

Deterministic multi-phonon entanglement between two mechanical resonators on separate substrates

Corresponding Author: Professor Andrew Cleland

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

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

In the manuscript entitled "Deterministic multi-phonon entanglement between two mechanical resonators on separate substrates," the authors demonstrated the entanglement generation of two SAWs on separate chips using superconducting qubits based on their flip-chip architecture. Although their scheme is straightforward, entangling phonons on separate chips represents a significant technological advance for scalable quantum memory. Additionally, entangling multi-phonon states is crucial for hardware-efficient bosonic encoding, leveraging the advantages of mechanical systems, including their small footprints and multimode nature. Therefore, I recommend publishing their work in Nature Communications, subject to addressing the following comments.

1. Could the authors elaborate on "The improvement is possibly due to a modified SAW resonator geometry"? Without a clear image of the SAWs, it is unclear what modifications were made.
2. Regarding joint state tomography, it would be helpful to explain the completeness of their measurement in reconstructing bipartite multi-phonon states. At a minimum, they should cite relevant papers. Specifically, is the measurement of simultaneous Rabi oscillations sufficient to determine the joint photon number distributions? What conditions are necessary to achieve this? What is the maximum photon number that can be characterized using the current parameter regime?
3. Considering all relevant transitions (ge, ef, SAW), are there any undesired mode crossings during the qubit frequency sweep in their protocols? If so, how do they mitigate occupation leakage to undesired states?
4. Although there is no population beyond the intended excitation quantum, why is there a finite population in $|03\rangle$? And why is there no population in $|30\rangle$?
5. Related to the previous questions, could the authors include error bars in the joint photon number distribution plot?
6. Could the authors provide the relaxation time and dephasing time for the f state?
7. Could the authors provide the relaxation time and dephasing time for the e and f states while each is interacting with SAWs?
8. As seen in Extended Fig. 2, why are the displacements saturated?
9. The authors should include paragraphs in the METHOD section to explain all the Extended Figures.

Reviewer #2

(Remarks to the Author)

The authors describe the creation of Bell states and two-phonon $N00N$ states in two surface acoustic modes on spatially separated chips. They use a deterministic protocol to create these entangled states and achieve impressive fidelities given the coherence of their system. In addition, the device used in this experiment is built in a modular and scalable way.

The authors utilize a system consisting of two surface acoustic wave resonators (SAWs) which are each coupled to their own superconducting qubit. The two qubits are capacitively coupled to each, allowing to first generate entanglement between the two qubits and subsequently swap the entangled state to the two acoustic modes in the two SAWs.

The work initially describes the device and characterizes the qubit-resonator interactions for both SAWs. Afterwards, the protocol for creating Bell states and their characterization by the means of Wigner tomography of the state is shown. In the last part of the paper, the protocol is extended to two-photon entangled N00N states.

Compared to related work (e.g von Lüpke et. al. Nature Physics 2024), the entanglement in this work is distributed over two modes that live in two spatially separated resonator substrates, a key ingredient to eventually close the locality loophole and perform a bell test with quantum states that can be assigned finite masses. This highlights the impact of the here presented results. In addition, the results are an important advancement in the field of cQAD as they show the ability to couple multiple acoustic modes in separate resonators with high fidelity coherent control and readout.

The authors furthermore claim to achieve fast entanglement gates and highlight the scalability of their system by using multiple modes in each SAW resonator.

While the experimental results fully support the claim of fast entanglement, the claim of scalability could be explained in more detail by elaborating a bit more on the device and parameters shown in Extended data Fig.4.

Overall, the paper is well written with a clear logical outline. The information presented is mostly self-contained or completed by the supplementary information. There are some small revisions we would suggest, listed below.

The novelty and impact of the results, as well as their presentation fulfill a high standard and are well within the scope of nature communications. We enthusiastically recommend the publication of this work after addressing minor comments.

