scientific reports

OPEN

Robust and fast post-processing of single-shot spin qubit detection events with a neural network

Tom Struck[1], Javed Lindner[1], Arne Hollmann[1], Floyd Schauer[2], Andreas Schmidbauer[2], Dominique Bougeard[2] & Lars R. Schreiber[1]✉

Establishing low-error and fast detection methods for qubit readout is crucial for efficient quantum error correction. Here, we test neural networks to classify a collection of single-shot spin detection events, which are the readout signal of our qubit measurements. This readout signal contains a stochastic peak, for which a Bayesian inference filter including Gaussian noise is theoretically optimal. Hence, we benchmark our neural networks trained by various strategies versus this latter algorithm. Training of the network with 10^6 experimentally recorded single-shot readout traces does not improve the post-processing performance. A network trained by synthetically generated measurement traces performs similar in terms of the detection error and the post-processing speed compared to the Bayesian inference filter. This neural network turns out to be more robust to fluctuations in the signal offset, length and delay as well as in the signal-to-noise ratio. Notably, we find an increase of 7% in the visibility of the Rabi oscillation when we employ a network trained by synthetic readout traces combined with measured signal noise of our setup. Our contribution thus represents an example of the beneficial role which software and hardware implementation of neural networks may play in scalable spin qubit processor architectures.

The fundamental information unit of a quantum computer is the quantum bit (qubit), a quantum mechanical two-level system. For quantum computation, the qubit needs to be initialised to a known state, be manipulated into an arbitrary superposition in Hilbert space and entangled with other qubits. Finally, qubit states need to be read out. In order to measure correlations between qubits, every single qubit-state readout must be performed in a single-shot, i.e. without averaging qubit evolution cycles or qubit ensembles[1]. Fast and high-fidelity single-shot readout of qubits is vital for the realisation of quantum information processing. Since quantum error correction schemes require frequent qubit readout[2], the qubit measurement time should not be much longer than the qubit manipulation time to avoid speed limitations. In this work we use a qubit defined by the two energetically split spin states of a single electron in a magnetic field. The readout scheme depends upon the specific spin qubit realisation and can be discriminated into two categories[3]: The measurement signal starts either immediately after the trigger of the detection process[4], or it is delayed randomly by a turn-on time[5]. While spin-to-charge conversion of a singlet-triplet spin readout by Pauli-spin blockade falls into the first category[6,7], single-spin detection by energy-dependent tunneling to a weakly tunnel-coupled reservoir falls into the second[5,8]. For the latter, the analog measurement signal is often post-processed by peak-signal filters to assign a binary qubit readout. Examples for peak-signal filter are wavelet edge detection[9], signal threshold[1,5] and slope threshold after filtering the signal with total variation denoising[10,11].

If only one single spin detection cycle is considered, a Bayesian inference filter capturing the tunneling constants and typical noise is optimal[3]. The readout speed with the Bayesian filter can be improved by adaptive decisions which allow to balance measurement time versus read-out fidelity[12]. As the signal-to-noise ratio (SNR) of the detection signal is lowered for qubits in a dense array[13] or for charge detectors operating at elevated temperature, post-processing robust to low SNR is essential for future quantum computing architectures and hot electron spin qubits[14], motivating the testing of alternatives to the theoretically optimal Bayesian method.

Here, we report on the performance of neural networks, which have been previously used to tune the electrostatics of devices[15–18], to post-process single spin readout by spin-to-charge conversion. We compare their robustness and post-processing time to a Bayesian inference filter. We find that a neural network can perform similarly to the Bayesian inference filter on synthetic data and slightly outperforms it on real data, if it is made

