optomechanics, the details are well described in *Aspelmeyer, et al. "Cavity optomechanics." Reviews of Modern Physics 86(4), 1391–1452 (2014)* which is cited in the sentence before the one identified by the reviewer.

To further emphasize that the NPSD is very closely related to the direct output of the RSA, we have modified the sentence in question to read:

“The optically detected noise power spectral density (NPSD) of thermal motion of the mechanical breathing mode at 4K**, recorded from a real-time spectrum analyzer (RSA),** is shown in Fig. 2b”

11. There are many instances where the authors refer to a supplementary section. It would be helpful if instead of just writing “(5.6)”, it would read “(see Sec. 5.6)”, to not confuse with other numbered items.

We thank the reviewer for this helpful suggestion, we have decided to use the wording “see SI section x.x” in all such instances.

12. Results: Nowhere in the paper is it mentioned what kind of gas is used for the tuning. Which gas is used? Does the gas form a liquid or a solid at these temperatures – there is both talk about condensation, which would indicate a liquid, but also ice, which would indicate a solid. Would that make a difference to the mechanical Q-factor? Previously, both nitrogen and xenon have been used for tuning photonic crystals. Would the authors expect a difference using a noble gas?

We thank the reviewer for the insightful questions regarding our gas based tuning approach.

As noted in section 5.2 of the SI, we observed that once the gas tube reaches a temperature in the range of 35-45K, the optical resonance of our device begins to red-shift, indicating gas deposition - even without the introduction of gas through the nozzle.

In our cryogenic gas tuning setup, we expect that frozen air on the gas tube is released when it is heated. We believe that the composition of the deposited gas is predominantly nitrogen since that is the major component of air with the lowest sublimation point at high vacuum (~22 K) [6].

To reflect this discussion, we have revised the following sentences in the Gas Tuning and ALD Cladding section (SI section 5.7) of our supplementary material:

“We observed that once the gas tube reaches a temperature in the range of 35-45K **at the heater (around the 4K stage)**, the optical resonance of our device begins to red-shift, indicating gas ~~condensation~~ **deposition. The material deposited is most likely Nitrogen, the component of air with the lowest sublimation temperature at high vacuum (~22 K [65], likely the temperature of the gas tube near the sample).**”

At our operating temperatures of 4K and below, the gas is expected to freeze solid. The referee is correct that condensation is not the correct term to use here, condensation is a phase transition from a liquid to a solid. The correct term for the process occurring here is deposition, meaning a phase transition from gas to solid. We thank the reviewer for pointing out this error and we have replaced all usage of “condensation” in the manuscript with “deposition”.

Regarding its effect on mechanical Q-factor, we hypothesize that deposited gases act as acoustic loss channels due to their relatively low speed of sound relative to diamond, as we discussed in this sentence of our manuscript:

"Since the ~~ice~~ **solid** deposited on the structure due to gas ~~condensation~~ **deposition** is expected to have significantly lower speed of sound relative to diamond [51], we expect gas ~~condensation~~ **deposition** to increase **the** acoustic radiative loss from the mechanical resonator into the bulk substrate, introducing excess damping."

Since diamond has an exceptionally high sound velocity, we expect that deposited inert gases (including nitrogen and noble gases like xenon) are generally detrimental to mechanical Q due to enhanced phonon leakage, but we have not tried using other types of gases and cannot comment on how they would compare with nitrogen.

While gas tuning was an enabling technique for our current experiment, future experiments aiming to preserve high mechanical Q factors would benefit from techniques like electrostatically-tunable cavities, as noted in our conclusion.

13. Results: It would be more correct to refer to the "ALD tuning" method as "ALD + gas tuning", both in the text as well as in Fig. 3b – legend.

We thank the reviewer for this comment. When referencing the spin decay data, we agree that it is more correct to refer to the tuning method as "ALD + gas tuning" rather than just "ALD tuning". Throughout this manuscript, we were generally very careful to refer specifically to "ALD tuning/cladding" or "ALD with minimal gas deposition" as was appropriate to the specific measurement under discussion. We've identified three locations in the manuscript where this can be clarified.

In the results section in the discussion of Fig. 2, we've revised this sentence to read:

"To compensate for the frequency difference between the SiV-center's optical transition and the optical cavity mode, which is caused by finite fabrication tolerances **(see SI section 5.1)**, we investigated two methods for cavity frequency tuning: in situ gas ~~condensation~~ **deposition** frequency tuning (gas tuning), and atomic layer deposition (ALD) of alumina **with minimal additional gas deposition (see SI section 5.7).**"

We have revised the caption of Fig. 3c to read:

"Spin population decay curves for the ALD**+gas**-tuned device with the spin transition frequency $\omega_s/(2\pi)$ is at a detuning $\Delta_{m,s}$ from the ~12 GHz mechanical mode …"

In the conclusion of our text we have modified the following sentence to read:

"In the present work, the cooperativity is primarily limited by excess mechanical damping (increased κ_m) from the material deposition (gas ~~condensation~~ **deposition** or ALD alumina **with minimal gas deposition**) required to tune the co-localized optical mode into resonance with the SiV."

14. Results: The sentence "While the former is commonly used in cryogenic nanophotonic cavity QED experiments [8], the latter technique allowed us to minimize reduction in the Q-factor of the mechanical resonator upon material deposition." implies that gas tuning is lowering the mechanical Q-factor, but it might be good explicitly mention this here.

We thank the reviewer for this suggestion, we have modified this sentence to explicitly state that gas deposition reduces the mechanical quality factor, it now reads:

"~~While~~ **Although** the former is commonly used in cryogenic nanophotonic cavity QED experiments [8], ~~the latter technique allowed us to minimize reduction in the~~ **it was observed to reduce the mechanical** Q-factor of the mechanical resonator upon material deposition. The latter technique allowed us to minimize this effect."

15. Results: "We initialise the SiV spin state by pumping on one of the Zeeman split transitions". Which one?

We thank the reviewer for this request for further specificity, we have modified this sentence to address this concern. This sentence now reads:

"We initialize the SiV spin state by optically pumping on ~~one of~~ the Zeeman-split transitions **using the repump and pump pulses shown in the inset of Fig. 3**"

16. Results: In "After initializing the spin into the $|\uparrow\rangle$ state, the population decay rate $\gamma s/(2\pi)$ is proportional to the acoustic density of states…" it would probably be good to add "at these temperatures".

We thank the reviewer for this comment, we have modified this sentence to read:

"**At millikelvin temperatures,** after initializing the spin into the $|\uparrow\rangle$ state, the population decay rate $\gamma_s/(2\pi)$ is proportional to the acoustic density of states of the surrounding structure at $\omega_s/(2\pi)$."