Comments

- Regarding potential applications in quantum sensing and high-precision measurements. The authors cite [42] Carney et al., who call specifically for kHz and milligram sensors of suspended mirror-like optomechanical structures. Also, Degen et al. 2017 [41] don't seem to mention integrated quantum acoustical systems like SAWs, BARs, or phononic crystals at all (please point out where if I am mistaken). For quantum sensing and high-precision measurements, I suggest instead citing literature more relevant to the acoustic devices and their capabilities, such as e.g the efforts in BAW based gravitational wave sensing (e.g Goryachev et.al PRL2021) or others to support the claimed about quantum sensing.
- What is the vertical line at ~3.3 and 3.5GHz in Fig. 2b?
- The vacuum Rabi oscillation in Fig. 2c seem to deviate slightly from a decaying cosine. Is it clear why?
- Regarding the improved mech. Q factors compared to previous devices, what loss mechanisms dominate these devices and how did the mentioned process improvements mitigate these?
- Why do the authors use this particular resonant Wigner tomography measurement rather than a dispersive one as in von Luepke et al nat phys 2022, Bild et al, science 2023, etc? In this context I would appreciate to quote the qubit coherences or dispersive cooperativity in the main text or similar to make the platform easier to compare to other cQAD approaches. Also, regarding the scalable architecture claim, does the presence of the SAW devices impact qubit coherence and how much, at which frequencies?
- Do you see a frequency detuning between the qubit phonon $g_1 - e_0$ transition and the $g_2 - f_0$ transition beyond the independently measured self-Kerr of the qubit that could be indicative of an inherited non-linearity of the phonon mode as analyzed in Yang et al., arXiv:2406.07360
- In SM eq. 2, you use g_{ef} for the qubit ef transition coupling to the resonator. Intuitively, I would expect this coupling to be larger than that for the ge transition g_{ge} due to the extra photon involved in the former. Could you comment why this is not the case here?
- For the N00N state tomography, you use $|\alpha| = 0.35$ and $|\alpha| = 0.5$ (SM, line 322). In the caption of Extended Figure 2 you note that all data in the main text has been taken in the linear regime of the displacement pulse power vs coherent state population relation. $|\alpha_A| = 0.5$ doesn't seem to be in this linear regime. Did you take this into account when reconstructing the density matrix? Can you elaborate on the impact on the fidelity of the coherent state you create with that amplitude? Does this affect the final fidelity of the entangled states?
- For the displacement calibration and the vacuum Rabi style phonon number measurement, there are several references, such as Chu et al., Nature 2018, Wollacks et al., Nature 2022, and von Luepke et al., nat phys 2022. While this technique is fairly common in the field, there seem to be a few key differences between your implementation and previous ones. An example is that you start the qubit in the g state before running the Rabi oscillations for the tomography. Thus, I was missing either a more detailed explanation and illustration of your tomography process, calculations on how the fidelity and its uncertainty are extracted from the measured data, or a reference to previous work.
- The fits shown in extended data figure 3 include 5 and 6 levels per resonator for the Bell and N00N state. Does this choice

influence the fit result? The highest level of resonator A seems to be at least somewhat populated in both cases (state 40 for the Bell state and state 50 for the N00N state). Is there a specific reason why the authors chose 5 and 6 levels rather than, for example 10 or 20 levels, to be sure the physics is accurately captured?