[1]JARA-FIT Institute for Quantum Information, Forschungszentrum Jülich GmbH and RWTH Aachen University, Aachen, Germany. [2]Institut für Experimentelle und Angewandte Physik, Universität Regensburg, Regensburg, Germany. ✉email: lars.schreiber@physik.rwth-aachen.de
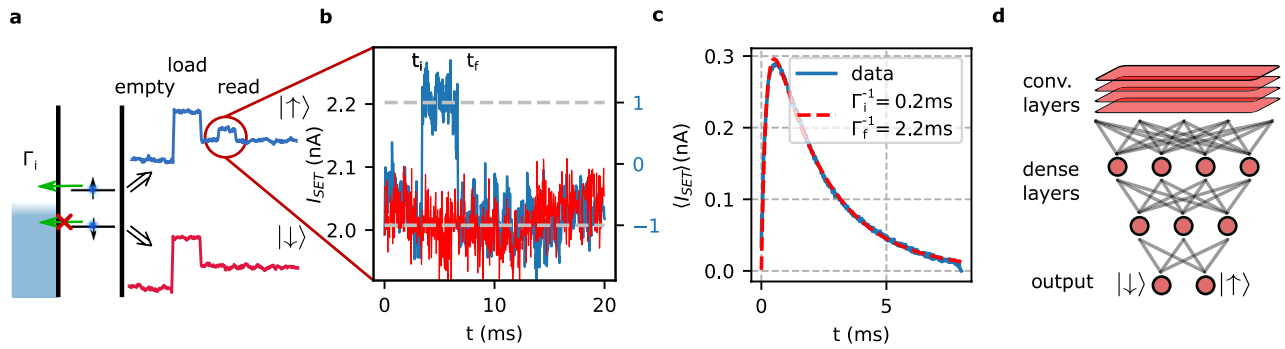
Figure 1. Single-shot measurement traces and architecture of the neural network. (**a**) Sketch of the Zeeman energy-resolved tunneling from the QD to the reservoir with a tunnel rate Γ_i. Two typical SET current traces showing the cycle of emptying and loading stage of the QD followed by the qubit readout stage are also shown. Only when the qubit state is $|\uparrow\rangle$-spin, the electron tunnels out of the QD at time t_i and a $|\downarrow\rangle$-spin tunnels on the QD at time t_f resulting in a peak in the SET current (blue trace). (**b**) Zoom-in of two measured SET current traces spanning the readout stage. The blue and red traces are classified to be $|\uparrow\rangle$ and $|\downarrow\rangle$, respectively. (**c**) Current signal spanning the readout stage averaged for many $|\downarrow\rangle$ and $|\uparrow\rangle$ traces (blue), fitted by Eq. (1) (dashed red line). (**d**) Architecture of the network used to classify peak traces. One trace spanning the readout time is used as input (top) and classification is output by two neurons (bottom), giving the probability of $|\downarrow\rangle$ and $|\uparrow\rangle$ trace.

more robust by variations in the training data. Considerably better performance is achieved on the classification of measured data, when the neural network is trained with synthetic traces combined with measured noise.

Results

Our qubit system is a single electron spin qubit trapped in an electrostatically defined Si/SiGe quantum dot (QD)[11]. The qubit is encoded by the electron spin states $|\uparrow\rangle$ and $|\downarrow\rangle$, which are energetically split by an in-plane magnetic field of 668 mT (see "Methods" section). Our detection signal consists of the single-shot readout of spin orientations via spin-to-charge conversion: Setting the chemical potential of the two-dimensional electron reservoir as plotted in Fig. 1a at time $t = 0$ s, energetically only an electron in the $|\uparrow\rangle$ state can tunnel into the reservoir after a time t_i following Poisson statistics[5]. At a time t_f, the empty QD is reinitialized by an electron in a $|\downarrow\rangle$-state from the reservoir. The two tunnel events are detected by the current I_{SET} of a capacitively coupled single-electron transistor (SET). Signal traces for $|\uparrow\rangle$ and $|\downarrow\rangle$ events are shown exemplarily in Fig. 1b. Averaging the I_{SET} traces of $3 \cdot 10^5$ $|\uparrow\rangle$-events (Fig. 1c), we fit the distribution by

$$\langle I_{SET}(t) \rangle \propto e^{-t\Gamma_f}(1 - e^{-t\Gamma_i}) \tag{1}$$

where $I_{SET}(t)$ is the current signal, meaning that $\langle I_{SET}(t) \rangle$ is proportional to the probability that the QD is unoccupied (Fig. 1c). Γ_i and Γ_f are the tunneling rates out and into the QD, respectively. For the plotted example, we find $\Gamma_i^{-1} = 0.20 \pm 0.07$ ms and $\Gamma_f^{-1} = 2.20 \pm 0.004$ ms.