17. Results, Fig. 3a: Can all 4 peaks be labelled?

We thank the reviewer for this suggestion, we have modified Fig. 3 to label all four peaks.

18. Results, Fig. 3b: Can the authors explain why there is such a large difference between the right and left sides of the peak in Fig. 3b. Also, since there is so much discussion about the

linewidths of these peaks, it would be nice to see how the linewidth was extracted (I am assuming it was a fit).

We appreciate the reviewer's observation regarding the peak asymmetry in Fig. 3b. We attribute this asymmetry to the presence of other mechanical modes near 12 GHz in frequency, which are expected due to the absence of a complete phononic bandgap in this frequency range. A higher-Q resonance would naturally result in a more symmetric peak.

Regarding the linewidth extraction, this was indeed performed via fitting. Due to the aforementioned asymmetry, we performed the fit on a truncated portion of the data - excluding part of the peak shoulder - to yield a representative estimate of the linewidth. For the ALD tuned case, a modified background with a linear shape on the right hand side of the peak was required to fit the observed asymmetric profile. The fit models used to extract the peak linewidths are now shown in the revised Fig. 3b panel. We have also added the following discussion of the linewidth extraction process to the caption of Fig. 3b:

"(b) Resonance in the spin decay rate measured when the spin qubit frequency, $\omega_s/(2\pi)$ is tuned across the ~12 GHz breathing mode due to the spin-phonon interaction (inset). The data is taken with a $\theta \approx 55^\circ$ magnetic field in two separate runs where the optical cavity is tuned into resonance with SiV center's optical transition using the gas tuning method (gray data points**, fit is performed to a truncated data set excluding the non-Lorentzian shape of the right-hand side of the peak**), and a combination of ALD and gas tuning (purple data points**, fit is performed to a truncated data set and uses a heuristic piecewise linear model of the background**), resulting in a fit resonance linewidth of ~200 MHz and ~35 MHz, respectively"

19. Results, Fig. 3c: Can the extracted decay times be added to the panel?

We thank the reviewer for this comment, we have revised the Fig. 3c panel to include the extracted decay rates.

20. Results, Fig. 3a caption: "to estimate the the spin qubit transition"

We thank the reviewer for catching this error, we have changed this part of the Fig. 3a caption to:

"To estimate the spin qubit transition"

21. Results, Fig. 3c caption: It would be good here to refer to the markers in Fig. 3b when describing the 0 GHz and -0.1 GHz datasets.

We thank the reviewer for this comment, we agree that referring to the marker locations in Fig. 3b would enhance the clarity of the caption of Fig. 3c. We have revised this caption to read:

"Spin population decay curves for the ALD**+gas**-tuned device with the spin transition frequency $\omega_s/(2\pi)$ is at a detuning, $\Delta_{m,s}$ from the ~12 GHz mechanical mode of 0 GHz (solid line**, corresponding to the peak in panel b**) and -0.1 GHz (dashed line**, corresponding to the baseline in panel b**)."

22. Results: "The difference between this maximum value and the experimentally measured value is well explained by the non-ideal conditions in the studied device, including the presence of static strain (5.2, now SI section 5.13) and the typical expected error in SiV placement." Was another SiV measured in this device? If yes, what is the quantitative difference between them?

We appreciate the reviewer's request for further details. In the specific device examined in this study, only a single SiV center was observed and characterized. Generally, distinct SiV centers exhibit spectral lines at unique frequencies due to local variations in crystal strain. Importantly, the sentence in question compares the theoretically calculated maximum possible spin-phonon coupling rate with the experimentally measured value for this particular SiV. It does not involve a comparison between two independently measured SiV centers.

23. Results: "…could attain a maximum value of 9 MHz (5.12)." The 9 MHz maximum spin-phonon coupling seems overly optimistic given (according to the SI section 5.12) it assumes there is vanishing static strain and a static magnetic field oriented perpendicular to the SiV symmetry axis. However, in these operating conditions, the splitting between the two spin sub-levels also vanishes and is ultimately limited by the Jahn-Teller effect. Therefore, it may not be possible to apply a sufficient magnetic field to tune the SiV splitting to the mechanical resonance. Ideally, a more conservative estimate based on experimentally achievable values would be better.

We appreciate the reviewer's insightful comment regarding the optimistic nature of the 9 MHz spin-phonon coupling estimate in the SI section 5.12. This value assumes the simultaneous realization of two idealized conditions - vanishing static strain and optimal positioning within the nanostructure - which are challenging to achieve concurrently in experimental settings.

While each assumption is individually plausible: SiV centers with near-zero static strain are routinely observed, and favorable positioning within the implantation aperture is reasonably likely. The 9 MHz figure was intended as a theoretical upper bound for the spin phonon coupling rate under idealized assumptions. In that context, we view our experimentally obtained, lower coupling rates to be a realistic and encouraging outcome.

Importantly, the concern about sub-level level degeneracy under zero static strain is mitigated by the presence of spin-orbit coupling, which preserves a finite splitting between sublevels even in the absence of static strain and the Jahn-Teller effect (crystal strain and the Jahn-Teller Effect have the same Hamiltonian making them spectroscopically indistinguishable).

In this idealized scenario, the maximum frequency accessible to our broadband spectroscopy method would be reduced to approximately 25 GHz - still more than sufficient for this experiment.

24. Results, Fig. 4a: What is the expected T1 vs B0 dependency? How does mixing with the orbital states enter here? Could the theoretically expected curve be added in the figure? And where does the jump at 22 GHz come from?

We appreciate the reviewer's thoughtful questions regarding Fig. 4a. The expected dependence of decay rate (1/T1) on spin transition frequency (proportional to B0) is modeled to the best of our ability and described in detail in the SI section 5.14. Good agreement with the measured 12 GHz breathing mode is seen and many other spectral features observed in the experimental data are qualitatively reproduced by our model. Due to the complexity of the modelling and its interpretation, we have chosen to not include the theoretical expectation in Fig. 4, instead we present and discuss it in the SI section 5.14, and reference this in the main text where appropriate.

Regarding mixing with the orbital states, spin-orbit coupling is critical to the spin-phonon coupling mechanism described in this work as spins do not couple directly to phonons. Off-axis magnetic fields create mixtures of orbital components in the spin sublevels, which enables acoustic transitions between spin-sublevels. We believe that this discussion is too detailed and technical to warrant inclusion in our manuscript, it is discussed at length in [3] which we cite frequently throughout the manuscript. On the separate issue of mixing between orbital levels, which the reviewer may be referring to, the magnetic field strengths used in this work are moderate and do not cause significant hybridization between orbital states. At higher frequencies, any potential influence of orbital mixing - through altered strain susceptibility - is convolved with the acoustic density of states, as probed via spin decay spectroscopy.

