

Simultaneous High-Fidelity Single-Qubit Gates in a Spin Qubit Array

Corresponding Author: Dr Leon Camenzind

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)

The authors made an exciting achievement by performing simultaneous single qubit gates on multiple qubits, well within the fault tolerance error required. No doubt this is a major milestone in development of quantum computers in the silicon community.

With a deep understanding of the underlying problem, The authors accurately identified MW driven AC Stark Shift between qubits as the major root of control error, and developed a protocol to calibrate and compensate, in a scalable way. The in-depth understanding of the system makes the manuscript a suitable for Nature Physics.

However, if I understand the experimental results correctly, there is a caveat. From the SRB results, it assumes

- a) The same single qubit gate operation must be applied to all qubits
- b) No idling the of any qubit (I operation) is allowed while other qubits are in non-Identity operations.

While there exist protocols which works under such assumption, no doubt in general it restricts the how much one can condense a quantum circuit. Authors please clarify if that is not the case.

While the experimental result this manuscript has achieved is exceptional, all the calibration steps the authors have followed in this manuscript are extensions of previous literature results. Therefore, beside a deeper understanding of AC stark shift and a more comprehensive protocol, is there any major obstacles that fundamentally prevented the demonstration of high fidelity parallel single qubit gates until now?

Furthermore, I have a few minor comments / questions on the manuscript:

In page 4, where "The QND readout is repeated twice to increase the initialization fidelity of qubit Q3", why is the initialisation fidelity of Q3 bad to begin with? Is it because of poor CROT between Q2 and Q3?

In page 6, "we replace the identity gate I in the gate set with $4 \times X$ or $0 \times X$." Is $0 \times X$ simply skipping any I gates entirely?

In page 6, "For qubits Q1 to Q4, the optimal gate time results in a fidelity close to 99.999%, while Q5 reaches an optimal fidelity of $\sim 99.995\%$ ". While I agree in Fig 5c and Ext Fig 1, Q5 has the worst gate fidelity / quality factor, Fig 2d shows Q5 on par, if not slightly better than Q1. Any comments on it?

In Fig 3f & 4g, can the authors elaborate or reference the optimisation protocol used? Also, in Fig 3f caption, please add a one line explanation for the 2D map in the subfigure caption.

In fig 3e, this scheme assumes z-error accumulate at every integer of $\pi/2$ pulse, but in z-error accumulates continuously, even while the other qubit is driving.

I know it's more complicated, but has the author thought about phase compensation in a non-discrete way?

In fig 3f & g, in the SRB characterization, are each of the green and red pulse must be gate of the same operation phase, eg. both $X \pi/2$? Or one of them can have an a different operation phase, eg one $X \pi/2$ and the other $Y \pi/2$? The latter would give the quantum circuit slightly more freedom.

In Fig 4f, is $\Delta \phi_{ij}$, should it be ϕ_{ij} instead? Also, recursive notation of $\phi_i^{(n)}$ on both left and right hand side of formula is confusing. Is the right $\phi_i^{(n)}$ an 'pre-phase-calibrated' phase one wish to put into the quantum circuit?

Overall, the demonstration of high fidelity simultaneous single qubit gates is an exceptional achievement to the development of a scalable silicon quantum computer, and the less-than-exponential calibration overhead make the protocol more appetizing on a scalable platform. No doubt this manuscript is a strong candidate for Nature Physics. After the authors address all the comments / questions I stated above, I would like to recommend publishing this manuscript to Nature Physics.

Reviewer #2

(Remarks to the Author)

Simultaneous control is critical for running quantum circuits in larger systems as it reduces idle times which can be a dominant source of error. The manuscript from Wu et al. investigates the simultaneous single qubit control of up to 5 spin qubits using frequency multiplexing on a single gate line combined with engineered micromagnet field gradients. This work represents the largest system for spin qubits where simultaneous control has been explored and the fidelities achieved are by far the highest reported for spin qubits. To achieve such high fidelities, they optimized their MM design/fab and device design to generate high driving field gradients across all 5 qubits while maintaining low decoherence gradients. This allows for high Rabi frequencies with respect to coherence times while staying below the power threshold where heating effects dominate gate errors. They then developed novel error amplification strategies to accurately calibrate out errors related to the ac stark shift and frequency multiplexing. In general, the manuscript is well written, the data is of the highest quality, and importantly they make a significant effort to understand their main error sources. Although there are still clear challenges related to scaling simultaneous control to a larger number of spin qubits (as clearly highlighted in this work when operating all 5 qubits at once), I still believe this is a significant advancement for spin qubit control and relevant to the wider quantum computing community. I recommend publication in Nature Physics after the following comments/questions have been addressed:

- (1) In figure 1(a,c) the phase correction is calibrated for the case when Q3 is idle while driving Q4. This same phase correction is then applied to the case for when both Q3 and Q4 are driven (figure 1e). Its not immediately clear to me if this correction is the most optimal even though it clearly shows an improvement in gate fidelity? Is it possible to show with simulation or analytically that the Z phase error for the two cases (idle Q3, drive Q4 vs drive Q3 drive Q4) is the same? If not how close is the approximation?
- (2) Given the metrics in 1g and 1h, it was surprising that Q5 in Figure 5c performs the worse with simultaneous driving. The reason becomes clearer after looking at Extended data figure 1b,1c but there is no clear discussion of this in the main text. It would be helpful for the reader to provide more discussion around Figure 5c and give more context to extended data figure 1. I think its important to highlight that there are decoherence mechanisms related to RF driving that aren't very well understood and that affect each qubit differently.
- (3) In the methods the authors say "In this device, spin-valley mixing occurs as the exchange interaction is increased". Can the authors give more information about this? There could be many ways to interpret this and more information might highlight a potential issue for SiGe qubits. I would be interested to know how it actually shows up in the experiment (eg. excited valley population?).
- (4) In extended data Figure 7, is there an explanation for the variation in alpha? Is it related to heating induced frequency shifts? I also could believe that frequency dependent signal transmission could also lead to variation in this experiment?
- (5) In extended data Figure 7, there is a typo in the in the caption. "e Fitted parameters from e" should be "e Fitted parameters from d"
- (6) Can the authors estimate the expected dot shift at the powers used in this experiment? This would be a useful metric when comparing this work to other experiments as it gives some information on the electrical coupling of the drive line and the quantum dots.

Reviewer #3

(Remarks to the Author)

Summary of the key results

In "Simultaneous High-Fidelity Single-Qubit Gates in a Spin Qubit Array" the authors demonstrate high-fidelity operation of a five-dot spin qubit array in Si/SiGe. Individually controlled single-qubit primitive gate fidelities exceed four nines (99.99%), and some approach five nines (99.999%). These fidelities are to my knowledge among the best reported in silicon spin qubits so far. The authors demonstrate that fidelities remain above the practical fault-tolerant threshold of three nines (99.9%) when operating all five qubits in parallel.

The main results of the work are the identification of microwave-drive-induced AC Stark shifts as the dominant source of fidelity loss when operating spin qubits in parallel, and that this can largely be compensated through pairwise calibration of the crosstalk phase. Especially the latter could constitute an important step towards large scale qubit control, as it shows that the exponential overhead of full crosstalk calibration can be avoided — at least for this separable, 5-uncoupled qubit system.

The authors go into great detail investigating the limiting sources of decoherence and optimizing the operation of the qubits to maximize fidelities. I very much appreciate the overall clarity and depth of the presented work.

Comments that should be addressed

The authors have shown record-high single qubit fidelities through careful calibration of qubit operation, and an extensive investigation of the limiting decoherence channels. Therefore, they have without doubt made significant steps towards optimally controlling single qubits. However, I am unsure as to the significance of demonstrating parallel operation of a limited number of single qubits towards tackling the scalability issue of silicon spin qubits in general. The authors show that the overhead of crosstalk compensation can be reduced from exponential to $N(N-1)$ for their uncoupled qubits, which does constitute a significant reduction for large-scale systems, were this to remain true for fully coherent, coupled systems. Could

the authors comment on this — whether any effect on entangling gates could be foreseen or a priori discarded, such that the methods shown can indeed be used for a relatively simple, quadratic-scaling compensation scheme in a 'real' quantum processor? In addition, one of the main remaining limitations is heating. As stated by the authors, a careful balance must be struck between gate times and applied microwave power. As such, extending the presented parallel operation from five to ten qubits would require a doubling of the gate times to maintain the same total microwave power (as stated on p.13). It therefore seems to me that parallel operation through a common drive line is only feasible for a limited number of qubits, and thus the gain in calibration overhead will be marginal when only applied to a small number of qubits.