- The fit results in extended data figure 3c and d suggest much higher average populations in qubit A vs qubit B, even though both are driven with the same amplitude and the N00N state is symmetric. Is this related to the different phonon number vs drive power calibrations in the previous figure? Is it clear why those calibrations are so different?
- A detail that we would like to understand better is the coherent displacement drives you use in the Wigner tomography sequence. From the wiring diagram in the appendix and Fig.1a), it looks like you have an individual drive line for both resonators A, B, rather than driving through the qubit. Is this correct? If so, it would be important to highlight this, since it is a significant improvement to the qubit mediated drives applied in v.Lüpke et. al 2022 and/or Bild et. al 2023. Assuming you can address the resonators individually through this line, by turning the coupler off while applying the displacement drives, the qubit state should not be affected by the coherent displacement in contrast to the residual driving of the qubit in other architectures (see ref. above). If this is true, I don't fully understand why your coherent state sizes start saturating at $n \sim 2$. Is the phonon mode intrinsically non-linear or is it still hybridized with the qubit during the coherent drive? It would be good to clarify this in the main text or the corresponding section in the appendix and add the tunable coupler pulses to the pulse sequence diagram in the resp. figures.
- In Line 143ff. you state that "There is thus a trade-off between the qubit-resonator coupling strength and this unwanted phonon emission." It would be instructive to mention (at least in the appendix) that transmon anharmonicities cannot be on the order of the ~ 600 MHz IDT bandwidth here, so that's why the IDT bandwidth is the limiting factor for phonon emission
- In Extended Fig.2 data d), the Fock 5 population seems to be higher than P3 and P4. Is this within the errorbars? Or a result from Hilbert space truncation? In case of the former, putting errorbars on the Fock state distribution would be helpful.
- Extended Data Fig.4: Can you give some more information on the device measured in Fig. part b)? What is the FSR here and what changes in design are involved?

Reviewer #3

(Remarks to the Author)

The manuscript by Chou, et al., reports deterministic multi-phonon entanglement between two surface acoustic wave (SAW) resonators controlled by two coupled superconducting transmon qubits. The quantum devices are carefully designed and experiments are conducted with a high-level of complexity and advancement. To my knowledge, this is the first demonstration of multi-phonon entanglement of SAW phonons. The results are solid and well organized. The achieved fidelities of entangled states are relatively high. Therefore, I will highly recommend its publication in Nature Communications.

I have one minor suggestion: it would be helpful if the authors can add description of how the numerical calculations of SAW resonators in Fig. 2a were performed. Although the authors may have explained it in their previous papers, the devices are likely different. If I understand, the SAW resonators in this work cannot be directly measured without coupling to a qubit. It is not straightforward to me how the authors chose parameters in their numerical calculations to match experimental results.

Reviewer #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

Thank you very much for the detailed and satisfactory answers to every comment from all the reviews. I have no further questions or comments on the revised manuscripts for publication in Nature Communications.

Reviewer #2

(Remarks to the Author)

I thank the authors for the clarifications in the response letter and for the changes to the manuscript. Repeating the original review, I think this is a great result and recommend publication.

Reviewer #3

(Remarks to the Author)

The authors have addressed all my questions. I support its publication in Nat. Commun.

Reviewer #4

(Remarks to the Author)

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

We thank all four reviewers for their positive reviews of the manuscript, and their thoughtful and very helpful comments. Below, we address each of the comments in detail. Any modifications made to the manuscript are indicated in red in a redline copy of the revised manuscript, which accompanies a clean copy of the revised manuscript.

Reviewer #1 (Remarks to the Author):

1. Could the authors elaborate on "The improvement is possibly due to a modified SAW resonator geometry"? Without a clear image of the SAWs, it is unclear what modifications were made.

In our SAW resonator design, a significant source of loss is due to scattering from the SAW mode into bulk modes, scattering from the acoustic mirrors and the transducer. To mitigate this, we increased the size of the SAW resonator cavity, which appears to reduce the bulk scattering rate and slightly increases the quality factor. Specifically, the distances between the two acoustic mirrors in our resonators are 74 and 70 μm , approximately 3.5 times longer than the device in Ref. [3] (K. Satzinger et al, Nature). Additionally, the metal thickness for the mirrors and transduce in the acoustic device was reduced to 10 nm, which lowers the reflection coefficient associated with each finger element, further improving the quality factor. We have added the SAW design parameters in Extend Data Table 2 in the revised manuscript.

2. Regarding joint state tomography, it would be helpful to explain the completeness of their measurement in reconstructing bipartite multi-phonon states. At a minimum, they should cite relevant papers. Specifically, is the measurement of simultaneous Rabi oscillations sufficient to determine the joint photon number distributions? What conditions are necessary to achieve this? What is the maximum photon number that can be characterized using the current parameter regime?