The sudden increase at 22 GHz is a reproducible feature consistent with a band edge resonance in the underlying acoustic bandstructure. Features like this are frequently seen in other techniques which map for example, photonic or electronic density of states [7], and we do not believe that this is a surprising feature.

25. Sec. 5.1 (now SI section 5.8): "…we get a conversion constant of 2.700±0.186 GHz/kG." Should this value not be 2.8 GHz/kG? How precisely was the sample mounted in the center of the superconducting magnet?

We thank the reviewer for this insightful question. 2.8 GHz/kG is the electron gyromagnetic ratio and is the typical proportionality constant between applied magnetic field and spin transition frequency for systems with simple S*B Hamiltonian's. Due to spin-orbit coupling in the SiV center, there is also an orbital contribution to the Zeeman effect which causes significant deviation of this proportionality constant from 2.8 GHz/kG. At zero strain the spin transition frequency is proportional to the on-axis component of the magnetic field to a first order approximation, while at high strain it is proportional to the total magnetic field magnitude [3], this

SiV under study falls in an intermediate regime where this proportionality constant depends on the orientation of the applied field and can be less than 2.8 GHz/kG. The result presented in the SI section 5.8 is simply an example calibration for a particular magnetic field orientation.

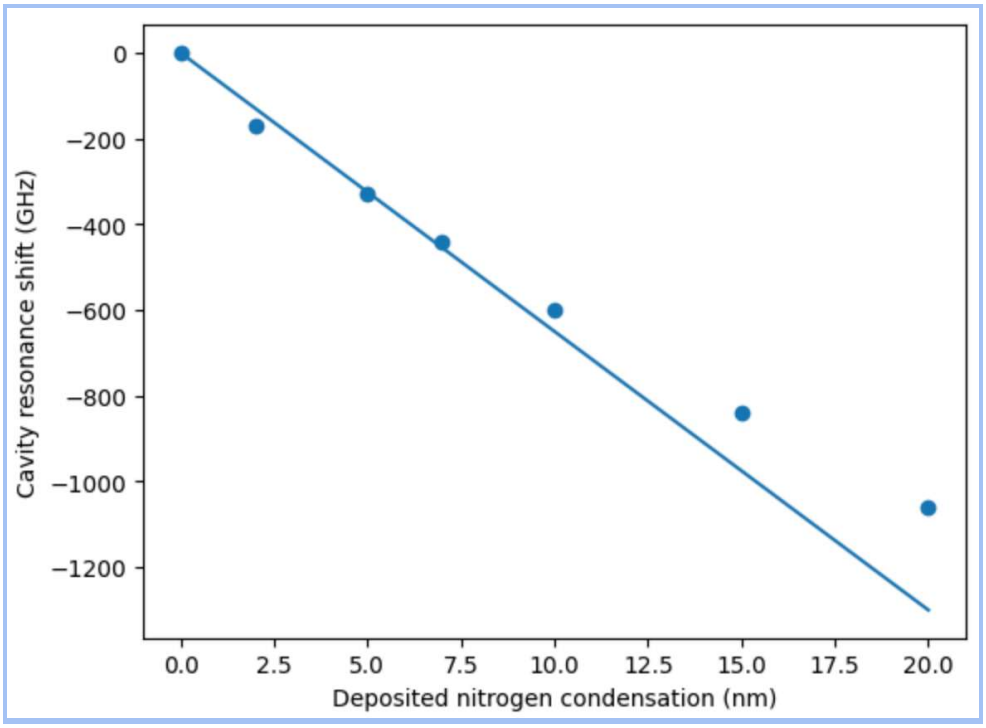
With regards to the precision of sample mounting. The sample mounting structure was designed to position the diamond sample at the center of the magnetic bore, machine precision will naturally cause deviations from this, likely on the millimeter scale. The sample itself is mounted in the center of a small copper piece by hand which could result in positional deviations from the center of the magnet bore by a few millimeters. The uniformity of the magnetic field within the bore for the z-axis coil is very good at ±0.5% over a 1 cm diameter spherical volume; we therefore do not expect significant deviations between the true and nominally applied magnetic field. Additionally, since we perform a calibration of spin transition on the nominally applied magnetic field for any given experimental run, the accuracy of the magnetic field applied to the sample is not particularly consequential to our experimental results.

26. Sec. 5.2 (now SI section 5.13): “The data in Fig. 6 (now Fig. 18) was taken after the device was cladding in ALD alumina”, “cladding” should be “clad” here.

We thank the reviewer for this astute observation, we have made this correction.

27. Sec. 5.3 (now SI section 5.7): “Further gas condensation was required ... but a comparatively small amount relative to the bare device.” Is there a way of quantifying this?

We thank the reviewer for this thoughtful question. The simplest way to quantify this is in terms of the resonance shift required to achieve resonance with the SiV center’s C-line. Without ALD alumina cladding the optical mode had to be frequency tuned by 3 THz with gas deposition. After ALD cladding this was reduced to 200 GHz. Assuming that the deposited gas is solid nitrogen, with a refractive index of 1.25 [6] , we can use FDTD modelling to estimate the resonance frequency shift as a function of the thickness of the deposited gas layer. Our simulations yield a resonance tuning rate of 0.12 nm/nm of deposited gas (65 GHz / nm of deposited gas) in the low nitrogen deposition regime. Therefore the deposited gas layer in the uncladded case is about 45 nm thick, whereas in the ALD alumina clad case it is around 4 nm thick.



To reflect this discussion, we have added the following sentences to the Gas Tuning and ALD Cladding section (SI section 5.7):

"FDTD simulations of our structure yield a frequency tuning rate of 65 GHz/nm of deposited gas, so the thickness of the deposited gas layer used to tune the bare cavity into resonance with the SiV center's C-line is ~45 nm."

28. Sec. 5.4, Figure 8 (now SI section 5.5, Figure 9): Please define RSA.

We thank the reviewer for this astute observation. We have revised the caption of Fig. 9 to define RSA as real-time spectrum analyzer.

29. Sec. 5.5 (now SI section 5.4) : The etch chemistry for the silicon nitride is not stated. Additionally, the manuscript states, "OMC pattern is transferred into the diamond by oxygen plasma RIE". Diamond is very difficult to etch in a standard RIE and typically a ICP-RIE is required. Similarly, it is mentioned "create the mask for the deterministic ion implantation process". Without saying how, I am assuming EBL but this should be clearly stated.