One of the main results of the manuscript are the high single-qubit primitive gate fidelities obtained through (variations of) randomized benchmarking. As it is such a prominent topic, not only as a direct measure of the fidelity, but also as a tool for analyzing the performance of the presented calibration techniques, I feel the analysis of the RB results requires slightly more detail and clarity in the manuscript.

- First, it is not clear to me what exact function was used to fit the data. In Methods/Randomized benchmarking (p.22), it is mentioned that the exponential function $F(n) = A n^p$ is used to obtain the depolarization parameter. However, in some plots of the normalized fidelities there are data points with a negative fidelity (e.g. main Figure 2.a (Q4), and main Figure 3.g (Q3)). It is not possible to obtain such data when normalizing only with the scale factor A . I assume a constant offset is also subtracted from these data to normalize, i.e. an exponential function of the form $F(n) = A n^p + C$. Although the decay slope of the data contains most of the information on the fidelity, the normalization is not entirely trivial and could affect the extracted fidelity. In some of the figures it is not obvious to me how the data should be normalized. For instance, in main Figure 2.a (Q5) it is not clear from the data alone whether the final data point is at zero (normalized) fidelity, or the data is still decaying. In general, I dislike the use of normalized fidelities as they also lead to individual fidelities above unity (which is also the case in the presented data). In any case, the authors should clearly state how data is normalized.

- Second, the authors concisely discuss the achievable accuracy of their single-qubit GST experiments in the Supplementary Information (p.3). I believe the rigor of the RB analysis would also benefit from such a back-of-the-envelope estimation of the achievable accuracy (not necessarily in the main text).

- Third, and building on my previous comment, I think some attention should be spent on the goodness of the fits. How confident are the authors that a five-nine fidelity can be claimed based on (I assume) a least-square fitting procedure performed on ~10 points?

The clarity and rigor of the RB analysis would greatly benefit from addressing these comments.

In Section 1.3 of the main text, the authors nicely demonstrate how the crosstalk phase can be compensated when operating two qubits in parallel. On p.9 it is argued that the discrepancy in phase shift magnitudes for Q3 and Q4 arises from the different power required to drive each qubit at the same operation speed, and the authors refer to Extended Data Fig. 1 to show that Q3 requires less power than Q4 to achieve the same operation speed. I agree with this. However, in the caption of Extended Data Fig. 1, the authors state that the data shown here corresponds to the data shown in main Fig. 1.h, where the amplitude axis is converted to power measured by the spectrum analyzer. I don't see how the data shown in main Fig. 1.h can correspond to the data shown in Extended Data Fig. 1 given this transformation of the axes. Is this indeed the same data? Are the data labelled incorrectly in one of the figures?

The authors mention the calibration of single-qubit gate parameters (Supplementary Information/Gate Set Tomography, p.3). Since this is not mentioned in the context of the presented crosstalk compensation, I assume the calibration of other gate parameters is meant here. Certain calibration schemes, such as Larmor frequency feedback, can have large effects on qubit parameters, as also stated by the authors: "Although we performed calibrations during the measurement process, slow drifts in qubit parameters could not be fully compensated, which becomes critical when targeting fidelities above 99.99%." (SI, p.4). What calibrations are performed during the experiments, and at what frequency? Are all reported metrics measured using these calibrations? How large is the effect of these regular calibrations on the reported metrics such as T_2^* ? Is any feedback used at all? Please elaborate and state concisely in the manuscript such that readers can evaluate and assess the methods used from a reproducibility perspective.

General comments, suggestions, and questions

Besides the comments that I think should be addressed by the authors, I have a few smaller remarks and questions that could help improve the quality of the manuscript.

- The five-qubit device is operated in the (3,1,1,3,1) charge configuration. As explained by the authors, the projection dots contain three electrons to enhance the readout windows. Why are dots Q1 and Q4 used as projection dots? I assume that, ideally, the configuration would be (1,3,1,3,1) to minimize screening to the charge sensors, but at least one pair must be flipped to allow for the readout of Q3. Is this fair to assume, or is there another (device-specific) reason for this exact configuration?

- The discussion of the PSB feedback initialization procedure could benefit from some more details. Firstly, does the PSB measurement used in the initialization circuit differ from the PSB measurement for readout in integration time, i.e. does it have to be faster? Also, I find this sentence somewhat ambiguous: "If an even parity is measured, a conditional X2 gate applies a pi flip to the qubits, bringing them into the up-down or down-up state." Are the dots pulsed back to the (3,1

configuration before applying the conditional gate (I assume they are)? Lastly, the sentence would be clearer if it were changed to: "If an even parity is measured, a conditional X2 gate applies a pi flip to one of the qubits."

- On p.4 of the main text, the authors report an individually controlled single-qubit Hahn-echo time of ~60us. This seems to be a relatively small improvement with respect to the reported T_2^* time of ~10us, suggesting that high-frequency noise is important. Can this be related to the micromagnet design discussed in the Supporting Information/Micromagnet for power-efficient control (p.2), where a significant dephasing gradient is already allowing e.g. high-frequency charge noise to couple to the spin qubits?
- On p.6 of the main text, the authors mention that the typical qubit frequency fluctuations in the device are on the order of 100kHz. Is this concluded from the same Ramsey measurements mentioned in Extended Data Fig. 6? If so, this figure could include an experimental time trace as well to refer to, and simultaneously serve as a comparison for the simulated time trace presented in panel b.
- In the caption to main Figure 2 it should be restated that the data presented in panel c was collected on Q3 for clarity.
- The authors mention several times that long AWG upload times up to minutes are a limiting factor in their experiments. This seems excessively long for the Keysight M3202A AWGs. What is causing such long upload times?
- Typo in the caption of Extended Data Fig 2: "sim30".
- Typo in the caption of Extended Data Fig 7: "Fitted parameters from e.", should presumably refer to another panel in the figure.
- Missing spaces on p.13 above main Fig. 5 after the equal signs.
- In Supplementary Figure 1 (SI, p.5), the authors directly compare fidelities obtained with a rectangular window and a Kaiser window, demonstrating that the Kaiser window — which is chosen since it minimizes spectral leakage — also performs better when driving individual qubits separately (if I remember correctly, this was also included in the ArXiv version of main Figure 2, which I personally appreciated). Throughout the manuscript, the authors argue that heating-induced dephasing associated with the pulse peak power is an important dephasing mechanism. Given that the Kaiser window has approximately twice the peak power of the rectangular window for the same pulse duration, do the authors have any intuition for when this increased peak power would be expected to induce stronger dephasing than for a rectangular pulse? Is there a gate time at which the rectangular window is expected to outperform the Kaiser window? In addition, since AC Stark shifts scale approximately as the square of the Rabi frequency, do the authors anticipate an optimal pulse shape where the effects of spectral filtering and AC Stark dephasing are balanced? Does this optimum correspond to the choice of $\beta = 8$?

Conclusion

In summary, after addressing the above comments, I believe the editors could consider whether this manuscript would fit best in Nature Physics or elsewhere. Although I am somewhat reserved as to the impact of the demonstration of high-fidelity (parallel) single-qubit operation on the overall challenge of scaling silicon spin qubits. I do appreciate, however, the depth to which the authors have analyzed and optimized their control, and the methodological advance in crosstalk compensation remains substantial. So in that sense, I would not object to publication of this work, which is overall well crafted and executed and for which the authors are to be commended.

Reviewer #4

(Remarks to the Author)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors addressed all of my feedback and questions in my previous review around, with modifications on the manuscript. I am satisfied with all of their responses.

While the restrictions on simultaneous single qubit operations remains a concern, I appreciate that the authors acknowledges that the SRB circuit laid out in Fig 3 subject to restrictions, especially when the different primitive gate requires different pulse time or amplitude.

However, although we have yet to conclude on a direct metric for total circuit fidelity, one shall remember that, in all quantum circuits applications, simultaneous single qubit operations, as an optimisation tool, will always performs at least same, if not better results, than otherwise, i.e. worst case is, the quantum circuit can simply remain sequential if simultaneous single

qubit gate shows no benefits.

I have remained my view from the previous round, and still would like to recommend publishing this manuscript to Nature Physics.

Reviewer #2

(Remarks to the Author)

The authors have answered most of the reviewers questions satisfactorily and the quality of the manuscript has consequently been improved. I still believe this manuscript should be published in Nature Physics but there are two questions that have not been adequately answered in my opinion:

1. Reviewers 1 first question was not something that I realized when I first reviewed the manuscript. To clarify, is the reason for doing the amplitude optimization in figure 4f due to the fact that the combined microwave amplitudes (going to the same MW drive line) do not add linearly? If that is the case then these MW amplitudes need to be recalibrated depending on which combination of qubits are being driven. So for example, driving qubits Q1, Q2, Q3, Q4 and leaving Q5 idle will have a different amplitude calibration for Q1/Q2/Q3/Q4 compared to driving all 5 qubits simultaneously. If true, this caveat brought up by reviewer 1 should be discussed in the manuscript.