Our method is based on the approach described in Ref. [52] (M. Hofheinz et al., Nature 2009, a newly added reference), where complex superposition states are created in a single resonator. This approach was extended to two superconducting resonators in Ref. [51] (H. Wang et.al, PRL). In our case, several factors limit the completeness of the measurements for reconstructing the multi-phonon states.

Our method for determining phonon number probability leverages the on-resonance Rabi swap between qubit and resonator. In this process, the Rabi swap frequency scales with \sqrt{n} , where n represents the phonon number in the resonator. Multi-phonon states are prepared by transferring the qubit excitation into the resonator, as described in the main text. To characterize our targeted states, such as Bell and NOON states, we perform Wigner

tomography to measure parity. This enables us to evaluate the phonon number probability and assess the fidelity of the final states.

Given this method, the lifetimes of the qubits and resonators play a critical role. The relatively long duration of the tomography process means that phonon energy decay significantly affects the state fidelities. When using the qubit f-state to swap excitations into the resonators, the resulting e-state in the qubits will also decay, lowering the NOON state fidelity. The qubit lifetime also limits the device performance in Ref. [52] (H. Wang et.al, PRL), while in addition here the resonator lifetime and unwanted acoustic emissions (such as from the qubit e state) further limit the fidelities in our experiment. Given our current parameters, the largest entangled state we can generate with reasonable fidelity is an N=2 NOON state.

We note that precise pulse control and calibration are crucial for accurate measurements. For instance, if the qubit and resonator interact off-resonance due to e.g. pulse distortion, the system oscillates at a higher frequency $\sqrt{n\omega^2 + \Delta^2}$, where Δ is the frequency detuning between the qubit and resonator. This detuning introduces fitting errors when extracting phonon numbers, ultimately reducing the state fidelity.

3. Considering all relevant transitions (ge, ef, SAW), are there any undesired mode crossings during the qubit frequency sweep in their protocols? If so, how do they mitigate occupation leakage to undesired states?

In our experiments, we did not observe any significant undesired mode crossings. One of the undesired modes we encountered were from strongly coupled two-level system (TLS) defects. These can be avoided by tuning the qubit frequency away from the defect resonances. This adjustment ensures that the qubits operate in a regime free from unwanted interactions. Another potential issue is the presence of an additional weakly-coupled acoustic mode approximately 250 MHz detuned from the primary mode used in our experiment (see response to reviewer #2 comment). The large detuning ensures that this SAW mode does not noticeably impact our measurements.

Further, we turn the qubit-acoustics coupling off when sweeping the qubit through undesired transitions with the SAW resonator.

4. Although there is no population beyond the intended excitation quantum, why is there a finite population in $|03\rangle$? And why is there no population in $|30\rangle$?

We believe the small $|30\rangle$ population in Fig. 4d is due to pulse distortion during the Rabi swap between qubit Q_A with the corresponding SAW resonator R_A . A slight mismatch in frequency between the qubit and SAW resonator could cause slight off-resonant Rabi

swaps compared to ideal resonant case, which occur at a faster swap rate, likely generating higher excitations such as in $|30\rangle$.

5. Related to the previous questions, could the authors include error bars in the joint photon number distribution plot?

We have added error bars to the joint phonon number excitations in Figs. 3 and 4. These are calculated by bootstrapping methods (resampling for 100 times with replacement, see Ref. [57]). We note the non-zero $|30\rangle$ population cannot be fully explained by error bars in the fitting, and as mentioned above, we attribute this population to pulse distortion.

6. Could the authors provide the relaxation time and dephasing time for the f state?

We provide the separately measured f state relaxation time at the qubit idle frequency in Extended Data Table. 1. We did not measure the f state dephasing time for this specific device, but expect this dephasing time is of the same order as for the e state.

7. Could the authors provide the relaxation time and dephasing time for the e and f states while each is interacting with SAWs?