We thank the referee for pointing this out - indeed we should have been more precise. According to the reviewer's comments regarding the gas species for the SiN and the diamond etch processes, we have revised the fabrication process section in the SI section 5.4. The revised sentences are shown below.

"The SiN layer is etched by inductively coupled plasma-reactive ion etching (ICP-RIE) **with sulfur hexafluoride (SF_6) and octafluorocyclobutane (C_4F_8) gases** [(Fig. 8(iii))]. After the EB

resist is removed[(Fig. 8(iv))], the OMC pattern is transferred into the diamond by oxygen plasma**-based ICP**-RIE[(Fig. 8(v))]."

We also have added details of the mask ion implantation technique we used in this work. The revised sentences are shown below.

"To implant SiV centers in the fabricated OMC devices, **we employ a deterministic ion implantation technique. The details of the technique can be found in our previous report [65]. Briefly,** we spin-coat a PMMA layer on the fabricated devices [(Fig. 8(xi))] and create the **implantation mask with apertures (62.5 nm x 62.5 nm) by a combination of EB lithography and resist development. The silicon ions (Si+) are implanted only in the apertures with an energy of 150 keV. The resulting mean ion range was ~70 nm from the surface, as simulated by the software package Stopping and Range of Ions in Matter. We choose a low ion implantation dose (1.0× 10^{12} ions per cm^2) intended to produce roughly 3 SiV centers per aperture on average.** After masked ion implantation [(Fig. 8(xii))], we remove the PMMA layer to complete the fabrication of OMC devices [(Fig. 8(xiii))]. The SiV centers in OMC devices are generated by annealing. ~~The details of the mask ion implantation technique can be found in our previous report[62].~~ We note that our fabricated devices typically have a few SiV centers, we use an OMC device with a single SiV center in this work."

30. Sec. 5.10 (now SI section 5.9): "at alignments close the SiV axis" should be "at alignments close to the SiV axis".

We thank the reviewer for the astute observation, we have made the suggested change to this sentence.

31. Sec. 5.12: "…and the components of the static strain tensor ϵ are in the coordinate frame of the SiV center." Equation 6 does not make sense unless this is the single phonon strain tensor of the resonator. Similarly, the stated resonance condition "\omega_s = \omega_m * g_{sm}" also does not make sense. Was \omega_s = \omega_m intended. Furthermore, the relation "gamma_s B_x = \omega_s / \sqrt{2}" is only an approximation, since tan55° ~ \sqrt{2}. The approximately equals sign should be used.

We thank the reviewer for this question. We guess that the referee is asking about the meaning of the strain tensor components used in equation 6. As is mentioned previously in the same section, these are the components of the static strain tensor in the coordinate frame of the SiV center, not the single phonon strain tensor of the resonator. These static strain components determine the susceptibility of the spin transition to an incoming acoustic wave or phonon.

With regards to the second equation "$\omega_s = \omega_m * g_{sm}$". The interpreted multiplication sign is in fact a period. The following sentence simply begins with g_{sm}. We recognize that this was confusing, we have revised this sentence to begin with "The spin-mechanical coupling rate" rather than g_{sm} for improved clarity,

With regards to the relation "$\gamma_s B_x = \omega_s/\sqrt{2}$", in the case being discussed of a magnetic field applied along the laboratory z-axis, 55 degrees is a decimal approximation for arccos(1/√3), the tangent of this angle is exactly √2. However, this relation assumes that the spin transition is unaffected by the off-axis magnetic field which is only a first order approximation even in the ideal zero static strain case. We have therefore revised this equation to use an approximately equals sign:

$$\gamma_s B_x \approx \omega_s/\sqrt{2}$$

We have also revised equation 6 to use an approximately equals sign, it now reads:

$$g_{SO} \approx \sqrt{2}\omega_m/\lambda_{SO}\sqrt{(d(\epsilon_{XX} - \epsilon_{YY}) + f\epsilon_{YZ})^2 + (-2d\epsilon_{XY} + f\epsilon_{XZ})^2}$$

Referee #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

We thank the reviewer for all of the questions and comments. Anything raised by this reviewer in the listed reports is addressed in that report.

Referee #5 (Remarks to the Author):

The manuscript entitled “Observation of the acoustic Purcell effect with a color-center and a nanomechanical resonator” reports experimental studies on the coupling between highly confined mechanical modes in a diamond optomechanical crystal and the spin states of a SiV color center. The main result is the demonstration of enhanced spin relaxation rate when the spin transition is tuned to a mechanical resonance, which can be viewed as the acoustic or mechanical version of the well-known Purcell effect.

For a SiV center, the direct spin-phonon coupling depends on the B-field induced spin-state mixing and thus depends on the magnitude of the B-field normal to the SiV axis (B_normal or Bx in the supplement). The strong dependence of the spin relaxation rate on the orientation of the magnetic field (for the 21 GHz mechanical mode) shown in 4b (and in Fig. 15 of the supplement, now Fig. 13) as well as the related resonance behavior provide a convincing demonstration of the acoustic Purcell effect.

Overall, the experimental realization of the acoustic Purcell effect represents a major milestone in the field of spin-mechanics and is a major step toward coherent coupling between spins and phonons at the level of single phonons and single spins, which can open a new frontier for pursuing spin-mechanics based quantum information processing.

We sincerely thank the reviewer for their thoughtful evaluation and constructive feedback on our manuscript. We are encouraged by the reviewer’s recognition of our key result and appreciate their acknowledgment of how the supporting data in Figure 4b strengthens our claim. We are especially honored by the reviewer’s characterization of this work as a major milestone in the field of spin-mechanics. In response, we have addressed each of the reviewer’s comments in detail and have revised the manuscript accordingly to strengthen its clarity and approachability.

I have a few detailed questions and comments in terms of the interpretation and analysis of the experimental results for the authors to consider:

1) For a SiV center, the spin relaxation can take place through the direct spin-phonon coupling (i.e., first order spin-coupling coupling) as well as high order spin-phonon coupling (such as two-phonon processes). The first order coupling is a resonant or nearly resonant process. The corresponding Purcell effect should thus depend on the resonance condition, as demonstrated in the manuscript for a few select mechanical modes). Furthermore, the corresponding enhanced spin relation rate is expected to be roughly proportional to the square of B_normal (due to B-field induced spin-state mixing). Both of these behaviors are nicely shown in Fig. 15 (now Fig. 13) of the supplement.