2. I think my (reviewer 2) first question was possibly misunderstood. My understanding of the AC-stark shift is that if the off-resonant MW pulse (used to drive the target qubit) is far away in frequency (as in this paper), then it effectively shifts the 'victim' qubit frequency, f_0 , by some frequency f_{AC} . This leads to an accumulation of phase on the 'victim' qubit as measured directly in Figure 3c,d in the manuscript which is easy to correct. However, if both qubits are being driven simultaneously, then the frequency of both qubits are shifted and the pulses applied to them are now both slightly off-resonant. Can this error due to applying a slightly off-resonant pulse to each qubit be fully described by the phase errors measured in Figure 3c,d which corresponds to the phase accumulation while one qubit is idle not driven? A clear discussion of this topic can be found section B1 in this NMR review <https://arxiv.org/pdf/quant-ph/0404064>.

Reviewer #3

(Remarks to the Author)

I thank the authors for addressing all my questions and comments in such detail. I have no further remarks.

The manuscript is clear, precise, and from my perspective ready to publish.

Reviewer #4

(Remarks to the Author)

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

The authors have answered all my questions and I recommend publication in its current form.

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/>

Reply to the Reviewers

We thank the reviewers for their insightful comments and questions, which have helped improve our manuscript. We have modified the main text and supplementary information in response to the reviewers' comments as follows:

- Rewrote the description of the PSB feedback initialization in Sec. 1.1 for clarity and added an explanation for the use of the (3, 1, 1, 3, 1) charge configuration.
- Added a discussion in Sec. 1.2 on the susceptibility of Q1 and Q5 to drive-induced decoherence due to their proximity to the charge sensors serving as reservoirs.
- Included a discussion of dynamical phase compensation within a single pulse in Sec. 1.3.
- Replotted Fig. 2d with connected markers to improve clarity.
- Added experimental details to the captions of Figs. 1, 2, 3, and 4.
- Corrected the notation for phase compensation used in Fig. 4f.
- Added a discussion in the conclusion section regarding fidelity in practical quantum circuits and a comparison of the calibration overhead required for intermediate-scale qubit devices.
- Included additional details for the RB experiments and a discussion of RB accuracy in the Methods section.
- Included additional details on the choice of β for the Kaiser pulse and a short discussion on AC Stark shifts with different pulses
- Added experimental details to the caption of Extended Data Fig. 1 explaining the conversion between applied voltage amplitude and power.
- Added a discussion regarding the variation of the fitted α in Extended Data Fig. 7

We address the comments from Reviewers #1, #2, and #3 point-by-point below.

Reviewer #1

The authors made an exciting achievement by performing simultaneous single qubit gates on multiple qubits, well within the fault tolerance error required. No doubt this is a major milestone in development of quantum computers in the silicon community.

With a deep understanding of the underlying problem, The authors accurately identified MW driven AC Stark Shift between qubits as the major root of control error, and developed a protocol to calibrate and compensate, in a scalable way. The in-depth understanding of the system makes the manuscript a suitable for Nature Physics.

Reply

We thank the referee for these positive and encouraging comments.

However, if I understand the experimental results correctly, there is a caveat. From the SRB results, it assumes

a) The same single qubit gate operation must be applied to all qubits

b) No idling the of any qubit (I operation) is allowed while other qubits are in non-Identity operations.

While there exist protocols which works under such assumption, no doubt in general it restricts the how much one can condense a quantum circuit. Authors please clarify if that is not the case.

Reply

We agree with the reviewer that in practical quantum circuits, overall fidelity depends on a combination of single-qubit and two-qubit operations, as well as idling, readout, and initialization steps. In the simultaneous driving scheme, the single-qubit gates being applied to all qubits can differ by a phase if the gate times for all qubits are identical. Thus, different gates are applied to individual qubits during the simultaneous drive. We have made this clearer in the caption of Fig. 3.

Idling, on the other hand, reduces the overall circuit fidelity because spin qubits exhibit shorter coherence times in the absence of driving. We note that inserting $4 \times X$ gates, corresponding to a 2π rotation, could be used both to adjust the sequence length and to avoid qubit idling. More generally, identity operations are in principle allowed, but they would also acquire an AC-Stark shift when other qubits are operated on; this shift can be compensated. At the same time, idling leaves the qubits more susceptible to decoherence.

While the RB fidelities reported here are a critical metric for gate performance, we acknowledge that improvements in RB fidelity do not necessarily translate to a one-to-one increase in total circuit fidelity. The overall circuit fidelity is ultimately limited by the component with the lowest fidelity; therefore, increasing the fidelities of all remaining operations remains a critical task for the advancement of practical quantum computing with silicon spin qubits. We have added this discussion to the conclusion section of the main text.

While the experimental result this manuscript has achieved is exceptional, all the calibration steps the authors have followed in this manuscript are extensions of previous literature results. Therefore, beside a deeper understanding of AC Stark shift and a more comprehensive protocol, is there any major obstacles that fundamentally prevented the demonstration of high fidelity parallel single qubit gates until now?

Reply

We believe that several improvements compared to previous demonstrations contributed to the high-fidelity parallel single-qubit operations reported here, although it is difficult to identify a single factor that prevented earlier demonstrations (e.g. Mills et al., Sci. Adv. 8, 2022).

Beyond the compensation of AC Stark shifts, we also optimize the microwave drive amplitude specifically for parallel operation (see Fig. 2). In addition, we employ Kaiser-shaped control pulses, which reduce spectral leakage compared to rectangular pulses. This suppresses unwanted excitation of neighbouring qubits during microwave driving and therefore mitigates qubit–qubit crosstalk. To our knowledge, the use of such pulse shaping to suppress crosstalk in this context has not previously been systematically explored experimentally. A separate manuscript discussing these crosstalk effects is in preparation and will be posted on arXiv.

Also, in the present device, the qubits are separated in frequency by approximately 150 MHz due to the magnetic field gradient generated by the micromagnet. This relatively large frequency spacing suppresses off-resonant driving and therefore reduces both microwave crosstalk and AC Stark shifts on neighbouring qubits. In systems with smaller frequency separations, off-resonant excitation could significantly perturb nearby qubits, making simultaneous operation more challenging.

In addition to these operational improvements, the present device exhibits improved stability compared to earlier devices, which is important for the overall qubit fidelities achievable in both isolated and parallel operation. In particular, the charge noise level is reduced, leading to qubit resonance frequency fluctuations of approximately 100 kHz, compared to ~200 kHz observed previously. This lower noise floor improves the fidelity of individual gates and therefore also benefits simultaneous operations.

A likely contributing factor is the transition from a device with a single metal gate layer to an overlapping-gate architecture, which can reduce sensitivity to charge noise through improved electrostatic screening. Furthermore, the PSB readout used in this work removes the need to couple the quantum dots to a reservoir, as required for Elzerman-type readout, which further isolates the system from external noise sources.

Finally, we optimized the drive amplitude by measuring gate fidelity as a function of driving speed in order to identify the optimal operating point before microwave-induced

decoherence begins to dominate. This optimization balances faster gate speeds against the increased decoherence rates associated with higher microwave power.

Overall, the improvement in isolated and parallel-operation fidelity compared to previous work on this platform results from the combination of improved device stability, favourable qubit frequency spacing, reduced spectral leakage through pulse shaping, optimized drive conditions, and a comprehensive calibration protocol that includes compensation for AC Stark shifts (see Fig. 2).

Furthermore, I have a few minor comments / questions on the manuscript:

In page 4, where “The QND readout is repeated twice to increase the initialization fidelity of qubit Q3”, why is the initialisation fidelity of Q3 bad to begin with? Is it because of poor CROT between Q2 and Q3?

Reply

We observe that when the Q2-Q3 exchange coupling is activated by pulsing the barrier gate voltage, hybridization occurs between the ground and excited valley states. During the CROT operation used for QND readout, this excited valley state is also driven, resulting in randomization of the spin states. Since we perform a feedback initialization of Q3 based on this QND readout, the initialization fidelity of Q3 is directly reduced by the poor performance of the CROT gate within the QND process. We expect that increasing the device valley splitting will improve the readout and initialization fidelities.

In page 6, “we replace the identity gate I in the gate set with $4 \times X$ or $0 \times X$.” Is $0 \times X$ simply skipping any I gates entirely?

Reply

The $0 \times X$ operation indeed consists of skipping the gate entirely; we have added this detail to the main text.

In page 6, “For qubits Q1 to Q4, the optimal gate time results in a fidelity close to 99.999%, while Q5 reaches an optimal fidelity of $\sim 99.995\%$ ”. While I agree in Fig 5c and Ext Fig 1, Q5 has the worst gate fidelity / quality factor, Fig 2d shows Q5 on par, if not slightly better than Q1. Any comments on it?