The relaxation and dephasing times of the e state while interacting with the SAWs are provided in Extended Data Table 1. These values were extracted by fitting the Rabi swap time traces shown in Fig. 2b. We do not have sufficient measurement data to extract the lifetime or dephasing time of the f state when interacting with the SAW resonators.

8. As seen in Extended Fig. 2, why are the displacements saturated?

We believe the saturation of displacement is caused by pulse distortions and inaccuracies in fitting the higher phonon number populations. Larger displacements lead to higher mean phonon number populations and faster Rabi oscillations. Pulse distortions during resonant qubit-phonon Rabi swaps can also lead to faster Rabi oscillations, as mentioned above, thus making the fitting less accurate for higher phonon number populations. Our resonator tomography experiment always operates in the linear regime, as in Extended Data Fig. 2.

9. The authors should include paragraphs in the METHOD section to explain all the Extended Figures.

The descriptions for Extended Data Figures have been added to the Supplementary Information section in the revised manuscript.

Reviewer #2 (Remarks to the Author):

• *Regarding potential applications in quantum sensing and high-precision measurements. The authors cite [42] Carney et al., who call specifically for kHz and milligram sensors of suspended mirror-like optomechanical structures. Also, Degen et al. 2017 [41] don't seem to mention integrated quantum acoustical systems like SAWs, BARs, or phononic crystals at all (please point out where if I am mistaken). For quantum sensing and high-precision measurements, I suggest instead citing literature more relevant to the acoustic devices and their capabilities, such as e.g. the efforts in BAW based gravitational wave sensing (e.g. Goryachev et al. PRL2021) or others to support the claimed about quantum sensing.*

We agree, and have added references 43-46 to the revised manuscript for a better representation of relevant quantum sensing applications.

• *What is the vertical line at ~3.3 and 3.5GHz in Fig. 2b?*

These vertical lines likely represent other Fabry-Pérot modes of the SAW resonator (one free spectral range away), which are weakly coupled to the qubit. These features are observed only when the couplers are turned on. We note that these suspected SAW modes do not influence the experiment, as when the qubit is scanned through these features, the coupler is turned off.

• *The vacuum Rabi oscillations in Fig. 2c seem to deviate slightly from a decaying cosine. Is it clear why?*

We believe the deviation between the data and the cosine model arises from pulse distortions occurring during the resonant Rabi swap between each qubit and its corresponding SAW resonator. Physically, a slight frequency mismatch between the qubit and the SAW resonator results in slightly off-resonant, faster (and incomplete) Rabi swaps compared to the ideal resonant case.

• *Regarding the improved mech. Q factors compared to previous devices, what loss mechanisms dominate these devices and how did the mentioned process improvements mitigate these?*

As in our response to referee 1: In our SAW resonator design, a significant source of loss is due to scattering from the SAW mode into bulk modes, scattering from the acoustic mirrors and the transducer. To mitigate this, we increased the size of the SAW resonator cavity, which reduces the bulk scattering rate and slightly increases the quality factor. Specifically, the distances between the two acoustic mirrors in our resonators are 74 and 70 μm , approximately 3.5 times longer than the device in Ref. [3] (K. Satzinger et al, Nature). Additionally, the metal thickness for the mirrors and transducer in the acoustic device was reduced to 10 nm, which lowers the reflection coefficient associated with each

finger element, further improving the quality factor. We have added the SAW design parameters in Extend Data Table 2 in the revised manuscript.