For the broadband spectrum shown in Fig. 4a, the spin relaxation rates vary with the B-field (i.e., the spin transition frequency). However, there is no overall strong increase in the spin relaxation rate with increasing B-field. This lack of an overall or global B-field dependence suggests possible contributions of higher order spin-phonon interactions, which have no or only

a weak dependence on the B-field induced state mixing. Specifically, the spin relaxation can take place via two-phonon spin-flip transitions [see Phys. Rev. B 111, 035421 (2025)]. The authors should give some consideration on how the nanomechanical system affects these two-phonon spin relaxation processes and whether the broad spectrum shown of Fig. 4a can be better understood by including the relevant two-phonon processes.

By the way, why the theoretical calculations in Fig. 19b do not show stronger increase with increasing spin transition frequency (i.e., the B-field)?

We thank the reviewer for this thoughtful question.

The spin decay rate from the spin up state can be expressed as:

$\gamma_s = DOS(\omega_s)*(n_{th} + 1)$

Where $DOS(\omega_s)$ denotes the acoustic density of states at the spin transition frequency, and n_{th} is the thermal phonon occupation number. In a conventional 3D continuum model, under the Wigner-Weisskopf approximation, the density of states scales as ω_s^2, implying a monotonic increase in the decay rate with spin transition frequency, or equivalently, with B-field strength. For a 1D system like the one used here, the overall trend in density of states with increasing frequency is expected to be constant, not increasing. Importantly, in a more complex structured nanomechanical system such as the one investigated in this work, the Wigner Weisskopf approximation does not strictly hold, the DOS must be numerically computed, and sharp features in the profile can be observed due to phononic bandgaps and cavity modes. As shown in the SI section 5.14, the density of states profile exhibits such non-monotonic behavior, consistent with similar well-established results in phononic and photonic crystal systems.

Regarding two-phonon spin-flip processes, we agree that such higher-order mechanisms could play a role in modifying the observed B-field dependence, especially in regimes where first-order spin-phonon coupling is suppressed - though this does not apply to our system. Importantly, nearly all of our spin relaxation measurements are conducted at ultra-low temperatures (millikelvin regime), where $k_B T \ll \hbar\omega_s$. Under these thermal conditions, multi-phonon processes are strongly suppressed and contribute minimally to spin dynamics. The notable exception is described in the SI section 5.11, where elevated temperatures and reduced spin transition frequencies are deliberately chosen to probe thermal effects.

2) The manuscript highlights the acoustic Purcell effect due to the 12 GHz mode. Figure 3b shows the expected resonance behavior. I wonder if the author can also show the dependence on B-normal (which they have shown for the 21 GHz mode). I understood that in this case, the spin transition frequency will vary considerably with the orientation of the B-field. Nevertheless, the dependence on B-normal can still be carried out by suitably adjusting the overall magnitude of the B-field (while varying B-normal).

Assuming that the same SiV is used for the results obtained for resonant coupling to the 12 and 21 GHz modes, it would be nice if the authors can show that the Purcell effects including the dependence on B-normal for the two modes can be described by the same set of SiV parameters.

We appreciate the reviewer's request for further details on the characterization of the 12 GHz mode.

As noted, the magnitude of the magnetic field required to reach a given spin transition frequency is strongly dependent on its orientation relative to the SiV center's high symmetry axis. Nonetheless, the calibration procedures outlined in the SI section 5.8 enable comparison of data across different field orientations for a fixed spin transition frequency range. This methodology has been employed in Fig. 4b and SI section 5.9 for the 21 GHz mode.

For the 21 GHz mode, Fig. 13 demonstrates that the same curve shape (on a logarithmic scale) is observed for both normal and 55-degree field orientations, with amplitude scaled slightly due to the increased strain susceptibility at normal orientations. A similar behavior is expected for the 12 GHz mode. However, under the measurement conditions for Fig. 4b and Fig. 15, data for the 12 GHz mode was acquired only at a 55-degree field alignment. Repeating this experiment to obtain the full angular dependence would be prohibitively challenging at this point in time (leading authors have left the group).

We believe that the essential physics concerning the influence of magnetic field alignment on spin relaxation is adequately captured in Fig. 4b and Supplementary Section 5.9. While a full angular characterization of the 12 GHz mode would be ideal, such measurements are particularly challenging near the SiV high-symmetry axis due to an optical transition crossing which interferes with spin initialization. These issues are significantly reduced at higher frequencies due to the reduced range of orientations at which this crossing is problematic, which is why the 21 GHz mode was selected for detailed angular dependence studies.

3) For the sample fabrication, can the authors provide the dosage used for the masked implantation, along with the estimated average number of SiV centers per implantation aperture?

To address this request from the reviewer, we have added the following sentence to the device fabrication section (SI section 5.4):

"We choose a low ion implantation dose (1×10^{12} ions per cm^2) intended to produce roughly 3 SiV centers per aperture on average."

4) The relation between spin transition frequency and B-normal used in section 5.13 (now SI section 5.12) (see the paragraph above Eq. 6) is only an approximation. The transition frequency might also depend on the Jahn-Teller coupling parameter as well as the spin-orbit coupling rate. Can the authors provide a reference for the limits of the approximation or provide some discussions on the approximation?

We thank the reviewer for the insightful question, this approximate nature of the relationship between the spin transition frequency and the $B_\perp$ (referred to as B_x in the SI section 5.12) was also raised by the previous reviewer (remark number 31) and the dependence on the Jahn-Teller coupling parameter was raised in remark number 23.

Importantly, the concern about sub-level level degeneracy under zero static strain is mitigated by the presence of spin-orbit coupling, which preserves a finite splitting between sublevels even in the absence of static strain and the Jahn-Teller effect (crystal strain and the Jahn-Teller Effect have the same Hamiltonian making them spectroscopically indistinguishable). In this idealized scenario, the maximum frequency accessible to our broadband spectroscopy method would be reduced to approximately 25 GHz - still more than sufficient for this experiment.

With regards to the spin transition frequency's dependence on the Jahn-Teller coupling parameter and the spin-orbit coupling rate, the spin transition frequency is determined by a competition between spin-orbit coupling and the combined effect of static crystal strain and the Jahn-Teller effect. In section 5.12 we are conducting a theoretical analysis to determine a realistic upper bound on the spin-phonon coupling rate which is maximized when the combined effect of static crystal strain and Jahn-Teller coupling is negligible compared to spin-orbit coupling. This assumption is not unrealistic as SiV centers with roughly the minimum possible ground state splitting of ~46 GHz are observed fairly routinely in experiments.