Reply

We noticed that the overlapping markers in Fig. 2 of the previous version made it difficult to distinguish the fidelities between different qubits. We have replotted the figure with connected markers for improved clarity. It is now clearer that the optimal error rates for Q1 through Q4 approach similar values, while the Q5 minimum exhibits a higher error rate.

In Fig 3f & 4g, can the authors elaborate or reference the optimisation protocol used? Also, in Fig 3f caption, please add a one line explanation for the 2D map in the subfigure caption.

Reply

In the optimization procedures shown in Figs. 3f and 4g, we used the standard minimizer from the SciPy Python library with the Nelder-Mead method. The 2D maps display the cost functions that the optimizer minimizes and are provided for illustrative purposes only. In practice, the optimizer does not require a full measurement of the 2D map; it only samples the specific points necessary for the algorithm to converge. We have added this explanation to the captions of Figs. 3f and 4g.

In fig 3e, this scheme assumes z-error accumulate at every integer of $\pi/2$ pulse, but in z-error accumulates continuously, even while the other qubit is driving. I know it's more complicated, but has the author thought about phase compensation in a non-discrete way?

Reply

The hardware used in our experiments is capable of performing both phase and amplitude modulation within a single $\pi/2$ pulse although this was not implemented in the current experiments. We anticipate that applying a dynamical phase shift proportional to the pulse amplitude could in principle provide more precise compensation than the discrete method used here. We have added this discussion to Section 1.3 of the main text.

In fig 3f & g, in the SRB characterization, are each of the green and red pulse must be gate of the same operation phase, eg. both $X \pi/2$? Or one of them can have an a different operation phase, eg one $X \pi/2$ and the other $Y \pi/2$? The latter would give the quantum circuit slightly more freedom.

Reply

In Fig. 3f and g, each of the pulses can have a different phase within the same simultaneous operation, i.e. $\phi_i^{(2)}$, $\phi_i^{(3)}$ can each take values of 0 or $\pi/2$. This enables the application of simultaneous XX, XY, YX, or YY gates. We added this information to the figure caption for clarity.

In Fig 4f, is Delta ϕ_{ji} , should it be ϕ_{ij} instead? Also, recursive notation of $\phi_i^{(n)}$ on both left and right hand side of formula is confusing. Is the right $\phi_i^{(n)}$ an 'pre- phase-calibrated' phase one wish to put into the quantum circuit?

Reply

In Fig. 4f, the notation has been corrected to ϕ_{ij} . We also replaced the right-hand side term $\phi_i^{(n)}$ with $\phi_{0,i}^{(n)}$ to represent the uncompensated phase value, ensuring notational clarity.

Overall, the demonstration of high fidelity simultaneous single qubit gates is an exceptional achievement to the development of a scalable silicon quantum computer, and the less-than-exponential calibration overhead make the protocol more appetizing on a scalable platform. No doubt this manuscript is a strong candidate for Nature Physics. After the authors address all the comments / questions I stated above, I would like to recommend publishing this manuscript to Nature Physics.

We thank the referee for the positive assessment of our manuscript and for the constructive comments, which helped improve this manuscript.

Reviewer #2

Simultaneous control is critical for running quantum circuits in larger systems as it reduces idle times which can be a dominant source of error. The manuscript from Wu et al. investigates the simultaneous single qubit control of up to 5 spin qubits using frequency multiplexing on a single gate line combined with engineered micromagnet field gradients. This work represents the largest system for spin qubits where simultaneous control has been explored and the fidelities achieved are by far the highest reported for spin qubits. To achieve such high fidelities, they optimized their MM design/fab and device design to generate high driving field gradients across all 5 qubits while maintaining low decoherence gradients. This allows for high Rabi frequencies with respect to coherence times while staying below the power threshold where heating effects dominate gate errors. They then developed novel error amplification strategies to accurately calibrate out errors related to the ac stark shift and frequency multiplexing. In general, the manuscript is well written, the data is of the highest quality, and importantly they make a significant effort to understand their main error sources. Although there are still clear challenges related to scaling simultaneous control to a larger number of spin qubits (as clearly highlighted in this work when operating all 5 qubits at once), I still believe this is a significant advancement for spin qubit control and relevant to the wider quantum computing community. I recommend publication in Nature Physics after the following comments/questions have been addressed:

Reply

We thank the referee for their positive and constructive feedback. Their thoughtful comments have helped us improve the clarity and presentation of the manuscript.

(1) In figure 1(a,c) the phase correction is calibrated for the case when Q3 is idle while driving Q4. This same phase correction is then applied to the case for when both Q3 and Q4 are driven (figure 1e). Its not immediately clear to me if this correction is the most optimal even though it clearly shows an improvement in gate fidelity? Is it possible to show with simulation or analytically that the Z phase error for the two cases (idle Q3, drive Q4 vs drive Q3 drive Q4) is the same? If not how close is the approximation?

Reply

As pointed out by the referee, the drive of Q3 itself can generate an additional phase shift, though this is distinct from the off-resonant AC Stark shift effect. Ideally, an on-resonance pulse does not generate a phase shift. We verify this using the Ramsey sequence shown in Extended Data Fig. 1, which confirms the pulses are on resonance.

However, a phase shift could still be induced by heating-related frequency shifts as shown in Extended Fig. 5. We note that this heating requires approximately $120\ \mu\text{s}$ (roughly $1,500\ \pi/2$ gates) to saturate at a 200 kHz level. We attempted to measure and compensate for this self-induced phase using the same protocol but observed no improvement in gate fidelities. This is likely because the phase shift accumulates and saturates over the number of applied gates, rather than being a constant offset per gate.

Consequently, we excluded this specific phase from the protocol and anticipate that our calibration provides a good approximation for shorter gate sequences. We acknowledge that this may not hold for long sequences where heating effects become significant. Applying dynamical phase compensation, where different phase corrections are applied to gates depending on their position in a sequence, would be a potential solution. This approach is currently limited by the software used for waveform loading rather than hardware capabilities. We have added this discussion to the main text.

(2) Given the metrics in 1g and 1h, it was surprising that Q5 in Figure 5c performs the worse with simultaneous driving. The reason becomes clearer after looking at Extended data figure 1b,1c but there is no clear discussion of this in the main text. It would be helpful for the reader to provide more discussion around Figure 5c and give more context to extended data figure 1. I think its important to highlight that there are decoherence mechanisms related to RF driving that aren't very well understood and that affect each qubit differently.

Reply

In our experiments, we observed that Q1 and Q5 are more susceptible to decoherence caused by RF driving. This may be attributed to their proximity to the charge sensors, which also act as reservoirs, located at the edges of the device, which makes these qubits more vulnerable to RF-induced heating or noise. We have added this information to Section 1.2.

(3) In the methods the authors say "In this device, spin-valley mixing occurs as the exchange interaction is increased". Can the authors give more information about this? There could be many ways to interpret this and more information might highlight a potential issue for SiGe qubits. I would be interested to know how it actually shows up in the experiment (eg. excited valley population?).

Reply

In our device, spin-valley mixing becomes relevant when the exchange interaction is increased to the MHz regime for two-qubit gate operation. In this regime, we observe hybridization between the valley states. The reviewer is correct that it effectively results in an excited valley state population due to valley state mixing. Experimentally, this leads

to reduced two-qubit gate fidelity, modified exchange oscillations, and a reduction in SPAM fidelity due to our initialization scheme relying on a two-qubit gate.

Under the single-qubit operating conditions used throughout this work, where only residual exchange is present, the device can be tuned such that spin-valley mixing is negligible and the electrons remain in their valley state during operation. Therefore, this effect does not influence the reported single-qubit fidelities or the simultaneous single-qubit results.

We have clarified this point in the Methods section. Improving valley splitting remains an active materials challenge in Si/SiGe platforms (e.g. Phys. Rev. B 108, 125405).

(4) In extended data Figure 7, is there an explanation for the variation in alpha? Is it related to heating induced frequency shifts? I also could believe that frequency dependent signal transmission could also lead to variation in this experiment?

Reply

We attribute the large variation in the fitted alpha to the quality of the fits rather than to an underlying physical origin. In Extended Figures 7b and 7c, the phase shifts at various driving amplitudes are measured using the Hahn echo sequence shown in Figure 3. The Hahn echo coherence time limits the number of gates that can be applied before the qubit decoheres. For qubit pairs with large frequency separation, which results in relatively small phase shifts, the sequence is insufficient to accumulate significant phase shifts. This leads to reduced fitting accuracy, as seen in the relationship between the fitted amplitude A and the scaling exponent alpha.