We consider a network implemented in Tensorflow[19] with Keras and investigate its post-processing performance to classify the I_{SET} traces into $|\uparrow\rangle$ and $|\downarrow\rangle$ events. The input layer is connected to four 1D convolutional layers with kernel sizes (101, 51, 25, 10), a filter depth of (32, 16, 16, 8) and ReLU[20] activation (Fig. 1d). Inbetween each convolutional layer, a maxpooling layer with size 3 is inserted. The convolutional layers feed into a dense network with three layers of size (64, 32, 2). The first two layers use ReLU activation and the last layer uses softmax, since the last two neurons categorize into 1 and 0 for a $|\uparrow\rangle$- and $|\downarrow\rangle$-event, respectively. This neural network architecture was selected after testing variations of the network, both by changing it by hand and with Bayesian optimisation of some of the network parameters: We optimised the kernel sizes, filter depths of the convolutional layers, as well as the number and size of the dense layers and the dropout rate after each dense layer using the Bayesian optimisation. We find that sufficiently large networks have a similar error rate. If the network is too small, e.g. three small dense layers and no convolutional layers, the achieved accuracy decreases by approx. 4-5%. A network without convolutional layers is also able to reach the same accuracy, but converges slower. Note, that too large networks become inefficient as far as training and evaluation time is concerned. The size of the architecture chosen here, represents the best compromise we found.

We train this network architecture by three qualitatively different sets of traces and will call theses trained networks *B*, *C* and *D*. In all cases, we employ the neural network architecture explained above with the Adam optimiser[21] and categorical cross-entropy as the loss function. For the training of network *B*, we synthesize $|\uparrow\rangle$ and $|\downarrow\rangle$ traces with Gaussian noise and $\Gamma = \Gamma_f / \Gamma_i = 1$. We express the SNR by r, where r is the power signal to-noise-ratio integrated over the average high signal time $\langle t_f - t_i \rangle^3$

$$r^{-1} = \int_0^{\langle t_f - t_i \rangle} dt \int_0^{\langle t_f - t_i \rangle} dt' \langle \delta \Psi(t) \delta \Psi(t') \rangle \tag{2}$$

where Ψ is the signal trace scaled between -1 and 1, $\delta \Psi(t) = \Psi(t) - \langle \Psi(t) \rangle$, the deviation of the signal from the noiseless signal. Examples of traces with three different r are plotted as inset in Fig. 2a. The synthetic traces are
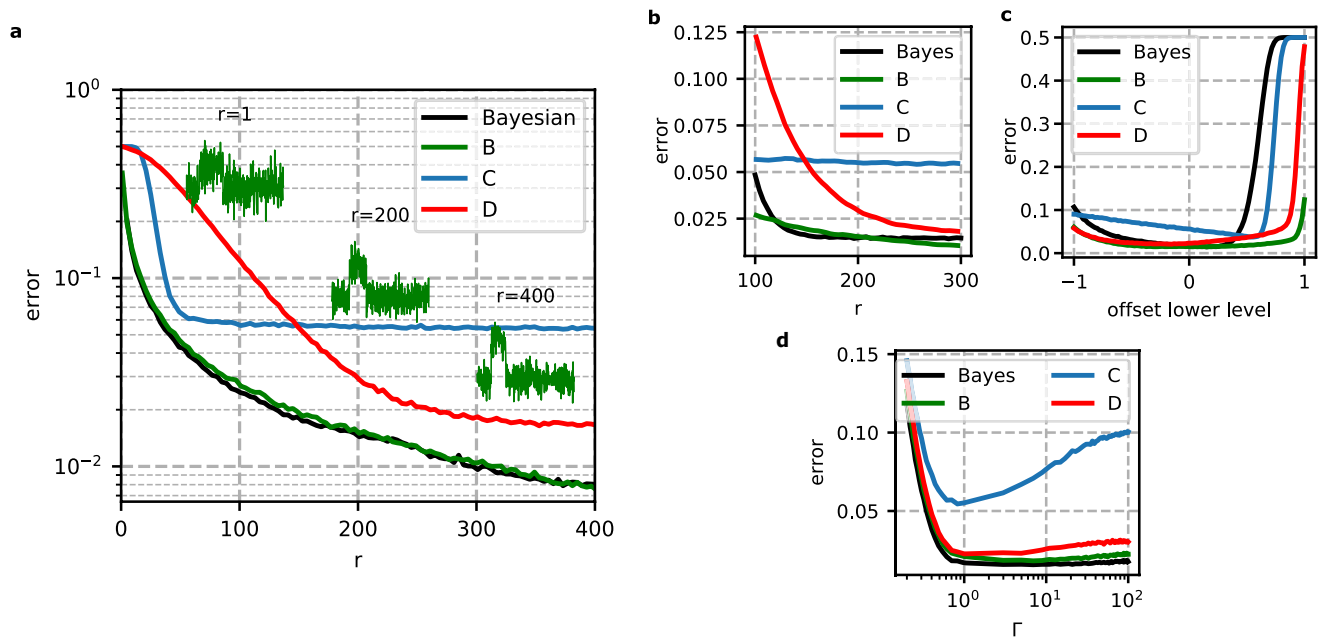
Figure 2. (**a**) Classification error of the neural network B, trained with synthetic $|\uparrow\rangle$ and $|\downarrow\rangle$ traces, compared to classification by the Bayesian inference filter, considering Gaussian noise, $\Gamma = 1$ and various r. The network C is trained on meausred data and network D is trained with synthetic traces superimposed with typical noise from the experimental setup. The green inset traces are example input traces for $|\uparrow\rangle$-events at a given r. (**b**–**d**) Robustness of the Bayesian inference filter and the neural networks B, C and D to three parameters of synthesized traces superimposed by Gaussian noise. (**b**) Robustness to the SNR r variations. (**c**) Robustness to an offset of the low-current level of I_{SET}. The x-axis here indicates the deviation of the lower level from the value of -1. (**d**) Robustness to variations in the tunnel rates Γ.