The approximate nature of the relationship in Sec. 5.12 was also addressed in response to the previous reviewer's remark number 31 and is copied below:

With regards to the relation "$\gamma_s B_x = \omega_s/\sqrt{2}$", in the case being discussed of a magnetic field applied along the laboratory z-axis, 55 degrees is a decimal approximation for $\arccos(1/\sqrt{3})$, the tangent of this angle is exactly $\sqrt{2}$. However, this relation assumes that the spin transition is unaffected by the off-axis magnetic field which is only a first order approximation even in the ideal zero static strain case. We have therefore revised this equation to use an approximately equals sign:

$$\gamma_s B_x \approx \omega_s/\sqrt{2}$$

We have also revised equation 6 to use an approximately equals sign, it now reads:

$$g_{SO} \approx \sqrt{2}\omega_m/\lambda_{SO}\sqrt{\left(d(\epsilon_{XX} - \epsilon_{YY}) + f\epsilon_{YZ}\right)^2 + \left(-2d\epsilon_{XY} + f\epsilon_{XZ}\right)^2}$$

[1] G. S. Agarwal, *Quantum Optics*. Cambridge University Press, 2012.
[2] C. Hepp *et al.*, "Electronic Structure of the Silicon Vacancy Color Center in Diamond," *Phys. Rev. Lett.*, vol. 112, no. 3, p. 036405, Jan. 2014, doi: 10.1103/PhysRevLett.112.036405.
[3] S. Meesala *et al.*, "Strain engineering of the silicon-vacancy center in diamond," *Phys. Rev. B*, vol. 97, no. 20, p. 205444, May 2018, doi: 10.1103/PhysRevB.97.205444.
[4] C. T. Nguyen *et al.*, "An integrated nanophotonic quantum register based on silicon-vacancy spins in diamond," *Phys. Rev. B*, vol. 100, no. 16, p. 165428, Oct. 2019, doi: 10.1103/PhysRevB.100.165428.

[5] “All-Optical Formation of Coherent Dark States of Silicon-Vacancy Spins in Diamond | Phys. Rev. Lett.” Accessed: Sep. 07, 2025. [Online]. Available: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.113.263601
[6] M. Á. Satorre, M. Domingo, C. Millán, R. Luna, R. Vilaplana, and C. Santonja, “Density of CH_4, N_2, and CO_2 ices at different temperatures of deposition,” *Planet. Space Sci.*, vol. 56, no. 13, pp. 1748–1752, Nov. 2008, doi: 10.1016/j.pss.2008.07.015.
[7] J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade, *Photonic Crystals: Molding the Flow of Light - Second Edition*. Princeton University Press, 2011. doi: 10.1515/9781400828241.

Manuscript Modifications: To meet *Nature's* 75 character title requirement, we have changed the title of our manuscript to "Purcell-enhanced spin-phonon coupling with a single color-center" and we have included references in the abstract. All changes to the manuscript since the last submitted version use red text.

Text Format: Original comments made by referees are shown in black text, our response is given in blue text, and any quotations and/or changes made to our manuscript are shown in red text.

Referees' comments:

Referee #1 (Remarks to the Author):

In this revised manuscript, the authors reply all my comments point to point.

1) For my previous comments on the acoustic Pucell effects being reported previously. I aggree with the authors's reply that the current study is based on a single spin in single phonon level, while the previouse ones only in classical regime. However, I donnot think that they are completely irrelevant. I strongly suggest the authors to cite and discuss them.

We appreciate the referee's recognition of the distinction between our single-spin, single-phonon regime work and the classical acoustic Purcell effects reported in the suggested references. We agree that these works share a broad conceptual lineage with ours, but the experimental regimes are substantially different.

Our introduction is structured to guide the reader from the original discovery and formulation of the Purcell effect to the most directly relevant developments in nanophotonics and nanomechanics, which provide the conceptual and technical foundations for our results. Incorporating a discussion of the classical acoustic Purcell effect, in our view, would not meaningfully enhance the reader's understanding of our work. Nonetheless, we appreciate the referee's perspective and are willing to include a brief clarifying discussion if the editor feels it would benefit the manuscript.

If we were to include this discussion in the introduction, it would be as follows:

"We observe the acoustic Purcell effect between the SiV spin qubit and the 12 GHz acoustic mode of the nanomechanical resonator, near its motional ground state at milliKelvin temperatures. **While the acoustic Purcell effect has been previously observed in macroscopic classical systems [Phys. Rev. Lett. 120, 114301, FUNDAMENTAL RESEARCH 4(1), pp. 57-62], this is the first such observation near the single-phonon level.** The resonator is created in an optomechanical crystal (OMC) [41] that can co-localize photons at the optical frequency linking the spin qubit with the color-center's excited state."

2) In present manuscript, the authors move the Fig. 5 to supplementary material as new Fig. 12. I think that this is because the quaulity of the data is not high enough. I donnot think it is proper.

The reordering of the Supplementary material was done solely to improve the logical flow: sections now appear in approximately the same order in which they are referenced in the main text, except where one section must naturally precede another for clarity. The relocation of the former Fig. 5 (now Fig. 12) was not motivated by concerns about data quality. We have full confidence in the quality of these measurements.

As referee #3 correctly noted, our original use of the optical full-width at half maximum (FWHM) as error bars in Fig. 12 was not the most appropriate representation of uncertainty. The spin transition frequency is estimated from a difference between peak optical frequencies, so the relevant uncertainty is that of the peak frequency, which is significantly smaller than the FWHM. We have therefore updated Fig. 12 to use the fit uncertainties as error bars. Note that these error bars are smaller than the data points and are therefore not visible in the figure.

The nuance that we were attempting to communicate here is that using this difference in peak frequencies of broad optical lines is not a direct measurement of the narrow spin transition frequency. While the statistical uncertainty in this estimator is very small, reflecting its high precision, this does not account for systematic errors like differential AC stark shifts between the various optical line measurements which limit its accuracy. In this work we use these calibrations as a coarse spin-transition frequency measurement and then use sharp identifiable features in the observed spin-decay spectra to line up and compare measurements under different conditions. Observation of the 12 GHz mechanical breathing mode (whose frequency is known to high precision from optomechanical measurements) then offers a useful reference to identify the frequency of other features with high confidence. We have modified the discussion in SI Sec. 5.8 to reflect this discussion and clarify this issue:

“In the case of a perpendicular field shown in Fig. 12, we get a conversion constant of 2.700 ± ~~0.186~~ **0.001** GHz/kG. ~~Due to the wide optical linewidths, the uncertainty in this conversion constant is somewhat large, reflecting the high precision but limited accuracy of the magnetic field magnitude as an estimator of the spin transition frequency.~~ **We note that using the difference in peak frequencies of these broad optical lines is not a direct measurement of the narrow spin transition frequency. While the statistical uncertainty in this estimator is very small, reflecting its high precision, this does not account for systematic errors like differential AC stark shifts between the various optical line measurements which limit its accuracy.** Here, we use these calibrations as a coarse spin-transition frequency measurement and then use sharp identifiable features in the observed spectra to line up and compare measurements under different conditions. Observation of the ~12 GHz mode (whose frequency is known to high precision from optomechanical measurements) then offers a useful reference to identify the frequency of other features with high confidence.”