Larger values of A correspond to alpha values closer to 2, where the data more clearly exhibits a quadratic increase. Conversely, for smaller fitted A values, the variation in alpha is significantly larger; this suggests that these variations are likely a result of data quality rather than the underlying driving physics. We have added this detail to the caption of Extended Figure 7.

(5) In extended data Figure 7, there is a typo in the in the caption. “e Fitted parameters from e” should be “e Fitted parameters from d”

Reply

We correct this typo in the caption.

(6) Can the authors estimate the expected dot shift at the powers used in this experiment? This would be a useful metric when comparing this work to other experiments as it gives some information on the electrical coupling of the drive line and the quantum dots.

Reply

We estimate the dot position shift using the simulated magnetic slanting field, which has a field gradient of approximately 1.25 mT/nm on average. By converting the Rabi frequency using $B_1 = \frac{\hbar f_R}{g \mu_B}$ with $f_R = 3$ MHz and $g \sim 2$. The driving field strength is then estimated to be $B_1 \simeq 0.107$ mT. This corresponds to a dot position shift of $\Delta x = \frac{0.107 \text{ mT}}{1.25 \text{ mT/nm}} \simeq 0.09$ nm in the gradient field. We have added this information to the Methods section to serve as a reference for comparison between experiments.

Reviewer #3

Summary of the key results

In “Simultaneous High-Fidelity Single-Qubit Gates in a Spin Qubit Array” the authors demonstrate high-fidelity operation of a five-dot spin qubit array in Si/SiGe. Individually controlled single-qubit primitive gate fidelities exceed four nines (99.99%), and some approach five nines (99.999%). These fidelities are to my knowledge among the best reported in silicon spin qubits so far. The authors demonstrate that fidelities remain above the practical fault-tolerant threshold of three nines (99.9%) when operating all five qubits in parallel.

The main results of the work are the identification of microwave-drive-induced AC Stark shifts as the dominant source of fidelity loss when operating spin qubits in parallel, and that this can largely be compensated through pairwise calibration of the crosstalk phase. Especially the latter could constitute an important step towards large scale qubit control, as it shows that the exponential overhead of full crosstalk calibration can be avoided — at least for this separable, 5-uncoupled qubit system.

The authors go into great detail investigating the limiting sources of decoherence and optimizing the operation of the qubits to maximize fidelities. I very much appreciate the overall clarity and depth of the presented work.

Reply

We thank the reviewer for their thoughtful comments. Their feedback provided a helpful perspective that allowed us to improve the manuscript's presentation and technical clarity.

Comments that should be addressed

The authors have shown record-high single qubit fidelities through careful calibration of qubit operation, and an extensive investigation of the limiting decoherence channels. Therefore, they have without doubt made significant steps towards optimally controlling single qubits. However, I am unsure as to the significance of demonstrating parallel operation of a limited number of single qubits towards tackling the scalability issue of silicon spin qubits in general. The authors show that the overhead of crosstalk compensation can be reduced from exponential to $N(N-1)$ for their uncoupled qubits, which does constitute a significant reduction for large-scale systems, were this to remain true for fully coherent, coupled systems. Could the authors comment on this — whether any effect on entangling gates could be foreseen or a priori discarded, such that the methods shown can indeed be used for a relatively simple, quadratic-scaling compensation scheme in a ‘real’ quantum processor?

Reply

It is important to distinguish microwave-drive-induced AC Stark shifts from phases associated with two-qubit gating. The Stark shifts discussed in this work arise from off-resonant microwave driving during single-qubit control in the regime where exchange is tuned close to zero. In this limit, the qubits are effectively independent and the resulting shifts correspond to additive single-qubit Z-rotations that can be compensated pairwise. Importantly, the microwave-induced Stark shifts studied here originate from single-qubit driving rather than from the two-qubit interaction itself, and therefore remain pairwise even in the presence of nearby coupled qubits as long as residual exchange coupling is negligible during the microwave drive. These phases are correctable within the single-qubit control layer such that qubits enter any subsequent two-qubit gate layer in a well-defined phase frame.

When entangling gates are applied, additional phases arise from the two-qubit interaction itself. For exchange-based CZ gates, no microwave tones are present during the interaction pulse and therefore no additional microwave-induced AC Stark shifts occur; however, exchange tuning can generate conditional and single-qubit phases that are calibrated as part of the two-qubit gate optimization.

For microwave-driven two-qubit gates (e.g., CROT, Nature 601, 338 (2022)), additional Stark shifts can in principle occur and would need to be characterized and compensated in a similar pairwise manner. In this case, however, the qubits are intentionally coupled during the microwave drive, and the resulting phase shifts are no longer limited to simple single-qubit Z phases but can also manifest as conditional phases due to the active two-qubit interaction. This makes the calibration of driving-induced phases more complex than in the uncoupled regime studied here. For this reason, baseband exchange-based CZ gates (Xue et al., Nature 601, 343 (2022)), which rely only on voltage pulses and do not involve microwave driving during the interaction, are expected to be preferable for large-scale devices. In that case, microwave-induced Stark shifts remain confined to the single-qubit control layer and can be compensated using the quadratic pairwise scheme demonstrated in this work.

In practice, quantum circuits can be compiled into alternating single-qubit and two-qubit layers, ensuring that microwave driving occurs while qubits are effectively isolated. Under these conditions, the Stark compensation demonstrated here remains pairwise and requires $N(N-1)$ calibration terms for N simultaneously driven qubits. Consequently, the quadratic scaling applies to the number of qubits driven from a shared control line rather than to the total number of qubits in a large-scale processor.

More generally, the calibration overhead associated with two-qubit gates depends strongly on the chosen architecture. In many proposed large-scale implementations, entangling operations are executed in designated interaction regions to which qubits are temporarily shuttled for two-qubit gates (see Nature Communications 15, 4977 (2024) for a recent proposal). Such regions can be calibrated independently, so the relevant calibration overhead scales with the number of interaction regions rather than with the total number of qubit pairs across the processor. In this operation mode, the quadratic compensation scheme demonstrated here remains applicable within each single-qubit control region, while the calibration of two-qubit interactions is localized to the corresponding interaction zones.

In addition, one of the main remaining limitations is heating. As stated by the authors, a careful balance must be struck between gate times and applied microwave power. As such, extending the presented parallel operation from five to ten qubits would require a doubling of the gate times to maintain the same total microwave power (as stated on p.13). It therefore seems to me that parallel operation through a common drive line is only feasible for a limited number of qubits, and thus the gain in calibration overhead will be marginal when only applied to a small number of qubits.

Reply

We agree with the referee that microwave-induced heating and decoherence ultimately limit the number of qubits that can be driven simultaneously within a single control region. The total microwave power applied to the control line increases with the number of driven qubits, and excessive power can lead to heating and drive-induced decoherence. To maintain acceptable power dissipation, the total microwave power must therefore remain limited, which requires reducing the drive power available per qubit when more qubits are operated in parallel. This can be compensated either by increasing the gate duration, which lowers the required Rabi frequency but increases exposure to decoherence, or by increasing the magnetic field gradient to enhance the electric-dipole spin resonance coupling, although stronger gradients can also increase sensitivity to charge noise. In practice, these trade-offs limit the number of qubits that can be driven simultaneously within a single control region.

However, this does not restrict the approach to only a small total number of qubits. Instead, we envision an architecture composed of multiple repeated control regions, each containing on the order of ten qubits that are driven simultaneously from a shared control line. The quadratic calibration scheme demonstrated here would then apply independently within each region, while the overall processor can still scale to a much larger number of qubits. Because the EDSR driving mechanism is electrically generated, multiple such regions could in principle be supplied from the same microwave source without increasing the required drive amplitude per control line, provided proper impedance matching is implemented. We added a short discussion to the Conclusions.

Even for a modest number of qubits per region, the difference between exponential and pairwise calibration scaling remains substantial. If all simultaneous driving combinations needed to be calibrated independently, the number of required calibrations would grow exponentially with the number of qubits. For example, for $N = 10$ qubits this corresponds to $2^N - N - 1 = 1013$ independent calibration terms, accounting for all possible multi-qubit driving combinations beyond single-qubit operations. In contrast, the pairwise model demonstrated here reduces the required calibrations to $N(N-1) = 90$ calibration terms. The reduction from 1013 to 90 illustrates that the quadratic scaling demonstrated in this work can significantly simplify calibration even for relatively small control regions.

Finally, the physical origin of the observed microwave-induced heating is not yet fully understood, and future device engineering may further mitigate these effects. For example, stronger magnetic field gradients from optimized micromagnet or magnetic gate-stack designs could reduce the required microwave power. Alternative pulse shapes, such as Gaussian-square envelopes, can reduce peak drive amplitudes while maintaining comparable gate times. In addition, future technological developments, for example, local heat sinks, may further mitigate microwave-induced heating. Such developments could increase the number of qubits that can be driven simultaneously in future devices.