generated such that the position of the lower level in the noiseless signal is at -1 and the high level is at 1. Then we add Gaussian noise to this trace. The network B is trained by synthetic traces with various r. It is trained on $5 \cdot 10^5$ traces with an equal distribution of traces synthesized with r ranging from 1 to 400. Since we synthesize the training data, the correct labeling of the event traces is ensured. This is in contrast to the network C, which we train with 10^6 measured traces collected over two months of continuously running experiments with one device. During the measurements we tried to keep Γ_i and Γ_f in the range specified in Fig. 1c. The measured data has to be classified, since we need labels for the training. We used a Bayesian inference filter for this classification[3]. The network D is trained by traces, which are synthesized similar to the traces for the network B, but instead of Gaussian noise we generate noise from the measured power spectrum of the experimental setup, which encompasses the qubit device and all the setup electronics representing our common noise sources. Since we synthesize the peak of these training traces as we do for the training data of the network B, labeling is 100% correct, while the labeling of the training data of network C is defective due to the error of the Bayesian inference filter.

Next, we compare the classification error of the neural networks B, C and D to the Bayesian inference filter from Ref.[3] (see equation C4 in the appendix therein) as a function of the SNR. In order to determine the classification error, defined by the sum of the $|\downarrow\rangle$-states labeled $|\uparrow\rangle$, we synthesized $100.000 \, |\uparrow\rangle$ and $100.000 \, |\downarrow\rangle$ traces with Gaussian noise, $\Gamma = 1$ and various r values. The classification error of the network B is nearly the same, compared to the Bayesian inference algorithm, within numerical uncertainties (Fig. 2a). This result is not surprising, since neural networks having a large enough size can emulate any function[22]. Remarkably, however, the network B classifies synthetic traces of various r, as it has been trained by various r values, meaning that the network B is made robust against r fluctuations. The neural networks C and D show a significant larger classification error than the network B and the Bayesian estimate. This is due to the fact that these networks are trained with at least partially experimental contribution, which are not captured by the synthetic traces used to determine the classification error in Fig. 2a. Specifically, the typical experimental noise is more involved than just Gaussian noise and the networks C and D were only trained by the measured SNR range. Note that the network C also contains the classification error of the Bayesian inference filter used for labeling the training data set. The network D is trained on synthetic traces superimposed with experimental noise (see "Methods" section). If we determine the classification error by such synthetic traces superimposed with the experimentally measured noise spectrum, the D network achieves already a lower classification error (2.7%) compared to the Bayesian inference filter (5.7%).

In experiments typically a SNR of $r \approx 400$ is achievable[23,24]. According to Fig. 2a, we can achieve a low qubit detection error rate of less than 1% with the network B as well as with the Bayesian inference filter. However, this result only refers to the ideal synthetic traces. Real signal traces contain noise with complicated noise spectral densities originating from e.g. interference noise, charge noise from an ensemble two-level fluctuators, and Johnson noise. Thermal excitation, spin relaxation or co-tunneling faster than the time-scale of t_i and t_f lead to defective signal traces as well. These latter effects can be suppressed by proper tuning of the ratio of the tunnel

rates to measurement bandwidth and the ratio of the Zeeman energy to the electron temperature. Challenging are slow variations of the SNR r, the current offset of the SET and variations of the ratio between tunnel-in and -out rate Γ, due to low-frequency charge-noise or uncompensated cross-capacitive couplings to other gates. Therefore, we investigate the robustness of the post-processing filters with respect to these parameters in the following paragraph.