• *Why do the authors use this particular resonant Wigner tomography measurement rather than a dispersive one as in von Luepke et al nat phys 2022, Bild et al, science 2023, etc? In this context I would appreciate to quote the qubit coherences or dispersive cooperativity in the main text or similar to make the platform easier to compare to other cQAD approaches. Also, regarding the scalable architecture claim, does the presence of the SAW devices impact qubit coherence and how much, at which frequencies?*

• *For the displacement calibration and the vacuum Rabi style phonon number measurement, there are several references, such as Chu et al., Nature 2018, Wollacks et al., Nature 2022, and von Luepke et al., nat phys 2022. While this technique is fairly common in the field, there seem to be a few key differences between your implementation and previous ones. An example is that you start the qubit in the g state before running the Rabi oscillations for the tomography. Thus, I was missing either a more detailed explanation and illustration of your tomography process, calculations on how the fidelity and its uncertainty are extracted from the measured data, or a reference to previous work.*

We respond to both questions below:

Our method is primarily based on the approach in Ref. [52] (M. Hofheinz et al., Nature 2009), where complex superposition states are created in a single resonator. This technique was later extended to analyze two entangled superconducting resonators in Ref. [51] (H. Wang et.al, PRL). Our protocol follows a similar framework but differs in that we use two couplers to control the coupling strengths of the individual qubit-resonator pairs. The detailed explanation about our method has been added to the manuscript.

The references the reviewer mentions here (Chu et al., Nature 2018; Wollack et al., Nature 2022; von Luepke et al., Nat Phys 2022, Bild et al, science 2023) employ a dispersive interaction to validate entangled phonon states, which is distinct from our approach. In the dispersive regime, the qubit frequency is shifted by $2\chi \sim \frac{g^2}{\Delta} \frac{\alpha_q}{\Delta - \alpha_q}$ per phonon, where the g is the coupling strength between qubit and phonon, Δ is the qubit-phonon detuning and α_q is the qubit anharmonicity. This results in a time scale $T_m \sim 2\pi/2\chi$ for resolving the phonon state. For dispersive measurements to be effective, both the qubit and phonon lifetimes must be longer than T_m to preserve high state fidelity. However, in our system, due to the interdigitated transducer (IDT) emission band of ~ 600 MHz, the qubit frequency cannot be left near the mechanical mode frequency, and the required large frequency offset results in a small dispersive shift. Given the lifetime of the SAW resonator states, resonant tomography methods, which enable fast Rabi swaps at maximum coupling strength, are a more suitable approach than a dispersive measurement.

The presence of the SAW devices does not impact qubit coherence in our tunable coupling architecture when the coupling is fully turned off. However, when the coupling is turned on for resonant Rabi swaps on a timescale of approximately 30 ns, we observe a reduction in qubit lifetime, likely due to the strong piezoelectric properties of lithium niobate. Qubit lifetimes during the Rabi swaps are provided in Extended Data Table 1. The reduction in qubit lifetime when the couplers are on occurs only during Rabi swaps with SAW modes and does not undermine our claim of a scalable architecture.

• *Do you see a frequency detuning between the qubit phonon $g_1 - e_0$ transition and the $g_2 - f_0$ transition beyond the independently measured self-Kerr of the qubit that could be indicative of an inherited non-linearity of the phonon mode as analyzed in Yang et al., arXiv:2406.07360*

We did not attempt to measure frequency detunings in our experiment. We believe that our SAW resonator is intrinsically linear, and the SAW resonator mode necessarily hybridizes with the qubit during the Rabi swap process, by an amount related to the swap rate, inducing the usual nonlinearity due to the qubit-resonator interaction. Experimentally, we fine-tune the swapping process to minimize the detuning between the qubit and SAW resonator transitions for resonant Rabi swaps.

• *In SM eq. 2, you use g_{ef} for the qubit ef transition coupling to the resonator. Intuitively, I would expect this coupling to be larger than that for the ge transition g_{ge} due to the extra photon involved in the former. Could you comment why this is not the case here?*

The maximum achieved g_{fe} is greater than g_{eg} . However, as mentioned in main text, during the f-e swap, the qubit suffers unwanted e state decay as it is inside the transducer bandwidth. We attempt to minimize this by operating with the qubit-SAW coupler set to provide a smaller coupling, resulting also in a smaller e state decay, thereby maximizing the achievable NOON state fidelity.