One of the main results of the manuscript are the high single-qubit primitive gate fidelities obtained through (variations of) randomized benchmarking. As it is such a prominent topic, not only as a direct measure of the fidelity, but also as a tool for analyzing the performance of the presented calibration techniques, I feel the analysis of the RB results requires slightly more detail and clarity in the manuscript.

- First, it is not clear to me what exact function was used to fit the data. In Methods/Randomized benchmarking (p.22), it is mentioned that the exponential function $F(n) = A n^p$ is used to obtain the depolarization parameter. However, in some plots of the normalized fidelities there are data points with a negative fidelity (e.g. main Figure 2.a (Q4), and main Figure 3.g (Q3)). It is not possible to obtain such data when normalizing only with the scale factor A . I assume a constant offset is also subtracted from these data to normalize, i.e. an exponential function of the form $F(n) = A n^p + C$. Although the decay slope of the data contains most of the information on the fidelity, the normalization is not entirely trivial and could affect the extracted fidelity. In some of the figures it is not obvious to me how the data should be normalized. For instance, in main Figure 2.a (Q5) it is not clear from the data alone whether the final data point is at zero (normalized) fidelity, or the data is still decaying. In general, I dislike the use of normalized fidelities as they also lead to individual fidelities above unity (which is also the case in the presented data). In any case, the authors should clearly state how data is normalized.*

Reply

In our RB analysis, we calculate the difference in return probabilities between two sequences, where the total sequence is compiled to be equivalent to either an identity or an X flip. This protocol was discussed in Muhonen et al., J. Phys.: Condens. Matter 27, 154205 (2015) and was later also used in Yoneda et al., Nat. Nanotechnol. 13 (2018) by our group. We then fit this difference to the function $A \cdot p^n$. The fitted A parameter is then used to normalize the plotted sequence fidelity. To validate our results, we performed an alternative fit using $A \cdot p^n + B$. We found that the fitted B values were all smaller than 0.015 in magnitude. The difference in gate fidelity between the fixed-B and varied-B models was an order of magnitude smaller than the significant figures we report, validating the use of the $A \cdot p^n$ model.

As noted by the reviewer, some normalized sequence fidelities appear slightly negative. This effect is not introduced by the normalization but arises from statistical fluctuations. For example, the values -0.007 ± 0.012 (Fig. 2a, Q4) and -0.012 ± 0.012 (Fig. 3g, Q3) are both consistent with zero within their error bars. We have clarified this point in the Methods section describing the RB experiments. We also added two references [Muhonen et al., J. Phys.: Condens. Matter 27 154205 (2015) and Harper et al., Phys. Rev. A 99 052350 (2019)] discussing this protocol.

We note that finite offsets in RB decay curves can occur in other qubit platforms, for example due to leakage in exchange-only qubits. In our system, however, we have not observed such offsets. While we cannot strictly exclude small leakage contributions, for example due to spin-valley excitation, we deliberately extended the Clifford sequence length to very long depths in the experiments reporting the highest fidelities in order to verify full saturation of the decay (qubit Q4 in Fig. 2a). The RB traces approach zero within experimental uncertainty, indicating the absence of a measurable offset. The sequence lengths used were limited only by practical waveform upload times rather than by measurement constraints.

• *Second, the authors concisely discuss the achievable accuracy of their single-qubit GST experiments in the Supplementary Information (p.3). ,I believe the rigor of the RB analysis would also benefit from such a back-of-the-envelope estimation of the achievable accuracy (not necessarily in the main text).*

Reply

We refer to [Epstein et al., Phys. Rev. A 89, 062321 (2014)], where the performance of RB experiments was numerically investigated. In that work, the accuracy of the RB-estimated average error rate $\hat{r}$ relative to the true average error rate r is defined by $\mu = \log_{10} \frac{\hat{r}}{r}$. For an RB experiment using $K = 35$ randomizations, the accuracy is $\mu \sim 0.1$ for $r =$

10^{-4} . This corresponds to an absolute accuracy of about 10^{-5} in the extracted error rate. Therefore, the RB experiments performed here provide sufficient accuracy to resolve an error rate on the order of 10^{-4} , but not reliably below 10^{-5} . We added this discussion to the Methods section on the RB experiments.

• *Third, and building on my previous comment, I think some attention should be spent on the goodness of the fits. How confident are the authors that a five-nine fidelity can be claimed based on (I assume) a least-square fitting procedure performed on ~ 10 points?*

Reply

While we observed gate fidelities approaching 99.999% in our experiments, we do note that the current RB setup could be insufficient to precisely resolve these values. To faithfully measure such high fidelities, a greater sequence length and additional data points would be required to observe the full decay curve. We have added this discussion to the Methods section regarding RB experiments for clarification.

The clarity and rigor of the RB analysis would greatly benefit from addressing these comments.

- *In Section 1.3 of the main text, the authors nicely demonstrate how the crosstalk phase can be compensated when operating two qubits in parallel. On p.9 it is argued that the discrepancy in phase shift magnitudes for Q3 and Q4 arises from the different power required to drive each qubit at the same operation speed, and the authors refer to Extended Data Fig. 1 to show that Q3 requires less power than Q4 to achieve the same operation speed. I agree with this. However, in the caption of Extended Data Fig. 1, the authors state that the data shown here corresponds to the data shown in main Fig. 1.h, where the amplitude axis is converted to power measured by the spectrum analyzer. I don't see how the data shown in main Fig. 1.h can correspond to the data shown in Extended Data Fig. 1 given this transformation of the axes. Is this indeed the same data? Are the data labelled incorrectly in one of the figures?*

Reply

We would like to clarify that Fig. 1.h and Extended Data Fig. 1.a are indeed the same data. However, we agree with the reviewer that the caption for Extended Data Fig. 1 requires a more detailed explanation of how the microwave power is converted into the driving field strength. We determine the power of the qubit drive by mapping the applied voltage amplitude to the spectrum analyzer measured power $P_{1\text{MHz}}$ at a driving with Rabi frequency $f_R = 1$ MHz. Because each signal channel may have a different impedance or

attenuation, we define a scaling factor $\alpha = \frac{V_{1\text{MHz}}}{\sqrt{50 \cdot 10^{-10} \frac{P_{1\text{MHz}}}{10}}}$. This allows us to convert the

applied voltages into a standardized power P using the relation $P(V) = 10 \log_{10} \left(\frac{\left(\frac{V}{\alpha}\right)^2}{50} \right)$.

We added these conversion details to the caption of Extended Data Fig.1.

- *The authors mention the calibration of single-qubit gate parameters (Supplementary Information/Gate Set Tomography, p.3). Since this is not mentioned in the context of the presented crosstalk compensation, I assume the calibration of other gate parameters is meant here. Certain calibration schemes, such as Larmor frequency feedback, can have large effects on qubit parameters, as also stated by the authors: "Although we performed calibrations during the measurement process, slow drifts in qubit parameters could not be fully compensated, which becomes critical when targeting fidelities above 99.99%." (SI, p.4). What calibrations are performed during the experiments, and at what frequency? Are all reported metrics measured using these calibrations? How large is the effect of these regular calibrations on the reported metrics such as T_2^* ? Is any feedback used at all? Please elaborate and state concisely in the manuscript such that readers can evaluate and assess the methods used from a reproducibility perspective.*

Reply

While we perform periodic recalibrations during Gate Set Tomography (GST) experiments due to their extended duration, we do not recalibrate during Randomized Benchmarking (RB) measurements. A GST experiment typically requires more than 24 hours to complete because of the large number of sequences and interleaved measurements, whereas an RB experiment is completed within approximately three hours.

During these measurements we observed that the qubit resonance frequency is subject not only to low-frequency noise but also to a slow monotonic drift of approximately 10 Hz/s, which we attribute to gradual magnetization loss in the external magnetic field (solenoid magnet), as discussed in our recently posted manuscript arXiv:2603.03051 (2026). Over the timescale of the RB experiments, this drift does not significantly affect the extracted fidelities and no in-experiment recalibration was required. However, for the substantially longer GST experiments, this slow drift becomes relevant and necessitates periodic recalibration within the measurement. We added a reference to arXiv:2603.03051 to the Supplementary Information.

All RB fidelities reported in the manuscript were obtained without mid-sequence recalibration and therefore reflect the intrinsic stability of the device over the duration of the RB experiment. We added this information to the Methods section.

General comments, suggestions, and questions

Besides the comments that I think should be addressed by the authors, I have a few smaller remarks and questions that could help improve the quality of the manuscript.