Figure 2b shows the robustness of the different approaches to variations in the r-parameter. Since the networks do not get r as an input parameter, the dependencies here are the same as in Fig. 2a. In contrast to Fig. 2a , now the Bayesian filter has a fixed r-parameter set to 200. Comparing the Bayesian inference filter to the network B, remarkably, the Bayesian inference filter performs generally worse than the neural network B, e.g. it shows up to 2% additional error at $r = 100$, and it deviates sharply from the optimum as the SNR decreases. Here the Bayesian algorithm seems to readily interpret single spikes in the noisy signal trace as peaks and labels them as $|\uparrow\rangle$ by error. Note that the error rate of the Bayesian inference filter saturates at an error above 1 % for an increasing SNR. This at first sight surprising observation is caused by $|\uparrow\rangle$ traces classified as $|\downarrow\rangle$, since the peak is mistaken to be Gaussian noise as the amplitude of Gaussian noise is much lower than the Bayesian filter expected by the implemented $r = 200$. Very important is the position of the overall I_{SET} level: The Bayesian estimate quickly fails at predicting the correct result (Fig. 2c), while the neural network B shows no increase of error in a wide offset range of -0.75 to 0.75. Here, the offset is to be understood as the error in assigning -1 to the lower signal level, while the total amplitude is kept at 2. Finally, we investigate the robustness of the different methods with respect to the Γ parameter (Fig. 2d). We observe that all post-processing approaches except for network C remain accurate for $\Gamma > 1$. If $\Gamma \ll 1$ and thus $\Gamma_f \ll \Gamma_i$, the classification error rises for every method, since the peak beginning starts to exceed the measurement window, i.e. the predefined length of the signal trace.

We now apply the neural network approach and the Bayesian inference filter to classify experimentally recorded datasets. In contrast to synthetic traces, here, we face the fundamental problem that we cannot know a priori whether traces correspond to a $|\uparrow\rangle$- or $|\downarrow\rangle$-state. We make use of a manipulation technique for qubits, during which the qubit is coherently driven between its two base states. This results in Rabi oscillations following the formula

$$P(|\uparrow\rangle) = \frac{V}{2} \exp\left(-\frac{t}{T_R}\right) \cos(2\pi \nu t) + P_0, \qquad (3)$$

where t is the Rabi driving time, ν the Rabi frequency, T_R a Rabi-specific spin decay. V and P_0 are the visibility and the offset of the Rabi oscillations, respectively.

Thus, we expect a continuous variation of probabilities to find the $|\uparrow\rangle$ state ($P(|\uparrow\rangle)$). As a result we can used the Rabi oscillation to set a well-controlled probability to measure either spin-up or spin-down state. The visibility V of the oscillation is used to benchmark the post-processing by the Bayesian inference and the neural network filters by comparing their averaged results to the expectation given by Eq. (3). Although V and P_0 are reduced due to initialisation errors of the qubit and manipulation errors during Rabi driving (e.g off-resonant driving), it is reasonable to assume that the classification error of post-processing the readout traces reduces V as well. Hence, we presume that a larger V corresponds to a lower classification error if the same set of readout data is post-processed. Before the data is analysed with the methods described above, it is rescaled and the lower level offset is removed. For rescaling we use the known height of the peak of 200pA and the lower level is estimated with the median of the trace. Each data point in Fig. 3a is the average of 250 traces classified to be either in $|\downarrow\rangle$ or $|\uparrow\rangle$ state. After fitting the data by Eq. (3) as shown in Fig. 3a, we find a visibility of $V = 0.664 \pm 0.004$ for the analysis using the Bayesian inference filter. The network C, which is trained on real data, has a lower visibility of $V = 0.644 \pm 0.004$ and thus does not perform better than the Bayesian inference filter. The neural network B, for which we find $V = 0.695 \pm 0.004$, slightly outperforms the Bayesian estimation mainly due to the superior robustness of the neural network to variations in the I_{SET} offset (Fig. 2c). The error in the lower level estimation can occur if a large portion of the signal is on the higher level, since in this case the median will estimate a wrong lower level current. The network D reveals a significant larger visibility of $V = 0.760 \pm 0.004$ compared to the Bayesian inference filter and all the other neural networks, thus its classification error is the lowest. Mainly the classification error of $|\uparrow\rangle$ traces is reduced (Fig. 3b). In contrast to network B and the Bayesian inference filter, network D is trained on the realistic noise spectrum, but does not suffer from the labeling problem of the real training data used for network C. Hence, this hybrid training approach outperforms all other training methods as well as the Bayesian inference filter based on a reasonably simple noise model.