We hope that the revised figure and clarified discussions address the referee’s concern.

Overall, I donnot think the revised manuscript could be published in Nature.

We thank the referee for their consideration, but we again respectfully disagree with the conclusion. We reiterate that this work demonstrates, for the first time, an **acoustic interface between a spin qubit and a nanomechanical resonator at the single phonon level** in a fully nanoscopic system, with record spin-phonon cooperativity. These results represent a meaningful advancement in spin-mechanical quantum systems and open new avenues for hybrid quantum technologies. We believe that our revised manuscript warrants consideration for publication in *Nature*.

Referee #2 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

We thank the referee for their comments. Anything raised by this referee in the listed reports is addressed in that report.

Referee #3 (Remarks to the Author):

The authors have provided very detailed replies to all questions, and we are very happy with their answers.
The only point of contention is maybe the question from Referee #1 on the error bars in Fig. 5 (now Fig. 12). For statistical errors, the error bars are way too large for the observed distribution of data points and are likely overestimated. It seems the authors are using the optical linewidth as the error bar which does not make sense, since from Fig. 3a, the transitions are spectroscopically resolvable, and the fit error would be a much better estimate.
In conclusion, this manuscript constitutes a major advancement, and we support publication in Nature.

We sincerely thank the referee for their positive assessment of our manuscript and for their strong endorsement of publication in *Nature*. We also appreciate their attention to the point regarding the error bars in Fig. 5 (now Fig. 12). This issue was raised by referee #1 and our response is reproduced here for completeness:

As the referee has correctly noted, our original use of the optical full-width at half maximum (FWHM) as error bars in Fig. 12 was not the most appropriate representation of uncertainty. The spin transition frequency is estimated from a difference between optical peak frequencies, so the relevant uncertainty is the fit uncertainty in the peak frequency, which is significantly smaller than the FWHM. We have therefore updated Fig. 12 to use the fit uncertainties as error bars.

The nuance that we were attempting to communicate here is that using this difference in peak frequencies of broad optical lines is not a direct measurement of the narrow spin transition frequency. While the statistical uncertainty in this estimator is very small, reflecting its high precision, this does not account for systematic errors like differential AC stark shifts between the

various optical line measurements which limit its accuracy. In this work we use these calibrations as a coarse spin-transition frequency measurement and then use sharp identifiable features in the observed spin-decay spectra to line up and compare measurements under different conditions. Observation of the 12 GHz mechanical breathing mode (whose frequency is known to high precision from optomechanical measurements) then offers a useful reference to identify the frequency of other features with high confidence. We have modified the discussion in SI Sec. 5.8 to reflect this discussion and clarify this issue:

"In the case of a perpendicular field shown in Fig. 12, we get a conversion constant of 2.700 ± ~~0.186~~ **0.001** GHz/kG. ~~Due to the wide optical linewidths, the uncertainty in this conversion constant is somewhat large, reflecting the high precision but limited accuracy of the magnetic field magnitude as an estimator of the spin transition frequency.~~ **We note that using the difference in peak frequencies of these broad optical lines is not a direct measurement of the narrow spin transition frequency. While the statistical uncertainty in this estimator is very small, reflecting its high precision, this does not account for systematic errors like differential AC stark shifts between the various optical line measurements which limit its accuracy.** Here, we use these calibrations as a coarse spin-transition frequency measurement and then use sharp identifiable features in the observed spectra to line up and compare measurements under different conditions. Observation of the ~12 GHz mode (whose frequency is known to high precision from optomechanical measurements) then offers a useful reference to identify the frequency of other features with high confidence."

We hope that the revised figure and clarified discussions address the referee's concern.

Referee #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

We thank the referee for their comments. Anything raised by this referee in the listed reports is addressed in that report.

Referee #5 (Remarks to the Author):

I thank the authors for their detailed response to my questions and comments. Most of these have been adequately addressed. The remaining one is with regard to the B_normal dependence of the acoustic Purcell effect for the 12 GHz mode.

The main reason I have asked for this additional data is that Fig. 19 shows good agreements between the experiment and simulation for the 12 GHz mode but does not show the strong Purcell effect observed for the 21 GHz mode. On the other hand, the B_normal dependence (see Fig. 13), which I consider as the most conclusive data for the acoustic Purcell effect, is demonstrated for the 21 GHz mode, but not for the 12 GHz mode.

I understood that the B_normal dependence is much harder to obtain for the 12 GHz mode than that for the 21 GHz mode and that the leading author has since left the group. In this context, I wonder if the authors can do the following:e

1) Revise/expand Fig. 19 to reproduce the strong Purcell effect observed for the 21 GHz mode in the simulation.
2) Include experimental data on the characterization of the 21 GHz mechanical mode as well as the 12 GHz mechanical mode in the supplement.

I would be glad to recommend the publication of the manuscript in Nature after the suggested revisions.

We are grateful for the referee's positive assessment of our manuscript and for their willingness to recommend publication following the requested clarifications. We also appreciate the referee's thoughtful comments regarding the $B_\perp$ dependence of the 12 GHz mode.

Regarding the $B_\perp$ dependence of the 12 GHz mode, we agree that an equivalent of Fig. 13 for the 12 GHz mode would have been a more ideal measurement due to the thorough characterization of that mode. As the referee notes, the full angular dependence measurement for the 12 GHz mode is technically challenging, particularly near the SiV high symmetry axis where the optical transitions used for spin initialization cross. These constraints ultimately led us to focus on the cleaner and more complete dataset for the 21 GHz mode presented in Fig. 13.

However, we do have measurements of the 12 GHz mode taken under gas-tuning-only conditions for both a 55° and 90° aligned field. This data was originally omitted from the manuscript and SI to avoid presenting extraneous experimental details that could distract from the main narrative. The measurements were performed by tuning the optical mode into resonance with the SiV C-Line using gas deposition, then boiling off the gas with a high power laser and the re-gas tuning. This repeated procedure further degraded the mechanical resonance - likely due to residual gas outside the cavity center - resulting in a mildly distorted resonance shape (see Fig. 14 below and now in SI Sec. 5.9). The mode quality was later restored after cleaning the chip and depositing ALD Alumina.