• *For the NOON state tomography, you use $|\alpha| = 0.35$ and $|\alpha| = 0.5$ (SM, line 322). In the caption of Extended Figure 2 you note that all data in the main text has been taken in the linear regime of the displacement pulse power vs coherent state population relation. $|\alpha_A| = 0.5$ doesn't seem to be in this linear regime. Did you take this into account when reconstructing the density matrix? Can you elaborate on the impact on the fidelity of the coherent state you create with that amplitude? Does this affect the final fidelity of the entangled states?*

This is a good point, and is related to an error in our original Extended Data Fig. 2. The horizontal axis is (now) the square of actual displacement pulse amplitude (dpa) in our experiment, in arbitrary units, with the vertical axis the mean phonon number extracted from the fitting. The displacement $|\alpha_A| = 0.5$ corresponds to a mean phonon number $\langle n \rangle_A =$

$|\alpha_A|^2 = 0.25$, which is within the linear regime of the displacement power vs phonon number calibration. The horizontal axis of that figure now displays $\pm |dpa_A|^2$ and $\pm |dpa_B|^2$ instead of the erroneous $\pm |Re(|\alpha|)|^2$ and $\pm |Re(|\beta|)|^2$.

- *The fits shown in extended data figure 3 include 5 and 6 levels per resonator for the Bell and NOON state. Does this choice influence the fit result? The highest level of resonator A seems to be at least somewhat populated in both cases (state 40 for the Bell state and state 50 for the NOON state). Is there a specific reason why the authors chose 5 and 6 levels rather than, for example 10 or 20 levels, to be sure the physics is accurately captured?*

We find that five and six levels are sufficient to extract the Bell and NOON state fidelities, respectively. The maximum total number of excitations for the experiments is one and two phonons for the N=1 and 2 NOON states, respectively. In the tomography, this is combined with a resonator displacement corresponding to a mean phonon number of 0.25 for each SAW resonator – the combination resulting in far less than five or six phonons, respectively, in the resonator states. The computational cost for performing the joint density matrix tomography is quite high, with a Hilbert space of $6 \times 6 \times 3 \times 3$ for the two SAW resonators combined with two 3-level qutrits. We need to fit the time evolution traces for this large Hilbert space for all 256 displacement combinations. We have verified that increasing the Hilbert space slightly, such as using seven levels for the NOON state analysis, yields consistent joint number populations.

- *The fit results in extended data figure 3c and d suggest much higher average populations in qubit A vs qubit B, even though both are driven with the same amplitude and the NOON state is symmetric. Is this related to the different phonon number vs drive power calibrations in the previous figure? Is it clear why those calibrations are so different?*

Qubit A exhibits a slightly higher average population due to its higher residual state population after preparing a NOON state. We believe this excess residual population arises from $|e\rangle$ state decay during the first swap from qubit's $|f\rangle$ state to SAW resonator. This nonideality will result in nonzero residual qubit state population after the second swap. We have accounted for this decay in our QuTiP model when fitting the tomography traces.

- *A detail that we would like to understand better is the coherent displacement drives you use in the Wigner tomography sequence. From the wiring diagram in the appendix and Fig.1a), it looks like you have an individual drive line for both resonators A, B, rather than driving through the qubit. Is this correct? If so, it would be important to highlight this, since it is a significant improvement to the qubit mediated drives applied in v.Lüpke et. al 2022 and/or Bild et. al 2023. Assuming you can address the resonators individually through this line, by turning the coupler off while applying the displacement drives, the qubit state should not be affected by the*

coherent displacement in contrast to the residual driving of the qubit in other architectures (see ref. above). If this is true, I don't fully understand why your coherent state sizes start saturating at $n \sim 2$. Is the phonon mode intrinsically non-linear or is it still hybridized with the qubit during the coherent drive? It would be good to clarify this in the main text or the corresponding section in the appendix and add the tunable coupler pulses to the pulse sequence diagram in the resp. figures.

The referee is correct, we use separate displacement drive lines for each SAW resonator for the tomography measurements. We do not observe residual effects due to these drives on the qubits, as during the displacement drive pulses, the qubit-resonator couplers are turned off and the qubits are detuned with respect to the resonator drive.