Apart from the classification error and robustness to variations in a real experiment, the time T required for the post-processing per trace is an important performance parameter. It adds up to the measurement time and can become critical for real-time feedback e.g. required during quantum error correction. In order to compare T of the different classification traces, we let the whole post-processing run as efficiently as possible on the same computer equipped with an Intel i9 9900K processor. The differential equations for the Bayesian algorithm are solved using a Runge-Kutta method, in a Python script using Numba[25] just-in-time compilation. The neural networks runs with the tensorflow package. We find that T for all neural networks is $\approx 50\,\mu s$ and $\approx 200\,\mu s$ for the Bayesian inference filter. Importantly, T is of the same order of magnitude as the fastest reported experimental measurement times[26], hence representing a relevant contribution to classification processes if it runs on a computer. Note that a peak finder algorithm used in Refs.[10,11] required approximately 100 times longer. T of the Bayesian inference filter might be boosted by hardware encoding of the algorithm e.g. in field programmable gate arrays. This is also possible with the neural network in dedicated neural network hardware chips. Hence, both methods present advantages for low-temperature and low-power control electronics in the future.
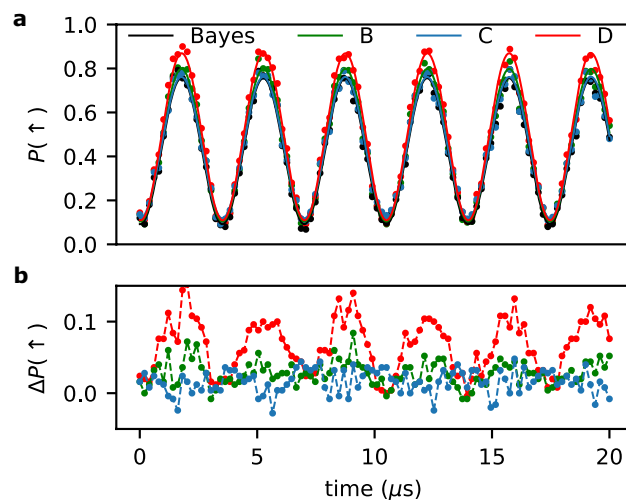
Figure 3. Detection of a Rabi-driven single electron spin analysed by different post-processing methods. (**a**) Spin $|\uparrow\rangle$-probability $P(|\uparrow\rangle)$ as a function of the driving time t. The same measurement traces are classified by the Bayesian inference filter, and the neural networks B, C and D. Solid lines are fits to Eq. (3). (**b**) $P(|\uparrow\rangle)$ from panel (**a**) classified by the neural networks B, C and D each subtracted by $P(|\uparrow\rangle)$ classified by the Bayesian inference filter (colors equal to panel **a**).

Discussion

In summary, we have shown that the neural network approach is a competitive alternative to post-processing of single-shot spin detection events by a Bayesian inference filter. The processing speeds are similar, with a slight advantage for the neural network. We have benchmarked the performance of the neural network versus a Bayesian inference filter on synthetic and experimental data, using different training methods for the network. Since the Bayesian filter is required to classify experimental traces, training the network with 10^6 experimental traces is of no advantage. On the synthetic data, we find a network trained with synthetic traces to yield a similar error rate than the Bayesian filter, while it slightly outperforms the latter in terms of robustness versus variations in experimental parameters such as SNR, the signal current offset and the tunnel couplings. This advantage is even more pronounced for the classification of real measurement data: our neural network trained with a combination of synthetic data and measured noise outperforms the Bayesian benchmark by 7%, as seen from the visibility of Rabi oscillations of the spin qubit. Here, the combination of an absence of labelling errors in the training data and the setup-specific noise proves to be particularly advantageous.