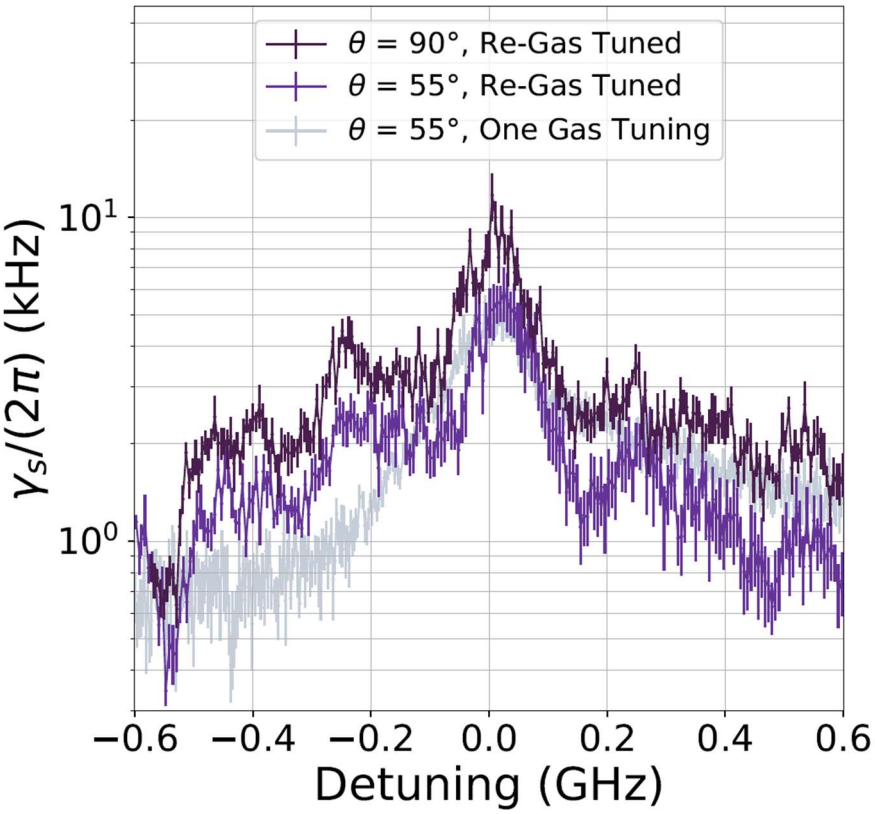


Figure 14 Spin decay spectra for the 12 GHz mechanical breathing mode under different gas-tuning-only experimental conditions. The gray curve is the data presented in Fig. 3b with a 55° aligned magnetic field after a single gas tuning. The lighter purple uses the same 55° aligned magnetic field but after boiling off the deposited gas with a high power laser and re-gas tuning. The darkest purple trace is under the same deposited gas conditions as the lighter purple trace, but with a 90° aligned magnetic field.

Despite these complications, the data clearly shows the theoretically expected angular dependence of the 12 GHz mode, more perpendicular magnetic field leads to faster spin-decay rates. This is consistent with the trend established in Fig. 13 for the 21 GHz mode, and with all other measurements performed throughout the course of this experiment. We have included this new figure and its caption in SI Sec. 5.9 along with the following text to reflect this discussion:

“In Fig. 14 we show measurements of the 12 GHz mode taken under gas-tuning-only conditions for both a 55° and a 90° aligned field. The gray curve is the data presented in Fig. 3b after a single gas tuning for a 55° aligned magnetic field. The other two measurements were performed by then boiling off the gas with a high power laser and re-gas tuning. This repeated procedure further degraded the mechanical resonance - likely due to residual gas outside the cavity center - resulting in a mildly distorted resonance shape. The mode quality was later restored after cleaning the chip and depositing ALD Alumina. The data clearly shows the theoretically expected angular dependence of the 12 GHz mode; a more perpendicular magnetic field leads to faster spin-decay rates. A full angular dependence for the comparatively low frequency 12 GHz mode is experimentally challenging due to a crossing of the optical transitions used for spin

initialization at alignments close to the SiV center's high symmetry axis, and was not performed."

Regarding the referee's two specific requests:

1) The absence of a strong simulated coupling to the ~21 GHz mode reflects current limitations in our modelling approach. In addition to the large sensitivity to the precise location of the SiV within the diamond cavity, the simulations do not fully capture real-world imperfections - such as Q-factor degradation from ALD alumina or residual gas deposition (as seen in Figs. 3b and 7), underside roughness and device non-uniformity resulting from the quasi-isotropic etching method. While a better physical model might better reproduce the observed data, developing such a model would require detailed knowledge of device specific imperfections which are challenging to measure.

In SI Sec. 5.14, we discuss these modelling challenges, we have included the additional factors mentioned above which were not previously mentioned.

"The highly local nature of spin-phonon coupling poses a modeling challenge due to the uncertain position of the SiV center within the implantation aperture, which is 50 nm wide in both the x and y directions. As seen in Fig. 20, even a few nanometer change of the in-plane location of the SiV center results in qualitatively distinct spin decay spectra. While this sensitivity could, in principle, be used to determine the precise SiV location, in practice, the complexity of the spectrum and the difficulty of modeling real-world imperfections—such as the aforementioned Q-factor degradation from ALD alumina and deposited gas (as shown in Fig. 3b and Fig. 11) **and device roughness and non-uniformity resulting from the quasi-isotropic etching method**, limit the accuracy of such simulations."

2) The ~21 GHz mode does not exhibit sufficiently strong optomechanical coupling to allow independent characterization using the same NPSD methods employed for the 12 GHz mode. Under identical measurement conditions, no detectable NPSD signal appears near 21 GHz. The 21 GHz mode is therefore characterized exclusively via spin-decay spectroscopy via the couple SiV center.

At the end of SI Sec. 5.2 we address the strong optomechanical coupling of the 12 GHz, and the negligible optomechanical coupling of all other modes.

"Finally, Fig. 6c shows the optomechanical coupling rate of the breathing mode (orange dot) as well as those of all simulated mechanical modes from 5 - 28 GHz, which covers the range of data measured as part of the spin T1 spectrum in Fig. 4. With the exception of modes around 12 GHz, nearly all modes display negligible optomechanical coupling to the TE-like mode due to the strict symmetry requirements for such coupling [62]."

We hope that the additional data of the 12 GHz mode and the clarifications above adequately address the referee's remaining concern.