We attribute the saturation effect to fitting inaccuracies induced by pulse distortions. As mentioned previously, qubit pulse distortions can cause slight frequency mismatches between the qubits and the SAW resonators, leading to faster (and incomplete) Rabi swaps. This effect combines with the faster Rabi swaps due to higher phonon numbers, resulting in fitting inaccuracies for the larger phonon number populations. We have incorporated these considerations into the discussion accompanying Extended Data Fig. 2. Additionally, we have included the tunable coupler pulse in this figure for further clarification.

• *In Line 143ff. you state that "There is thus a trade-off between the qubit-resonator coupling strength and this unwanted phonon emission." It would be instructive to mention (at least in the appendix) that transmon anharmonicities cannot be on the order of the ~ 600 MHz IDT bandwidth here, so that's why the IDT bandwidth is the limiting factor for phonon emission*

We believe that as long as half of the IDT bandwidth (~ 300 MHz) is engineered to be smaller than the qubit anharmonicity, this will sufficiently suppress unwanted decay from the e state during the f state swap. We have added to the revised manuscript in Line 148 the sentence, "This issue could be alleviated by reducing the IDT bandwidth so that the transition from the e state to the g state is outside the IDT emission bandwidth due to qubit anharmonicity."

• *In Extended Fig.2 data d), the Fock 5 population seems to be higher than P3 and P4. Is this within the errorbars? Or a result from Hilbert space truncation? In case of the former, putting errorbars on the Fock state distribution would be helpful.*

We have now included error bars on the Fock state distributions in Extended Data Fig. 2d with standard bootstrapping methods (resampling with replacement). The Fock $n = 5$ population fitting results are within these error bars.

- *Extended Data Fig.4: Can you give some more information on the device measured in Fig. part b)? What is the FSR here and what changes in design are involved?*

We have added this information to the caption of Extended Data Fig. 4 in the revised manuscript. For this design, the distance between the two acoustic mirrors is about 130 μm , increasing the cavity size and reducing the FSR to 44 MHz.

Reviewer #3 (Remarks to the Author):

The manuscript by Chou, et al., reports deterministic multi-phonon entanglement between two surface acoustic wave (SAW) resonators controlled by two coupled superconducting transmon qubits. The quantum devices are carefully designed and experiments are conducted with a high-level of complexity and advancement. To my knowledge, this is the first demonstration of multi-phonon entanglement of SAW phonons. The results are solid and well organized. The achieved fidelities of entangled states are relatively high. Therefore, I will highly recommend its publication in Nature Communications.

I have one minor suggestion: it would be helpful if the authors can add description of how the numerical calculations of SAW resonators in Fig. 2a were performed. Although the authors may have explained it in their previous papers, the devices are likely different. If I understand, the SAW resonators in this work cannot be directly measured without coupling to a qubit. It is not straightforward to me how the authors chose parameters in their numerical calculations to match experimental results.

When fabricating the SAW resonator used in this work, we fabricated several similar SAW devices with identical geometry to that used in this manuscript. These devices were used for cryogenic characterization, allowing us to extract the necessary data for building the electrical circuit models (see K.J. Satzinger, PhD thesis, 2018). Using these data, along with material properties provided by the manufacturer of our lithium niobate wafers, we constructed a P-matrix model for the devices, based on the coupling-of-modes (COM) model (see e.g. D. Morgan, Surface Acoustic Wave Filters, 2nd ed., Academic Press (2007)). This P-matrix model provides the basis for our numerical simulations, enabling us to generate Fig. 2a. The parameters used in the P-matrix model for these simulations are provided in Extended Data Table 2 in the revised manuscript.

Reviewer #4 (Remarks to the Author):

I co-reviewed this manuscript with one of the reviewers who provided the listed reports. This is part of the Nature Communications initiative to facilitate training in peer review and to provide appropriate recognition for Early Career Researchers who co-review manuscripts.