Given that that our results should be representative for qubit types with stochastic readout schemes and that the realtime performance of the neural network can be further optimized by running it on dedicated hardware, neural networks can represent an important building block for cryoelectronics yielding high-fidelity readout in scalable qubit architectures.

Methods

The experimental data was measured on an electrostatically defined quantum dot device, which consists of an undoped ^{28}Si/SiGe heterostructure forming a quantum well in the strained ^{28}Si layer. Two-layers of metallic gates patterned by electron-beam lithography fabricated on top of the semiconductor heterostructure form a quantum dot. The occupation number can be controlled down to a single electron and is detected by a proximal single electron transistor. The two spin-states of the singly occupied quantum dot define the qubit. Its state can be read out in a single-shot by spin-to-charge conversion[5]. The device is cooled down in a dilution refrigerator with a base temperature of 30 mK and an electron temperature of 114 mK. The details of the device including a study of the spin-splitting noise and the single-shot detection method can be found in Refs.[10,11].

The network architecture used in this work was implemented using the Keras API for Tensorflow. It is a sequential model with four 1D convolutional layers that have a max pooling layer in between them, followed by three dense layers. The convolutional layers use kernel sizes of 101, 51, 25 and 10 and filter depths of 32, 16, 16 and 8, respectively. The dense layers have the sizes 64, 32 and 2. All layers except from the last dense layer use the ReLU activation function. The last layer is used for classification and therefore uses a softmax activation function. The loss-function is the categorical cross-entropy and we employ the Adam optimizer for training.

For all calculations we use a computer equipped with an Intel i9 9900K CPU and a Nvidia 1050Ti GPU. For the Bayesian interference filter three differential equations are solved by a Runge-Kutta method realised in a Python script using Numba[25] just-in-time compilation. The neural networks run with the Tensorflow package[19]. As explained in the main text, we use three different training procedures of the neural network architecture labelled B, C and D. Networks B and D are trained on synthetic data each consisting of 495 equally spaced data points spanning a time of 8 ms. For a $|\uparrow\rangle$ trace, the initial and final times t_i and t_f are generated from an exponential distribution. The noiseless trace is set to 1 between t_i and t_f and the remainder of the trace is set to -1. For a $|\downarrow\rangle$ trace, all data points are set to -1. Noise is added by two different methods: (I) The Network B is trained
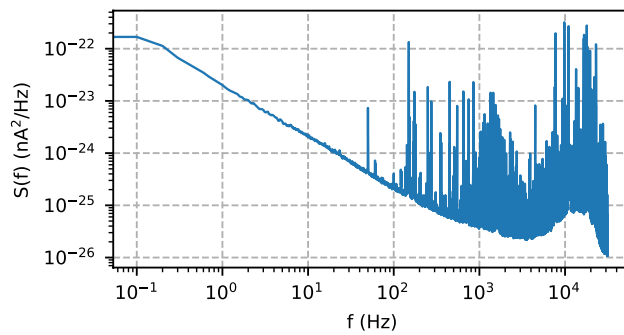
Figure 4. Experimentally measured power spectral density used to generate the colored noise for the synthetic data to train network *D*.

with synthetic traces with Gaussian noise, the σ of which is given by the signal-to noise ratio r[3]. We generate $2.5 \cdot 10^5 |\uparrow\rangle$ and $2.5 \cdot 10^5 |\downarrow\rangle$ traces. (II) For the training of network *D*, we generate $10^5 |\uparrow\rangle$ and $10^5 |\downarrow\rangle$ traces. We derive noise from the power spectral density $S(f)$ as a function of the frequency f (see Fig. 4) measured from the noise of the SET current under normal device operation conditions, following

$$\delta_c(t) = \mathcal{F}(\mathcal{F}(\delta_g(t))\sqrt{S(f)}), \tag{4}$$

where $\delta_c(t)$ and $\delta_g(t)$ is a signal trace containing coloured noise and Gaussian noise, respectively, and $\mathcal{F}$ denotes the Fourier transform. The network *C* is trained by measured spin detection-signal traces recorded in single-shot fashion as described above. The labels $|\uparrow\rangle$ and $|\downarrow\rangle$ for these data are assigned by the Bayesian filter[10].