- *The five-qubit device is operated in the **(3,1,1,3,1)** charge configuration. As explained by the authors, the projection dots contain three electrons to enhance the readout windows. Why are dots Q1 and Q4 used as projection dots? I assume that, ideally, the configuration would be (1,3,1,3,1) to minimize screening to the charge sensors, but at least one pair must be flipped to allow for the readout of Q3. Is this fair to assume, or is there another (device-specific) reason for this exact configuration?*

Reply

We use the (3,1,1,3,1) charge configuration and make the PSB projection of Q2 (Q5) onto Q1 (Q4) to maximize the PSB readout windows of both readouts, see Philips et al., Nature 609 (2022). We added this information to Sec. 1.1 of the main text.

- *The discussion of the PSB feedback initialization procedure could benefit from some more details. Firstly, does the PSB measurement used in the initialization circuit differ from the PSB measurement for readout in integration time, i.e. does it have to be faster? Also, I find this sentence somewhat ambiguous: "If an even parity is measured, a conditional X2 gate applies a pi flip to the qubits, bringing them into the up-down or down-up state." Are the dots pulsed back to the (3,1) configuration before applying the conditional gate (I assume they are)? Lastly, the sentence would be clearer if it were changed to: "If an even parity is measured, a conditional X2 gate applies a pi flip to one of the qubits."*

Reply

The Pauli spin blockade (PSB) readouts used for feedback initialization and standard readout are both performed with a 10 us integration time. In the feedback initialization sequence, the measurement is only used to determine whether the system occupies an even-parity state ($\uparrow\uparrow$ or $\downarrow\downarrow$), so that a conditional bit-flip can be applied if necessary. Any additional dephasing or phase accumulations during the integration time do not affect this operation. In principle, T1 relaxation effects ($\uparrow$ to $\downarrow$) would affect the quality of this initialisation. However, we have not observed any T1 relaxation effects in our experiments, as expected, since T1 times are typically much longer than the initialization and gate operation times. Consequently, the PSB integration time does not need to be shorter. We have added the integration time of the PSB readouts to the caption of Fig. 1.

The feedback initialization X^2 pulses are applied in the (3,1) charge configuration. Prior to the feedback pulse, the two qubits may occupy one of the two even parity states (i.e., $\uparrow\uparrow$ or $\downarrow\downarrow$). By flipping Q1 or Q5 with the X^2 pulse, the system is transitioned into an odd-parity state ($\downarrow\uparrow$ or $\uparrow\downarrow$). The odd-parity state is then initialized into the $\uparrow\downarrow$ state with an adiabatic PSB ramp. This initialisation method was reported in Philips et al., Nature 609 (2022). We have revised the text in Sec. 1.1 for clarity.

• On p.4 of the main text, the authors report an individually controlled single-qubit Hahn-echo time of $\sim 60\text{us}$. This seems to be a relatively small improvement with respect to the reported $T2^*$ time of $\sim 10\text{us}$, suggesting that high-frequency noise is important. Can this be related to the micromagnet design discussed in the Supporting Information/Micromagnet for power-efficient control (p.2), where a significant dephasing gradient is already allowing e.g. high-frequency charge noise to couple to the spin qubits?

Reply

As pointed out by the reviewer, the relatively modest improvement between $T2^*$ and $T2$ Hahn indicates that the qubit noise spectrum contains a significant high-frequency component that is not efficiently suppressed by a single refocusing pulse. The micromagnet gradient can indeed enhance the coupling of charge noise to the spin qubit, since fluctuations of the electron position translate into fluctuations of the Zeeman energy.

At the same time, the sensitivity of the qubit to charge noise depends strongly on the electrostatic tuning conditions, which determine the electron position within the micromagnet-induced magnetic field gradient. In the same device we have observed Hahn-echo coherence times up to about 200 us under different gate-voltage configurations. This indicates that the micromagnet design itself is not the fundamental origin of the relatively short $T2$ Hahn times observed in the present measurements.

In the present experiment, however, the device was tuned to enable high-fidelity operation of all five qubits simultaneously. This required careful adjustment of the electrostatic configuration, for example to suppress spin-valley mixing effects, achieve minimal exchange interaction during single-qubit operations, and ensure stable multi-qubit operation. Under these operating conditions, the qubits are not tuned to the transverse position where the micromagnet-induced dephasing gradient is minimized. As a result, charge-noise-induced position fluctuations couple more strongly to the qubit frequency, leading to the shorter Hahn-echo times observed here. The reported $T2$ Hahn values therefore reflect the tuning conditions required for stable multi-qubit operation rather than the maximum coherence times achievable for an individually optimized qubit.

• *On p.6 of the main text, the authors mention that the typical qubit frequency fluctuations in the device are on the order of 100kHz. Is this concluded from the same Ramsey measurements mentioned in Extended Data Fig. 6? If so, this figure could include an experimental time trace as well to refer to, and simultaneously serve as a comparison for the simulated time trace presented in panel b.*

Reply

Yes, the quoted ~ 100 kHz frequency fluctuations are derived from the same Ramsey-based time traces used to compute the power spectral density (PSD) shown in Extended Data Fig. 6(a). The PSD in panel (a) is obtained directly from experimentally measured frequency fluctuations over time. Panel (b) shows a representative time trace generated from the fitted PSD in order to illustrate the statistical properties of the extracted noise spectrum.

Importantly, the effective sampling time of the Ramsey noise measurement differs from the timescale relevant for the randomized benchmarking (RB) experiments. While the Ramsey experiment itself is fast, estimating the qubit frequency requires averaging over approximately 100 Ramsey points, which results in an effective sampling interval of about 150 ms when operating and reading all five qubits (see arXiv:2603.03051 (2026)). In contrast, the RB experiments are repeated with a repetition rate of a few milliseconds.

As a consequence, the raw Ramsey time traces probe a narrower range of noise frequencies than those that directly affect RB performance. By generating time traces from the fitted PSD using sampling intervals matched to the RB experiment, we obtain fluctuations representative of the noise experienced during the RB sequences. For this reason, we chose to present the spectral representation and the corresponding resampled time traces rather than a raw Ramsey time trace, since a direct visual comparison between traces obtained on different effective timescales could otherwise be misleading.

• *In the caption to main Figure 2 it should be restated that the data presented in panel c was collected on Q3 for clarity.*

Reply

We add this information to Fig. 2 caption for clarity.

• *The authors mention several times that long AWG upload times up to minutes are a limiting factor in their experiments. This seems excessively long for the Keysight M3202A AWGs. What is causing such long upload times?*

Reply

In our RB experiments, the waveforms are loaded individually for each randomization. Within a single randomization, we load waveforms for all sequences of varying Clifford

lengths (typically 10 to 15 points). This results in a total of approximately $3.25 \times 10 \times (2 + 4 + 8 + \dots + 2^{15})$ (about 2 million) primitive gate pulses. While this approach reduces fluctuations within a single measurement, it requires a longer waveform loading time of approximately five minutes. We added this information to the Method section regarding RB experiment.

- *Typo in the caption of Extended Data Fig 2: “sim30”.*
- *Typo in the caption of Extended Data Fig 7: “Fitted parameters from e.”, should presumably refer to another panel in the figure.*
- *Missing spaces on p.13 above main Fig. 5 after the equal signs.*

Reply

We corrected these typos.

• *In Supplementary Figure 1 (SI, p.5), the authors directly compare fidelities obtained with a rectangular window and a Kaiser window, demonstrating that the Kaiser window — which is chosen since it minimizes spectral leakage — also performs better when driving individual qubits separately (if I remember correctly, this was also included in the ArXiv version of main Figure 2, which I personally appreciated). Throughout the manuscript, the authors argue that heating-induced dephasing associated with the pulse peak power is an important dephasing mechanism. Given that the Kaiser window has approximately twice the peak power of the rectangular window for the same pulse duration, do the authors have any intuition for when this increased peak power would be expected to induce stronger dephasing than for a rectangular pulse? Is there a gate time at which the rectangular window is expected to outperform the Kaiser window? In addition, since AC Stark shifts scale approximately as the square of the Rabi frequency, do the authors anticipate an optimal pulse shape where the effects of spectral filtering and AC Stark dephasing are balanced? Does this optimum correspond to the choice of $\beta = 8$?*

Reply

The comparison between rectangular and Kaiser pulses was included in an earlier arXiv version but was removed from the main manuscript to keep the focus on simultaneous control as the central result.

Initially, we expected that the isolated single-qubit gate fidelity would not depend strongly on spectral leakage. Subsequent experiments clarified that the observed fidelity differences were not intrinsic to the addressed single-qubit gates, but instead originated from crosstalk to neighbouring qubits, in particular the qubit used for PSB readout. Because PSB readout effectively measures the parity of two spins (a ZZ-type measurement), it is sensitive to the joint two-qubit state. Spectral leakage from the control pulse can therefore weakly drive the neighbouring qubit and modify the readout

outcome, which can appear as a reduction of the extracted single-qubit fidelity even if the addressed qubit rotation itself is unchanged. Using an alternative readout approach (combined ZZ and IZ measurements), we can disentangle these contributions and reconstruct the relevant two-qubit Hilbert-space readout channels. With this method we find that the bare single-qubit fidelities obtained with rectangular and Kaiser pulses can be comparable, indicating that the previously observed differences were dominated by crosstalk rather than by the intrinsic gate performance. These results will be reported in a separate manuscript; therefore, we do not discuss this analysis in detail in the present manuscript.