Data availability
The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References
1. Nowack, K. C. *et al.* Single-shot read-out of an individual electron spin in a quantum dot. *Science* **333**, 1269–1273 (2011).
2. Fowler, A. G. *et al.* Surface codes: Towards practical large-scale quantum computation. *Phys. Rev. A* **86**, 032324 (2012).
3. D'Anjou, B. & Coish, W. A. Optimal post-processing for a generic single-shot qubit readout. *Phys. Rev. A* **89**, 012313 (2014).
4. Robledo, L. *et al.* High-fidelity projective read-out of a solid-state spin quantum register. *Nature* **477**, 574–578 (2011).
5. Elzerman, J. M. *et al.* Single-shot read-out of an individual electron spin in a quantum dot. *Nature* **430**, 431–435 (2004).
6. Petta, J. R. *et al.* Coherent manipulation of coupled electron spins in semiconductor quantum dots. *Science* **309**, 2180–2184 (2005).
7. Maune, B. M. *et al.* Coherent singlet-triplet oscillations in a silicon-based double quantum dot. *Nature* **481**, 344–347 (2012).
8. Morello, A. *et al.* Single-shot readout of an electron spin in silicon. *Nature* **467**, 687–691 (2010).
9. Prance, J. R. *et al.* Identifying single electron charge sensor events using wavelet edge detection. *Nanotechnology* **26**, 215201 (2015).
10. Struck, T. *et al.* Low-frequency spin qubit energy splitting noise in highly purified ^{28}Si/SiGe. *npj Quantum Inf.* **6**, 40 (2020).
11. Hollmann, A. *et al.* Large, tunable valley splitting and single-spin relaxation mechanisms in a Si/Six Ge1-x quantum dot. *Phys. Rev. Appl.* **430**, 034068 (2020).
12. D'Anjou, B. *et al.* Maximal adaptive-decision speedups in quantum-state readout. *Phys. Rev. X* **6**, 011017 (2016).
13. Li, R. *et al.* A crossbar network for silicon quantum dot qubits. *Sci. Adv.* **4**, eaar3960 (2018).
14. Vandersypen, L. M. K. *et al.* Interfacing spin qubits in quantum dots and donors—hot, dense, and coherent. *Phys. Rev. Appl.* **3**, 34 (2017).
15. Lennon, D. T. *et al.* Efficiently measuring a quantum device using machine learning. *npj Quantum Inf.* **5**, 79 (2020).
16. Kalantre, S. S. *et al.* Machine learning techniques for state recognition and auto-tuning in quantum dots. *npj Quantum Inf.* **5**, 6 (2019).
17. Nguyen, V. *et al.* Deep reinforcement learning for efficient measurement of quantum devices. *npj Quantum Inf.* **7**, 100 (2021).
18. Zwolak, P. *et al.* Autotuning of double-dot devices in situ with machine learning. *Phys. Rev. Appl.* **13**, 034075 (2020).
19. Abadi, M., *et al.* TensorFlow: Large-scale machine learning on heterogeneous systems (2015).
20. Nair, V. *et al.* Rectified linear units improve restricted Boltzmann machines Vinod Nair. *Proc. ICML* **27**, 807–814 (2010).
21. Kingma, D. P. *et al.* Adam: A method for stochastic optimization. arXiv:1412.6980 (2017).
22. Leshno, M. *et al.* Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Netw.* **6**, 861–867 (1993).
23. Yoneda, J. *et al.* A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%. *Nat. Nanotechnol.* **13**, 102–106 (2018).
24. Simmons, C. B. *et al.* Tunable spin loading and T1 of a silicon spin qubit measured by single-shot readout. *Phys. Rev. Lett.* **106**, 156804 (2011).
25. Lam, S. K. *et al. Numba: A LLVM-Based Python JIT Compiler* (Association for Computing Machinery, 2015). https://doi.org/10.1145/2833157.2833162.
26. Vink, I. T. *et al.* Cryogenic amplifier for fast real-time detection of single-electron tunneling. *Appl. Phys. Lett.* **91**, 123512 (2007).

Author contributions

T.S. and J.L. developed the network architecture. T.S. did the computation and analysis of the data assisted by F.S. and A.S. and supported by L.R.S. T.S. and A.H. measured noise spectra and experimental signal traces on a sample fabricated by F.S., A.S. and D.B. L.R.S. conceived and supervised the study and all authors discussed the results. T.S., A.S., D.B. and L.R.S. wrote the manuscript, which all other authors reviewed.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to L.R.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.