Regarding peak power and AC Stark shifts, the qubit rotation angle scales with the time integral of the Rabi frequency, whereas AC Stark phase accumulation scales with the square of the instantaneous drive amplitude. Pulse shapes that produce the same qubit rotation can therefore still lead to different Stark phases if their peak amplitudes differ. In this sense, pulse envelopes that remain closer to a rectangular shape can reduce peak amplitudes and therefore may reduce heating and Stark-shift effects, but they suffer from stronger spectral leakage and therefore increased crosstalk, as discussed above.

This also clarifies when a rectangular pulse could outperform a Kaiser pulse: As the gate time increases, the spectral width of the control envelope decreases, and off-resonant excitation of neighbouring qubits becomes negligible. In that regime, the lower peak amplitude of a rectangular pulse could reduce both heating and Stark-phase accumulation. In our present devices and frequency spacings we have not yet observed such a regime.

To address this trade-off, we are currently exploring Gaussian-square pulses, which suppress spectral leakage while maintaining a lower peak amplitude than Gaussian or Kaiser envelopes. Such pulse shapes may therefore provide a practical compromise between spectral confinement and peak amplitude.

Finally, the choice of $\beta = 8$ for the Kaiser pulses was based on practical waveform considerations rather than an optimization of Stark effects. For $\beta = 8$ the pulse amplitude approaches zero at the boundaries, enabling smooth concatenation without sharp voltage transitions, whereas smaller β leaves finite boundary amplitudes and larger β increases the peak amplitude (and the AC Stark effect magnitude).

We added additional details on the choice of β and the expected reduction of AC Stark shifts to the Methods section describing the Kaiser-window pulse.

Conclusion

In summary, after addressing the above comments, I believe the editors could consider whether this manuscript would fit best in Nature Physics or elsewhere. Although I am somewhat reserved as to the impact of the demonstration of high-fidelity (parallel) single-

qubit operation on the overall challenge of scaling silicon spin qubits. I do appreciate, however, the depth to which the authors have analyzed and optimized their control, and the methodological advance in crosstalk compensation remains substantial. So in that sense, I would not object to publication of this work, which is overall well crafted and executed and for which the authors are to be commended.

We thank the referee for their careful evaluation and balanced assessment of our work. We appreciate the recognition of the methodological advances and depth of analysis, and we have addressed the raised concerns to clarify the broader relevance and impact of the results, which has improved the manuscript.

Reply to the Reviewers

We thank the reviewers for their continued positive assessment and for their careful comments, which helped us further clarify the manuscript.

In response, we made the following changes:

- In Sec. 1.3, we clarified that simultaneous operations can introduce residual driven-gate errors beyond crosstalk phase shifts, and that both residual rotation errors and z-phase errors must be calibrated.
- In Sec. 1.4, we clarified that idle-qubit AC-Stark phase measurements do not fully describe residual errors during simultaneous driven gates, discussed the role of simultaneous-drive amplitude calibration in the five-qubit experiment, and noted that direct frequency-shift compensation may provide a more systematic future calibration strategy.

Reviewer #1 (Remarks to the Author):

The authors addressed all of my feedback and questions in my previous review around, with modifications on the manuscript. I am satisfied with all of their responses.

While the restrictions on simultaneous single qubit operations remains a concern, I appreciate that the authors acknowledge that the SRB circuit laid out in Fig 3 subject to restrictions, especially when the different primitive gate requires different pulse time or amplitude.

However, although we have yet to conclude on a direct metric for total circuit fidelity, one shall remember that, in all quantum circuits applications, simultaneous single qubit operations, as an optimisation tool, will always performs at least same, if not better results, than otherwise, i.e. worst case is, the quantum circuit can simply remain sequential if simultaneous single qubit gate shows no benefits.

I have remained my view from the previous round, and still would like to recommend publishing this manuscript to Nature Physics.

We thank the reviewer for maintaining their positive view of our work and for their recommendation to publish the manuscript.

Reviewer #2 (Remarks to the Author):

The authors have answered most of the reviewers questions satisfactorily and the quality of the manuscript has consequently been improved. I still believe this manuscript should be published in Nature Physics but there are two questions that have not been adequately answered in my opinion:

1. Reviewers 1 first question was not something that I realized when I first reviewed the manuscript. To clarify, is the reason for doing the amplitude optimization in figure 4f due to the fact that the combined microwave amplitudes (going to the same MW drive line) do not add linearly? If that is the case then these MW amplitudes need to be recalibrated

depending on which combination of qubits are being driven. So for example, driving qubits Q1, Q2, Q3, Q4 and leaving Q5 idle will have a different amplitude calibration for Q1/Q2/Q3/Q4 compared to driving all 5 qubits simultaneously. If true, this caveat brought up by reviewer 1 should be discussed in the manuscript .

We agree that subset-dependent amplitude corrections would become a scalability concern if they were significant. However, the amplitude optimization in Fig. 4f should not be interpreted as evidence that the combined microwave amplitudes on the shared drive line do not add linearly. Instead, we use this calibration as an empirical correction for residual driven-gate errors during simultaneous operation. In the five-qubit simultaneous operation experiment, applying the simultaneous-drive amplitude calibration does not lead to a significant change in the extracted fidelities, suggesting that such residual errors remain small at the lower driving strengths used there.

When simultaneous driving pulses are applied, the optimal driving amplitudes can differ slightly from those calibrated for individual single-qubit driving. This could arise from AC-Stark-induced shifts of the qubit frequencies, which make the driven pulses slightly off-resonant and thereby introduce residual rotation errors. Technical nonlinearities in the microwave chain at higher output powers could also contribute to small amplitude shifts (see e.g. recent discussion in arXiv:2603.08577). Such effects would be hardware-specific rather than fundamental to simultaneous driving. While the exact mechanism responsible for these amplitude differences has not yet been conclusively identified in our current experiments, the resulting driven-gate errors are accounted for phenomenologically through simultaneous-drive amplitude calibration. If these residual errors are dominated by AC-Stark-induced frequency shifts, directly shifting the driving frequency would provide a more precise calibration method.

In summary, the five-qubit experiment suggests that, at the lower driving strengths used there, the residual driven-gate errors addressed by the amplitude calibration remain small. We have added this discussion to the end of Sec. 1.4.

2. I think my (reviewer 2) first question was possibly misunderstood. My understanding of the AC-stark shift is that if the off-resonant MW pulse (used to drive the target qubit) is far away in frequency (as in this paper), then it effectively shifts the ‘victim’ qubit frequency, f_0 , by some frequency f_{AC} . This leads to an accumulation of phase on the ‘victim’ qubit as measured directly in Figure 3c,d in the manuscript which is easy to correct. However, if both qubits are being driven simultaneously, then the frequency of both qubits are shifted and the pulses applied to them are now both slightly off-resonant. Can this error due to applying a slightly off-resonant pulse to each qubit be fully described by the phase errors measured in Figure 3c,d which corresponds to the phase accumulation while one qubit is idle not driven? A clear discussion of this topic can be found section B1 in this NMR review <https://arxiv.org/pdf/quant-ph/0404064> .

We apologize for the misunderstanding of this point in the previous review. The phase shifts measured for idle qubits do not fully describe all possible errors during simultaneous driven gates. For an idle qubit, an AC-Stark-induced frequency shift appears as phase accumulation. During simultaneous driven gates, however, the same

frequency shift can also detune the applied pulse and thereby cause residual rotation errors, consistent with the mechanism discussed in the NMR review cited by the reviewer.

These residual driven-gate errors are minimized empirically by the simultaneous-drive amplitude calibration. Other microscopic mechanisms could, in principle, also contribute, such as variations in electron displacement within the micromagnet gradient field under different simultaneous driving conditions. Importantly, after phase correction and simultaneous-drive amplitude calibration, the two-qubit simultaneous-drive fidelities are close to the isolated-drive fidelities. This empirically shows that the relevant residual errors are compensated for or reduced to a sufficiently small level by the protocol.

We have added this discussion to Sec. 1.4.

Reviewer #3 (Remarks to the Author):

I thank the authors for addressing all my questions and comments in such detail. I have no further remarks.

The manuscript is clear, precise, and from my perspective ready to publish.

We thank the reviewer for the positive assessment of our work and for recommending the manuscript for publication.